TD9. Grammaires algébriques

Exercice 1. Exemples de dérivations

Soit $\Sigma = \{a, b\}$, $V = \{S, D, T, X\}$ et G la grammaire contenant les règles de production :

$$S \to XSX \mid D$$
$$D \to aTb \mid bTa$$
$$T \to XTX \mid X \mid \varepsilon$$
$$X \to a \mid b$$

1. Parmi les dérivations suivantes, lesquelles sont vraies ?

a) $T \Rightarrow aba$ b) $T \Rightarrow^* aba$ c) $T \Rightarrow T$ d) $T \Rightarrow^* T$

2. Parmi les dérivations suivantes, lesquelles sont vraies ?

a) $XXX \Rightarrow^* aba$
b) $X \Rightarrow^* aba$
c) $T \Rightarrow^* XX$
d) $T \Rightarrow^* XXX$
e) $D \Rightarrow^* \varepsilon$

3. Donner trois mots de $L(G)$. Justifier par une dérivation.
4. Donner trois mots de Σ^* qui ne sont pas dans $L(G)$. On ne demande pas de justifier.
5. Donner une description en français de $L(G)$. On ne demande pas de le prouver.

Exercice 2. Arbres de dérivation

Soit $\Sigma = \{a, b, c, ..., z, +, (,)\}$, $V = \{S, P, F, T\}$, et la grammaire G de symbole initial S et ayant les règles de production suivantes :

$$S \to S + P \mid P$$
$$P \to PF \mid F$$
$$F \to (F) \mid T$$
$$F \to a \mid b \mid \cdots \mid z$$

1. Donner un arbre de dérivation de $ab + bc$.
2. Donner un arbre de dérivation de $b + (a + b)c + a(bc)$.
3. Que représente cette grammaire, i.e. que représentent les mots engendrés par G ?

Exercice 3. Construction de grammaires

1. Soit $\Sigma = \{a, b\}$. Donner une grammaire G_i pour chaque langage L_i ci-dessous.

- $L_0 = \{w \in \Sigma^* \mid |w| \geq 3\}$
- $L_1 = \{w \in \Sigma^* \mid |w|_a \geq 3\}$
- $L_2 = \{w_1 \dots w_n \in \Sigma^+ \mid w_1 = w_n\}$
- $L_3 = \emptyset$
- $L_4 = \{w \in \Sigma^* \mid |w| \equiv 1[2]\}$
- $L_5 = \{w \in \Sigma^* \mid w \text{ est un palindrome}\}$
- $L_6 = L_5^c$
- $L_7 = \{a^n b^n \mid n \in \mathbb{N}\}$
- $L_8 = \{a^n b^m \mid m \geq n\}$
- $L_9 = \{a^n b^m \mid m \geq 2n\}$
- $L_{10} = \{w \in \Sigma^* \mid |w|_a \geq |w|_b\}$

2. On considère l'alphabet $\Sigma = \{$ `[` , `]` , `;` , `true` , `false` $\}$. Donner une grammaire dont le langage est l'ensemble des listes `OCaml` de booléens.

Exercice 4. Stabilité des langages algébriques

1. Montrer que la classe des langages algébriques est stable par l'union.
2. Montrer que la classe des langages algébriques est stable par concaténation.

Exercice 5. Preuves par induction
Soit $\Sigma = \{a, b\}$, $V = \{S\}$ et la grammaire G contenant les règles de production :

$$S \to aS \mid Sb \mid a \mid b \mid \varepsilon$$

1. Que vaut $L(G)$? Le démontrer.

Exercice 6. Ambiguïté
Soit $\Sigma = \{a, b\}$, $V = \{S\}$ et la grammaire G contenant les règles de production :

$$S \to aS \mid aSbS \mid \varepsilon$$

1. Montrer que G est ambiguë.
2. Montrer que $L(G) = \{v \in \Sigma^* \mid \text{pour tout } u \text{ préfixe de } v, |u|_a > |u|_b\}$.
3. Donner une grammaire non ambiguë reconnaissant $L(G)$.

Exercice 7. Un lemme d'itération
Dans cet exercice, on s'intéresse à un équivalent du lemme de l'étoile pour les langages non contextuels. On pourra ainsi montrer que certains langages ne peuvent pas être engendrés par des grammaires non contextuelles, car ils ne vérifient pas certaines propriétés.

Soit une grammaire $G = (V, \Sigma, P, S)$. On suppose dans la suite que $P \neq \emptyset$. On pose

$$p = \max\big(\{1\} \cup \{\, |w| \mid V \to w \in P \,\}\big),$$

c'est-à-dire la longueur du plus long mot apparaissant à droite d'une règle de production, ou 1 si tous les mots à droite des règles sont le mot vide.

1. Donner et prouver une majoration sur la longueur des mots de $(\Sigma \cup V)^*$ admettant un arbre de dérivation de hauteur h.
2. En déduire qu'il existe un entier N tel que, si w est un mot de $L(G)$ de longueur supérieure ou égale à N, alors tout arbre de dérivation de w admet une branche contenant au moins deux fois le même symbole non terminal.
3. En déduire qu'il existe un entier N tel que, pour tout mot $w \in L(G)$ vérifiant $|w| \geq N$, il existe des mots $u, x, c, y, v \in \Sigma^*$ tels que :

(i) $w = uxcyv$

(ii) $|xy| \geq 1$

(iii) $|xcy| \leq N$

(iv) pour tout $n \in \mathbb{N}$, on a $ux^n cy^n v \in L(G)$

4. En déduire que le langage $L = \{\, a^n b^n c^n \mid n \in \mathbb{N} \,\}$ n'est pas engendré par une grammaire non contextuelle.
5. En déduire que l'ensemble des langages non contextuels n'est stable ni par intersection, ni par complémentaire.