# Langages Formels

## TD n°1

Emilie Grienenberger
`emilie.grienenberger@lsv.fr`

28 janvier 2021

C'est pour l'instant une correction partielle, sans les représentations d'automates.

**Exercise 1 : Construction d'automates**

On note $\Sigma = \{0, 1\}$.

1. Donner un automate qui reconnaît les langages :
   (a) $L_1 = \{u \in \Sigma^* :$ toute occurence de 1 est suivie de deux occurences de 0$\}$
   (b) $L_2 = \{u \in \Sigma^* :$ $u$ ne contient pas deux occurences successives de 0$\}$
   (c) $L_3 = \{u \in \Sigma^* :$ le nombre d'occurences de 1 est pair$\}$
2. Donner un automate qui reconnaît le langage des multiples de 3 en base 2 où la représentation des entiers est "big-endian" (i.e. les bits sont rangés du plus fort au plus faible).

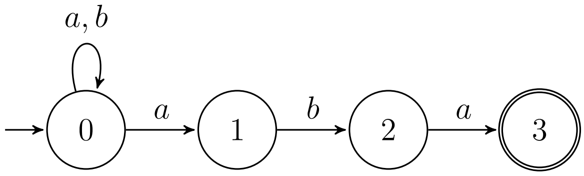
**Exercise 2 : Déterminisation**

1. Pour chacun des automates $\mathcal{A}_1$ et $\mathcal{A}_2$, déterminer le langage reconnu et donner un automate déterministe équivalent.
2. Si $\Sigma = \{a, b\}$, montrer que le langage $L = \Sigma^* a \Sigma^{n-1}$ ne peut être reconnu par un automate déterministe de moins de $2^n$ états.

   Soit $\mathcal{A}$ un automate déterministe reconnaissant $L$ d'état initial $i$, et de fonction de transition étendue $\delta^*$.

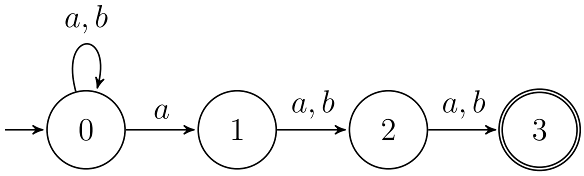
   Soient $u$ et $v$ deux mots distincts de $\Sigma^n$. Sans perte de généralité, il existe $u', v', w$ tels que $u = u'aw$ et $v = v'bw$. Supposons $\delta^*(i, u) = \delta^*(i, v)$. Puisque $\delta^*(i, ua^{n-1-|w|}) = \delta^*(i, va^{n-1-|w|})$ et $ua^{n-1-|w|} \in L$, on a que $va^{n-1-|w|} \in L$ ce qui est contradictoire.

   Ainsi, il y a au moins autant d'états dans $\mathcal{A}$ que de mots de $\Sigma^n$, c'est-à-dire au moins $2^n$.
3. Conclure sur le nombre d'états d'automates déterminisés par rapport à leur version non déterministe.




(a) Automate $\mathcal{A}_1$

Le langage reconnu est $\Sigma^* aba$.




(b) Automate $\mathcal{A}_2$

Le langage reconnu est $\Sigma^* a \Sigma^2$.

Il est au plus exponentiel par définition de l'algorithme de déterminisation, et cette borne est atteinte.

**Exercise 3 : Des étoiles**

On considère les énoncés suivants, de plus en plus forts.

Soit $L \subseteq \Sigma^*$, il existe $N > 0$ tel que pour tout $x \in L$ :

(a) Si $|x| \geq N$ alors il peut s'écrire $x = u_1 u_2 u_3$ avec $u_2 \neq \varepsilon$ et $u_1 u_2^* u_3 \subseteq L$.

(b) Si $x = w_1 w_2 w_3$ avec $|w_2| \geq N$ alors $w_2 = u_1 u_2 u_3$ avec $u_2 \neq \varepsilon$ et $w_1 u_1 u_2^* u_3 w_3 \subseteq L$.

(c) Si $x = u v_1 v_2 \dots v_M w$ avec $|v_i| \geq 1$ et $M \geq N$, alors il existe $0 \leq j < k \leq M$ tel que $u v_1 \dots v_j (v_{j+1} \dots v_k)^* v_{k+1} \dots v_M w \subseteq L$.

1. Montrer que tout langage reconnaissable vérifie l'énoncé (a).

Soit $\mathcal{A} = \langle Q, i, F, \delta \rangle$ un automate déterministe econnaissant un langage $L$. Soit $N = |Q| + 1$ et $x \in L$ tel que $|x| \geq N$.

Par principe des tiroirs, il existe $u_1, u_2, u_3$ tels que $u = u_1 u_2 u_3$ et $\delta^*(i, u_1) = \delta^*(i, u_1 u_2)$. Notons $q$ cet état.

On a que $\delta^*(q, u_2) = q$, et donc par récurrence immédiate : pour tout $k \in \mathbb{N}$, $\delta^*(q, u_2^k) = q$. Nous savons de plus que $\delta^*(q, u_3) \in F$ car $x \in L$.

