# IFT-2002 Informatique Théorique
# Devoir 2

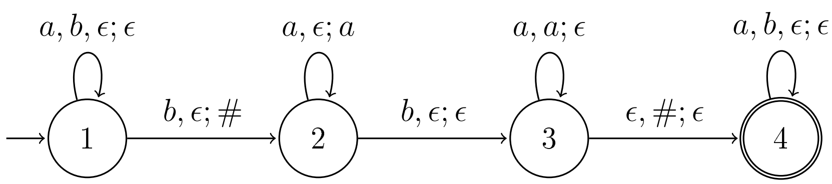
111 124 564
Raphaël Sylvain

**Question 1.**

Montrez que les langages ci-dessous sont non-contextuels. Vous pouvez le faire soit en donnant une grammaire non-contextuelle qui génère le langage, soit un automate à pile qui accepte le langage. (Notez que si vous donnez une grammaire et un automate, les deux seront corrigés mais seul le premier des deux sera noté)

1. $L = \{uba^nba^nbv : n \in \mathbb{N} \wedge u, v \in \{a,b\}^*\}$.
2. Soit $K$ l'ensemble des mots de la forme $x\$y$ où $x$ et $y$ sont des mots de longueur impaire sur l'alphabet $\{a,b\}$ et la lettre médiane de $x$ est égale à la lettre médiane de $y$. Par exemple $abb\$b$ et $aaaaa\$bab$ sont des mots dans $K$ mais $aa\$aaa$ n'est pas dans $K$ (le mot à gauche est de longueur paire), $aaa\$aab\$$ n'est pas dans $K$ (pas de la forme $x\$y$) et $aba\$abaab$ n'est pas dans $K$ (pas la même lettre médiane).

**Réponse 1:**




**Réponse 2:**

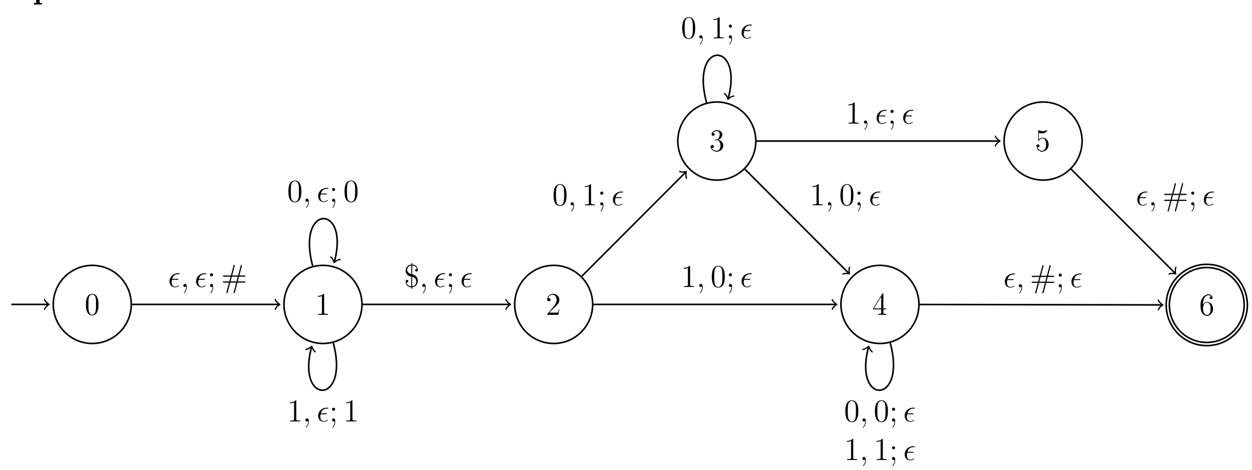
$S \to A\$A$; $S \to B\$B$;
$A \to CAC$; $A \to a$;
$B \to CBC$; $B \to b$;
$C \to a$; $C \to b$;

**Question 2.**

Montrez que le langage suivant est non-contextuel. (*Je pense qu'il est plus facile de construire un automate à pile pour ce langage plutôt que de fournir une grammaire non-contextuelle mais vous pouvez faire l'un ou l'autre.*)

