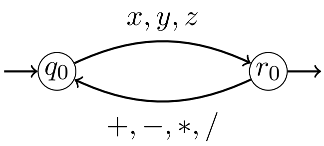
TD 6 - Grammaires hors-contexte

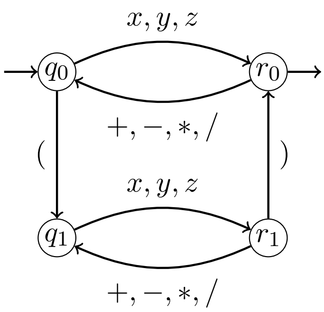
Licence 3 - Université de Bordeaux

Solution de l'exercice 1 :

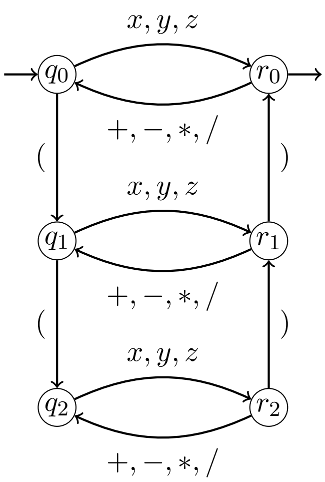
1. On donne l'automate :



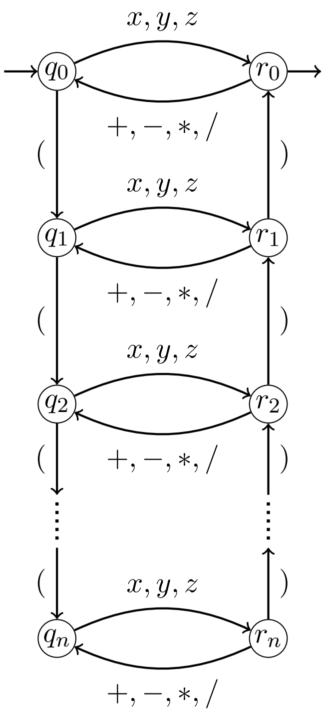
2. On autorise maintenant un niveau de parenthèses :



3. On autorise maintenant deux niveaux de parenthèses :



4. On autorise maintenant n niveaux de parenthèses pour n fixé :



5. Le langage $L = \cup_{n \geq 0} A_n$ n'est par contre pas régulier. On applique la contre-posée du lemme de pompage.
   - Soit $N \geq 0$ quelconque.
   - Nous choisissons le mot $z = a^N b^N \in L$.
   - Soit $z = uvw$ une décomposition quelconque telle que $v \neq \epsilon$ et $|uv| < N$. Les conditions sur les longueurs impliquent que $u = a^i$ et $v = a^j$ avec $j \neq 0$.
   - Nous choisissons $k = 0$: $uv^0w = uw = a^{N-j}b^N \notin L$ (car $j \neq 0$).

   On a donc démontré que L ne satisfait pas le lemme de pompage, on peut conclure que L n'est pas régulier.

Solution de l'exercice 2 :

1. *Les mots comportant exactement 42 occurrences de a et 42 occurrences de b.*
   Pour l'alphabet $A = \{a, b\}$ ce langage est fini, donc régulier. Or on a vu en cours que tout langage régulier est hors-contexte.
2. $L_1 = \{a^{2n}b^n \mid n \geq 0\}$.
   Le langage L_1 n'est pas régulier. On le montre avec la contre-posée du lemme de pompage.
   - Soit $N > 0$ quelconque.
   - Nous choisissons $z = a^{2N}b^N \in L_1$.
   - Soit $z = uvw$ une décomposition quelconque qui satisfait $v \neq \epsilon$ et $|uv| < N$. Les conditions sur u, v impliquent que $u = a^i$ et $v = a^j$ avec $j \neq 0$ et $i + j < N$.
   - Nous choisissons $k = 0$: $uw = a^{2N-j}b^N \notin L_1$.

