

# Devoir maison n°8

À faire pour le lundi 26/01

\*\*\*

## Théorème de Chomsky-Schützenberger

### 1 Langage de Dyck

Pour  $n \in \mathbb{N}^*$ , on définit un **alphabet à  $n$  paires de parenthèses**, noté  $\Sigma_n$ , comme un alphabet à  $2n$  lettres  $\Sigma_n = \{a_1, \overline{a_1}, a_2, \overline{a_2}, \dots, a_n, \overline{a_n}\}$ . Les lettres  $a_i$  seront appelées **parenthèses ouvrantes** et les  $\overline{a_i}$  sont les **parenthèses fermantes**.

Pour  $n \in \mathbb{N}^*$ , on appelle **langage de Dyck d'ordre  $n$** , noté  $D_n$ , le langage engendré par la grammaire hors-contexte  $G_n = (\Sigma_n, \{S\}, S, P)$ , où  $P$  contient les règles de production :

$$S \rightarrow a_1 S \overline{a_1} S \mid a_2 S \overline{a_2} S \mid \dots a_n S \overline{a_n} S \mid \varepsilon$$

Pour  $u \in \Sigma^*$ , on note  $\text{Pref}(u)$  l'ensemble de ses préfixes.

**Question 1** Représenter graphiquement un arbre de dérivation de  $u = a_1 a_2 \overline{a_2} a_3 \overline{a_3} a_1$  pour  $G_3$ .

**Question 2** On note  $\Sigma_1 = \{a, \overline{a}\}$ . Montrer que  $D_1 = \{u \in \Sigma_1 \mid |u|_a = |u|_{\overline{a}} \text{ et } \forall v \in \text{Pref}(u), |v|_a \geq |v|_{\overline{a}}\}$ .

**Question 3** En déduire que  $D_1$  n'est pas rationnel.

**Question 4** Pour  $n > 1$ , l'égalité suivante est-elle vraie ?

$$D_n = \{u \in \Sigma_n \mid \forall i \in \llbracket 1, n \rrbracket, |u|_{a_i} = |u|_{\overline{a_i}} \text{ et } \forall v \in \text{Pref}(u), \forall i \in \llbracket 1, n \rrbracket, |v|_{a_i} \geq |v|_{\overline{a_i}}\}$$

Justifier.

On définit les types suivants :

```
type lettre = Ouv of int | Fer of int
type mot = lettre list
```

tel qu'une lettre  $a_i$  sera représentée par `Ouv i` et une lettre  $\overline{a_i}$  par `Fer i`, et un mot de  $\Sigma_n^*$  comme une liste de lettres.

**Question 5** Écrire une fonction `dyck : mot -> bool` qui détermine si un mot  $u$  est un mot d'un langage de Dyck  $D_n$  pour un certain  $n \in \mathbb{N}$ .

### 2 Propriétés sur les langages algébriques

Soient  $\Sigma$  et  $\Gamma$  deux alphabets. On appelle **morphisme de mots** une fonction  $\varphi : \Sigma \rightarrow \Gamma$  telle que pour tout  $u, v \in \Sigma^*$ ,  $\varphi(uv) = \varphi(u)\varphi(v)$ .

**Question 6** Montrer que si  $\varphi$  est un morphisme de mots, alors  $\varphi(\varepsilon) = \varepsilon$ .

**Question 7** Soit  $\varphi$  un morphisme de mots. Montrer que si  $L$  est algébrique, alors  $\varphi(L)$  est algébrique. La réciproque est-elle vraie ? Justifier.

Soit  $G = (\Sigma, V, P, S)$  une grammaire hors-contexte. On dit que  $G$  est en **forme normale de Chomsky** si toutes les règles de production sont de l'une des formes suivantes :

- $X \rightarrow a$ , avec  $a \in \Sigma$  ;
- $X \rightarrow YZ$ , avec  $Y, Z \in V$ .

On admet que si  $G$  est une grammaire quelconque, alors il existe une grammaire  $G'$  en forme normale de Chomsky telle que  $L(G') = L(G) \setminus \{\varepsilon\}$ .

