TD : Grammaires non contextuelles

Exercice 1

Soit $\Sigma = \{a, b\}$ et L_1 le langage des mots sur Σ contenant autant de a que de b. On note G la grammaire hors contexte suivante :

$$S \to XS \mid YS \mid \varepsilon$$

$$X \to aXbX \mid \varepsilon$$

$$Y \to bYaY \mid \varepsilon$$

1. Le langage L est-il régulier ?
2. Justifier que si $X \Rightarrow^* u$ alors $|u|_a = |u|_b$.
3. En déduire $\mathcal{L}(G) \subset L_1$
4. Soit $u \in L_1$ tel que u commence par un a. Montrer qu'il existe deux mots $v, w \in \Sigma^*$ tels que $u = vw$ et v est un mot de Dyck et $w \in L_1$.
5. Montrer $L_1 \subset \mathcal{L}(G)$.
6. Soit L_2 le langage des mots contenant plus de a que de b. Montrer que L_2 est non contextuel.

Exercice 2

Proposer des grammaires algébriques qui engendrent les langages suivants.

1. L_1 le langage dénoté par $a * b*$.
2. $L_2 = \{a^n b^n \mid n \in \mathbb{N}\}$.
3. $L_3 = \left\{a^{2n} b^n \mid n \in \mathbb{N}\right\}$.
4. $L_4 = \left\{a^n b^n c^m d^m \mid (n, m) \in \mathbb{N}^2\right\}$.
5. $L_5 = \left\{a^n c^m b^n \mid (n, m) \in \mathbb{N}^2\right\}$.
6. $L_6 = \{a^m b^n \mid m \geq n\}$.
7. $L_7 = \{u \mid u = \overline{u}\}$ (palindromes) sur $\{a, b\}$.
8. L_8 mots contenant un nombre impair de a sur $\{a, b\}$.
9. $L_9 = \{a^p b^q c^n \mid p + q = n\}$.
10. $L_{10} = \{uc^n \mid u \in \{a, b\}^* \text{ et } |u| = n\}$.

Exercice 3

Stabilités des langages non contextuels

1. Démontrer que l'union de deux langages hors contexte est un langage hors contexte.
2. Démontrer que la concaténation de deux langages hors contexte est un langage hors contexte.
3. Démontrer que l'étoile d'un langage hors contexte est un langage hors contexte.
4. Démontrer que si L est non contextuel alors $\tilde{L}$ (miroir de L) est non contextuel
5. Démontrer que $L_1 = \{a^n b^n c^m \mid (n, m) \in \mathbb{N}^2\}$ et $L_2 = \{a^n b^m c^m \mid (n, m) \in \mathbb{N}^2\}$ sont non contextuels. Exprimer $L_1 \cap L_2$. On peut montrer (hors programme) que ce langage n'est pas un langage non contextuel.

Exercice 4

On considère la grammaire G suivante :

$$S \to S + S \mid S \times S \mid C \mid (S)$$

$$C \to 0 \mid 1 \mid 0C \mid 1C$$

1. Décrire intuitivement le langage engendré par G.
2. Montrer que G est ambiguë.
3. Construire une grammaire non ambiguë G' équivalente à G.

Exercice 5

Mots de Motzkin

On se place dans le plan $\mathbb{N}^2$ et on considère les déplacements élémentaires suivants :

- $a = (+1, +1)$,
- $b = (+1, -1)$,
- $c = (+1, 0)$

Un chemin est un mot sur Σ, il représente une trajectoire débutant en $(0, 0)$ et définie par la succession de déplacements élémentaires correspondant aux lettres du mot. Un chemin est valide s'il mène de $(0,0)$ à $(n,0)$ en restant bien dans l'espace $\mathbb{N}^2$.

1. Donner une grammaire pour décrire l'ensemble des chemins valides.
2. Démontrer que l'ensemble des chemins valides n'est pas régulier.

Exercice 6

Mots de Lukasiewicz

On appelle *mot de Lukasiewicz* sur $\{a,b\}$ un mot u vérifiant :
- $|u|_b = |u|_a + 1$
- Pour tout préfixe propre v de u : $|v|_a \geq |v|_b$

Soit G la grammaire algébrique :

$$S \to aSS \mid b$$

1. Démontrer que si $S \Rightarrow^* u \in \{a,b\}^*$ alors u est un mot de Lukasiewicz.
2. Réciproquement montrer que tout mot de Lukasiewicz se dérive à partir de S.

Exercice 7

Ambiguïté et dérivations gauches

Soit $G = (\Sigma, V, S, \mathcal{R})$ une grammaire algébrique, justifier les propositions suivantes :
1. Si $u \in \mathcal{L}(G)$ alors $S \Rightarrow_g^* u$.

2. Si G est non ambiguë et $u \in \mathcal{L}(G)$ alors il existe **une unique** suite de dérivations gauches $S \Rightarrow_g^* u$.

3. En déduire que $S \to \varepsilon \mid SS \mid aSb$ est ambiguë.

Remarque : on obtient exactement le même résultat pour des dérivations droites.

⋆Exercice 8

Grammaires linéaires

On dit qu'une grammaire est linéaire si chaque règle ne contient qu'un seul symbole non terminal au plus. On dit qu'elle est linéaire à droite si le symbole non terminal se situe nécessairement à la fin du mot produit.
1. Démontrer que tout langage régulier est engendré par une grammaire linéaire à droite.
2. Démontrer que tout langage engendré par une grammaire linéaire à droite est régulier.
3. Justifier que le résultat reste valide pour des grammares linéaires à gauche.

⋆Exercice 9

Forme normale de Chomsky

On dit qu'une grammaire est sous *forme normale de Chomsky* (aussi appelée *forme normale quadratique*) si elle n'utilise que des règles de production de la forme :
- $S \to \varepsilon$
- $X \to a$
- $X \to YZ$

où $a \in \Sigma$ est un symbole terminal et X, Y, Z sont des symboles non terminaux.
1. Mettre sous forme normale de Chomsky la grammaire $S \to aSbS \mid \varepsilon$
2. Démontrer que toute grammaire est faiblement équivalente à une grammaire sous forme normale de Chomsky
3. Quelle est la complexité de la mise sous forme normale de Chomsky ?