

## A.1 Grammaires formelles

- Donner une grammaire algébrique qui reconnaisse chacun des langages suivants (alphabet  $X = \{a, b, c\}$ ).
  - $L_1 = \{w \in X^* / w = a^n b^n c^p; n > 0 \text{ et } p > 0\}$
  - $L_2 = \{w \in X^* / w = a^n b^n a^m b^m; n, m \geq 1\}$
  - $L_3 = \{w \in X^* / |w|_a = 2|w|_b\}$
  - $L_4 = \{w \in X^* / \exists x \in X^* \text{ tq } w = x\bar{x}\}$
  - $L_5 = \{w \in X^* / w = \bar{w}\}$
- Soient les deux grammaires suivantes. Pour chacune d’entre elles, donnez le langage engendré et indiquez le type de la grammaire dans la classification de Chomsky. Commentez brièvement.

$S \rightarrow S_1 S_2$	$S \rightarrow a S B C$
$S_1 \rightarrow a S_1 b$	$S \rightarrow a B C$
$ab$	$C B \rightarrow B C$
$S_2 \rightarrow c S_2$	$a B \rightarrow ab$
$c$	$b B \rightarrow bb$
	$b C \rightarrow bc$
	$c C \rightarrow cc$

- Les deux grammaires suivantes reconnaissent un sous-ensemble des expressions arithmétiques habituelles (par exemple, le mot  $a + b \times c$ ). Quelle est la différence entre les deux ?

$E \rightarrow E + E \mid E \times E$	$E \rightarrow E + T \mid T$
$E \rightarrow (E) \mid a \mid b \mid c$	$T \rightarrow T \times F \mid F$
	$F \rightarrow (E) \mid a \mid b \mid c$

Proposez trois nouvelles grammaires en vous inspirant des deux précédentes :

- Une qui engendre le même langage mais donne une description structurelle où l’addition est prioritaire par rapport à la multiplication ;
  - Une qui reconnaît un sous-langage du langage précédent : seules les expressions arithmétiques bien parenthésées sont reconnues ;
  - Une qui reconnaît les expressions arithmétiques post-fixées (notation dite *polonaise*, où  $3 \times 7 + 2$  est écrit  $3 7 \times 2 +$ )
- Un groupe nominal est soit un déterminant suivi d’un nom soit un groupe nominal suivi d’un groupe prépositionnel. Un groupe prépositionnel est une préposition suivie d’un groupe nominal.
    - Ecrire la grammaire  $G$  du groupe nominal
    - Ecrire les deux arbres de dérivation que  $G$  associe au mot D N P D N P D N où D est le symbole associé à la catégorie “déterminant”, N le symbole associé à la catégorie “nom” et P celui associé à la catégorie “préposition”.
    - Combien d’arbres de dérivation admet le mot D N P D N P D N P D N ?
    - Parmi les deux arbres de la question (b), lequel choisiriez vous pour représenter la structure du groupe nominal “le chat de la voisine de la concierge” ?
    - Modifiez la grammaire  $G$  de sorte qu’elle n’associe au mot D N P D N P D N que l’analyse que vous avez identifiée dans la question précédente.
  - Règles simples (ou productions singulières). Transformer la grammaire suivante en grammaire sans règles simples.

$S \rightarrow AB \mid A ; A \rightarrow aB \mid bA \mid aSb ; B \rightarrow S \mid b$

6. Soit la grammaire hors contexte suivante :

$S \rightarrow p$   
 $p \rightarrow \text{gn } v1 \text{ que } p \mid \text{gn } v2$   
 $\text{gn} \rightarrow \text{np} \mid \text{det } \text{nc}$   
 $\text{np} \rightarrow \text{Léa} \mid \text{Luc} \mid \text{Ève} \mid \text{Max}$   
 $\text{nc} \rightarrow \text{femme} \mid \text{homme} \mid \text{étudiante} \mid \text{étudiant} \mid \text{fille} \mid \text{garçon}$   
 $\text{det} \rightarrow \text{le} \mid \text{la} \mid \text{l'}$   
 $v1 \rightarrow \text{pense} \mid \text{croit} \mid \text{voit} \mid \text{sait} \mid \text{dit} \mid \text{raconte}$   
 $v2 \rightarrow \text{se promène} \mid \text{marche} \mid \text{part}$

- (a) Donner quatre phrases distinctes reconnues par cette grammaire, contenant respectivement 0, 1, 2 et 3 fois le mot *que*.
  - (b) Pour quelles raisons ces phrases ne sont-elles pas toutes correctes en français? Comment modifier la grammaire pour corriger cela?
  - (c) Donner l'arbre de dérivation de *Luc sait que la femme croit que Léa part*.
7. Donner une grammaire régulière (vocabulaire  $V = \{a, b, c, d, \dots, y, z\}$ ) pour le langage qui contient l'ensemble des mots, de deux lettres minimum, composés en alternance d'une consonne et d'une voyelle et débutant par une consonne et finissant par une voyelle. Donner une grammaire algébrique engendrant le même langage.
8. Proposer une grammaire algébrique qui engendre le langage  $\{a^n db^p da^n, n > 0, p \geq 0\}$ .
9. Soit l'alphabet  $X = \{+, =, a\}$ . (1) Donner une grammaire algébrique pour le langage  $L$  dont chaque mot représente une addition correcte de deux suites de caractères  $a$ . Par exemple  $L$  contient le mot  $aa + aaaa = aaaaaa$ .
10. Soit le langage  $L_1$  sur le vocabulaire  $V = \{l', \text{homme}, \text{ours}, \text{qui}, a, \text{vu}\}$  formé de l'ensemble des phrases finies de la forme *l'homme qui a vu l'ours, l'homme qui a vu l'homme qui a vu l'ours, l'homme qui a vu l'homme qui a vu l'homme qui a vu ... qui a vu l'ours*.
- (a) Donner une grammaire algébrique (*context-free*) engendrant  $L_1$ .
  - (b) Donner une grammaire régulière engendrant  $L_1$ .

Soit  $L_2$  le langage engendré par la grammaire  $\mathcal{G}_2$  :

$S$	$\rightarrow$	NP Rel
NP	$\rightarrow$	l'homme
	$ $	l'ours
Rel	$\rightarrow$	qui a vu NP Rel
	$ $	qui a vu l'ours

- (c) Mettre la grammaire  $\mathcal{G}_2$  sous forme normale de Chomsky.
- (d) Proposer une grammaire  $\mathcal{G}_3$  qui engendre le langage  $L_3$ , sur-ensemble de  $L_2$  mais dans lequel les symboles *ours* et *homme* sont **strictement** interchangeables.
- (e) Décrire informellement les différences entre les langages  $L_1$  et  $L_2$ .
- (f) Donner une grammaire algébrique du langage  $L_2 \setminus L_1$ .

11. Soit la grammaire régulière suivante :

$$G = \langle \{a, b, c, d\}, \{S, T, V\}, S, \left\{ \begin{array}{l} S \rightarrow aS|bS|cT, \\ T \rightarrow aV|bV|aT|bT|cT|dT, \\ V \rightarrow \varepsilon \end{array} \right\} \rangle.$$

- a. Dessiner un automate (non déterministe) reconnaissant  $L_G(S)$ .
- b. Déterminiser cet automate.