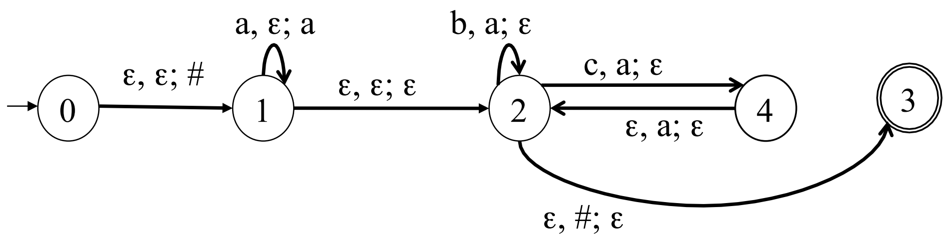
$T = \{x\$y^r : x$ est la représentation en binaire d'un entier $n_x$ et $y$ représente l'entier $n_x + 1\}$
Rappelons que $y^r$ est le mot $y$ renversé. Par exemple 10010\$11001 $\in T$ parce que 10010 représente l'entier 18 et que le renversé du mot qui suit le \$ est 10011 qui représente l'entier 19. De la même façon 11011\$00111 est dans le langage puisque 11011 représente 27 et 11100 représente 28.

**Réponse:**



**Question 3.**

Considérez l'automate à pile suivant:




a) Dites si les séquences suivantes sont acceptées par l'automate à pile ci-dessus: *ab*, *acc*, *aac*, *aaacb*.
b) Décrivez le langage accepté par cet automate. (Par une expression ensembliste ou de toute autre façon qui est suffisamment claire et précise pour que le correcteur vous comprenne)
c) Donnez une grammaire non-contextuelle qui génère le même langage.

**Réponse:**

a)

1) ab -> Oui
2) acc -> Non
3) aac -> Oui
4) aaacb -> Oui

b)

C'est le langage débutant par une série de $a$ et suivi par des $b$ et/ou $c$. Le nombre de $a$ est égal est égal au nombre de $b$ plus deux fois le nombre de $c$.

c)

$S \rightarrow aaSc$; $\quad S \rightarrow aSb$; $\quad S \rightarrow \epsilon$;

**Question 4.**

Considérez la grammaire $G$ donnée par les règles ci-dessous. (L'alphabet terminal est $\{a, b, c\}$.)

$S \to abPR$;  $S \to QTba$;
$P \to abPR$;  $P \to c$;
$T \to QTba$;  $T \to c$;
$R \to Ra$;  $R \to ba$;
$Q \to Qb$;  $Q \to ab$;

a) Décrivez le langage généré par cette grammaire. (Par une expression ensembliste ou de toute autre façon qui est suffisamment claire et précise pour que le correcteur vous comprenne. Je suggère d'illustrer votre explication avec quelques exemples bien choisis de séquences qui appartiennent au langage et de séquences qui n'appartiennent pas au langage)

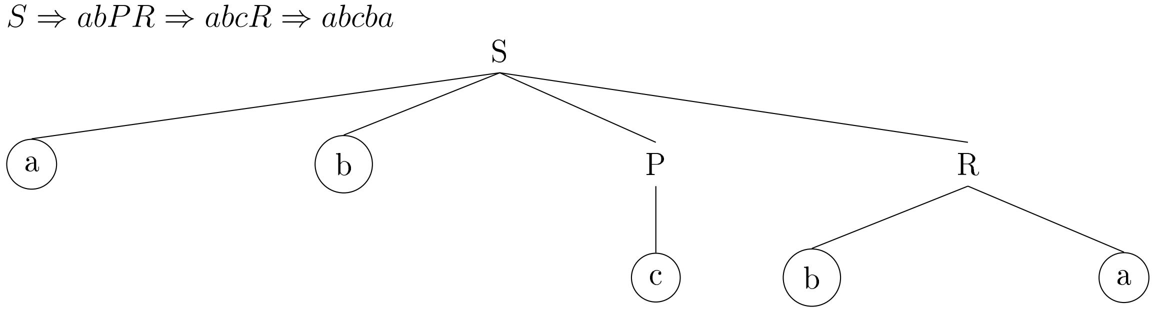
b) Montrez que $G$ est ambigüe en donnant deux arbres de dérivation distincts pour un même mot.

**Réponse: a)**
$\{\ \{(ab)^i c(x)^i | x = ba^j \wedge i \in \mathbb{N} \wedge j \in \mathbb{N}\} \cup \{(x)^i c(ba)^i | x = ab^j \wedge i \in \mathbb{N} \wedge j \in \mathbb{N}\}\ \}$
C'est un langage dont le mot peut avoir deux formes:
1) débutant par une suite de $ab$ suivie d'un $c$ suivi d'autant de groupe ayant la forme $ba^n \wedge n > 0$
2) finissant par une suite de $ba$ précédée d'un $c$ précédé d'autant de groupe ayant la forme $ab^n \wedge n > 0$

