Université Bordeaux 1
Licence d'Informatique 1, 2007/2008

INFOT2 : DS
Une solution possible

Exercice 1 [5 points]
1- L est engendré par la grammaire $G := \langle\{a,b\},\{S,T_1,T_2,U\},P,S\rangle$ où P est composé des règles suivantes :
$S \to aSd + T_1 + T_2$; $T_1 \to aT_1c + U$; $U \to bUc + \varepsilon$; $T_2 \to bT_2d + U$.
2- D' après la construction vue en cours, L est donc reconnu, par pile vide, par l'automate à pile

$$\mathcal{A} := \langle\{a,b,c,d\},\{a,b,c,d,S,T_1,T_2,U\},\{q\},S,q,\lambda\rangle$$

dont l'ensemble λ est composé des transitions

$$(\varepsilon,q,S,q,dSa),(\varepsilon,q,S,q,T_1),(\varepsilon,q,S,q,T_2),(\varepsilon,q,T_1,q,cT_1a),(\varepsilon,q,T_1,q,U),(\varepsilon,q,U,q,cUb),(\varepsilon,q,U,q,\varepsilon),$$

$$(\varepsilon,q,T_2,q,dT_2b),(\varepsilon,q,T_2,q,U),(a,q,a,q,\varepsilon),(b,q,b,q,\varepsilon),(c,q,c,q,\varepsilon),(d,q,d,q,\varepsilon).$$

Exercice 2 [7 points]

1. Première phase : éliminer ϵ.

$$R = \begin{cases} S & \to & T + aTb + ab + U + V \\ T & \to & aTb + ab \\ U & \to & SV + V + c \\ V & \to & Vc + c \end{cases}$$

Deuxième phase : éliminer les réécritures directes.

$$R = \begin{cases} S & \to & aTb + ab + SV + Vc + c \\ T & \to & aTb + ab \\ U & \to & SV + Vc + c \\ V & \to & Vc + c \end{cases}$$

Troisième phase : éliminer les variables inutiles. Toutes les variables engendrent un langage non-vide. La variables U n' est pas accessible à partir de S.

$$R = \begin{cases} S & \to & aTb + ab + SV + Vc + c \\ T & \to & aTb + ab \\ V & \to & Vc + c \end{cases}$$

Quatrième phase : ajouter un nouvel axiome, σ.

$$R = \begin{cases} \sigma & \to & aTb + ab + SV + Vc + c \\ S & \to & aTb + ab + SV + Vc + c \\ T & \to & aTb + ab \\ V & \to & Vc + c \end{cases}$$

Cinquième phase : mettre en forme quadratique les membres droits.

$$R = \begin{cases} \sigma & \rightarrow & WB \,+\, AB \,+\, SV \,+\, VC \,+ c \\ S & \rightarrow & WB \,+\, AB \,+\, SV \,+\, VC \,+ c \\ T & \rightarrow & WB \,+\, AB \\ V & \rightarrow & VC \,+ c \\ W & \rightarrow & AT \\ A & \rightarrow & a \\ B & \rightarrow & b \\ C & \rightarrow & c \end{cases}$$

La grammaire engendre le langage $\{a^n b^n c^m, n \geq 0, m \geq 0\}$.

2. – Le langage K_1 est algébrique : il est engendré par la grammaire $G_1 := \langle \{a, b\}, \{S, T\}, P_1, S\rangle$ où l'ensemble P_1 est composé des règles $S \rightarrow aSc + T$; $T \rightarrow bT + b$. Il satisfait donc le lemme d'itération.
   – Le langage K_2 est algébrique comme concaténation d'un régulier (a^+) et d'un algébrique ($\{b^n c^n, n \geq 0\}$). Il satisfait donc le lemme d' itération.
   – Le langage K_3 ne satisfait pas le lemme d'itération (considérer tous les mots du langage pour lesquels n est supérieur à la constante d'itération supposée), il n'est donc pas algébrique.

