

Cursus: M1, computer-science
Code UE: JEIN8602
Subject: Formal languages theory
Date: 11 July 2020
Duration: 3H
Documents: authorized
Lectures by: Mr Géraud Sénizergues

The exercises are independant one from each other.
Thus each mathematical symbol (G, L, \dots) has a single definite meaning *inside* each exercise, but might have different meanings from an exercise to another.
It is *not* required to solve *all* the exercises. Every correct solution to an exercise will give (around) 4 points. A question marked with a (*) is difficult. One can admit the result of such a question and, nevertheless, *use* it in subsequent questions.

Exercise 1 [/4] Let us consider the regular expression:

$$e := ((bc)^*(cb)^*) \cup ((ca)^*b(ccb)^*)$$

Construct, by Glushkov's method, a finite automaton recognizing L_e .

Exercise 2 [/5] We consider the language

$$L = \{a, b\}^* \{abb, baa\} \{a, b\}^*$$

- 1- Construct a non-deterministic finite automaton recognizing L .
- 2- Give a *complete deterministic* finite automaton recognizing L .
- 3- Compute the *minimal* complete deterministic finite automaton recognizing L .
- 4- Give a finite automaton recognizing $\{a, b\}^* \setminus L$ (i.e. the complement of L in $\{a, b\}^*$).
- 5- Give a *regular expression* for the language $\{a, b\}^* \setminus L$.

Exercise 3 [/4]

For each of the following languages over the terminal alphabet $\{a, b, c, d\}$, construct a context-free grammar that generates the language L_i :

$$\begin{aligned} L_1 &:= \{a^n b^n \mid n \geq 0\} \\ L_2 &:= \{a^p b^q c^r \mid p \geq 0, q \geq 0, r \geq 0, p = q \text{ or } q = r\} \\ L_3 &:= \{a^p b^q c^r \mid p \geq 0, q \geq 0, r \geq 0, p + q = r\} \\ L_4 &:= \{a^p b^q c^r \mid p \geq 0, q \geq 0, r \geq 0, p + r = q\} \\ L_5 &:= \{a^p b^q c^r d^s \mid p \geq 0, q \geq 0, r \geq 0, s \geq 0, p + r = q + s\} \end{aligned}$$

Exercise 4 [/6] We recall that the language

$$L_0 = \{a^n b^n \mid n \geq 0\}$$

is non-regular.

The exercise consists in showing that various similar languages are non-regular. The main tool for doing so is to use the *closure properties* of the family of regular languages: it is closed under boolean operations, under union, product, star, and under regular substitutions.

1- Let $L_1 = \{a^n \# b^n \mid n \geq 0\}$. Let $h : \{a, b, \#\}^* \rightarrow \{a, b\}^*$ be the homomorphism such that $h(a) = a, h(b) = b, h(\#) = \varepsilon$.

1.1 show that $h(L_1) = L_0$

1.2 show that L_1 is not regular.

2- Let $L_2 = \{u\#v \mid u, v \in \{a, b\}^*, |u| = |v|\}$. Show that L_2 is not regular.

Hint: Find a regular set R such that $L_2 \cap R = L_1$.

3- Let $L_3 = \{a^n \# a^n \mid n \geq 0\}$. Show that L_3 is not regular.

Hint: Find a finite substitution $\sigma : \mathcal{P}(\{a, \#\}^*) \rightarrow \mathcal{P}(\{a, b, \#\}^*)$ such that $\sigma(L_3) = L_2$.

4- Let $L_4 = \{a^{n_1} \# a^{n_2} \dots \# a^{n_k} \dots \# a^{n_p} \mid p \geq 1, \exists i, j \in [1, p], (i \neq j \text{ and } n_i = n_j)\}$.

Show that L_4 is not regular.

Let us consider now the language

$$M_0 = \{a^p b^q \mid p \geq 0, q \geq 0, p \neq q\}$$

5- Show that M_0 is not regular.

Hint: express L_0 as a boolean combination of some regular languages with M_0 .

