TD 06 – Régulièrement hors-contexte

Définition. Soit Σ et Γ deux alphabets. Un *homomorphisme* est une fonction h de Σ dans $\Gamma^\star$ (une lettre de Σ est envoyée sur un mot de Γ). On étend le domaine de définition des homomorphismes à $\Sigma^\star$ en posant $h(a_1 \cdots a_n) = h(a_1) \cdots h(a_n)$. L'image d'un langage $L \subseteq \Sigma^\star$ par h est $\{h(w) \in \Gamma^\star : w \in L\}$.

Définition. Soit L, L' deux langages réguliers, et a un terminal de L. On note $L[a \leftarrow L']$ le langage constitué des mots de L dans lesquels chaque occurence de la lettre a est remplacée par un mot de L'. Par exemple, si $abbaba \in L$ et $cc, dd \in L'$, alors $ccbbddbcc \in L[a \leftarrow L']$.

Exercice 1. *Stabilité*

1. Montrer que la famille des langages réguliers est close par homomorphisme.
2. Soit G et G' définies respectivement par $\{S \to bA, A \to aA|aB, B \to bC|a, C \to b\}$ et $\{S' \to cX, X \to bX|cS'|c\}$. Construire une grammaire G'' qui génère $L(G)[a \leftarrow L(G')]$. Trouver une grammaire linéaire à droite équivalente à G''.
3. Montrer que si L et L' sont deux langages réguliers, $L[a \leftarrow L']$ est régulier.
4. Soit L un langage régulier, et L^T son langage miroir : $L^T = \{x^T : x \in L\}$ où $(a_1 \ldots a_n)^T = a_n \ldots a_1$. Montrer que L^T est également régulier.
5. Les langages hors-contextes vérifient-ils les propriétés précédentes ?

Définition. Soit $G = (V, \Sigma, P, S)$ une grammaire hors-contexte. La grammaire est dite *autoenchâssante*[1] s'il existe $X \in V$ tel que $X \overset{+}{\Rightarrow} \alpha X \beta$ avec $\alpha, \beta \in (V \cup \Sigma)^+$ ($\alpha, \beta \neq \epsilon$).
Un langage hors-contexte L est dit *autoenchâssant* si toute grammaire hors-contexte le générant est autoenchâssante.

Exercice 2. *Autoenchâssements*
On souhaite montrer le théorème suivant :

Théorème. *Un langage hors-contexte est régulier si et seulement s'il n'est pas autoenchâssant.*

1. Montrer l'implication directe.

On veut montrer l'implication inverse. Soit donc un langage L hors-contexte non autoenchâssant. Soit $G = (V, \Sigma, P, S)$ une grammaire hors-contexte le générant telle que tout symbole non terminal est *accessible*[2] et sans règle de la forme $X \to Y$, $X, Y \in V$.

2. Supposons que pour tout X, il existe une dérivation $X \overset{\star}{\Rightarrow} \gamma S \gamma'$. Montrer qu'alors, G est nécessairement de type 3.

1. Ou *self-embedding* en anglais.
2. Pour tout $X \in V$, il existe une dérivation $S \overset{\star}{\Rightarrow} \alpha X \beta$.

3. On suppose maintenant qu'il existe $X \in V$ tel que $X \stackrel{*}{\not\Rightarrow} \gamma S \gamma'$. Montrer par induction sur $|V|$ que G est de type 3. Indication : vous pouvez utiliser les résultats de l'exercice 1.

Exercice 3. *dangling else*

$Stmt \to$ **if** $expr$ **then** $Stmt$ | **if** $expr$ **then** $Stmt$ **else** $Stmt$ | $stmt$

1. Montrer que cette grammaire est ambiguë.
2. Proposer une grammaire non ambiguë pour le même langage.

Définition. Une grammaire contextuelle $G = (V, \Sigma, P, S)$ est sous *forme normale de Kuroda* si toutes ses règles sont de la forme (i) $S \to SB$, (ii) $CD \to EF$, (iii) $G \to H$, (iv) $A \to a$ où $A, B, C, D, E, F, H \in V \setminus \{S\}$, $G \in V$, $A \in \Sigma$.

Exercice 4. *Forme normale de Kuroda*

✎ Montrer que toute grammaire contextuelle est équivalente à une grammaire sous forme normale de Kuroda. **Indication.** Partant d'une grammaire $G = (V, \Sigma, P, S)$ dont les productions sont soit de la forme $\alpha \to \beta$ avec $2 \geq |\beta| \geq |\alpha| \geq 0$ et $\alpha, \beta \in V^\star$ soit de la forme $A \to a$ avec $A \in V$ et $a \in \Sigma$, construire une nouvelle grammaire $G' = (V \cup \{S', Q\}, \Sigma, P', S')$ sous forme normale de Kuroda, où Q sert à compléter les règles $\alpha \to \beta$ qui ne sont d'aucune des formes (i) à (iii).