Option Informatique MP, DS n°1

I) Quelques exercices

Q1) Quels sont les langages reconnus par ces automates ? Donnez une expression régulière correspondante.

a)

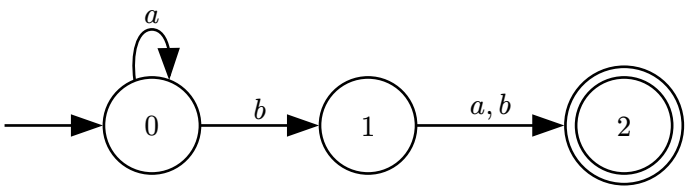


Figure 1: A_1

b)

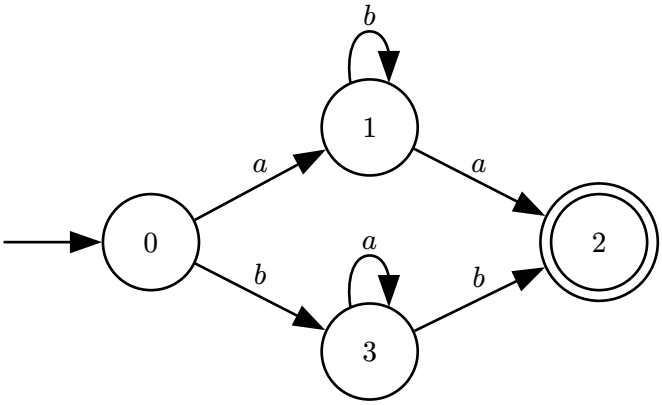


Figure 2: A_2

a) A_1 lit un nombre quelconque de a, puis un b, puis une lettre quelconque : $L(A_1) = a^\star b(a|b)$.

b) Depuis 0, un a mène en 1 où on lit des b avant de finir par un a ; un b mène en 3 où on lit des a avant de finir par un b :

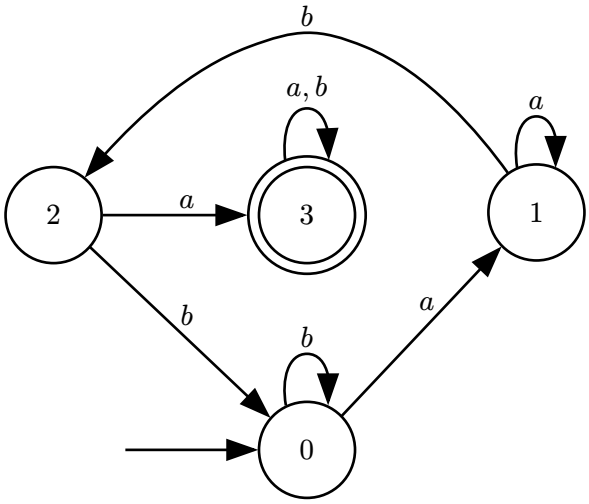
$$L(A_2) = ab^\star a \mid ba^\star b.$$

Q2) Donnez un automate déterministe correspondant aux langages suivants :