Exercice 3 [5 points]
1- Posons $C := \{a^n b^n \mid n > 0\}$. Le langage C est connu pour être algébrique.
Soit $h : \{a, b\}^* \rightarrow \{a, b\}^*$, l'homomorphisme de monoïdes défini par $h(a) = a, h(b) = bb$. On a

$$h(C) = \{a^n b^{2n} \mid n > 0\}.$$

Comme la famille des langages algébriques est close par homomorphisme, $h(C)$ est algébrique.
Comme $L = h(C) \cdot c^*$, L est le produit d'un langage algébrique et d'un langage rationnel, donc L est algébrique.
2- L est engendré par la grammaire $G := \langle \{a, b\}, \{S, U, V\}, P, S\rangle$ où P est composé des règles suivantes : $S \rightarrow UV$; $U \rightarrow aUbb + abb$; $V \rightarrow cV + \varepsilon$.

Exercice 4 [15 points]
1- Les trois affirmations sont vraies. Voici des dérivations gauches qui le prouvent :

$$S \rightarrow_G TS \rightarrow_G abS \rightarrow_G abTS \rightarrow_G ababS \rightarrow_G abab.$$

$$T \rightarrow_G aSb \rightarrow_G aTSb \rightarrow_G aabSb \rightarrow_G aabTSb \rightarrow_G aababSb \rightarrow_G aababb$$

$$S \rightarrow_G TS \rightarrow_G abS \rightarrow_G abTS \rightarrow_G ababS \rightarrow_G ababTS \rightarrow_G ababaSbS \rightarrow_G ababaTSbS \rightarrow_G ababaabSbS \rightarrow_G ababaabTSbS$$

$$\rightarrow_G ababaababSbS \rightarrow_G ababaababbS \rightarrow_G ababaababb.$$

2- Un arbre de dérivation, dans la grammaire G, de S en $ababaababb$ est dessiné sur la figure 1.
3.1 $\pi(\varepsilon) = 0$ donc $\varepsilon \in L_1$.
$\pi(ababab) = 0$ et les préfixes de $ababab$ sont $\varepsilon, a, ab, aba, abab, ababa, ababab$, qui ont comme images par π : $0, 1, 0, 1, 0, 1, 0$. Donc $ababab \in L_1$.
$\pi(aababb) = 0$ et les préfixes de $aababb$ sont $\varepsilon, a, aa, aab, aaba, aabab, aababb$, qui ont comme