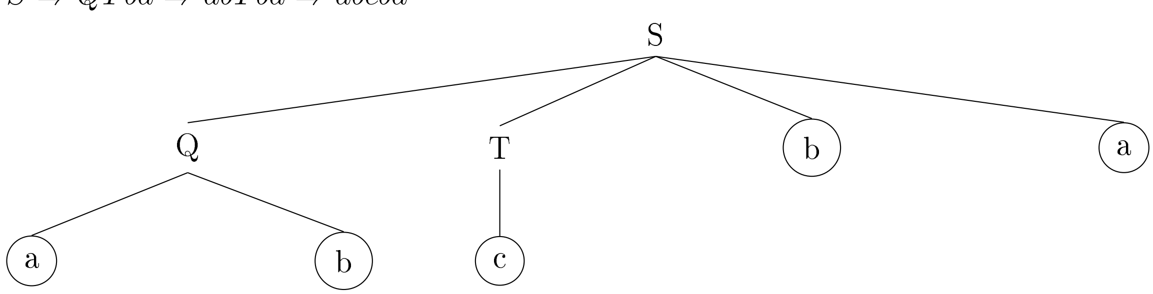
Mots acceptés par la forme 1: *abcba*, *abcbaa*, *ababcbaaba*
Mots acceptés par la forme 2: *abcba*, *abbcba*, *ababbcbaba*
Mots refusées: *abbcbba aabcbba*

**Réponse: b)**

$S \Rightarrow abPR \Rightarrow abcR \Rightarrow abcba$




$S \Rightarrow QTba \Rightarrow abTba \Rightarrow abcba$



**Question 5.**
Utilisez le lemme de pompage pour les langages non-contextuels pour montrer que le langage suivant n'est pas non-contextuel.

$L = \{a^\ell b^n c^m | \ell, n, m \in \mathbb{N} \wedge \ell < n < m\}$

**Réponse:**

Montrons que $L$ n'est pas non-contextuel. Il satisfait donc le corrolaire de pompage pour tout $p > 0$.

Quel que soit $p$, considérons le mot $z = a^p b^{p+1} c^{p+2}$. Ce mot est de longeur $3p+3$ et appartient à $L$ il devraut donc être "pompé".

Plus précisément, il existe $z = svuwt$ avec $|vw| \geq 1$ et $|vuw| \leq p$ et $\{\forall i | sv^i uw^i t \in L\}$

Il reste à montrer qu'en fait, peut importe le découpage, il existe un i tel que $sv^i uw^i t \notin L$

1) $v$ contient des $a$

Donc $uw$ ne peut pas contenir de $c$ car sinon $|vuw| > p$. Alors $sv^5 uw^5 t$ n'est pas dans le langage, car il y aura plus de $a$ que de $c$

2) $v$ contient des $b$ mais pas de $a$

Alors aucun $a$ est pompable et $sv^0 uw^0 t$ n'est plus dans le langage, car l'on enlève au moins un $b$ et le nombre de $a$ sera plus grand ou égal au nombre de $b$

3) $v$ contient des $c$ mais pas de $b$

Alors aucun $b$ est pompable et $sv^0 uw^0 t$ n'est plus dans le langage, car l'on enlève au moins un $c$ et le nombre de $b$ sera plus grand ou égal au nombre de $c$

4) Puisque $v$

Ne peut contenir de $c$ sans contenir de $b(3)$
Ne peut contenir de $b$ sans contenir de $a(2)$
Ne peut contenir de $a(1)$

alors $v = \epsilon$.

5) $w$ contient des $c$ mais pas de $b$

Alors aucun $b$ est pompable puisque $v = \epsilon(4)$ et $sv^0 uw^0 t$ n'est plus dans le langage, car l'on enlève au moins un $c$ et le nombre de $b$ sera plus grand ou égal au nombre de $c$

6) $w$ contient des $b$ mais pas de $a$

Alors aucun $a$ est pompable puisque $v = \epsilon(4)$ et $sv^0 uw^0 t$ n'est plus dans le langage, car l'on enlève au moins un $b$ et le nombre de $a$ sera plus grand ou égal au nombre de $b$

7) $w$ contient des $a$ Donc $w$ ne peut pas contenir de $c$ car sinon $|vuw| > p$. Alors $sv^5 uw^5 t$ n'est pas dans le langage, car il y aura plus de $a$ que de $c$

8) Puisque $w$

Ne peut contenir de $c$ sans contenir de $b(5)$
Ne peut contenir de $b$ sans contenir de $a(6)$
Ne peut contenir de $a(7)$

alors $w = \epsilon$.

9) Nous avons que $v = \epsilon$ et $w = \epsilon$. Ceci est impossible car $|vw| > 1$

$L$ n'est pas non-contextuel.