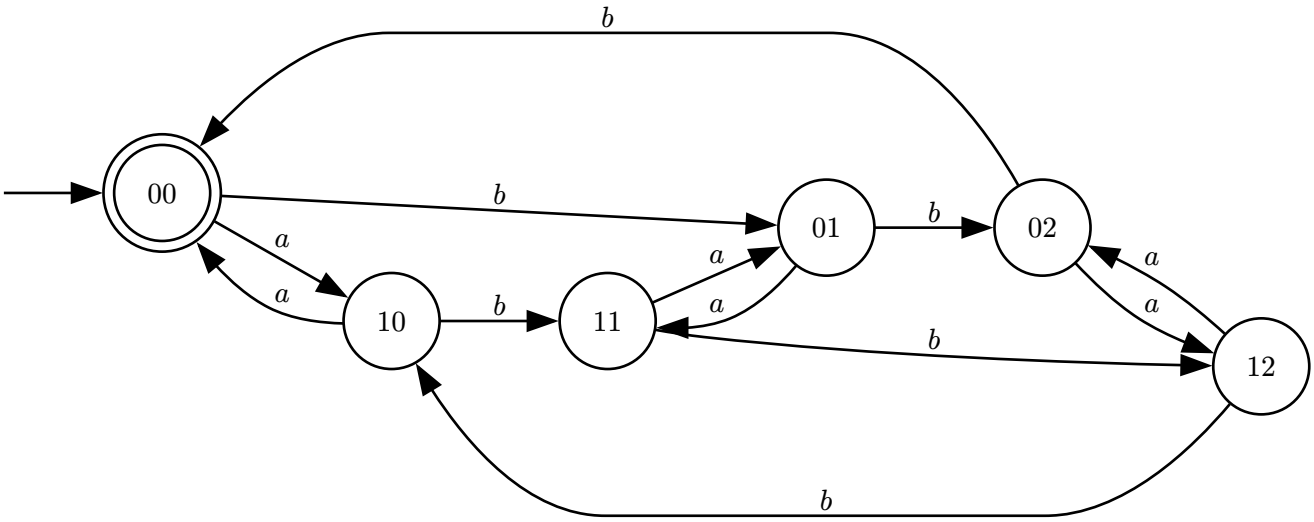
a) Le langage des mots contenant le facteur `aba`.

b) Le langage des mots avec un nombre pair de `a` et un nombre multiple de 3 de `b`.

a)



b) Automate : on retient $(|w|_a \bmod 2, |w|_b \bmod 3)$, soit 6 états (i, j) ; a envoie (i, j) sur $(i + 1 \bmod 2, j)$ et b sur $(i, j + 1 \bmod 3)$.



II) Racine carrée d'un langage (CCP 2019)

II.1) Définitions

Définition 1 [Carré d'un langage]

> Soit L un langage. On note L^2 l'ensemble $\{uu |\ u \in L\}$

Définition 2 [Racine carrée d'un langage]

> Soit L un langage. On note $\sqrt{L}$ l'ensemble $\{u \mid uu \in L\}$

Q3) Décrire $\sqrt{L}$ lorsque $\Sigma = \{a, b\}$ et L est décrit par l'expression rationnelle $a^\star b^\star$

Je note $\sim$ pour "est décrit par l'expression rationnelle".

$\sqrt{L} \sim a^\star \mid b^\star$

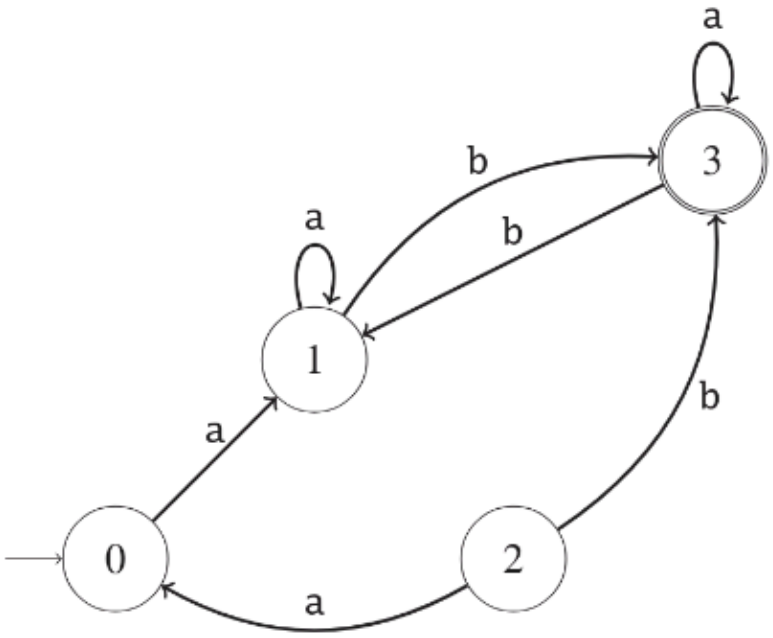
Q4) Décrire $\sqrt{L}$ lorsque $\Sigma = \{a, b\}$ et L est décrit par l'expression rationnelle $b^\star a^\star b^\star$

