Grammaires et automates finis

Corrigé partiel de la feuille de travaux dirigés n^{o}8

30 mars 2009

1. On peut envisager plusieurs démarches pour répondre aux questions posées :

i) En représentant graphiquement l'automate, on reconnait qu'il s'agit de l'automate non déterministe construit de manière algorithmique à partir de l'expression rationnelle $(a+b)^\star abb$. Ceci nous permet d'obtenir la grammaire :

$N = \{S\}, T = \{a,b\}, S, P = \left\{ \ S \to aS \mid bS \mid abb \ \right\}$

ii) construction de la grammaire à partir de l'automate :

$$N = \{S_0, S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9, S_{10}\}, T = \{a,b\}, S_0, P = \left\{ \begin{array}{ll} S_0 \to S_1 \mid S_7 & S_5 \to S_6 \\ S_1 \to S_2 \mid S_4 & S_6 \to S_1 \mid S_7 \\ S_2 \to aS_3 & S_7 \to aS_8 \\ S_3 \to S_6 & S_8 \to bS_9 \\ S_4 \to bS_5 & S_9 \to bS_{10} \\ S_{10} \to \varepsilon & \end{array} \right\}$$

Après suppression des renommages et des ε-productions :

$N = \{S_0, S_3, S_5, S_8, S_9\}, \quad T = \{a,b\}, \quad S_0,$

$$P = \left\{ \begin{array}{l} S_0 \to aS_3 \mid bS_5 \mid aS_8 \\ S_3 \to aS_3 \mid bS_5 \mid aS_8 \\ S_5 \to aS_3 \mid bS_5 \mid aS_8 \\ S_8 \to bS_9 \\ S_9 \to b \end{array} \right\}$$

Observons que les membres droits de S_0, S_3 et S_5 sont identiques et peuvent être identifiées. D'où la grammaire (sous forme normale de Greibach) :

$N = \{S_0, S_8, S_9\} \quad T = \{a,b\}, \quad S_0$

$$P = \left\{ \begin{array}{l} S_0 \to aS_0 \mid bS_0 \mid aS_8 \\ S_8 \to bS_9 \\ S_9 \to b \end{array} \right\}$$

Il est maintenant facile de conclure que le langage engendré est $(a+b)^\star abb$.

iii) troisième approche : déterminisation de l'automate de l'énoncé.

On obtient :

δ	a	b
$\to \{q_0, q_1, q_2, q_4, q_7\}$	$\{q_1, q_2, q_3, q_4, q_6, q_7, q_8\}$	$\{q_1, q_2, q_4, q_5, q_6, q_7\}$
$\{q_1, q_2, q_3, q_4, q_6, q_7, q_8\}$	$\{q_1, q_2, q_3, q_4, q_6, q_7, q_8\}$	$\{q_1, q_2, q_4, q_5, q_6, q_7, q_9\}$
$\{q_1, q_2, q_4, q_5, q_6, q_7\}$	$\{q_1, q_2, q_3, q_4, q_6, q_7, q_8\}$	$\{q_1, q_2, q_4, q_5, q_6, q_7\}$
$\{q_1, q_2, q_4, q_5, q_6, q_7, q_9\}$	$\{q_1, q_2, q_3, q_4, q_6, q_7, q_8\}$	$\{q_1, q_2, q_4, q_5, q_6, q_7, q_{10}\}$
$\leftarrow \{q_1, q_2, q_4, q_5, q_6, q_7, q_{10}\}$	$\{q_1, q_2, q_3, q_4, q_6, q_7, q_8\}$	$\{q_1, q_2, q_4, q_5, q_6, q_7\}$

après renommage :

δ	a	b
$\to q_0$	q_1	q_2
q_1	q_1	q_3
q_2	q_1	q_2
q_3	q_1	q_4
$\leftarrow q_4$	q_1	q_2

Ainsi la minimisation donne :