   On peut conclure que L_1 ne satisfait pas le lemme de pompage, donc L_1 n'est pas régulier.
   Une grammaire pour L_1 : $G_1 = (\{S\}, \{a, b\}, R, S)$ avec $R = S \longrightarrow aaSb \mid \epsilon$.

3. $L_2 = \{a^{p+q}b^pc^q \mid p, q \geq 0\}$.
   Le langage L_2 n'est pas régulier.
   En effet supposons que L_2 est régulier, alors le langage $L \cap a^*b^* = \{a^nb^n \mid n \geq 0\}$ est lui aussi régulier, ce qu'on sait être faux. Le langage est par contre hors-contexte, on peut le définir grâce à la grammaire suivante : $G_2 = (\{S, T\}, \{a, b, c\}, R, S)$ avec R :

$$\begin{array}{lcl} S & \to & \epsilon \\ S & \to & aSc \\ S & \to & T \\ T & \to & \epsilon \\ T & \to & aTb \end{array}$$

4. L'ensemble L_3 des mots dont la longueur est un multiple de 2^{64}.
   L_3 est régulier. Considérons K le langage des mots de taille 2^{64}, K étant fini il est régulier. On a $L_3 = K^*$, donc L_3 est régulier.
5. $L_4 = \{w \in \{a, b\}^* \mid |w| = 2^k, k \geq 0\}$.
   Donc L_4 consiste des mots dont la longueur est une puissance de 2.
   Le langage L_4 n'est pas régulier. Il se trouve que L_4 n'est même pas hors-contexte. On le montre avec la contre-posée du lemme de pompage pour les langages hors-contexte.
   — Soit $N > 0$ quelconque.
   — Nous choisissons $z = a^{2^N} \in L_4$.
   — Soit $z = uvwxy$ une décomposition quelconque saisfaisant $|vx| < N$ et $vx \neq \epsilon$. On a donc $u = a^i, v = a^j, w = a^m, x = a^p, y = a^r$ tels que $2^N = i + j + m + p + r$ et $0 < j + p < N$.
   — Nous choisissons $k = 2$.
   $uv^2wx^2y = a^{2^N+j+p}$. Comme $0 < j + p < N$ on déduit que $2^N < 2^N + j + p < 2^N + N < 2^{N+1}$, donc $uv^2wx^2y \notin L_4$.

   On a montré que L_4 ne satisfait pas le lemme de pompage pour les langages hors-contexte, il n'est donc pas hors-contexte.

Solution de l'exercice 3 :

1. On commence par le cas où w est de longueur impaire qui est trivial puisque L ne contient que des mots de longueur paire.
   Maintenant, soit $w \notin L$ de longueur paire : $w = uv$ avec $|u| = |v|$ et $u \neq v$. Il existe donc un plus petit indice i, $1 \leq i \leq |u|$, tel que la i-ème lettre de u diffère de la i-ème lettre de v. Sans perte de généralité on peut supposer que cette lettre soit un a dans u et un b dans v. On a donc $u = u_1au_2$ et $v = v_1bv_2$ avec $|u_2| = |v_2|$. Alors $w = uv = u_1au_2v_1bv_2$ et si on pose $u_1 = x$, $v_2 = z$ et $y = u_2v_1$ on a bien $|xz| = |y|$.
   Réciproquement, si $w = xaybz$ avec $|xz| = |y|$: on décompose $y = y_1y_2$ avec $|y_1| = |z|$ et $|y_2| = |x|$. On peut donc réécrire $w = xay_1y_2bz$ et on a $|xay_1| = |y_2bz|$ avec $xay_1 \neq y_2bz$ donc $w \notin L$. Analoguement si $w = xbyaz$.
2. On donne une grammaire pour chacun des deux langages.
   Pour L_a on donne $G_a = (\{T\}, \{a, b\}, R_a, T)$:

$$R_a : \left\{ \begin{array}{lcl} T & \to & a \mid aTb \mid bTa \mid aTa \mid bTb \end{array} \right\}$$

