École normale supérieure **2015-2016**
Département d'informatique

Langages formels, calculabilité, complexité
Partiel
Durée : 3 heures
Documents autorisés

Exercice 1. Posons $L = \{a\}^* \cup \{b\}^* \cup \{c\}^*$.

1. Donner l'automate fini déterministe complet minimal $\mathcal{A}$ tel que $\mathcal{L}(\mathcal{A}) = L$.
2. Donner le monoïde syntaxique $\mathcal{M}$ de L.
3. Montrer que $\mathcal{M}$ n'est isomorphe à aucun monoïde syntaxique d'un langage régulier à deux lettres.

Exercice 2. Un langage L sur un alphabet fini Σ est dit *local* s'il existe deux parties $P, S \subseteq \Sigma$ et une partie N de Σ^2 telles que

$$L \setminus \{\epsilon\} = (P\Sigma^* \cap \Sigma S) \setminus \Sigma^* N \Sigma^*.$$

La terminologie s'explique ainsi : pour tester si un mot appartient à L, il suffit de vérifier que sa première lettre est dans P, sa dernière lettre est dans S, et que ses facteurs de longueur 2 ne sont pas dans N. Toutes ces vérifications sont « locales ».

1. Montrer, en donnant les ensembles P, S, N appropriés, que $\{abc\}^*$ est un langage local sur $\Sigma = \{a, b, c\}$.
2. Montrer qu'un langage local L sur Σ est reconnu par un automate fini $\mathcal{A} = (Q, \Sigma, \delta, I, F)$ vérifiant $\#Q = 1 + \#\Sigma$ et tel que $\{(p, a, q) | p, q \in Q\}$ est un singleton pour tout $a \in \Sigma$.
3. Montrer que si L est local, alors L^* est local.
4. Montrer que si L_1 et L_2 sont deux langages locaux sur Σ_1 et Σ_2 respectivement, avec $\Sigma_1 \cap \Sigma_2 = \emptyset$, alors $L_1 \cup L_2$ et $L_1 L_2$ sont des langages locaux.
5. Montrer qu'un langage L sur Σ est local si et seulement si, pour tout $a \in \Sigma$, l'ensemble $\{(ua)^{-1}L | u \in \Sigma^*\}$ contient au plus un élément.
6. On appelle automate *local* un automate fini déterministe (mais pas néssairement complet) $\mathcal{A} = (Q, \Sigma, \delta, I, F)$ tel que pour toute lettre $a \in \Sigma$, l'ensemble $\{(p, a, q) | p, q \in Q\}$ contient au plus un élément.
   Montrer qu'un langage reconnaissable est local si et seulement si il est reconnu par un automate local.
7. Montrer que pour tout langage régulier $L \subseteq \Sigma^*$, il existe un alphabet Σ', un langage local L' sur Σ' et un morphisme $\sigma : \Sigma'^* \to \Sigma^*$ tel que $\sigma(L') = L$.
8. Soit L un langage local sur Σ. Montrer que si xyz et xy^2z sont des éléments de L, alors $xy^+z \subseteq L$.

Exercice 3 (Grammaires algébriques)**.** Le but de cet exercice est d'étudier les langages L_i pour $i \in \mathbb{N}$ définis sur les alphabets $\Sigma_i = \{0, 1, 2, \ldots, i-1\}$ par :

$$L_i = \{w = xyyz, \text{ avec } x, y, z \in \Sigma_i^*, y \neq \epsilon\} \subset {\Sigma_i}^*.$$

Les langages L_i pour $i \in \mathbb{N}$ sont donc les langages formés des mots qui contiennent un carré sur l'alphabet à i lettres Σ_i.

1. Montrer que L_1 et L_2 sont des langages réguliers.
2. Rappelons que le mot (binaire) de Thue-Morse

$$\tau = \tau(0)\tau(1)\tau(2)\tau(3)\cdots = 0110100110010110100101100110100 1\ldots$$

est défini par $\tau(n) = 0$ si et seulement le nombre de 1 dans le développement binaire de l'entier n est pair.
Considérons le morphisme $\sigma : \Sigma_3^* \longrightarrow \Sigma_2^*$ défini par $\sigma(0) = 0$, $\sigma(1) = 01$ et $\sigma(2) = 011$ (et étendu par compatibilité avec le produit de Σ_3^*). Montrer que le mot infini ternaire $\sigma^{-1}(\tau)$ est sans carré (i.e. que les facteurs de $\sigma^{-1}(\tau)$ n'appartiennent pas à L_3).

Nous allons maintenant démontrer que L_i n'est pas un langage algébrique pour $i \geq 3$. Soit $G = (V, \Sigma, P, S)$ une grammaire algébrique sous forme normale de Chomsky (pour un alphabet Σ) et soit $\mathcal{L}(G) \subset \Sigma^*$ le langage engendré par G. Soient m et n deux entiers avec $2 \leq m \leq n$.

