TD 8 : Grammaires algébriques

Exercice 1 (**À préparer** - Théorème du feuillage)**.**

1. Soit L un langage régulier d'arbres. Montrer que $\mathrm{Fr}(L)$ est algébrique.
2. Soit L un langage algébrique propre, c'est-à-dire tel que $\varepsilon \notin L$. Montrer qu'il existe un langage régulier d'arbres L' tel que $\mathrm{Fr}(L') = L$.

Exercice 2 (Propriétés de fermeture)**.** Montrer que la famille des langages algébriques est fermée par :

1. Concaténation et itération
2. Substitution algébrique

Montrer que la famille des langages algébriques ainsi que la famille des langages linéaires sont fermées par :

1. Union et image miroir
2. Intersection avec un langage rationnel
3. Morphisme
4. Projection inverse
5. Morphisme inverse

Exercice 3 (Clôture ascendante et descendante)**.** On dit qu'une formule MSO $\varphi(x, y)$ définit une relation binaire R si, pour tout arbre t et pour tous noeuds u, v de t, $t, x \mapsto u, y \mapsto v \models \varphi$ ssi $(u, v) \in R$.

1. Montrer que la relation $R_1 = \{(x, y) \mid x \text{ est un descendant de } y\}$ est définissable par une formule logique MSO.
2. Soit une relation binaire R définie par une formule MSO $\varphi(x, y)$. Montrer qu'il existe une formule MSO $\varphi^{\uparrow^*}(x, y)$ qui définit la clôture reflexive et ascendante $R^{\uparrow^*}$ de R, c'est-à-dire, $(x, y) \in R^{\uparrow^*}$ si et seulement si, soit :
   - $x = y$.
   - $\exists z_1, \ldots, z_n, z_1 = x, z_n = y$ et, $\forall i < n, (z_i, z_{i+1}) \in R \cap R_1$.
3. Même question pour la clôture reflexive et descendante $R^{\downarrow^*}$ de R.
4. Qu'en est-il de la clôture reflexive et transitive R^* de R ?

Exercice 4 (Langages de Dyck)**.** Soit $\Sigma_n = \{a_1, \ldots, a_n\} \cup \{\bar{a}_1, \ldots, \bar{a}_n\}$ l'alphabet formé de n paires de parenthèses. Soit $G_n = (\Sigma_n, V, P_n, S)$ la grammaire définie par $S \to a_1 S \bar{a}_1 S + \ldots + a_n S \bar{a}_n S + \varepsilon$. Le langage $D_n^* = \mathcal{L}_{G_n}(S)$ est appelé langage de Dyck sur n paires de parenthèses.

1. Montrer que $D_1^* = \{w \in \Sigma_1^* \mid |w|_{a_1} = |w|_{\bar{a}_1} \text{ et } |v|_{a_1} \geq |v|_{\bar{a}_1} \text{ pour tous } v \leq w\}$
2. On considère le système de réécriture (de type 0) $R_n = (\Sigma_n, P'_n)$ dont les règles sont $P'_n = \{(a_i \bar{a}_i, \varepsilon \mid 1 \leq i \leq n\}$.
   Montrer que $D_n^* = \{w \in \Sigma_n^* \mid w \xrightarrow{*} \varepsilon \text{ dans } R_n\}$.

3. Soit Γ un alphabet disjoint de Σ_n, $\Sigma = \Sigma_n \cup \Gamma$ et $L \subset \Sigma^*$ un langage.
   On définit la clôture $\mathrm{clot}(L) = \{v \in \Sigma^* \mid \exists w \in L, w \xrightarrow{*} v \text{ dans } R_n\}$.
   Montrer que si L est reconnaissable, alors $\mathrm{clot}(L)$ l'est aussi.
   On définit la réduction $\mathrm{red}(L) = \{v \in \mathrm{clot}(L) \mid v \not\to \text{ dans } R_n\}$.
   Montrer que si L est reconnaissable, alors $\mathrm{red}(L)$ l'est aussi.

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.