Lemme de pompage ultime: Ehrenfeucht, Parikh et Rozenberg

On propose deux versions du lemme de l'étoile pour les langages réguliers. Soit L un langage régulier. Alors L vérifie les deux lemmes suivants:

Lemme 1

Il existe $N \geq 0$ tel que, pour tout mot $x \in L$ avec $|x| \geq N$, il existe une décomposition $x = u_1 u_2 u_3$ avec $u_2 \neq \varepsilon$ telle que $\forall n \in \mathbb{N}, u_1 u_2^n u_3 \in L$

Lemme 2

Il existe $N \geq 0$ tel que, pour tout mot $x \in L$, pour toute décomposition $x = uv_1v_2...v_Nw$ avec $\forall 1 \leq i \leq N, |v_i| \geq 1$, il existe $0 \leq j < k \leq N$ tel que pour tout $n \geq 0$:

$$uv_1...v_j\left(v_{j+1}...v_k\right)^n...v_Nw \in L$$

Q0) Soit $L = \{u \in \{a, b\}^* \mid u$ a le même nombre de a que de b$\}$. Montrez que L vérifie le lemme 1 mais pas le lemme 2.
On ne s'attend pas à des preuves lourdement formelles pour cette question

On dit qu'un langage L vérifie la propriété σ_k (respectivement ${\sigma_k}'$) si pour tout mot $f \in \Sigma^*$ et toute factorisation $f = uv_1...v_kw$ dans laquelle chaque mot v_i est non vide, il existe deux indices i, j avec $0 \leq i < j \leq k$ tels que :

$$\forall n \geq 0 \quad \left[f \in L \Leftrightarrow uv_1...v_i\left(v_{i+1}...v_j\right)^n v_{j+1}...v_kw \in L\right] \tag{σ_k}$$

$$f \in L \Leftrightarrow uv_1...v_iv_{j+1}...v_kw \in L \tag{${\sigma_k}'$}$$

Nous allons montrer le résultat suivant, attribuée à Ehrenfeucht, Parikh et Rozenberg.

Proposition 3

Il existe une équivalence entre:
- (1) L est régulier
- (2) Il existe $k > 0$ tel que L vérifie σ_k
- (3) Il existe $k > 0$ tel que L vérifie ${\sigma_k}'$

Q1) Sans reprouver la dernière version du lemme de l'étoile, justifier rapidement que (1) implique (2) (penser au langage complémentaire) et que (2) implique (3).

Q2) On note, pour tout mot $v \in \Sigma^*$, $v^{-1}L = \{u \mid u.v \in L\}$. Montrer que si L vérifie ${\sigma_k}'$, alors, pour tout mot $v \in \Sigma^*$, $v^{-1}L$ vérifie aussi ${\sigma_k}'$.

On cherche à montrer que pour tout entier non nul k, il existe un nombre fini de langages vérifiant ${\sigma_k}'$. Soient L_1 et L_2 deux tels langages. Pour ce faire, nous invoquons le théorème de Ramsey qui sera admis ici. On note $P_p(E)$ les parties à p élèments de E.

Théorème 4 (Théorème de Ramsey)

Pour tout triplet d'entiers (p, m, r), il existe un entier $N(p, m, r)$, tel que pour tout :

- ensemble E tel que $|E| \geq N(p, m, r)$,
- ensemble C tel que $|C| = m$,
- fonction $f : P_p(E) \to C$,

il existe $F \subset E$ tel que :

- $|F| \geq r$,
- $|f(P_p(F))| \leq 1$.

Q3) À quoi vous fait penser ce théorème pour les valeurs $p = 1$ et $r = 2$ (avec m quelconque) ? Donner une telle valeur $N(1, m, 2)$ minimale.

Q4) Montrer qu'il existe un entier que l'on notera N tel que pour tout ensemble P de paires de $[\![1, N]\!]$, il existe un sous-ensemble F_P de $[\![1, N]\!]$ de cardinal au moins $k + 1$ dont ses paires sont soit toutes dans P , soit toutes hors de P.

Q5) En utilisant la question précédente, montrer que pour tout mot f de taille au moins N, il existe une factorisation $f = uv_1...v_kw$, où tous les mots v_i sont non vides, telle que pour tous $0 \leq i < j \leq k$:

$$f \in L_1 \Leftrightarrow uv_1...v_iv_{j+1}...v_kw \in L_1$$

Q6) Montrer par récurrence sur la taille des mots que, si les mots de taille au plus N de L_1 sont exactement les mots de taille au plus N de L_2, alors $L_1 = L_2$.

Q7) En déduire que, pour un k donné, il existe un nombre fini de langages vérifiant ${\sigma_k}'$.

Q8) Conclure que L est régulier et achever la preuve du lemme de pompage ultime, à l'aide du théorème suivant:

Théorème 5 (Théorème de Myhill-Nerode)

On note, pour un langage L, l'ensemble de ses résiduels $E = \{u^{-1}L \mid u \in \Sigma^*\}$. Alors:

$$L \text{ est régulier} \Leftrightarrow |E| < \infty$$

Q9) Démontrer le sens direct du théorème de Myhill-Nerode.

Q10) Démontrer le sens indirect.