$\sqrt{L} \sim a^\star \mid b^\star$

Définition 3 [Automates]

> Soient $A = (Q, \Sigma, q_0, F, \delta)$ un automate fini déterministe, q' un élément de Q et F' une partie de Q. L'automate $(Q, \Sigma, q', F', \delta)$ est noté $A_{q',F'}$. Si on note L le langage reconnu par A, $L_{q',F'}$ désigne le langage reconnu par $A_{q',F'}$

Ici, L désigne le langage reconnu par l'automate A suivant :



Q5) Construire un automate reconnaissant $L_{3,\{1\}}$ en modifiant légèrement A.

- l'état initial est 3
- le seul état acceptant est 1

II.2) Propriétés de la racine carrée d'un langage

Ici, on fixe L un langage rationnel sur un alphabet Σ et $A = (Q, \Sigma, q_0, F, \delta)$ un automate fini reconnaissant celui-ci.

Q6) Soit $u \in \Sigma^\star$. Montrez que $u \in \sqrt{L}$ si et seulement s'il existe $q \in Q$ tel que $u \in L_{q_0,\{q\}}$ et $u \in L_{q,F}$.

($\Leftarrow$) Si $u \in L_{q_0,\{q\}} \cap L_{q,F}$, alors $\delta^\star(q_0, u) = q$ et $\delta^\star(q, u) \in F$.

Donc $\delta^\star(q_0, u^2) \in F$, et $u \in \sqrt{L}$.

($\Rightarrow$) Si $u \in \sqrt{L}$, alors $u^2 \in L$. Soit $q_0 \to q_1 \to \ldots \to q_{2|u|} \in F$ un chemin acceptant pour u^2.

Alors, avec $q = q_{|u|}$, $u \in L_{q_0,\{q\}}$, et $u \in L_{q,\{q_{2|u|}\}} \subset L_{q,F}$.

Q7) En déduire que $\sqrt{L}$ est un langage rationnel.

Par définition, les langages $L_{q',F'}$ sont rationnels.

Or, on a montré que $u \in \sqrt{L} \Leftrightarrow \exists q, u \in L_{q_0,\{q\}} \cap L_{q,F}$.

Donc $\sqrt{L} = \bigcup_{q \in Q} \left(L_{q_0,\{q\}} \cap L_{q,F} \right)$: c'est une union finie d'intersections de langages rationnels, donc un langage rationnel.

Q8) Montrer qu'on a $\sqrt{L}^2 \subset L$.

Soit $w \in \sqrt{L}^2$: il existe $u \in \sqrt{L}$ tel que $w = uu$. Par définition de $\sqrt{L}$, $uu \in L$, donc $w \in L$.

III) Automates et reconnaissance de motifs (CCP 2024)

Un alphabet A est un ensemble fini non vide d'éléments appelés lettres. Un mot défini sur A est une suite finie d'éléments de A. Le mot vide sera noté ε. L'ensemble des mots sur l'alphabet A est noté $A^\star$. La longueur d'un mot x, notée $|x|$, est le nombre de lettres de ce mot. Pour i allant de 0 à $|x| - 1$, x_i désignera la lettre en position i. Ainsi, aab est un mot de longueur 3, sur tout alphabet contenant les lettres a et b.

Un mot x est un *facteur* d'un mot y s'il existe 2 mots u et v tels que $y = uxv$. Dans ce cas, on dit qu'il y a une *occurrence* de x dans y. Quand $u = \varepsilon$, x est un *préfixe* de y et quand $v = \varepsilon$, x est un *suffixe* de y.

Soit x un mot non vide. Un entier $0 < p \leq |x|$ est une *période* de x si $x_i = x_{i+p}$ pour $0 \leq i \leq |x| - p - 1$. On définit *la période* de x comme étant la plus petite des périodes.

