1 Grammaires hors-contexte

Exercice 1

Montrer que les langages suivants sur $\Sigma = \{a, b\}$ sont algébriques en exhibant une grammaire hors-contexte les engendrant.

1. L'ensemble des mots de taille paire.
2. Les mots ne contenant pas bba comme facteur.
3. $\{a^m b^n \mid n \neq m\}$
4. Le langage de Dyck, c'est-à-dire l'ensemble des mots correspondant à un parenthésage correct d'une expression arithmétique, en assimilant a à une parenthèse ouvrante et b à une parenthèse fermante (par exemple $aabaabbb$ mais pas $baab$ ou aab).
5. $\{a^m b^n \mid m \leqslant n \leqslant 2m\}$
6. Les mots qui ne sont pas des palindromes.
7. Le complémentaire de $\{a^n b^n \mid n \in \mathbb{N}\}$.

Exercice 2

Décrire en français ou formellement les langages engendrés par les grammaires suivantes :

1. $S \to aSa \mid aSb \mid \varepsilon$;
2. $S \to bSbb \mid A$, $A \to Aa \mid \varepsilon$;
3. $S \to XY \mid a \mid b$, $X \to YS \mid a \mid b$, $Y \to a \mid b$.

Exercice 3

Montrer que les langages suivants sont algébriques.

1. $\{a^i b^j c^k \mid i = j + k\}$;
2. $\{a^i b^j c^k \mid j = i + k\}$;
3. $\{u\#v \mid u \in \{a, b\}^*, |u| \neq |v|\}$;
4. $\{u\#v \mid u \in \{a, b\}^*, u \neq v\}$.

Exercice 4

On pose $L_a = \{uav \mid u, v \in \{a, b\}^*, |u| = |v|\}$ et $L_b = \{ubv \mid u, v \in \{a, b\}^*, |u| = |v|\}$. On pose également $L = \{uv \mid u, v \in \{a, b\}^*, |u| = |v|, u \neq v\}$.

1. Montrer que $L = L_a L_b \cup L_b L_a$.
2. En déduire que ces trois langages sont algébriques.

Exercice 5

Déterminer une grammaire engendrant le langage $L = \{u \in \{a, b\}^* \mid |u|_a \geqslant |u|_b\}$ et justifier rigoureusement sa correction.

2 Dérivation et ambiguïté

Exercice 6

Construire un arbre de dérivation, ou deux lorsque c'est possible, pour :

1. $u = aabbba$ et G définie par $S \to XS \mid \varepsilon$ et $X \to aa \mid ab \mid ba \mid bb$;

2. $v = abaabb$ et G définie par $S \to aSbS \mid bSaS \mid \varepsilon$;
3. $w = ababa$ et G définie par $S \to XY \mid a \mid b$, $X \to YS \mid a \mid b$, $Y \to a \mid b$.

Exercice 7

Soit G une grammaire hors-contexte sans ε-production (c'est-à-dire sans règle $X \to \varepsilon$). Montrer que si $u \in L(G)$ et $S \Rightarrow^k u$, alors u possède un arbre de dérivation à moins de $|u| + k$ nœuds.

Exercice 8

On considère G définie par $S \to SS \mid aSb \mid \varepsilon$.

1. Décrire le langage $L(G)$.
2. Montrer que G est ambiguë.
3. Déterminer une grammaire G' non ambiguë telle que $L(G) = L(G')$.

Exercice 9

Une grammaire régulière peut-elle être ambiguë ? Un langage rationnel peut-il être intrinsèquement ambigu ? Justifier.

3 Propriétés de clôture

Exercice 10

Montrer que les langages suivants ne sont pas algébriques en utilisant le lemme de pompage algébrique.

1. $\{a^n b^{2n} a^n \mid n \in \mathbb{N}\}$;
2. $\{uu \mid u \in \Sigma^*\}$ si $|\Sigma| > 1$;
3. $\{u\#v \mid u, v \in \{a,b\}^*, |u| = |v|, u \neq v\}$;
4. L'ensemble des mots u sur $\Sigma = \{(,),[,]\}$ tels que les projections de u sur $\{(,)\}$ et sur $\{[,]\}$ sont chacune bien parenthésées (par exemple, $u = ([)]$ est un tel mot) ;
5. $\{a^p \mid p \in \mathbb{P}\}$;
6. $\{a^n \mid n \notin \mathbb{P}\}$, en admettant le théorème de Dirichlet : si u_0 et r sont deux entiers premiers entre eux, alors la suite $(u_0 + kr)_{k\in\mathbb{N}}$ contient une infinité de nombres premiers.

Exercice 11

Soit L un langage algébrique sur $\Sigma = \{a\}$. On cherche à montrer que L est rationnel. On note n la longueur de pompage de L.