On construit aisément l'exécution acceptante pour tout mot de $u_1 u_2^* u_3$ dans $\mathcal{A}$.

2. Montrer que $\{ \, a^n b^n \ : \ n \in \mathbb{N} \, \}$ n'est pas reconnaissable.

Par l'absurde : on applique le critère (a) sur $x = a^N b^N$, et on effectue une disjonction de cas : $u_2 = a^i$, $u_2 = a^i b^j$ ou $u_2 = b^j$. Les trois cas mènent à des absurdités.

On remarque qu'on s'embêterait moins avec le critère (b) en choisissant $w_1 = a^N$, $w_2 = b^N$, $w_3 = \varepsilon$.

On admet que les langages reconnaissables satisfont les trois propriétés (a), (b), et (c).

3. Montrer que le langage $\{ a^{n^2} : n \in \mathbb{N} \}$ n'est pas reconnaissable.

Par l'absurde : on applique le critère (b) sur $x = a^{N^2}$ et $w_1 = a^{N^2-N}$, $w_2 = a^N$, $w_3 = \varepsilon$. On conclut à une contradiction : $a^{N^2+m}$ n'est pas dans le langage considéré si $0 < m \leq N$ car $N^2 + m < N^2 + 2N + 1 = (N + 1)^2$ donc $N^2 + m$ n'est pas un carré.

4. Montrer que $\{ u : u = \widetilde{u} \}$, c'est à dire le langage des palindromes, n'est pas reconnaissable.

Par l'absurde : on applique le critère (b) sur $x = a^N ba^N$ et $w_1 = a^N b$, $w_2 = a^N$, $w_3 = \varepsilon$. On conclut à une contradiction.

5. De même pour $\{ (ab)^n(cd)^n \} \cup (\Sigma^*\{ aa, bb, cc, dd, ac \}\Sigma^*)$.

Par l'absurde : on applique le critère (c) sur $x = (ab)^N(cd)^N$ avec $u = w = \varepsilon$, $v_1 = \cdots = v_N = ab$, et $v_{N+1} = \cdots = v_{2N} = cd$. On conclut à une contradiction de manière similaire à la question 2.

6. Montrer que le langage suivant satisfait (c) mais n'est pas reconnaissable : $\{ udv : u, v \in \{a, b, c\}^*,\ u \neq v$ ou bien l'un des deux contient un carré $\}$. On pourra admettre l'existence d'une infinité de mots sans carrés sur l'alphabet $\{a, b, c\}$.

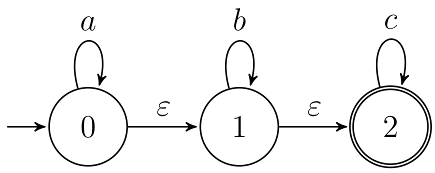
Supposons qu'il soit reconnaissable par un automate déterministe $\mathcal{A}$ à $N$ états. Il existe deux mots $w_1, w_2 \in \{ a, b, c \}^*$ de taille supérieure à $N$ que $\mathcal{A}$ ne sait différencier, qui ne contiennent pas de carrés. Or, $w_1dw_2$ est dans le langage mais pas $w_2dw_2$, ce qui est absurde.

On peut noter $x = udv$. On prend une décomposition selon (c), avec $M \geq 4$. Au moins deux $v_i$ adjacents sont dans $u$ ou $v$, on suppose qu'ils sont dans $u$ sans perte de généralité, et on note $u = u_1v_iv_{i+1}u_2$. On ne peut avoir $u_1v_iu_2 = u_1u_2 = v$. On choisit le bon facteur à itérer ne rendant pas les composantes égales pour 0, laissant le mot identique pour 1 et créant un carré ensuite.

## Exercise 4 : $\varepsilon$-moves

Un automate fini avec $\varepsilon$-transitions sur un alphabet $\Sigma$ est un 5-uplet $\langle Q, I, F, \delta\rangle$ où $I, F \subseteq Q$ et $\delta \subseteq Q \times \Sigma \cup \{\varepsilon\} \times Q$. Une $\varepsilon$-transition (une transition étiquetée par $\varepsilon$) peut être empruntée sans consommer de lettres du mot d'entrée. Les notions de chemin et d'acceptation sont étendues de manière naturelle.

1. On se place sur l'alphabet $\Sigma = \{a, b, c\}$. Quel est le langage reconnu par l'automate $\mathcal{A}_0$ ?




Un automate $\mathcal{A}_0$ avec $\varepsilon$-transitions

Le langage reconnu est $\{ a^n b^m c^k \ : \ n, m, k \in \mathbb{N} \}$.