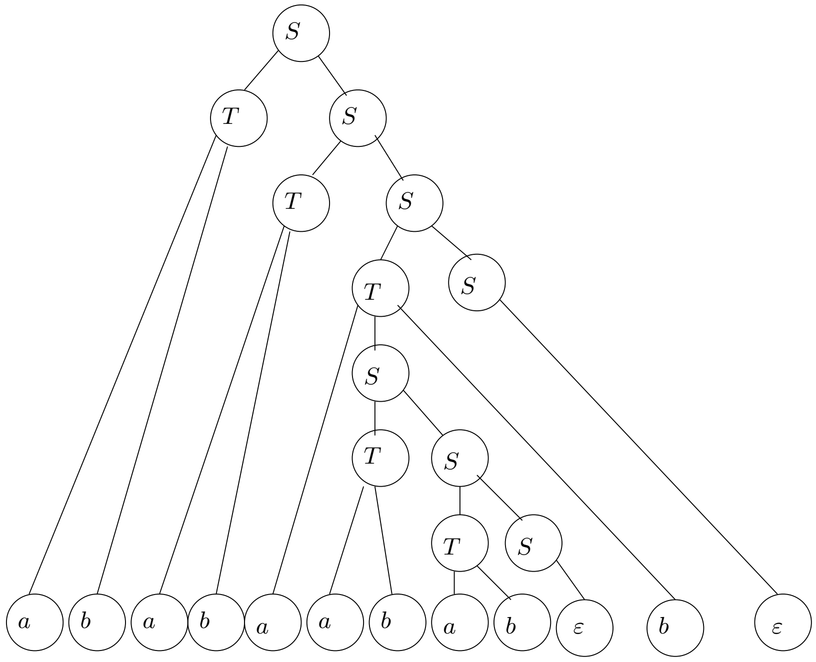

FIG. 1 – Un arbre de dérivation

images par π : $0, 1, 2, 1, 2, 1, 0$. Donc $aababb \in L_1$.
3.2 $ab \in A^+$ et $\pi(ab) = 0$. L'ensemble $\{v \in \{a,b\}^+, v \prec ab\}$ vaut $\{a\}$ et $\pi(a) = 1 > 0$. Donc $ab \in L_2$.
$aababb \in A^+$ et $\pi(aababb) = 0$. L'ensemble $\{v \in \{a,b\}^+, v \prec aababb\}$ vaut $\{a, aa, aab, aaba, aabab\}$ et les valeurs correspondantes de π sont $1, 2, 1, 2, 1$. Donc $aababb \in L_2$.
3.3 $\varepsilon \notin L_2$ car $\varepsilon \notin A^+$.
$ababab \notin L_2$ car le mot $v = ab \in A^+$ vérifie que $v \prec ababab$ mais $\pi(v) = 0$.
$aabbab \notin L_2$ car le mot $v = aabb \in A^+$ vérifie que $v \prec aabbab$ mais $\pi(v) = 0$.
4- Pour tout langage L, on a $L \subseteq L^*$, donc $L_1 \subseteq L_1^*$.
On vérifie que $\varepsilon \in L_1$ et $L_1 \cdot L_1 \subseteq L_1$. On en déduit, par récurrence sur l'entier $n \geq 0$ que $L_1^n \subseteq L_1$ et par suite que $\bigcup_{n\geq 0} L_1^n \subseteq L_1$ i.e. que

$$L_1^* \subseteq L_1.$$

Donc $L_1 = L_1^*$.
5-
5.1 Montrons que tout mot $u \in L_1$ a une unique décomposition de la forme :

$$u = u_1 \cdot u_2 \cdots u_n$$

avec $n \geq 0$ et chaque mot u_i appartient à L_2.
Existence
Nous prouvons l' existence de cette décomposition, par récurrence sur la longueur de u.
Base : $|u| = 0$.
La décomposition définie par $n = 0$ convient.
Induction : $|u| = \ell + 1$, $\ell \geq 0$.

Soit $u_1 := min_{\preceq}\{v \in L_1 \mid \varepsilon \prec v \preceq u\}$ et soit u' telque $u = u_1 \cdot u'$.
La minimalité de u_1 entraîne que $u_1 \in L_2$. D'autre part, comme $u \in L_1$ et $u_1 \in L_1$, on peut montrer que $u' \in L_1$. Comme $|u_1| \geq 1$, $|u'| \leq \ell$ et par hypothèse de récurrence, u' admet une décomposition de la forme

$$u' = u'_1 \cdots u'_m$$

avec $m \geq 0$ et chaque mot u'_i appartient à L_2.
Posons $n := m + 1$ et, pour tout $i \in [2, m+1], u_i := u'_{i-1}$. On a alors

$$u = u_1 \cdot u_2 \cdots u_n$$

avec $n \geq 0$ et chaque mot u_i appartient à L_2.
Unicité
Remarquons que L_2 est "libre de préfixe" i.e. si $v, u \in L_2$ et $v \preceq u$, alors $u = v$. Supposons qu'un même mot u admet deux décompositions de la forme demandée :

$$u = u_1 \cdot u_2 \cdots u_n = v_1 \cdot v_2 \cdots v_m$$

avec $n \geq 0, m \geq 0$ et tous les mots u_i, v_j appartiennent à L_2.
Montrons par récurrence sur $m + n$ que ces décompositions sont égales. Si $m + n = 0$, elles sont clairement égales.
Si $m+n \geq 1$, on ne peut avoir $n = 0$ ou $m = 0$, car les mots de L_2 sont non-vides. Donc $n \geq 1$ et $m \geq 1$. Les mots u_1, v_1 appartiennent à L_2 et sont comparables pour l'ordre préfixe, donc $u_1 = v_1$ (puisque L_2 est "libre de préfixe"). On est ramené à l'égalité des décompositions

$$u_2 \cdots u_n = v_2 \cdots v_m$$

Ces décompositions ont une somme de longueurs $< m+n$, donc, par hypothèse de récurrence, elles sont égales.
5.2 On vérifie que $L_2 = aL_1b$. L'existence et l' unicité des décompositions demandées pour les mots de L_2 résultent de l'existence et l' unicité des décompositions établies pour les mots de L_1.
6- D'après la question 5.1 $L_1 = L_2^*$. On en déduit que

$$L_1 = L_2L_1 \cup \{\varepsilon\}. \tag{1}$$

D'après les questions 5.1,5.2 $L_2 = aL_1b$. Comme $\varepsilon \in L_1$, $L_1 = L_1 \cup \{\varepsilon\}$, donc

$$L_2 = aL_1b \cup \{ab\}. \tag{2}$$

7- Montrons que $L_G(S) = L_1$ et $L_G(T) = L_2$.
7.1 Montrons, par induction sur l'entier n que : $\forall u \in \{a, b\}^*$

$$[S \xrightarrow{n} u \Rightarrow u \in L_1] \text{ et } [T \xrightarrow{n} u \Rightarrow u \in L_2].$$