$\sim_0 \{q_0, q_1, q_2, q_3\} \ \{q_4\}$
$\sim_1 \{q_0, q_1, q_2\} \ \{q_3\} \ \{q_4\}$
$\sim_2 \{q_0, q_2\} \ \{q_1\} \ \{q_3\} \ \{q_4\}$

et donc l'automate

δ	a	b
$\to q_0$	q_1	q_0
q_1	q_1	q_2
q_2	q_1	q_3
$\leftarrow q_3$	q_1	q_0

On peut résoudre le système d'équations suivant :

$$\left\{\begin{array}{l} Z_0 = \varepsilon + Z_0 b + Z_3 b \\ Z_1 = Z_0 a + Z_1 a + Z_2 a + Z_3 a \\ Z_2 = Z_1 b \\ Z_3 = Z_2 b \end{array}\right. \qquad \text{d'où} \qquad \left\{\begin{array}{l} Z_1 = b^\star a + Z_1 b^\star a \\ Z_3 = Z_1 bb \end{array}\right.$$

D'où l'expression $(b^\star a)^+ bb$ pour le langage. A partir de cette expression, on peut proposer la grammaire :

$$N = \{S, A, B, C\}, \quad T = \{a, b\}, \quad S, \quad P = \left\{\begin{array}{l} S \to Abb \\ A \to AC \mid C \\ C \to a \mid Ba \\ B \to b \mid bB \end{array}\right\}$$

Nous pouvons remarquer que nous obtenons des grammaires très différentes.
Malheureusement, dès que le langage n'est pas rationnel, nous ne disposons pas d'outil pour vérifier que deux grammaires engendrent le même langage. Un outil de ce genre ne peut pas exister ! Pour plus de détails sur le sujet, nous vous orientons vers les ouvrages de la théorie de la calculabilité ou vers le cours optionnel de calculabilité en SI4.

2.

a) Comme toute sous-chaîne de longueur au plus 5 doit contenir au moins un 0, on déduit que la distance entre deux 0 consecutifs est au plus 4 d'où l'expression rationnelle
$L = (\varepsilon + 1 + 11 + 111 + 1111)\,(0(\varepsilon + 1 + 11 + 111 + 1111))^\star$ et la grammaire :

$$N = \{S, A, B, C\}, \quad T = \{0, 1\}, \quad S, \quad P = \left\{\begin{array}{ll} S \to AB & B \to BC \mid \varepsilon \\ A \to \varepsilon \mid 1 \mid 11 \mid 111 \mid 1111 & C \to 0A \end{array}\right\}$$

Autre manière de construire une grammaire pour ce langage : on considère le complémentaire du langage, i.e. le langage des mots ayant 11111 comme facteur qui a comme automate fini (minimal). En passant au complémentaire, on obtient l'automate fini :

δ	a	b
$\rightarrow q_0$	q_0	q_1
q_1	q_0	q_2
q_2	q_0	q_3
q_3	q_0	q_4
q_4	q_0	q_5
$\leftarrow q_5$	q_5	q_5

δ	a	b
$\longleftrightarrow q_0$	q_0	q_1
$\leftarrow q_1$	q_0	q_2
$\leftarrow q_2$	q_0	q_3
$\leftarrow q_3$	q_0	q_4
$\leftarrow q_4$	q_0	

Et ceci permet de déduire la grammaire :

$$N = \{S_0, S_1, S_2, S_3, S_4\}, \quad ,T = \{0, 1\}, \quad S, \quad P = \left\{\begin{array}{l} S_0 \to 0S_0 \mid 1S_1 \mid \varepsilon \\ S_1 \to 0S_0 \mid 1S_2 \mid \varepsilon \\ S_2 \to 0S_0 \mid 1S_3 \mid \varepsilon \\ S_3 \to 0S_0 \mid 1S_4 \mid \varepsilon \\ S_4 \to 0S_0 \mid \varepsilon \end{array}\right\}$$

b) De la définition du langage, on déduit :

$$N = \{S, V\}, \quad T = \{a, b, c\}, \quad S, \quad P = \left\{\begin{array}{l} S \to aVb \mid aVc \mid bVa \mid bVc \mid cVa \mid cVb \\ V \to \varepsilon \mid aV \mid bV \mid cV \end{array}\right\}$$

On pouvait aussi passer par l'automate fini (déjà rencontré en début de semestre ...), d'où la grammaire :