Pour L_b on donne $G_a = (\{U\}, \{a,b\}, R_b, U)$

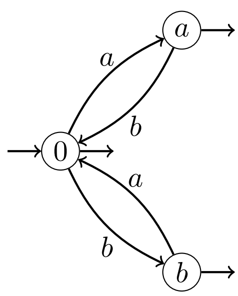
$$R_b : \left\{ \begin{array}{lcl} U & \rightarrow & b \mid aUb \mid bUa \mid aUa \mid bUb \end{array} \right\}$$

3. Si $w \notin L$ et $|w|$ paire : on a établi au point 1. que $w = xaybz$ ou $w = xbyaz$ avec $|xz| = |y|$. Si $w = xaybz$: on peut décomposer $w = xax'z'bz$ avec $|x'| = |x|$ et $|z'| = |z|$ donc $w \in L_aL_b$. Analoguement si $w = xbyaz$ on montre que $w \in L_bL_a$.

   Réciproquement, si $w \in L_aL_b$ on peut décomposer $w = xax'z'bz$ avec $|x'| = |x|$ et $|z'| = |z|$ donc $|x'z'| = |xz|$ et si on pose $y = x'z'$ on satisfait les hypothèses du point 1. et on peut conclure que $w \notin L$. La preuve est analogue si $w \in L_bL_a$.
4. Soit $w \notin L$: si $|w|$ est impair on a donné en 2. les grammaires pouvant le générer (suivant que la lettre centrale de w est un a ou un b). Si $|w|$ est paire, on construit une grammaire à partir des grammaires G_a et G_b, en rajoutant la variable initiale S et les règles $S \longrightarrow TU \mid UT$.

Solution de l'exercice 4 :

1. L'ensemble des mots super équilibrés est régulier, on donne ci-dessous un automate :



2. Soit u un mot équilibré non-vide. Deux cas :
   — soit u a un préfixe v non-vide et différent de u qui est équilibré : donc $u = vw$ avec v, w équilibrés et non-vides ;
   — ou ce n'est pas les cas, ce qui signifie que u est soit de la forme avb tel que la différence entre le nombre de a et de b dans tout préfixe de av est au moins 1 ; ou u est la forme bva tel que la différence entre le nombre de b et de a dans tout préfixe de bav est au moins 1. Dans les deux cas, v est équilibré (parce que u l'est).
3. D'après le point 2. on peut donner la grammaire $G = (S, \{a,b\}, R, S)$ avec $S \longrightarrow SS \mid aSb \mid bSa \mid \epsilon$.

Solution de l'exercice 5 :

On donne une grammaire pour chaque langage. Dans tous les cas on se contente de donner l'ensemble de règles (S étant la variable initiale).

1. On donne :

$$R : \left\{ \begin{array}{lcl} S & \rightarrow & TaT \\ T & \rightarrow & aTbT \mid bTaT \mid aT \mid \epsilon \end{array} \right.$$

   Le langage généré par T est $\{w \in \{a,b\}^* \mid |w|_a \geq |w|_b\}$. La règle initiale pour S assure qu'on a forcément un a en plus.

2. On construit la grammaire comme union de grammaires pour les langages $\{w \in \{a,b\}^* \mid |w|_a < |w|_b\}$ et $\{w \in \{a,b\}^* \mid |w|_a > |w|_b\}$.

$$R : \left\{ \begin{array}{lcl} S & \rightarrow & TaT \mid UbU \\ T & \rightarrow & aTbT \mid bTaT \mid aT \mid \epsilon \\ U & \rightarrow & aUbU \mid bUaU \mid bU \mid \epsilon \end{array} \right.$$

Solution de l'exercice 6 :

On commence par prouver que tout mot généré est dans $L(G)$. Soit w un mot généré par G et soit t son arbre de dérivation. On remarque que dans t chaque feuille étiquetée par b a un frère à sa gauche étiqueté par un a. Donc on peut associer de manière unique à chaque b dans w, un a qui est à sa gauche. On a donc bien $w \in L(G)$.