Un *bord* d'un mot non vide x est un facteur de x, différent de x et qui est à la fois un préfixe et un suffixe de la chaîne de caractères. Par exemple, ε, a, aa, $aabaa$ sont les bords du mot $aabaabaa$. On note $\text{Bord}(x)$ le plus long bord de x. On constate que lorsqu'il est défini, le bord d'un bord d'un mot x est aussi un bord de x. On le note $\text{Bord}^2(x)$. On définit ainsi récursivement $\text{Bord}^n(x)$ avec n un entier naturel.

Q9) Justifier le fait que tout mot non vide possède au moins une période.

Soit x non vide. Pour $p = |x|$, la condition porte sur les i tels que $0 \leq i \leq -1$: il n'y en a aucun, donc elle est vérifiée. Ainsi $|x|$ est une période de x.

Q10) Déterminer la période du mot $x = aabaabaa$ défini sur l'alphabet $A = \{a, b\}$.

- $p = 1$ n'est pas une période : $x_1 = a \neq b = x_2$.
- $p = 2$ n'est pas une période : $x_0 = a \neq b = x_2$.
- $p = 3$ est une période : $x = aab \cdot aab \cdot aa$, et on vérifie $x_i = x_{i+3}$ pour $0 \leq i \leq 4$.

La période de x est donc 3.

Q11) Soit x une chaîne de caractères non vide et p un entier naturel tel que $0 < p \leq |x|$. Montrer que p est une période de x si et seulement s'il existe 3 chaînes de caractères u, v, w telles que $x = uw = wv$ et $|u| = |v| = p$.

Notons $n = |x|$.

($\Rightarrow$) Supposons que p est une période. Posons $u = x_0 \ldots x_{p-1}$, $w = x_p \ldots x_{n-1}$ et $v = x_{n-p} \ldots x_{n-1}$. On a bien $x = uw$ et $|u| = |v| = p$. De plus, le préfixe de longueur $n - p$ de x est $x_0 \ldots x_{n-p-1} = x_p \ldots x_{n-1} = w$ car $x_i = x_{i+p}$ pour $0 \leq i \leq n - p - 1$. Donc $x = wv$.

($\Leftarrow$) Supposons $x = uw = wv$ avec $|u| = |v| = p$. Alors $|w| = n - p$, et w est à la fois le suffixe de longueur $n - p$ de x (car $x = uw$), c'est-à-dire $w_i = x_{i+p}$, et le préfixe de longueur $n - p$ (car $x = wv$), c'est-à-dire $w_i = x_i$. Donc pour $0 \leq i \leq n - p - 1$, $x_i = w_i = x_{i+p}$: p est une période.

Q12) Soit x une chaîne de caractères non vide et n le plus grand entier tel que $\text{Bord}^n(x)$ est bien défini. On a donc $\text{Bord}^n(x) = \varepsilon$. Montrer par récurrence sur la longueur des mots que $\text{Bord}(x), \text{Bord}^2(x), ..., \text{Bord}^n(x)$ est la suite des bords de x dans l'ordre décroissant de leur longueur et que $|x| - |\text{Bord}(x)|, |x| - |\text{Bord}^2(x)|, ..., |x| - |\text{Bord}^n(x)|$ est la suite des périodes de x dans l'ordre croissant.

Lemme. Soit y un bord de x et z un mot. Alors z est un bord de y si et seulement si z est un bord de x avec $|z| < |y|$.

En effet, si z est un préfixe de y, qui est un préfixe de x, alors z est un préfixe de x ; de même pour les suffixes, et $|z| < |y|$ car $z \neq y$. Réciproquement, si z est un préfixe de x avec $|z| < |y|$, comme y est aussi un préfixe de x, z est un préfixe de y ; de même pour les suffixes, et $z \neq y$.

Bords. Montrons par récurrence forte sur $|x| \geq 1$ que les bords de x sont exactement $\text{Bord}(x), ..., \text{Bord}^n(x)$, de longueurs strictement décroissantes.

