

# Module Langages Formels

## TD 6 : Lemme de l'étoile et Lemme d'Ogden

**Exercice 1** Point d'Étoile noire, le lemme de l'étoile nous suffit !

Montrer que les langages suivants ne sont pas algébriques.

1.  $L_1 = \{a^i b^j c^k, i < j < k\}$ ;
2.  $L_2 = \{a^n b^n c^m, n \leq m \leq 2n\}$ ;
3.  $L_3 = \{a^{2^n}, n \geq 0\}$ ;
4.  $L_4 = \{a^{n^2}, n \geq 0\}$ ;

**Exercice 2** Lemme d'Ogden

L'objectif de cet exercice est de montrer une version plus forte du lemme de l'étoile pour les langages algébriques :

**Lemme (Ogden) :**

Soit  $L$  un langage algébrique. Il existe un entier  $N$  tel que pour tout mot  $z \in L$  dans lequel on marque au moins  $N$  positions distinctes, il est possible de décomposer  $z$  sous la forme  $z = uxv y w$  avec

- $x$  ou  $y$  contient au moins une position marquée,
- $xv y$  contient au plus  $N$  positions marquées,
- pour tout  $i \geq 0$ ,  $ux^i v y^i w \in L$ .

**2.1.** On se donne une grammaire algébrique propre  $G$  engendrant un langage  $L$ . Montrer qu'il existe une grammaire sous forme normale de Chomsky (CNF) reconnaissant le langage  $L - \{\varepsilon\}$ .

► Rappel : Une grammaire CNF est une grammaire dont toutes les productions sont de la forme

$$A \rightarrow BC \quad \text{ou} \quad A \rightarrow a$$

**Définition :**

Soit  $T$  un sous-arbre d'un arbre de dérivation selon une grammaire CNF. 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.

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

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

**2.3.** On considère une grammaire CNF  $G$  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.

**2.4.** En déduire le lemme d'Ogden.

**2.5.** On s'intéresse au langage  $L_5 = \{a^i b^j c^k d^l \mid i = 0 \text{ ou } j = k = l\}$ .

**2.5. 1.** Montrer que pour tout  $N \in \mathbb{N}^+$  et tout mot  $z \in L_5$  avec  $|z| \geq N$ , il existe une décomposition  $z = uxv y w$  telle que

- $|xy| \geq 1$
- $|xvy| \leq N$
- pour tout  $i \geq 0$ ,  $ux^i v y^i w \in L_5$ .

**2.5. 2.** Montrer que  $L_5$  n'est pas algébrique.

**Exercice 3** Couper la poire en deux n'est pas toujours rationnel.

Pour tout langage  $L$  sur  $\Sigma$ , on définit

$$\frac{1}{2}L = \{x \in \Sigma^* \mid \exists y \in \Sigma^*, |x| = |y| \text{ et } xy \in L\}$$

**3.1.** Montrer que le langage  $L_6 = \{a^n b^n c^m d^{3m} \mid n, m \geq 1\}$  est algébrique.

**3.2.** Calculer  $\frac{1}{2}L_6$ .

**3.3.** Montrer que  $\frac{1}{2}L_6$  n'est pas algébrique.