**Question facultative**

Soit $\{a, b\}$ un alphabet et définissons la relation $\sim$ sur les mots de $\{a, b\}^*$ par $x \sim y$ si $x$ et $y$ ont le même nombre de $a$ et ont le même nombre de $b$. Par exemple $abaa \sim baaa$ et $ababb \sim bbaab$. Si $L$ est un langage sur $\{a, b\}^*$ on peut maintenant définir

$$K_L = \{x : \exists y \; x \sim y \wedge y \in L\}.$$

Donc un mot $x$ appartient à $K_L$ si ses lettres peuvent être réordonnées pour former un mot de $L$.

1. Donnez un exemple de langage régulier $L$ tel que $K_L$ n'est pas régulier.

   **Réponse:**

   $L = (ab)^*$ Le langage de suites de paires $ab$ (ex: $abababab$)

   Supposons que $K_L$ est régulier. Il satisfait donc le lemme de pompage pour une certaine longueur de pompage $p > 0$

   Quel que soit $p$, considérons le mot $w = a^p b^p$. Ce mot est de longeur $2p$ et appartient à $K_L$, car les symboles peuvent être réorganisés pour former le mot $(ab)^p$ qui est acceptés par $L$. Donc le mot peut être "pompé".

   Plus précisément, il existe $w = xyz$ tel que

   (a) $|xy| \leq p$

   (b) $|y| > 0$

   (c) $xy^i z \in K_L$ pour tout $i \in \mathbb{N}$.

   Puisque $|xy| \leq p$ et puisque les $p$ premiers symboles de $w$ sont des $a$, on sait que $y$ ne contient que des $a$.

   On ne sait pas combien de $a$ sont contenus dans $y$, mais on sait qu'il en a $j$ pour un certain $j$ compris entre 1 et $p$.

   Prenons $i = 0$. On a $xy^0 z = xz = a^{p-j} b^p$. Ce mot ne fait pas partie du language, puisque $j > 0$ et que $p - j \neq p$ alors que le mot doit avoir autant de $a$ et de $b$ pour être une série de $ab$ et appartenir à $L$.

   $K_L$ n'est pas régulier.

2. Montrez que si $L$ est un langage régulier alors le langage $K_L$ est toujours non-contextuel.

   **Réponse:**

   Si $L$ est régulier, alors il peut être écrit par une grammaire régulière.

   Appliquons ces transformations aux règles:

   Si une règle à la forme $X \Rightarrow aY$, ajouter aussi une règle $X \Rightarrow Ya$
   Si une règle à la forme $X \Rightarrow bY$, ajouter aussi une règle $X \Rightarrow Yb$
   Garder les règles de la formes $X \Rightarrow a$, $X \Rightarrow b$ et $X \Rightarrow \epsilon$

   Avec de tels règles, nous gardons la consistence du nombre de caractère du langage régulier, puisque nous n'ajoutons aucun nouveau symboles non-terminal et que nous ne modifions pas les symboles terminales déjà existant dans les règles. Avec l'ajout de nos règles, nous pouvons ordonner les symboles dans n'importe quel ordre.

Par exemple, pour avoir un suite de $a$ suivi d'une suite de $b$ nous utilisons les formes de transormation dont le $a$ est avant le symbole non terminal et que le $b$ est après le symbole non terminal.

Pour avoir les $b$ avant les $a$, nous faisons l'inverse en prenant les règles dont le $b$ sont avant le symbole non terminal et que les $a$ sont après le symbole non terminal.

3. Montrez si l'alphabet est maintenant $\{a, b, c\}$ alors on peut trouver un langage régulier $L$ tel que $K_L$ n'est pas non-contextuel. (Évidemment, on ajuste la définition de $\sim$, c'est-à-dire $x \sim y$ si $x$ et $y$ ont le même nombre de $a$, le même nombre de $b$ et le même nombre de $c$.)

   **Réponse:**

   Prenons le langage régulier $ccb(abc)^*$, soit le langage des mots commençant par deux $c$, suivi d'un $b$ suivi d'un nombre indéterminée de triplets $abc$.

   Prenons le mot $a^p b^{p+1} c^{p+2}$ avec $p > 0$. Ce mot peut être réodonnée pour former un mot appartenant à $L$ et appartient donc à $K_L$.

   Dans la question 5 de ce présent document, nous avons montrer qu'aucun langage non-contextuel peut former des mots de la forme $\{a^\ell b^n c^m | \ell, n, m \in \mathbb{N} \wedge \ell < n < m\}$.

   La réapplication de la preuve est laissée comme exercice pour le lecteur.