Langages algébriques

Exercice 1. Lemme d'Ogden.

Notations. Soit T un arbre, on note $y_1, y_2, \ldots, y_m$, $m \geq 1$, les feuilles dans l'ordre habituel de gauche à droite. Si y est un fils de x, on note $x \to y$. La fermeture transitive de la relation $\to$ est notée $\to^+$ et la fermeture réflexo-transitive est notée $\to^*$.
On note $x \hookrightarrow y$ si et seulement si :

- x et y ne sont pas sur la même chaîne de l'arbre,
- $\exists i,\ 1 \leq i < m,\ x \to^* y_i$ et $y \to^* y_{i+1}$.

Les fermetures transitive et réflexo-transitive de la relation $\hookrightarrow$ sont dénotées respectivement par $\hookrightarrow^+$ et $\hookrightarrow^*$.

1 - On considère un arbre T d'arité au plus r (chaque nœud de l'arbre a au plus r fils) dont certaines feuilles sont marquées. On note M l'ensemble des feuilles marquées.
On pose

$$D = \{x \in T \mid \exists\, y \in M,\ x \to^* y\}$$

$$B = \{x \in T \mid \exists\, y_1, y_2 \in D,\ y_1 \neq y_2 \text{ avec } x \to y_1 \text{ et } x \to y_2\}$$

a - Montrer que si toute chaîne de T a au plus i nœuds de type B alors $|M| \leq r^i$.

b - Montrer que si $|M| \geq r^{2v+3}$, il existe une chaîne de T ayant au moins $2v+3$ nœuds de type B.

2 - On suppose maintenant que $|M| \geq r^{2v+3}$. On considère une chaîne s de la racine s_0 de l'arbre à une feuille s_t contenant au moins $2v+3$ sommets de type B. On note C_s l'ensemble des $2v+3$ sommets de s de type B les plus éloignés de s_0.

a - On définit alors

$$C_L = \{x \in C_s \mid x \to\ y \text{ et } y \hookrightarrow^+\ s_t \text{ pour un } y \in D\}$$

$$C_R = \{x \in C_s \mid x \to\ y \text{ et } s_t \hookrightarrow^+\ y \text{ pour un } y \in D\}.$$

Montrer que $C_s = C_L \cup C_R$.

b - En déduire que $|C_L| \geq v+2$ ou $|C_R| \geq v+2$.

3 - Si $w = a_1 a_2 \ldots a_n$, une position est un indice i, $1 \leq i \leq n$. Si K est un ensemble de positions de w, toute factorisation de w en $v_1 v_2 \ldots v_l$ induit une partition de K en $\{K_1, K_2, \ldots, K_l\}$. Montrer le théorème d'itération qui s'énonce :

Soit L un langage algébrique et $G = (V, T, P, S)$ une grammaire algébrique engendrant L. Montrer qu'il existe une constante $p(G)$ telle que pour tout ensemble K de positions avec $|K| \geq p(G)$ alors il existe une factorisation de w, $w = v_1 v_2 v_3 v_4 v_5$ vérifiant :

1 - Il existe $A \in V$ tel que

- $S \Longrightarrow^* v_1 A v_5$,
- $A \Longrightarrow^* v_2 A v_4$,
- $A \Longrightarrow^* v_3$.

2 - Pour tout $q \geq 0,\ v_1 v_2^q v_3 v_4^q v_5 \in L$.
3 - Si $K = \{K_1, K_2, K_3, K_4, K_5\}$ alors

- $K_1, K_2, K_3 \neq \emptyset$ ou $K_3, K_4, K_5 \neq \emptyset$,
- $|K_2 \cup K_3 \cup K_4| \leq p(G)$.

4 - En déduire que le langage $L = \{a^n b^n c^n,\ n \geq 1\}$ n'est pas algébrique.

Exercice 2
En utilisant le résultat de l'exercice précédent montrer que les langages suivants ne sont pas algébriques.
$L_1 = \{a^i b^j c^k,\ i < j < k\}$.
$L_2 = \{a^i b^j,\ j = i^2\}$.
$L_3 = \{a^i,\ i \text{ premier}\}$.
$L_4 = \{a^n b^n c^m,\ n \leq m \leq 2n\}$.
$L_5 = \{a^{n^2},\ n \geq 0\}$.

Exercice 3
Montrer que le langage $L = a^*bc \cup \{a^p b a^n c a^n \mid p \text{ premier et } n \geq 0\}$ n'est pas algébrique mais qu'on ne peut pas le prouver en utilisant le *lemme* d'itération.

Exercice 4 On veut prouver le résultat suivant :

Tout langage algébrique sur $\{a\}$ est rationnel.

a - En appelant N_0 la constante donnée par le lemme d'itération et en choisissant pour p le plus petit commun multiple de $\{1, 2, \ldots, N_0\}$, montrer que :

$$w \in L \cap \Sigma^{N_0}\Sigma^+ \implies w(a^p)^* \subseteq L.$$

b - On pose $A_i = a^{N_0+i}(a^p)^* \cap L,\ 1 \leq i \leq p$, et y_i est le mot le plus court de A_i (si $A_i = \emptyset$, y_i n'est pas défini). Montrer que

$$L = (L \cap (\bigcup_{i=0}^{N_0} a^i)) \cup (\bigcup_{i=1}^{p} y_i(a^p)^*).$$

c - Conclure.