- Si $|x| = 1$, le seul bord de x est $\varepsilon = \text{Bord}(x)$, et $n = 1$.
- Sinon, soit $y = \text{Bord}(x)$. Un bord de x est soit y, soit (par maximalité de y) de longueur $< |y|$, donc d'après le lemme un bord de y. Réciproquement, les bords de y sont des bords de x. Donc les bords de x sont y et les bords de y.
  - Si $y = \varepsilon$, alors $n = 1$ et y est le seul bord de x.
  - Sinon, $1 \leq |y| < |x|$, et par hypothèse de récurrence, les bords de y sont $\text{Bord}(y), ..., \text{Bord}^{n-1}(y)$, c'est-à-dire $\text{Bord}^2(x), ..., \text{Bord}^n(x)$, de longueurs strictement décroissantes, toutes $< |y|$.

Périodes. D'après la question précédente, $p \in [\![1, |x|]\!]$ est une période si et seulement s'il existe w de longueur $|x| - p$ qui est à la fois préfixe et suffixe de x, c'est-à-dire (comme $p > 0$, $w \neq x$) un bord de longueur $|x| - p$. L'application $p \mapsto |x| - p$ est donc une bijection décroissante des périodes de x vers les longueurs des bords de x. Les périodes de x sont donc $|x| - |\text{Bord}(x)| < |x| - |\text{Bord}^2(x)| < ... < |x| - |\text{Bord}^n(x)|$.

IV) Étude d'une famille d'automates (CCP 2023)

Dans cette partie, l'alphabet Σ désigne l'ensemble $\{0, 1\}$, le symbole ε désigne le mot vide et on rappelle que $\Sigma^\star$ désigne l'ensemble des mots sur l'alphabet Σ.

Étant donné un mot w, on rappelle que $|w|$ désigne la longueur du mot w et l'indexation des lettres de w commence par 0. La première lettre de w est donc w_0.

La notation $\text{Card}(E)$ désigne le cardinal d'un ensemble E.

Étant donnés un entier n et un entier non nul m, la notation $n \bmod m$ désigne le reste de la division euclidienne de n par m.

IV.1) Définitions

Définition 4 [Automate $A_{k,p}$, fonction indicatrice $L_{k,p}$]

Soient p et k deux entiers vérifiant $0 \leq k \leq p-1$. L'automate $A_{k,p}$ est défini par :

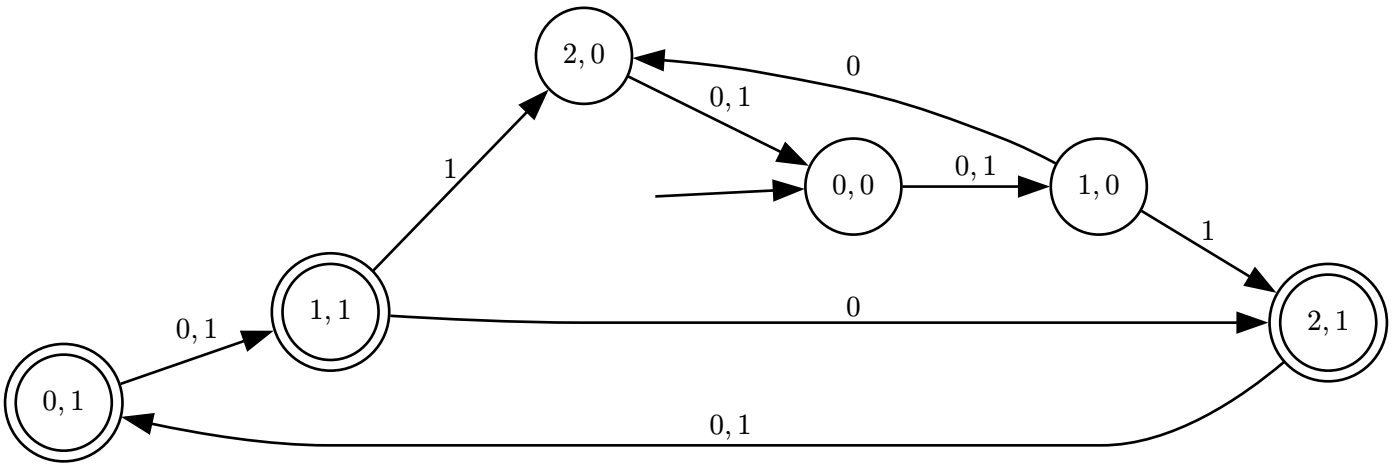
- $Q = \{0,1,...,p-1\} \times \{0,1\}$,
- $\Sigma = \{0,1\}$,
- $\forall (c,e) \in Q, \delta((c,e),0) = ((c+1) \bmod p, e)$,
- $\forall (c,e) \in Q, \delta((c,e),1) = \begin{cases} ((c+1) \bmod p, 1-e) & \text{si } c = k \bmod p \\ ((c+1) \bmod p, e) & \text{sinon.} \end{cases}$
- $q_0 = (0,0)$,
- $F = \{0,1,...,p-1\} \times \{1\}$.

