Module Langages Formels
TD 5 : Lemmes d'itération et Automates à Pile

Exercice 1 `Lemme de Bar-Hillel, Perles et Shamir`

Lemme (Bar-Hillel, Perles et Shamir) :
Soit L un langage algébrique. Il existe un entier N tel que tout mot $w \in L$ de longueur supérieure ou égale à N se factorise en $w = \alpha u \beta v \gamma$, où $\alpha, u, \beta, v, \gamma \in \Sigma^*$, avec :
- $|uv| > 0$,
- $|u\beta v| \leqslant N$,
- pour tout $n \geqslant 0$, $\alpha u^n \beta v^n \gamma \in L$.

Montrer que les langages suivants ne sont pas algébriques.

1. $L_1 = \{a^i b^j c^k,\ i < j < k\}$;
2. $L_2 = \{a^{n^2},\ n \geqslant 0\}$;

Exercice 2 `Lemme d'Ogden`

Soit G une grammaire algébrique sous forme normale de Chomsky (FNC) engendrant un langage L.

Définition :
Soit T un sous-arbre d'un arbre de dérivation selon une grammaire FNC. On suppose marquées certaines feuilles de T. On appelle *embranchement* un nœud de T ayant deux fils, tel que chacun de ses fils contienne au moins une feuille marquée.

2.1. Soit T un sous-arbre d'un arbre de dérivation selon une grammaire FNC. On suppose qu'au moins 2^h feuilles distinctes de T ont été marquées.

Justifier qu'il existe un chemin, de la racine à une feuille, passant par au moins h embranchements.

2.2. On considère une grammaire FNC G engendrant le langage L. Montrer qu'il existe un entier N tel que :

— pour tout mot $w \in L$ dans lequel on marque au moins N positions,

— pour tout arbre de dérivation de w,

il existe deux embranchements b_1 et b_2 tels que

- b_1 est un ancêtre de b_2,
- b_1 est un ancêtre d'au plus N feuilles marquées,
- b_1 et b_2 sont étiquetés par la même variable.

2.3. En déduire le lemme d'Ogden.

Lemme (Ogden) :
Soit L un langage algébrique. Il existe un entier N tel que tout mot $w \in L$ ayant au moins N positions distinguées *(ou lettres marquées selon les ouvrages)*, se factorise en $w = \alpha u \beta v \gamma$, où $\alpha, u, \beta, v, \gamma \in \Sigma^*$, avec :

— u ou v contient au moins une position distinguée,

— $u\beta v$ contient moins de N positions distinguées,

— pour tout $n \geqslant 0$, $\alpha u^n \beta v^n \gamma \in L$.

On trouve aussi la formulation suivante, par exemple dans le livre de Carton, page 92 :

Lemme (Ogden) :
Pour toute grammaire $G = (\Sigma, V, P)$ et pour toute variable $S \in V$, il existe un entier N tel que tout mot $w \in \widehat{L_G}(S)$ ayant au moins N positions distinguées se factorise en $w = \alpha u \beta v \gamma$, où $\alpha, u, \beta, v, \gamma \in (\Sigma \cup V)^*$, avec :

— $S \xrightarrow{*} \alpha T \gamma$ et $T \xrightarrow{*} uTv \mid \beta$,

— soit α, u, β, soit β, v, γ contiennent des positions distinguées,

— $u\beta v$ contient moins de N positions distinguées.

2.4. On s'intéresse au langage $L = \{a^i b^j c^k d^l \quad | \quad i = 0 \quad \text{ou} \quad j = k = l\}$.

2.4.1. Montrer que pour tout $N \in \mathbb{N}^*$ et tout mot $w \in L$ avec $|w| \geq N$, il existe une décomposition $w = \alpha u \beta v \gamma$ telle que

— $|uv| > 0$

— $|u\beta v| \leqslant N$

— pour tout $n \geqslant 0$, $\alpha u^n \beta v^n \gamma \in L$.

2.4.2. Montrer que L n'est pas algébrique.

Exercice 3 `Un langage intrinsèquement ambigu`
On considère le langage $L = \{a^l b^m c^n | l = m \vee m = n\}$.

3.1. Montrer que ce langage est algébrique.

3.2. Soit G une grammaire reconnaissant L. Montrer qu'il existe $u \in L$ tel qu'il existe deux arbres de dérivation de G distincts menant à u. Conclure.
► *On pourra considérer les mots $a^N b^N c^{N+N!}$ et $a^{N+N!} b^N c^N$ pour un N bien choisi.*

Exercice 4 `Construction d'automates.`

Donner des automates à pile reconnaissant les langages suivants :

1. $L_1 = \{u \in \{a,b\}^* | \; |u|_a = |u|_b\}$
2. $L_2 = \{u \in \{a,b\}^* | \; |u|_a \geqslant |u|_b\}$
3. $L_3 = \{u \in \{a,b\}^* | \; |u|_a = 2 \cdot |u|_b\}$
4. $L_4 = \{u\#\tilde{v} | \; u, v \in \{0,1\}^* \; \exists i \in \mathbb{N}, \; u = \bar{i}_2)$ et $v = \overline{(i+1)}_2\}$ où $\overline{n}_2$ est l'écriture de n en binaire.