TD 8 : Formes normales et complexité (2)

Exercice 1 (Intersection avec un langage reconnaissable et algorithme de Cocke, Kasami et Younger)**.**

1. Soit G une grammaire algébrique et A un automate fini. Contruire G' algébrique telle que $L(G') = L(G) \cap L(A)$.
2. Modifier la construction pour calculer *au vol* si $L(G') = \emptyset$.
3. Quelle est la complexité de la construction d'une grammaire algébrique pour le langage $L(G) \cap \{w\}$?

   Un problème avec cet algorithme est qu'il génère en général beaucoup de règles inutiles. On voudrait filtrer la génération des règles de la grammaire pour n'ajouter une règle avec un non terminal de la forme $[q, A, q']$ (où q et q' sont des états de l'automate reconnaissant w, et A un non terminal de G) que s'il existe réellement un facteur u de w dans $L_G(A) \cap L_{q,q'}$.

   Proposer un algorithme avec cet invariant. Quelle borne de complexité obtenez-vous ?

Exercice 2 (Formes normales de Greibach.)**.** Mettre les deux grammaires suivantes sous forme normale de Greibach :

1. $$\begin{aligned} A_1 &\to A_2A_3|a \\ A_2 &\to A_3A_1|A_1b \\ A_3 &\to A_1A_2|A_2A_1|b \end{aligned}$$
2. $$\begin{aligned} A &\to AB|a \\ B &\to BC|b \\ C &\to CA|c \end{aligned}$$

Exercice 3. Montrer que l'intersection d'un langage algébrique et d'un langage rationnel est algébrique.

Exercice 4 (Langage local d'arbres)**.** Soit Σ un alphabet et t un terme de $T(\Sigma)$. On note $\mathsf{r}(t)$ pour le symbole qui étiquette la racine de t et $b(t)$ l'ensemble des *branchements* de t, définis par

$$\begin{aligned} \mathsf{r}(a) &= a & \mathsf{b}(a) &= \emptyset & &\text{si } a \in \Sigma_0 \\ \mathsf{r}(f(t_1,\dots,t_n)) &= f & \mathsf{b}(f(t_1,\dots,t_n)) &= \{f(\mathsf{r}(t_1),\dots,\mathsf{r}(t_n))\} \cup \bigcup_{i=1}^{n} \mathsf{b}(t_i) & &\text{si } f \in \Sigma_n \end{aligned}$$

Par exemple $\mathsf{b}(\ f(g(a), f(a,b))\ ) = \{f(g,f), g(a), f(a,b)\}$.

Un langage d'arbres L est *local* si et seulement s'il existe deux ensembles $R \subseteq \Sigma$ de symboles racines et $B \subseteq \mathsf{b}(T(\Sigma))$ de branchements tels que $t \in L$ ssi $\mathsf{r}(t) \in R$ et $\mathsf{b}(t) \subseteq B$. Posons

$$L(R,B) = \{t \in T(\Sigma \mid \mathsf{r}(t) \in R \text{ et } \mathsf{b}(t) \subseteq B\}$$

on a alors qu'un langage L est local si et seulement si $L = L(\mathsf{r}(L), \mathsf{b}(L))$.

1. Montrer que le langage $\{ f(g(a), g(b)) \}$ n'est pas local.
2. Montrer que tout langage local d'arbres est reconnaissable.
3. Soit $\Sigma = \Sigma_0 \uplus \Sigma_{>0}$. Montrer qu'un langage d'arbres sur Σ est local avec $|R| = 1$ si et seulement s'il est l'ensemble des arbres de dérivation d'une grammaire algébrique.
4. Montrer que tout langage reconnaissable d'arbres est l'image d'un langage local d'arbres par un morphisme d'arbre alphabétique. En déduire que le langage $\mathsf{Fr}(L)$ des frontières d'un langage reconnaissable d'arbres est algébrique.