On note $L_{k,p}$ la fonction indicatrice de l'ensemble des mots reconnus par $A_{k,p}$. Soit autrement :

$$\forall u \in \Sigma^\star, L_{k,p}(u) = \begin{cases} 1 \text{ si } A_{k,p} \text{ reconnaît } u \\ 0 \text{ sinon.} \end{cases}$$

IV.2) Exemples et propriétés élémentaires des $A_{k,p}$

Q13) Soit $w \in \Sigma^\star$. Expliciter sans démonstration l'état $\delta^\star(q_0, w)$, la lecture du mot étant effectuée dans l'automate $A_{1,3}$. On pourra s'aider d'une représentation graphique de l'automate.



La première composante compte les lettres lues modulo 3, et la seconde change à chaque 1 lu en position $i \equiv 1[3]$:

$$\delta^\star(q_0, w) = (|w| \bmod 3, \text{Card}(\{i \in [\![0, |w|-1]\!] \mid i \bmod 3 = 1 \text{ et } w_i = 1\}) \bmod 2).$$

Dans les questions suivantes, p et k désignent des entiers tels que $p > 2$ et $k \in \{0,1,...,p-1\}$.

Q14) Soit $w \in \Sigma^\star$. Expliciter l'état $\delta^\star(q_0, w)$, la lecture du mot étant effectuée dans l'automate $A_{k,p}$. On ne demande pas de démonstration.

$$\delta^\star(q_0, w) = \big(|w| \bmod p, \text{Card}\big(\{m \in \mathbb{N} \mid pm + k \leq |w| - 1 \text{ et } w_{pm+k} = 1\}\big) \bmod 2\big).$$

Un corollaire direct du résultat de la question précédente est que l'ensemble des mots reconnus par l'automate $A_{k,p}$ est égal à :

$$\big\{w \in \Sigma^\star \mid \text{Card}\big(\{m \in \mathbb{N} \mid pm + k \leq |w| - 1 \text{ et } w_{pm+k} = 1\}\big) \text{ est impair}\big\}.$$

Dans la suite du problème, on admet ce résultat.

Q15) Soit w un mot reconnu par un automate $A_{k,p}$. Montrer que w est reconnu par au moins un autre automate parmi $A_{0,2}$, $A_{1,2}$, $A_{l,p}$ avec $l \neq k$.

Pour $r \geq 1$ et $l \in [\![0, r-1]\!]$, notons $N_{l,r}(w)$ le nombre de 1 de w en une position $i \equiv l[r]$. Ainsi $L_{l,r}(w) = N_{l,r}(w) \bmod 2$. Notons aussi $N(w)$ le nombre total de 1 dans w. On a :

$$N(w) = \sum_{l=0}^{p-1} N_{l,p}(w) = N_{0,2}(w) + N_{1,2}(w).$$

Par hypothèse, $N_{k,p}(w)$ est impair. S'il existe $l \neq k$ tel que $N_{l,p}(w)$ est impair, w est reconnu par $A_{l,p}$. Sinon, $N(w)$ est une somme d'un nombre impair et de nombres pairs, donc est impair. Alors $N_{0,2}(w)$ ou $N_{1,2}(w)$ est impair, et w est reconnu par $A_{0,2}$ ou $A_{1,2}$.

