TD 9 : Grammaires algébriques

Exercice 1 (**À préparer** - Forme normale de Greibach)**.** Mettre les grammaires suivantes en forme normale de Greibach quadratique :

1.
$$G_1 : \begin{cases} x_1 \to x_1 b + a \\ x_2 \to x_1 b + a x_2 \end{cases}$$

2.
$$G_2 : \begin{cases} x_1 \to x_1(x_1 + x_2) + x_2(a + b) \\ x_2 \to x_1 x_2 + x_2 x_1 + a \end{cases}$$

Soit G la grammaire suivante :

$$\begin{aligned} E &\to E + T \mid T \\ T &\to T \times F \mid F \\ F &\to (E) \mid a \end{aligned}$$

Mettre G en forme normale de Chomsky, puis en forme normale de Greibach quadratique. Quelles sont les conséquences de ces deux transformations sur les arbres de dérivations de la grammaire ?

Exercice 2 (**À préparer** - Grammaires algébriques et linéaires)**.** Pour chacun des langages suivants, déterminer s'il est algébrique ou non. Si oui, donner une grammaire algébrique qui l'engendre, et linéaire quand c'est possible.

1. $L_1 = \{w\bar{w} \mid w \in \{a,b\}^*\}$ où $\bar{w}$ représente l'image miroir de w.
2. $L_2 = \{ww \mid w \in \{a,b\}^*\}$
3. $L_3 = \{a^n b^{n^2} \mid n \geq 0\}$
4. $L_4 = \{a^n b^{2n} \mid n \geq 0\}$
5. $L_5 = ab^*$
6. $L_6 = (ab^*)^*$
7. $L_7 = \{a^n b^m \mid n \leq m\}$
8. $L_8 = L_7^*$

Exercice 3 (Ambiguïté)**.**

1. Soit $L = \{a^m b^m c^n \mid m,n \geq 0\} \cup \{a^m b^n c^n \mid m,n \geq 0\}$. Montrer que L est algébrique. Montrer que L est inhéremment ambigu. (Indication : on pourra se servir du Lemme d'Ogden appliqué à $a^k b^k c^{k+k!}$ pour un k bien choisi.)
2. Soit G la grammaire suivante :

$$\begin{aligned} S &\to \text{if } c \text{ then } S \text{ else } S \\ S &\to \text{if } c \text{ then } S \\ S &\to a \end{aligned}$$

Montrer que G est ambiguë, mais que le langage engendré par G ne l'est pas.

Exercice 4 (Ambiguïté et indécidabilité)**.** Soit $f, g : A^* \to B^*$ deux morphismes. Soit \$ un symbole qui n'apparaît pas dans Σ. On note $\overline{w}$ pour l'image miroir de $w \in \Sigma^*$ et on pose $L_f = \{w\$\overline{f(w)} \mid w \in \Sigma^+\}$ et $L_g = \{w\$\overline{g(w)} \mid w \in \Sigma^+\}$.

1. Montrer qu'on ne peut pas décider si $L_f \cap L_g = \emptyset$.
2. Montrer que L_f et L_g ne sont pas ambigus. En déduire qu'on ne peut pas décider si une grammaire algébrique est ambiguë.
   Soit $L_1 = \{a^n b^n c^m \mid m, n \geq 0\}$ et $L_2 = \{a^n b^m c^m \mid m, n \geq 0\}$.
3. On rappelle que $L_1 \cup L_2$ est algébrique et inhéremment ambiguë. En déduire qu'on ne peut pas décider si un langage algébrique est inhéremment ambigu.

Exercice 5 (Théorème de Chomsky et Schützenberger)**.** Le théorème dit qu'un langage L est algébrique si et seulement s'il existe un entier n, un langage rationnel R et un morphisme π tels que $L = \pi(D_n^* \cap R)$, où D_n^* dénote l'ensemble des mots bien parenthésés sur un alphabet à n paires de parenthèses.

L'intuition derrière ce théorème est que l'on peut séparer les aspects de structure (le parenthésage, c'est-à-dire la structure d'arbre) et de contrôle (le langage rationnel) d'un langage algébrique – ce dont on verra une autre interprétation avec les automates à pile.

1. Soit G une grammaire algébrique sur Σ. Proposer une grammaire algébrique G', qui explicite la structure des dérivations de G au moyen d'un alphabet Σ_n de n sortes de parenthèses, telle que
$$L(G') \subseteq D_n^* \text{ et } L(G) = \pi(L(G'))$$
   avec π une projection de $\Sigma_n^* \to \Sigma^*$.
2. Il faut maintenant trouver un langage rationnel R tel que
$$L(G') = D_n^* \cap R\ .$$
   Proposer un tel langage.
3. Montrer qu'il existe un morphisme μ de $\Sigma_n \to \Sigma_2$ tel que
$$D_n^* = \mu^{-1}(D_2^*)\ .$$
   En déduire une autre formulation du théorème.