δ	a	b	c
$\rightarrow q_0$	q_1	q_2	q_3
q_1	q_1	q_4	q_4
q_2	q_5	q_2	q_5
q_3	q_6	q_6	q_3
$\leftarrow q_4$	q_1	q_4	q_4
$\leftarrow q_5$	q_5	q_2	q_5
$\leftarrow q_6$	q_6	q_6	q_3

$$
\begin{array}{l}
N = \{S_0, S_1, S_2, S_3, S_4, s_5, s_6\} \\
T = \{a, b, c\} \\
S \\
P = \left\{ \begin{array}{l}
S_0 \rightarrow aS_1 \mid bS_2 \mid cS_3 \\
S_1 \rightarrow aS_1 \mid bS_4 \mid cS_4 \\
S_2 \rightarrow aS_5 \mid bS_2 \mid cS_5 \\
S_3 \rightarrow aS_6 \mid bS_6 \mid cS_3 \\
S_4 \rightarrow aS_1 \mid bS_4 \mid cS_4 \mid \varepsilon \\
S_5 \rightarrow aS_5 \mid bS_2 \mid cS_5 \mid \varepsilon \\
S_6 \rightarrow aS_6 \mid bS_6 \mid cS_3 \mid \varepsilon
\end{array} \right\}
\end{array}
$$

3. On cherche d'abord les variables productives, et on trouve $P = \{X, Z\}$. On doit donc supprimer S et Y. On obtient la grammaire :

$$
\begin{array}{l}
N = \{X, Z\} \\
T = \{a, b, c\} \\
- \\
P \left\{ \begin{array}{l}
X \rightarrow aX \mid c \\
Z \rightarrow aZ \mid c
\end{array} \right.
\end{array}
$$

Dans cette grammaire nous n'avons plus d'axiome, c.a.d. qu'aucune variable n'est accessible, donc on ne peut pas dériver aucun mot, donc $L = \emptyset$.

4. a) La grammaire n'a pas l'air d'être linéaire. La première chose à faire est de la nettoyer. Nous supprimons la variable improductive F. On supprime A non-accessible. On obtient :

$$
N = \{S, B, D, E, G\}, \quad T = \{0, 1\}, \quad S, \quad P = \left\{ \begin{array}{l}
S \rightarrow 1B \mid 0E \\
B \rightarrow 0D \mid 1S \mid 0 \mid 1 \\
D \rightarrow G \mid 1E \\
E \rightarrow 0S \mid 1D \mid 0 \mid 1 \\
G \rightarrow 0B
\end{array} \right\}
$$

La substitution de G rend la grammaire linéaire, d'où l'automate fini :

δ	0	1
$\rightarrow q_0$	q_1	q_2
q_1	$\{q_0, q_4\}$	$\{q_3, q_4\}$
q_2	$\{q_3, q_4\}$	$\{q_0, q_4\}$
q_3	q_2	q_1
$\leftarrow q_4$		

b) Pour trouver l'expression régulière on peut résoudre un système associé à l'automate. Cependant ce système s'avère être lourd à résoudre. En effet, il est autrement plus simple de déterminiser et minimiser d'abord l'automate fini, pour obtenir :

δ	0	1
$\rightarrow q_0$	q_1	q_1
q_1	q_2	q_2
$\leftarrow q_2$	q_1	q_1

Ainsi il est facile d'obtenir l'expression : $L = ((0+1)^2)^+$.

5. a) Cette grammaire n'est pas ambiguë. Il est facile de montrer que le langage engendré est $a^\star b(a+b)^\star$. La seule manière d'obtenir un mot du langage est d'utiliser la première règle de manière à couper le mot lors du premier b. La partie qui précède (et qui contient que des a ne peut être dérivée que d'une seule manière. De même pour la deuxième partie.

b) Cette grammaire est ambiguë. En effet on peut avoir la dérivation : $S \rightarrow XaSbY \rightarrow XaXaSbYbY \rightarrow XaXabYbY$, qui permet d'obtenir par la suite le mot $aaabbb$ de quatre manières (selon si le a manquant est obtenu à partir du premier ou du second X, est de même pour le b manquant.