Comme $p > 2$, ces automates sont bien différents de $A_{k,p}$.

Définition 5 [Ou exclusif étendu aux mots binaires]

On rappelle que le *Ou exclusif*, qu'on note $\oplus$, est une opération définie sur $\{0,1\}$ par :

$$0 \oplus 0 = 1 \oplus 1 = 0 \text{ et } 0 \oplus 1 = 1 \oplus 0 = 1.$$

Soient $n \in \mathbb{N}$, u et v deux éléments de $\{0,1\}^n$. On définit le *Ou exclusif* de u et v, noté $u \oplus v$, le mot de longueur n défini par :

$$\forall i \in \{0, 1, ..., n-1\}, (u \oplus v)_i = u_i \oplus v_i.$$

Q16) Soient u et v deux mots de $\Sigma^\star$ de même longueur. Montrer que :

$$L_{k,p}(u \oplus v) = L_{k,p}(u) \oplus L_{k,p}(v).$$

Notons $n = |u| = |v|$ et $I = \{i \in [\![0, n-1]\!] \mid i \equiv k[p]\}$. Sur $\{0,1\}$, $a \oplus b = (a+b) \bmod 2$. Donc :

$$L_{k,p}(u \oplus v) = \left(\sum_{i \in I} (u \oplus v)_i\right) \bmod 2 = \left(\sum_{i \in I} (u_i + v_i)\right) \bmod 2 = \left(\sum_{i \in I} u_i + \sum_{i \in I} v_i\right) \bmod 2$$

$$= L_{k,p}(u) \oplus L_{k,p}(v).$$

Q17) Soit w un mot binaire vérifiant :

$$L_{0,2}(w) = L_{1,2}(w) = 0 \text{ et } \forall k \in \{1, 2, ..., p-1\}, L_{k,p}(w) = 0.$$

1. Montrer que $L_{0,p}(w) = 0$.
2. En déduire que pour tout mot $w' \in 0^\star \cdot w$, on a :

$$L_{0,2}(w') = L_{1,2}(w') = 0 \text{ et } \forall k \in \{1, 2, ..., p-1\}, L_{k,p}(w') = 0.$$

1. Avec les notations précédentes, $N(w) = N_{0,2}(w) + N_{1,2}(w)$ est pair. Or $N(w) = N_{0,p}(w) + \sum_{k=1}^{p-1} N_{k,p}(w)$, et les $N_{k,p}(w)$ pour $k \geq 1$ sont pairs. Donc $N_{0,p}(w)$ est pair, soit $L_{0,p}(w) = 0$.
2. Soit $w' = 0^j w$. Les 1 de w' sont exactement en les positions $i + j$ pour i position d'un 1 de w. Donc, pour $r \in \{2, p\}$ et $l \in [\![0, r-1]\!]$, $N_{l,r}(w') = N_{l',r}(w)$ où $l' = (l-j) \bmod r$. Or, d'après la question précédente et l'hypothèse, tous les $N_{l',p}(w)$ et $N_{l',2}(w)$ sont pairs. Donc tous les $N_{l,r}(w')$ le sont, ce qui donne le résultat.

Q18) Montrer que pour tout $w \in \Sigma^\star$ et $w' \in w \cdot 0^\star$, on a $L_{k,p}(w) = L_{k,p}(w')$.

Soit $w' = w0^j$. Les positions $i < |w|$ de w' portent les mêmes lettres que dans w, et les positions $i \geq |w|$ portent des 0. Donc w et w' ont les mêmes 1 aux mêmes positions, et $N_{k,p}(w') = N_{k,p}(w)$, d'où $L_{k,p}(w) = L_{k,p}(w')$.

Remarque 6

Ces égalités permettent la construction d'une relation d'équivalence sur les mots qui est utilisée pour montrer que deux mots de longueur N peuvent être séparés par un automate de la forme $A_{k,p}$ ayant $O\left(\sqrt{N}\ln(N)\right)$ états.