Pour $u \in L$ tel que $|u| \geqslant n$, on note $u = vwxyz$ une décomposition donnée par le lemme de pompage algébrique (décomposition choisie arbitrairement s'il y en a plusieurs). Pour $q \in [\![1, n]\!]$, on note $L_q = \{u \in L \mid |u| \geqslant n, |wy| = q\}$, w et y faisant référence à la décomposition précédente. Enfin, on note $M_q = \{a^{p+kq} \mid a^p \in L_q, k \in \mathbb{N}\}$.

1. Montrer que pour $q \in [\![1, n]\!]$, M_q est rationnel.
2. En déduire que L est rationnel.

Exercice 12

Soit Σ et Σ' deux alphabets et $\varphi : \Sigma^* \to \Sigma'^*$ un morphisme de mots. Soit $G = (\Sigma, V, P, S)$ une grammaire hors-contexte.

1. Montrer que $\varphi(L(G))$ est un langage algébrique.

   Indication : on rajoutera une variable X_a pour chaque $a \in \Sigma$.
2. En déduire que $L_0 = \{uvu \mid u \in \{a,b\}^*, v \in \{c,d\}^*\}$ n'est pas algébrique.
3. Soit L un langage tel que $\varphi(L)$ est algébrique. Peut-on conclure que L est algébrique ?

Exercice 13

Soit L un langage algébrique et R un langage rationnel sur Σ. On cherche à montrer que $L \cap R$ est algébrique. Pour cela, on considère $G = (\Sigma, V, P, S)$ une grammaire hors-contexte en forme normale de Chomsky et $A = (Q, \Sigma, \delta, q_0, F)$ un AFD tels que $L = L(G)$ et $R = L(A)$.

Montrer qu'il existe $G' = (\Sigma, V', P', S')$ une grammaire hors-contexte en forme normale de Chomsky telle que $V' = Q \times V \times Q$ et $L(G') = L \cap R$.

Exercice 14

Soit L un langage non algébrique et F un langage fini sur Σ. Montrer que $L \cup F$ n'est pas algébrique.

Exercice 15

Prouver la propriété suivante ou exhiber un contre-exemple : « si L^* est algébrique, alors L est algébrique. »

Exercice 16

Pour L un langage sur Σ, on pose $\frac{1}{2}L = \{u \in \Sigma^* \mid \exists v \in \Sigma^*, uv \in L \text{ et } |u| = |v|\}$.

1. Montrer que si L est rationnel, alors $\frac{1}{2}L$ est rationnel.
2. Sur $\Sigma = \{a, b, c\}$, montrer qu'il existe L un langage algébrique tel que $\frac{1}{2}L$ n'est pas algébrique.

4 Analyse syntaxique

Exercice 17

> **Définition**
>
> On dit qu'une grammaire hors-contexte $G = (\Sigma, V, P, S)$ est en **forme normale de Greibach** si ses règles sont de l'une des formes suivantes :
>
> – $S \to \varepsilon$;
>
> – $X \to aX_1X_2\ldots X_n$, avec $a \in \Sigma$, $X, X_1, \ldots, X_n \in V$ et $S \notin \{X_1, \ldots, X_n\}$.

1. Soit G une grammaire hors-contexte et $X \in V$. On suppose que les règles de production de X sont $X \to X\alpha_1 \mid X\alpha_2 \mid \ldots \mid X\alpha_p$ et $X \to \beta_1 \mid \beta_2 \mid \ldots \beta_q$ où les β_j ne commencent pas par X. Montrer qu'on peut modifier ces règles pour que si $X \to \alpha$ est une règle de production, alors α ne commence par par X.

   Indication : on rajoutera une variable Y et on considèrera $X \to \beta_j \mid \beta_j Y$.
2. Montrer que toute grammaire hors-contexte est faiblement équivalente à une grammaire en forme normale de Greibach.

Indication : on partira d'une grammaire en forme normale de Chomsky et on donnera la construction.

3. Montrer que si G est en forme normale de Greibach et $u \in L(G) \setminus \{\varepsilon\}$, alors toute dérivation de u est de taille exactement $|u|$.
4. Quel est l'intérêt d'une grammaire en forme normale de Greibach pour une analyse syntaxique top-down ?

Exercice 18

On souhaite écrire un analyseur syntaxique pour la grammaire G définie par $S \to (S)S \mid \varepsilon$.

1. Décrire en pseudo-code un analyseur syntaxique récursif descendant (top-down) de G.

On souhaite déterminer l'imbrication des parenthèses de manière arborescente : on utilise pour cela le type OCaml :

```
type arbre_abstrait = Par of arbre_abstrait list;;
```

Avec ce type, on souhaite transformer une chaîne de caractères de la forme `(())(()())` en `[Par [Par []]; Par [Par []; Par []]]`.

2. Écrire une fonction `analyseur_syntaxique : string -> arbre_abstrait list` qui fait cette analyse syntaxique. On lèvera une exception le cas échéant.