Grammaires

Corrigé partiel de la feuille de travaux dirigés n^{o}6

1. Numérotons les règles :

$$\begin{array}{ll} S \to_1 aB & A \to_5 bAA \\ S \to_2 bA & B \to_6 b \\ A \to_3 a & B \to_7 bS \\ A \to_4 aS & B \to_8 aBB \end{array}$$

a) On a trois dérivations gauches. Toutes trois commencent par $S \to_1 aB \to_8 aaBB \to_8 aaaBBB$ mais diffèrent ensuite

$$aaaBBB \to_6 aaabBB \to_7 aaabbSB \to_1 aaabbaBB$$
$$aaaBBB \to_7 aaabSBB \to_2 aaabbABB \to_3 aaabbaBB$$
$$aaaBBB \to_6 aaabBB \to_6 aaabbB \to_8 aaabbaBB$$

puis se terminent de la même manière :$aaabbaBB \to_6 aaabbabB \to_7 aaabbabbS \to_2 aaabbabbbA \to_3 aaabbabbba$

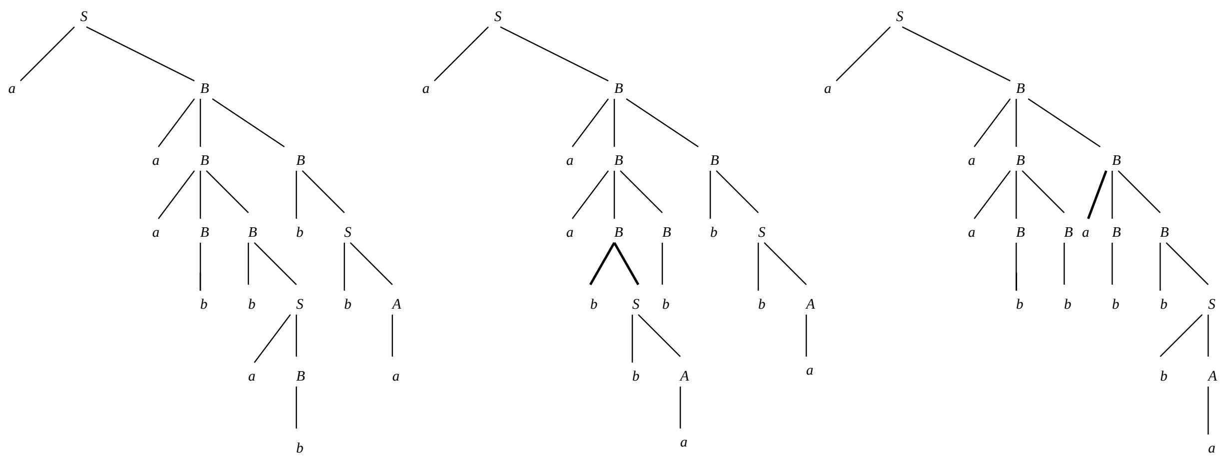
b) Comme pour la dérivation gauche, on a trois dérivations droites qui commencent de la même manière$S \to_1 aB \to_8 aaBB$ mais diffèrent ensuite

$aaBB \to_7 aaBbS \to_2 aaBbbA \to_3 aaBbba \to_8 aaaBBbba \to_7 aaaBbSbba \to_1 aaaBbaBbba \to_6 aaaBbabbba \to_6 aaabbabbba$

$aaBB \to_7 aaBbS \to_2 aaBbbA \to_3 aaBbba \to_8 aaaBBbba \to_6 aaaBbbba \to_7 aaabSbbba \to_2 aaabbAbbba \to_3 aaabbabbba$

$aaBB \to_8 aaBaBB \to_7 aaBaBbS \to_2 aaBaBbbA \to_3 aaBaBbba \to_6 aaBabbba \to_8 aaaBBabbba \to_6 aaaBbabbba \to_6 aaabbabbba$

c) Un arbre syntaxique qui engendre le mot $aaabbabbba$:



Un parcours en profondeur gauche de cet arbre donne une dérivation gauche associée, idem pour la droite. Ce mot a en fait trois arbres syntaxiques, ce qui implique que la grammaire est ambiguë.

