TD 4 – Grammaires algébriques

1 Grammaires algébriques (non contextuelles)

On rappelle qu'étant donné un mot $w \in \Sigma^*$, on définit le *mot renversé de* w, noté $w^{\mathcal{R}}$, comme étant $w^{\mathcal{R}} = \varepsilon$ si $w = \varepsilon$ et $w^{\mathcal{R}} = a_n \cdots a_1$ si $w = a_1 \cdots a_n$ pour $n \in \mathbb{N}_{>0}$.

Exercice 1. Donner une grammaire algébrique pour chacun des langages suivants (sans trop justifier).

1. $\{w \in \{a,b\}^* \mid w = w^{\mathcal{R}}\}$ (langage des palindromes sur $\{a,b\}$).
2. $\{w \in \{a,b\}^* \mid w \neq w^{\mathcal{R}}\}$ (langage des non palindromes sur $\{a,b\}$).
3. $\{w \in \{(,),0,1,\ldots,9,*,+\}^* \mid w$ correspond à une expression arithmétique sur $\mathbb{N}$ valide$\}$.
4. $\{a^i b^j c^k \mid i \neq j \vee j \neq k\}$.
5. $\{w \in \{a,b\}^* \mid |w|_a \geq |w|_b\}$.
6. $\{a^{n_0} b a^{n_1} b \cdots a^{n_k} b \mid k \in \mathbb{N} \wedge \exists j \in \mathbb{N}, n_j \neq j\}$.
7. $\{ww' \in \{a,b\}^* \mid |w| = |w'| \wedge w \neq w'\}$.

Exercice 2. Trouver une grammaire algébrique non ambiguë pour le langage

$$\{w \in \{(,),0,1,\ldots,9,*,+\}^* \mid w \text{ correspond à une expression arithmétique sur } \mathbb{N} \text{ valide}\}$$

telle que dans chaque arbre de dérivation (« parse tree »), $*$ a priorité sur $+$.

Exercice 3. Soit $\mathcal{G} = (\{S\}, \{(,)\}, R, S)$ et $\mathcal{G}' = (\{B,R\}, \{(,)\}, R', B)$ des grammaires algébriques où R contient les règles

$$S \to SS \mid (S) \mid \varepsilon$$

et R' contient les règles

$$B \to (RB \mid \varepsilon$$
$$R \to) \mid (RR \, .$$

Montrer que $\mathcal{G}$ et $\mathcal{G}'$ engendrent le même langage, mais que l'une est ambiguë et pas l'autre.

2 Langages algébriques (non contextuels) et langages rationnels

Étant donné une grammaire algébrique $\mathcal{G} = (V, \Sigma, R, S)$, on dit qu'elle est *linéaire droite* lorsque dans toute règle $X \to u$ de R, soit $u \in \Sigma^*$, soit $u = u'Y$ avec $u' \in \Sigma^*$ (le corps de chaque règle contient au plus une variable tout à droite).

Exercice 4. Montrer qu'un langage est rationnel si et seulement s'il est engendré par une grammaire linéaire droite.

Exercice 5. Soit la grammaire algébrique $\mathcal{G} = (\{S\}, \{a,b,c\}, R, S)$ où R contient les règles $S \to aSSb \mid c$. Montrer que tout langage rationnel inclus dans $\mathcal{L}(\mathcal{G})$ est fini.

3 Forme normale de Chomsky

On rappelle qu'une grammaire algébrique $\mathcal{G} = (V, \Sigma, R, S)$ est en forme normale de Chomsky (FNC) quand toutes les règles de R sont de la forme :

$$A \to BC \quad \text{ou} \quad A \to a \quad \text{ou} \quad S \to \varepsilon$$

avec $A \in V$, $B, C \in V \setminus \{S\}$ et $a \in \Sigma$.

Exercice 6. Proposer une grammaire algébrique en FNC équivalente à la grammaire algébrique $\mathcal{G} = (\{B, C, S\}, \{a, b\}, R, S)$ où R contient les règles :

$$
\begin{aligned}
S &\to CSC \mid aB \\
C &\to B \mid S \\
B &\to b \mid \varepsilon \ .
\end{aligned}
$$

Exercice 7. Proposer un algorithme en temps polynomial (en la taille cumulée du mot et de la grammaire) qui reconnaît si un mot appartient au langage engendré par une grammaire algébrique en FNC.