*Attention, cette définition est légèrement différente de celle donnée dans le cours.*

**Question 8** Déterminer, en justifiant succinctement, le langage engendré par la grammaire  $G_0$  en forme normale de Chomsky définie par les règles suivantes :

- $S \rightarrow AX \mid AB$  ;
- $X \rightarrow SB$  ;
- $A \rightarrow a$  ;
- $B \rightarrow b$ .

**Question 9** Soit  $L \subseteq \Sigma^*$  un langage algébrique et  $R \in \text{Rat}(\Sigma)$ . Montrer que  $L \cap R$  est algébrique. On pourra partir d'une grammaire en forme normale de Chomsky engendrant  $L \setminus \{\varepsilon\}$  et d'un automate fini reconnaissant  $R$  et construire une grammaire dont les variables sont des triplets état  $\times$  variable  $\times$  état.

**Question 10** Si  $L$  et  $L'$  sont deux langages algébriques, peut-on conclure que  $L \cap L'$  est algébrique ? Justifier.

### 3 Théorème de Chomsky-Schützenberger

Dans cette partie, on cherche à montrer le théorème de Chomsky-Schützenberger, dont l'énoncé est le suivant :

#### Théorème 3.1

Soit  $L \subseteq \Sigma^*$  un langage. Alors les deux propriétés suivantes sont équivalentes :

1.  $L$  est algébrique ;
2. il existe  $n \in \mathbb{N}^*$ ,  $R \in \text{Rat}(\Sigma)$  et  $\varphi : \Sigma_n \rightarrow \Sigma$  un morphisme de mots tels que  $L = \varphi(D_n \cap R)$ .

**Question 11** Montrer l'implication  $2 \Rightarrow 1$  du théorème.

Soit  $G = (\Sigma, V, P, S)$  une grammaire hors-contexte en forme normale de Chomsky. On numérote les règles de production de la forme  $X \rightarrow YZ$  par  $r_1, r_2, \dots, r_k$ . On pose  $G' = (\Sigma', V, P', S)$  où :

- $\Sigma' = \Sigma \cup \{\bar{a} \mid a \in \Sigma\} \cup \bigcup_{i=1}^k \{a_i, \bar{a}_i, b_i, \bar{b}_i, c_i, \bar{c}_i\}$  ;
- $P' = \{X \rightarrow a_i b_i Y \bar{b}_i c_i Z \bar{c}_i \bar{a}_i, \text{ pour } i \in \llbracket 1, k \rrbracket \text{ et } r_i = X \rightarrow YZ\} \cup \{X \rightarrow a\bar{a}, \text{ pour } X \rightarrow a \in P\}$ .

**Question 12** Déterminer la grammaire  $G'_0$ , pour  $G_0$  la grammaire de la question 8.

**Question 13** Montrer que  $L(G')$  est inclus dans un langage de Dyck d'ordre  $n$ , pour  $n$  bien choisi.

On pose  $\varphi : \Sigma'^* \rightarrow \Sigma^*$  le morphisme de mots défini par :

- pour  $a \in \Sigma$ ,  $\varphi(a) = a$  et  $\varphi(\bar{a}) = \varepsilon$  ;
- pour  $i \in \llbracket 1, k \rrbracket$ ,  $\varphi(a_i) = \varphi(\bar{a}_i) = \varphi(b_i) = \varphi(\bar{b}_i) = \varphi(c_i) = \varphi(\bar{c}_i) = \varepsilon$ .

**Question 14** Montrer que  $L(G) = \varphi(L(G'))$ .

Pour  $L$  un langage sur un alphabet  $\Sigma$ , on note  $P(L) \subseteq \Sigma$  l'ensemble des premières lettres des mots de  $L$ ,  $F(L) \subseteq \Sigma^2$  l'ensemble des facteurs de taille 2 des mots de  $L$  et  $N(L) = \Sigma^2 \setminus F(L)$ .

**Question 15** On pose  $R = P(L(G'))\Sigma'^* \setminus \Sigma'^*N(L(G'))\Sigma'^*$ . Montrer que  $R \in \text{Rat}(\Sigma')$ .

**Question 16** Montrer l'implication  $1 \Rightarrow 2$  du théorème.

\*\*\*