Base :$n = 0$.
La propriété est trivialement vraie car, pour chacune des deux implications, aucun mot $u \in \{a, b\}^*$ ne vérifie le membre gauche de l'implication.
Induction :$n = m + 1$.
Supposons que $S \rightarrow TS \xrightarrow{m} u$. Par le lemme fondamental, u admet une décomposition $u = u_1u_2$

avec $T \overset{m_1}{\to} u_1, S \overset{m_2}{\to} u_2$ et $m = m_1 + m_2$.
Par hypothèse de récurrence, $u_1 \in L_2$, $u_2 \in L_1$ et en utilisant l'égalité (1), $u \in L_1$.
Supposons que $T \to ab \overset{m}{\to} u$. Alors $ab = u$ et d'après l'égalité (2), $u \in L_2$.
Supposons que $T \to aSb \overset{m}{\to} u$. Par le lemme fondamental, u admet une décomposition $u = au_1b$ avec $S \overset{m}{\to} u_1$.
Par hypothèse de récurrence, $u_1 \in L_1$ et en utilisant l' égalité (2), $u \in L_2$.
7.2 Montrons, par induction sur l'entier n que : $\forall u \in \{a,b\}^*$,

$$[(u \in L_1 \text{ et } |u| \leq n) \Rightarrow S \overset{*}{\to} u] \text{ et } [(u \in L_2 \text{ et } |u| \leq n+1) \Rightarrow T \overset{*}{\to} u].$$

Base :$n = 0$.
La propriété est vraie car, $S \to \varepsilon$ est une règle de G et aucun mot de longueur 0 ou 1 n'appartient à L_2.
Induction :$n = m + 1$.
Supposons que $u \in L_1$. D'après l'égalité (1),$u \in L_2L_1 \cup \{\varepsilon\}$. Si $u = \varepsilon$, alors $S \to \varepsilon = u$. Si $u \in L_2L_1$, u a une decomposition de la forme $u = u_1 \cdot u_2$ avec $u_1 \in L_2, u_2 \in L_1$. Comme $u_1 \leq m+1$ et $u_2 \leq m$, par hypothèse de récurrence, $T \overset{*}{\to} u_1$ et $S \overset{*}{\to} u_2$, donc

$$S \to TS \overset{*}{\to} u_1u_2 = u.$$

Supposons que $u \in L_2$. D'après l'égalité (2),$u \in aL_1b \cup \{ab\}$. Si $u = ab$ alors $T \to ab$. Sinon, u a une decomposition de la forme $u = au_1b$ avec $u_1 \in L_1$. Comme $u_1 \leq m - 1$, par hypothèse de récurrence, $S \overset{*}{\to} u_1$, donc

$$S \to TS \to aSbS \to aSb \overset{*}{\to} au_1b = u.$$

8- La grammaire G est ambiguë car les deux dérivations gauches suivantes engendrent le mot ab à partir de l'axiome S :

$$S \to_G TS \to_G abS \to_G ab$$

$$S \to_G TS \to_G aSbS \to_G abS \to_G ab$$

On remarquera (mais ce n'était pas demandé au candidat) que la grammaire G' obtenue en supprimant la règle $T \to ab$, a les propriétés suivantes : $L_{G'}(S) = L_1, L_{G'}(T) = L_2$. De plus G' est non-ambiguë. L'unicité des décompositions énoncées à la question 5 permet de montrer que :
si $u \in L_1$ alors, soit $u = \varepsilon$ (cas 1) soit u se décompose en $u = u_1u_2$ (cas 2) avec $u_1 \in L_2$, $u_2 \in L_1$. Les cas1 et cas 2 sont disjoints et, dans le cas 2, les mots u_1, u_2 sont uniques.
Si $u \in L_2$ alors u se décompose en $u = au_1b$ avec $u_1 \in L_1$ et le mot u_1 est unique.
On en déduit, par récurrence sur la longueur des mots, que pour tout mot u de L_1 (resp. de L_2), il existe une unique dérivation gauche de S (resp. de T) en u, dans la grammaire G'.