Remarque : On observe que :
- la grammaire donnée engendre l'ensemble des mots ayant autant de a que de b,
- la variable A engendre l'ensemble des mots ayant exactement 1 a de plus que de b,
- la variable B engendre l'ensemble des mots ayant exactement 1 b de plus que de a.

2.

a) On peut partir du fait que $L_1 = \{a^n b^n | n > 0\} c^+$. Il suffit alors de trouver une variable A qui engendre $\{a^n b^n | n > 0\}$, et une variable C qui engendre c^+ pour engendrer L_1 par une variable S avec la règle : $S \to AC$.

Or on peut engendrer le langage (rationnel) c^+ via les 2 règles suivantes : $C \to c \mid cC$ et le langage $\{a^n b^n | n > 0\}$ via les 2 règles suivantes : $A \to ab \mid aAb$. On obtient :

$N = \{S, A, C\}; T = \{a, b, c\}; S$ et $P = \{S \to AC, A \to ab \mid aAb, C \to c \mid cC\}$ qui engendre le langage algébrique L_1.

b) L'idée est de dériver des mots dont le début est identique à la fin, puis à un moment donné d'engendrer une différence. A partir de ce moment, la seule contrainte à respecter est la parité de la longueur. On obtient donc :

$N = \{S, T, U\}; T = \{a, b\}; S$ et $P = \{S \to aSa \mid bSb \mid aTb \mid bTa, T \to \varepsilon \mid aU \mid bU, U \to aT \mid bT\}$

c) Le langage L_3 est l'union de 2 langages, $L_{31} = \{a^i b^j \mid i \neq j\} c^+$ et $L_{32} = a^+ \{b^i c^j \mid i \neq j\}$. Soient les variables S_{31} et S_{32} qui engendrent respectivement L_{31} et L_{32}, et la variable S_3 qui permet d'engendrer le langage union L_3 via la règle : $S_3 \to S_{31} \mid S_{32}$.

$N = \{S_3, S_{31}, S_{32}, T_{31}, T_{32}, A, B, C\}, T = \{a, b, c\}, S_3$

$$P \left\{ \begin{array}{ll} S_3 \to S_{31} \mid S_{32} & T_{32} \to bT_{32}c \mid bB \mid cC \\ S_{31} \to T_{31}C & A \to aA \mid \varepsilon \\ T_{31} \to aT_{31}b \mid aA \mid bB & B \to bB \mid \varepsilon \\ S_{32} \to AT_{32} & C \to cC \mid \varepsilon \end{array} \right.$$

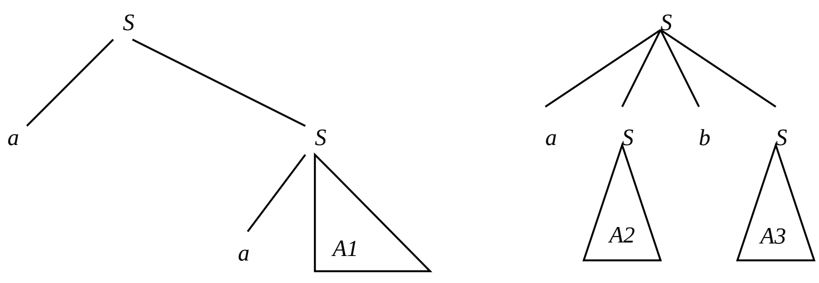
d) On peut utiliser le langage et la grammaire de l'exercice 1. Il suffit de l'adapter, pour engendrer à la place des mots contenant autant de a que de b les mots contenant soit plus de a que de b, soit plus de b que de a. Ainsi A (resp. B) engendre les mots ayants un excédent de a (resp. b) et E engendre les mots équilibrés. On obtient : $N = \{S, A, B, E, C, D\}, T = \{a, b\}, S,$ $P \left\{ \begin{array}{ll} S \to A \mid B & E \to aD \mid bC \mid \varepsilon \\ A \to aE \mid aA \mid EA & C \to a \mid aE \mid bCC \\ B \to bE \mid bB \mid EB & D \to b \mid bE \mid aDD \end{array} \right.$

3. Notons L_S le langage engendré par S et L le langage des mots w vérifiant la propriété $P(w)$:

$P(w)$: tout préfixe de w a au moins autant de a que de b.

Il s'agit de montrer que $L_S = L$.