Soit $\mathcal{A} = \langle Q, I, F, \delta \rangle$ un automate avec $\varepsilon$-transitions. Pour tout état $q \in Q$, on note $Cl(q)$ la clôture par $\varepsilon$-transitions de l'état $q$ :

$$q' \in Cl(q) \quad \Leftrightarrow \quad \text{il existe } n \in \mathbb{N} \text{ tel que } q \xrightarrow{\varepsilon}{}^{n}_{\delta} q'$$

2. Quelles sont les clôtures des états de $\mathcal{A}_0$ ?

$Cl(0) = \{ 0, 1, 2 \}$, $Cl(1) = \{ 1, 2 \}$, et $Cl(2) = \{ 2 \}$.

3. Pour tout $\mathcal{A} = \langle Q, I, F, \delta \rangle$ un automate avec $\varepsilon$-transitions, construire un automate sans $\varepsilon$-transitions $\hat{\mathcal{A}} = \langle Q, I, \hat{F}, \hat{\delta} \rangle$ qui reconnait le même langage.

On va construire un automate $\hat{\mathcal{A}} = \langle Q, I, \hat{F}, \hat{\delta} \rangle$ sans $\varepsilon$-transitions qui reconnait le même langage que $\mathcal{A}$. On définit $\hat{\delta}$ et $\hat{F}$ de la manière suivante. Pour $a \in \Sigma$ :

$$\begin{aligned} q \in \hat{F} \quad &\Leftrightarrow \quad \text{ou} \begin{cases} q \in F \\ \text{il existe } q' \in Cl(q) \text{ tel que } q' \in F \end{cases} \\ (q, a, q') \in \hat{\delta} \quad &\Leftrightarrow \quad \bigcup_{p \in P_{q,a}} Cl(p) \\ &\qquad\ \text{where } P_{q,a} = \{ p \ : \ \exists r \in Cl(q).\ (r, a, p) \in \delta \} \end{aligned}$$

On opère une induction sur un mot $w$, avec pour hypothèse d'induction que s'il existe un chemin dans $\mathcal{A}$ étiqueté par $w$ de la forme $c = q_1 \xrightarrow{a_1}_{\delta} \dots \xrightarrow{a_{n-1}}_{\delta} q_n$, alors il existe un chemin dans $\hat{\mathcal{A}}$ étiqueté par $c$ de la forme $s_1 \xrightarrow{b_1}_{\hat{\delta}} \dots \xrightarrow{b_{k-1}}_{\hat{\delta}} s_k$ tel que $s_1 = q_1$ et $q_n \in Cl(s_k)$.

**Si** $w = \varepsilon$ **:** $a_1 = \dots = a_{n-1} = \varepsilon$, donc $q_n \in Cl(q_1)$ et le chemin $q_1$ statisfait les hypothèses d'induction.

**Si** $w = aw'$ **:** Il existe $i \leq n$ tel que pour tout $j < i$, $a_j = \varepsilon$, et $a_i = a$. On applique l'hypothèse d'induction à $w'$ et au chemin $q_{i+1} \xrightarrow{a_{i+1}}_{\delta} \dots \xrightarrow{a_{n-1}}_{\delta} q_n$. On obtient un chemin $c = q_{i+1} = s_1 \xrightarrow{b_1}_{\delta} \dots \xrightarrow{b_k}_{\delta} s_k$ dans $\hat{\mathcal{A}}$ étiqueté par $w'$, commençant par $q_{i+1}$ et terminant par un état $s_k$ tel que $q_n \in Cl(s_k)$.
On a de plus que $q_i \in Cl(q_1)$, et $(q_i, a, q_{i+1}) \in \delta$. Ainsi, $(q_1, a, q_{i+1}) \in \hat{\delta}$. En juxtaposant le chemin $q_1 \xrightarrow{a}_{\hat{\delta}} q_{i+1}$ et $c$, on obtient un chemin de la bonne forme.

On conclut. Soit $w$ un mot du langage reconnu par $\mathcal{A}$. Alors il existe un chemin $i = q_1 \xrightarrow{a_1} \dots \xrightarrow{a_{n-1}} q_n \in F$. Alors il existe un chemin $i = s_1 \xrightarrow{b_1} \dots \xrightarrow{b_{k-1}} s_k$ dans $\hat{\mathcal{A}}$ étiqueté par $w'$ tel que $q_n \in Cl(s_k)$. Donc $s_k \in \hat{F}$ et ce chemin est acceptant.

4. Donner un automate déterministe qui reconnait le même langage que l'automate $\mathcal{A}_0$.

**Exercise 5 : Opérations et problèmes : questions de cours**

1. Étant donné un langage reconnaissable, montrer que son complémentaire est encore reconnaissable.
2. Étant donnés deux langages reconnaissables montrer que leur union, intersection et différence sont encore des langages reconnaissables.
3. Étant donné un langage reconnaissable $L$, montrer que l'ensemble des suffixes des mots de $L$ est encore reconnaissable.
4. L'appartenance du mot vide $\varepsilon$ à un langage reconnaissable est-elle décidable ?
5. La vacuité d'un langage reconnaissable est-elle décidable ?
6. L'universalité d'un langage reconnaissable est-elle décidable ?
7. Étant donné un langage reconnaissable $L$ non vide, montrer que l'on peut calculer le plus long préfixe commun à tous les mots de $L$.