6- Let

$$M_1 = \{a^p \# b^q \mid p \geq 0, q \geq 0, p \neq q\}, \quad M_2 = \{u\#v \mid u, v \in \{a, b\}^*, |u| \neq |v|\},$$

$$M_3 = \{a^p \# a^q \mid p \geq 0, q \geq 0, p \neq q\},$$

$$M_4 = \{a^{n_1} \# a^{n_2} \dots \# a^{n_k} \dots \# a^{n_p} \mid p \geq 1, \forall k \in [1, p], n_k \geq 0, \exists i, j \in [1, p], (i \neq j \text{ and } n_i \neq n_j)\}.$$

Show that M_1, M_2, M_3, M_4 are not regular.

7- Are the languages L_4, M_4 context-free ?

Give c.f. grammars for those languages (among L_4, M_4) which are context-free.

Exercise 5 [/8] Let us consider the language

$$L_0 = \{w \in \{a, b\}^* \mid |w|_a = |w|_b\}.$$

We have considered (course 7) the following c.f. grammar $G = \langle A, N, P, S \rangle$ with

$$A = \{a, b\}, \quad N = \{S\}, \quad S \longrightarrow aSbS \mid bSaS \mid \varepsilon$$

and we have proved that

$$L(G, S) = L_0.$$

Let us consider the language

$$L_1 = \{w \in \{a, b\}^* \mid |w|_a \geq |w|_b\}.$$

1- Show that $L_1 = L_0(aL_0)^*$.

2- Build a context-free grammar G_1 generating the language L_1 .

We have given (course 7) a non-ambiguous grammar H generating L_0 :
 $H = \langle A, N, R, S \rangle$ with

$$\begin{aligned} A &= \{a, b\}, \quad N = \{S, S_a, S_b, D_a, D_b\}, \\ S &\longrightarrow D_aS \mid D_bS \mid \varepsilon, \\ S_a &\longrightarrow D_aS_a \mid \varepsilon, \quad S_b \longrightarrow D_bS_b \mid \varepsilon, \\ D_a &\longrightarrow aS_ab, \quad D_b \longrightarrow bS_ba. \end{aligned}$$

We recall the notation: for every word $u, v, w \in \{a, b\}^*$

$$\|w\| = |w|_a - |w|_b$$

$u \preceq v$ means that u is a prefix of v .

3- Prove, by induction over the length of words, that, for every $w \in \{a, b\}^$:

$$w \in L(H, S_a) \Leftrightarrow \left(\|w\| = 0 \text{ and } (\forall v \in \{a, b\}^*, (v \preceq w \Rightarrow \|v\| \geq 0)) \right).$$

4- Prove that

$$L_1 = L(H, S)(aL(H, S_a))^*.$$

*5- Prove that: if $\left(n \geq 0, m \geq 0, \quad u_0, u'_0 \in L(H, S), \forall i \in [1, n], v_i \in L(H, S_a), \quad \forall j \in [1, m], v'_j \in L(H, S_a) \text{ and } u_0av_1av_2 \cdots av_n = u'_0av'_1av'_2 \cdots av'_m \right)$ then

$$n = m \text{ and } \forall i \in [1, n], \quad v_i = v'_i.$$

6- Give a *non-ambiguous* context-free grammar generating L_1 .

Exercise 6 [/4] We consider the context-free grammar $G := \langle A, N, R, S_1 \rangle$ where $A = \{a, b, c\}$, $N = \{S_1, S_2, S_3, S_4, S_5, S_6\}$ and R consists of the following 13 rules:

$$\begin{aligned}
 S_1 &\rightarrow aS_3S_5 & S_1 &\rightarrow cS_6 & S_1 &\rightarrow S_2S_3 \\
 S_1 &\rightarrow aS_4 & S_1 &\rightarrow aS_1 \\
 S_2 &\rightarrow aS_1 & S_2 &\rightarrow bS_2S_4 \\
 S_3 &\rightarrow bS_3S_5 & S_3 &\rightarrow aS_4S_4 \\
 S_4 &\rightarrow aS_3S_6 & S_4 &\rightarrow aS_3S_3 \\
 S_5 &\rightarrow a & S_5 &\rightarrow S_3S_4 \\
 S_6 &\rightarrow bS_5S_5
 \end{aligned}$$

The start symbol of G is S_1 .

- 1- What are the *productive* non-terminals of G ?
- 2- What are the *useful* non-terminals of G ?
- 3- Transform the grammar G into an equivalent grammar G' where every non-terminal is productive and useful.
- 4- Is the language $L(G, S_1)$ empty ?
- 5- Is the language $L(G, S_1)$ finite ?
- 6- Is the language $L(G, S_1)$ regular?