- On montre d'abord que tout mot w de L_S vérifie $P(w)$, c.a.d. $L_S \subseteq L$, par induction sur la hauteur des arbres syntaxiques. Pour les mots de L_S obtenus avec un arbre syntaxique de hauteur 1 (i.e. ε) ou de hauteur 2 (i.e. a ou ab) la propriété $P(w)$ est vraie. Soit n un entier, supposons $P(w)$ vraie pour tous les mots de L_S obtenus avec un arbre syntaxique de hauteur inférieure ou égale à n. Soit w un mot de L_S obtenu avec un arbre syntaxique de hauteur $n + 1$. Les arbres pouvant donner w sont de l'un des 2 types suivants :



  où $A1$ est un arbre syntaxique de hauteur n, et $A2$, $A3$ sont 2 arbres syntaxiques dont au moins 1 est de hauteur n. Par hypothèse d'induction, les mots obtenus à partir de ces arbres vérifient $P(w)$. Il est alors immédiat de vérifier que w vérifie aussi $P(w)$. Ainsi tous les mots de L_S vérifient $P(w)$ et donc $L_S \subseteq L$.
- Réciproquement, on montre que $L \subseteq L_S$. Les mots de L sont exactement les mots w tels que pour tout préfixe w' de w on a : $|w'|_a - |w'|_b \geq 0$. Remarquons que tout mot de L commence par a et qu'il y a 2

types de mots dans L :

1. les mots w tels que pour tout préfixe w' de w on a : $|w'|_a - |w'|_b > 0$ - ces mots s'écrivent am, où m est un mot de L
2. les mots w tels que pour au moins 1 préfixe w' on a : $|w'|_a - |w'|_b = 0$. Soit v le plus court des préfixes w'. $w = vv'$, où :
   - $v \in L$ (tout préfixe d'un mot de L est aussi un mot de L) qui se termine par b (car $|v|_a - |v|_b = 0$), et donc $v = aub$ où $u \in L$ (car v étant le plus petit mot du type 2., au est du type 1.)
   - $v' \in L$, car $|v|_a - |v|_b = 0$ et que vv' (i.e. w) est un mot de L.

On raisonne alors par induction sur la longueur des mots de L. Les mots de L de longueur 0 ou 1 sont bien engendrés par la grammaire. Soit n un entier, supposons que tous les mots de L de longueur inférieure ou égale à n sont obtenus par la grammaire. Soit w un mot de longueur $n+1$:

1. soit il est du type 1., i.e. $w = am$ où m est un mot de L de longueur n et donc engendré par la grammaire, d'où en appliquant $S \to aS$, puis en dérivant S pour obtenir m, on obtient w à partir de S.
2. soit il est du type 2., i.e. $w = aubv$ où $u, v \in L$ sont de longueur inférieure à n et donc engendrés par la grammaire, en appliquant $S \to aSbS$, puis en dérivant le premier S pour obtenir u et le deuxième S pour obtenir v, on obtient w à partir de S.

Donc tous les mots de L sont engendrés par la grammaire et $L \subseteq L_S$.

D'où le résultat.

4. 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 :

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

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

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

Après suppression des renommages et des ε-productions :

$N = \{S_0, S_3, S_5, S_8, S_9\}$

$T = \{a, b\}$

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

S_0

Observons que les membres droits de S_0, S_3 et S_5 sont identiques et ainsi peuvent être identifiées. D'où la grammaire :

$N = \{S_0, S_8, S_9\}$

$T = \{a, b\}$

$$P = \begin{cases} S_0 \to aS_0 \mid bS_0 \mid aS_8 \\ S_8 \to bS_9 \\ S_9 \to b \end{cases}$$

S_0

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 :

$$\begin{array}{l} \sim_0 \{q_0, q_1, q_2, q_3\}\ \{q_4\} \\ \sim_1 \{q_0, q_1, q_2\}\ \{q_3\}\ \{q_4\} \\ \sim_2 \{q_0, q_2\}\ \{q_1\}\ \{q_3\}\ \{q_4\} \end{array}$$