Réciproquement, soit $w \in L(G)$. On va montrer que w est générable par la grammaire. Si $w = \epsilon$ alors w est générable en utilisant la troisième règle. Sinon, on prouve par induction sur $|w|$ que $S \xrightarrow{*} wS$ (il suffit ensuite d'utiliser une fois la troisième règle pour conclure).

Si $|w| = 1$ alors $w = a$ (le mot b contient plus de b que de a). Donc on a $S \xrightarrow{*} aS$ par la première règle.

Supposons maintenant $|w| > 1$. On distingue deux cas : si il existe $w_1, w_2 \in L \setminus \{\epsilon\}$ tels que $w = w_1w_2$ alors par induction on sait que $S \xrightarrow{*} w_1S$ et $S \xrightarrow{*} w_2S$, en combinant les deux on obtient $S \xrightarrow{*} wS$.

Dans le cas contraire, on sait que pour tout couple $w_1, w_2 \neq \epsilon$ tel que $w = w_1w_2$, on a $w_2 \notin L$. On peut donc en déduire que $w = aw'b$ (la dernière lettre est forcément un b puisque ce n'est pas un mot de L). On montre que $w' \in L$. En effet si $w' = uv$, on sait par hypothèse sur w que $vb \notin L$ et contient donc un préfixe v' tel que $|v'|_b > |v'|_a$, or puisque $w \in L$, on a $|auv'|_b \leq |auv'|_a$. On en déduit que $|au|_b < |au|_a$ et en enlevant le a initial $|u|_b \leq |u|_a$. Autrement dit tout préfixe u de w' contient bien plus de a que de b, donc $w' \in L$. Par induction on sait donc que $S \xrightarrow{*} w'S$. On a donc :

$$S \rightarrow aSbS \xrightarrow{*} aw'SbS \rightarrow aw'bS = wS$$

Solution de l'exercice 7 :

Il suffit de transformer la grammaire en remplaçant tous ses membres droits de règles par leur image miroir. Formellement, on définit $G' = \{V', A, R', S'\}$ pour le langage miroir comme suit : on fixe $V = V'$ et $S = S'$. Pour toute règle $r : X \rightarrow a_1 \cdots a_k$ ($X \in V$ et $a_1, \ldots, a_k \in A \cup V$), on a $r \in R'$ si et seulement si la règle $X \rightarrow a_k \cdots a_1$ est dans R.

Solution de l'exercice 8 :

1. On donne :

$$R : \left\{ \begin{array}{lcl} S & \rightarrow & aSbS \mid bSaS \mid cS \mid \epsilon \end{array} \right.$$

2. On donne :

$$R : \left\{ \begin{array}{lcl} S & \rightarrow & T_1T_2 \mid U_1U_2 \\ T_1 & \rightarrow & \epsilon \mid aT_1b \\ T_2 & \rightarrow & \epsilon \mid cT_2 \\ U_1 & \rightarrow & \epsilon \mid aU_1 \\ U_2 & \rightarrow & \epsilon \mid bU_2c \end{array} \right.$$

La grammaire est construite comme union des grammaires pour les langages $\{a^p b^q c^r \mid p = q\}$ et $\{a^p b^q c^r \mid q = r\}$.

3. Commençons par écrire une grammaire pour le langage $\{v\#w \mid v^{\text{rev}} = w\}$:

$$R : \left\{ \begin{array}{lcl} S & \rightarrow & aSb \mid bSa \mid \# \end{array} \right\}$$

On généralise maintenant cette grammaire au langage qui nous intéresse $\{v\#w \mid v^{\text{rev}} \text{ est facteur de } w\}$:

$$R : \left\{ \begin{array}{lcl} S & \rightarrow & TU \\ T & \rightarrow & aTa \mid bTb \mid V \\ U & \rightarrow & aU \mid bU \mid \epsilon \\ V & \rightarrow & \# \mid Va \mid Vb \end{array} \right.$$