3. Montrer que pour tout mot $z \in \mathcal{L}(G)$ tel que $|z| \geq m$, il existe une variable $A \in V$ et une dérivation

$$S \overset{*}{\Rightarrow} wAy \overset{*}{\Rightarrow} wxy = z$$

avec $m/2 < |x| \leq m$.

4. Pour tout ensemble $R \subseteq \mathcal{L}(G) \cap \Sigma^n$, toute variable $A \in V$ et tout couple d'entiers (n_1, n_2) tels que $0 \leq n_1, n_2 \leq n$, nous posons

$$Q_{n,R}(n_1, A, n_2) = \left\{ z \in R \,\middle|\, \begin{array}{l} \text{Il existe une dérivation } S \overset{*}{\Rightarrow} wAy \overset{*}{\Rightarrow} wxy = z \\ \text{avec } |w| = n_1 \text{ et } |y| = n_2 \end{array} \right\}.$$

Montrer que pour tout ensemble R, il existe deux entiers n_1 et n_2 tels que $0 \leq n_1, n_2 \leq n$ avec $m/2 < n - n_1 - n_2 \leq m$ et une variable $A \in V$ tel que

$$\#Q_{n,R}(n_1, A, n_2) \geq \frac{\#R}{(\#V)(n+1)^2}.$$

5. Montrer qu'il existe une constante $c > 0$ telle que pour tout ensemble $R \subseteq \mathcal{L}(G) \cap \Sigma^n$ et pour tout couple d'entiers (m, n) avec $2 \leq m \leq n$, il existe un sous-ensemble $Z = \{z_1, \ldots, z_k\} \subseteq R$ avec $k > \#R/(c(n+1)^2)$ et des décompositions $z_i = w_i x_i y_i$ tels que
   — $|w_1| = |w_2| = \cdots = |w_k|$;
   — $|y_1| = |y_2| = \cdots = |y_k|$;
   — $m/2 < |x_1| = |x_2| = \cdots = |x_k| \leq m$;
   — $w_i x_j y_i \in L$ pour tout couple d'entiers (i, j) vérifiant $1 \leq i, j \leq k$.
6. Nous allons montrer que L_6 n'est pas algébrique.
   (a) Supposons par l'absurde que L_6 est algébrique et soit N un entier. Montrer qu'il existe un mot $v = v_1 \ldots v_N$ de $\{3, 4, 5\}^*$ sans carré et considérer le langage

$$R = \{0t0t | t = d_0 v_1 d_1 v_2 \ldots v_N d_N, d_i \in \{2, 3\}\} \subset \Sigma_6^{4(N+1)}.$$

   Montrer que tout mot de R est un carré mais ne contient pas de facteur carré propre.

(b) Montrer qu'il existe une constante c et un sous-ensemble $Z = \{z_1, \ldots, z_k\} \subseteq R$ avec $k > 2^{N+1}/16c(N+1)^2$ vérifiant les conditions de la question 5.

(c) Montrer que $k \leq 2^{(N+1)/2}$ et conclure.

7. Nous admettons qu'il existe un morphisme $\theta : \Sigma_6^* \to \Sigma_i^3$ pour lequel $w \in \Sigma_6^*$ est sans carré si et seulement si $\theta(w) \in \Sigma_3^*$ est sans carré. Montrer que L_3 n'est pas algébrique
8. En déduire que L_i n'est pas algébrique pour $i \geq 3$.
9. En considérant le morphisme $\psi : \Sigma_3^* \longrightarrow \Sigma_2^*$ défini par $\psi(0) = 0$, $\psi(1) = 1$ et $\psi(2) = 1$ (et étendu par compatibilité avec le produit de Σ_3^*). Montrer que l'ensemble des mots ternaires qui contiennent un chevauchement (*i.e.* un facteur de la forme $uvuvu$ avec $u, v \in \Sigma^*$ et $u \neq \epsilon$) n'est pas algébrique.
10. * Considérons le morphisme $\theta : \Sigma_6^* \to \Sigma_3^*$ défini par

$$\begin{array}{rcl}
\theta(0) & = & 010201202201210201\,0212 \\
\theta(1) & = & 0102012022201210120212 \\
\theta(2) & = & 0102012101202101210212 \\
\theta(3) & = & 0102012101202120121012 \\
\theta(4) & = & 0102012102010210120212 \\
\theta(5) & = & 0102012102120210120212
\end{array}$$

(et étendu par compatibilité avec le produit de Σ_6^*). Montrer qu'un mot $w \in \Sigma_6^*$ est sans carré si et seulement si $\theta(w) \in \Sigma_3^*$ est sans carré.