Et on a donc l'automate minimal

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

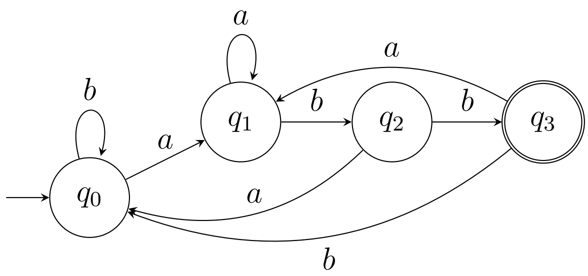


FIGURE 1 – L'automate minimal

On peut résoudre le système d'équations suivant :

$$\left\{\begin{array}{l} Z_0 = \varepsilon + Z_0 b + Z_3 b \\ Z_1 = Z_0 a + Z_1 a + Z_2 a + Z_3 a \\ Z_2 = Z_1 b \\ Z_3 = Z_2 b \end{array}\right.$$

d'où

$$\left\{\begin{array}{l} Z_1 = b^\star a + Z_1 b^\star a \\ Z_3 = Z_1 bb \end{array}\right.$$

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

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

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.

5.

a) Comme toute sous-chaîne de longueur au plus 5 doit contenir au moins un 0, on déduit que la distance entre deux 0 consécutifs 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 :

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

Une 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) :

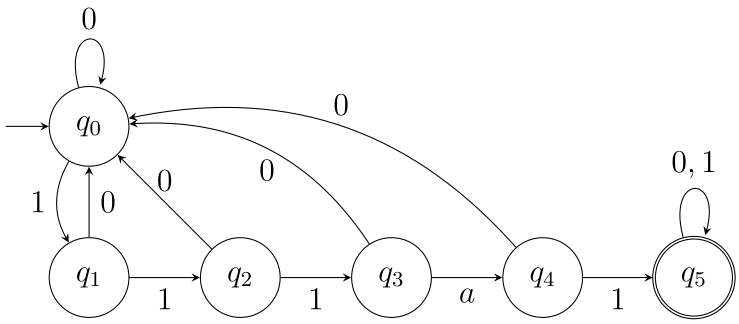


FIGURE 2 – L'automate minimal

En passant au complémentaire, on obtient l'automate fini :

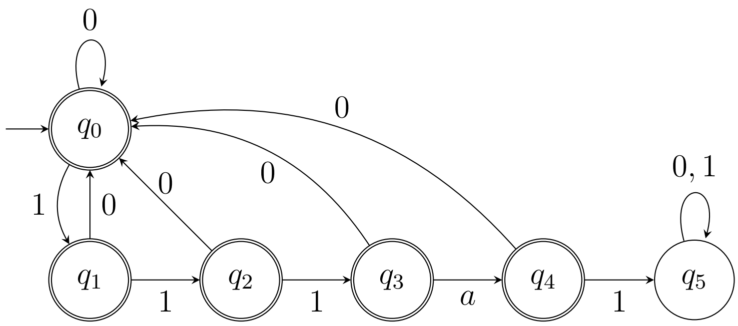


FIGURE 3 – L'automate minimal après complémentation

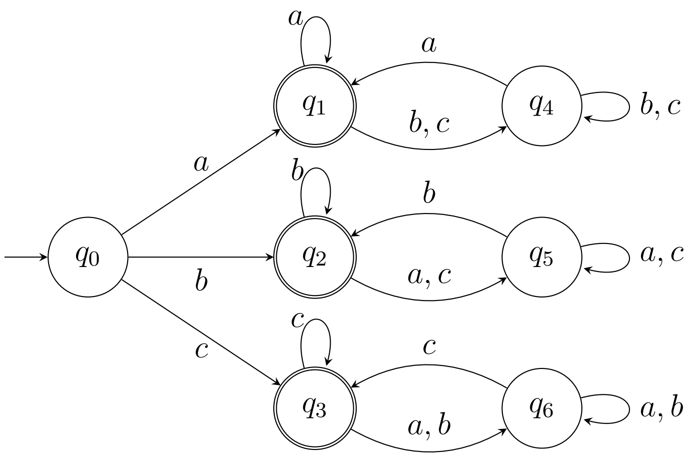
Et ceci permet de déduire la grammaire :

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

b) De la définition du langage, on déduit :

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

On pouvait aussi passer par l'automate fini (déjà rencontré en début de semestre ...),



d'où la grammaire :

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