TD 3 – Rationalité, minimisation et monoïdes

1 Rationalité

Exercice 1. Les langages suivants sont-ils rationnels ? Justifier.

1. $\{a^n b^n \mid n \in \mathbb{N}\}$.
2. $\{a^m b^n \mid n \equiv m \pmod d\}$ pour un $d \in \mathbb{N}_{>0}$ donné.
3. $\{a^p \mid p \text{ premier}\}$.
4. $\{a^{P(n)} \mid n \in \mathbb{N}\}$ pour P un polynôme à coefficients dans $\mathbb{N}$.
5. $\{w \in \{a,b,c\}^* \mid (|w|_a = 0) \Rightarrow (|w|_b = 0)\}$.
6. $\{w \in \{a,b\}^* \mid |w|_a < |w|_b\}$.
7. $\{w \in \{a,b\}^* \mid 7 \text{ divise } |w|_a, 3 \text{ divise } |w|_b\}$.
8. $\{w \in \{(,)\}^* \mid w \text{ est bien parenthésé}\}$.

Solution 1.

1. Non, car d'après le lemme de pompage pour m suffisamment grand, on aurait que $a^m b^m$ dans le langage se décomposerait comme $a^m b^m = uvw$ avec $v \neq \varepsilon$, $|uv| \leq m$ et $uv^k w$ dans le langage pour tout k. On aurait donc $v = a^d$ avec $d > 0$, ce qui impliquerait que $uw = a^{m-d} b^m$ est dans le langage.
2. Oui, c'est le langage $\mathcal{L}\left(\sum_{r=0}^{d-1} a^r (a^d)^* b^r (b^d)^*\right)$.
3. Non, car d'après le lemme de pompage pour p suffisamment grand, on aurait que a^p dans le langage se décomposerait comme $a^p = uvw$ avec $v \neq \varepsilon$ et $uv^x w$ dans le langage pour tout x. On aurait donc $v = a^d$ avec $d > 0$, ce qui impliquerait que $uv^{p+1} w = a^{p-d+(p+1)d} = a^{p(d+1)}$ est dans le langage, sachant que $p(d+1)$ n'est pas premier.
4. Si P est un polynôme de degré 0, de la forme $P(X) = c_0$, alors le langage est rationnel, puisqu'il s'agit de $\mathcal{L}(a^{c_0})$.

   Si P est un polynôme de degré 1, de la forme $P(X) = c_1 X + c_0$, alors le langage est également rationnel, puisqu'il s'agit de $\mathcal{L}((a^{c_1})^* a^{c_0})$.

   Si P est un polynôme de degré $k \geq 2$, de la forme $P(X) = c_k X^k + \cdots + c_1 X_1 + c_0$, alors le langage n'est pas rationnel. Sinon, d'après le lemme de pompage pour n suffisamment grand, on aurait que $a^{P(n)}$ dans le langage se décomposerait comme $a^{P(n)} = uvw$ avec $v \neq \varepsilon$, $|uv| \leq n$ et $uv^k w$ dans le langage pour tout k. On aurait donc $v = a^d$ avec $0 < d \leq n$, ce qui impliquerait que $uv^2 w = a^{P(n)+d}$ est dans le langage, sachant que $P(n+1) - P(n) > n \geq d$.
5. Oui, c'est le langage $\mathcal{L}(c^* + (b+c)^* a (a+b+c)^*)$.
6. Non, car d'après le lemme de pompage pour n suffisamment grand, on aurait que $a^n b^{n+1}$ dans le langage se décomposerait comme $a^n b^{n+1}$ avec $v \neq \varepsilon$, $|uv| \leq n$ et $uv^k w$ dans le langage pour tout k. On aurait donc $v = a^d$ avec $d > 0$, ce qui impliquerait que $uv^2 w = a^{n+d} b^{n+1}$ est dans le langage, sachant que $n + d \geq n + 1$.
7. Oui, c'est le langage $\mathcal{L}\big(b^*((ab^*)^7)^*\big) \cap \mathcal{L}\big(a^*((ba^*)^3)^*\big)$.
8. Non, car d'après le lemme de pompage pour n suffisamment grand, on aurait que $(^n)^n$ dans le langage se décomposerait comme $(^n)^n = uvw$ avec $v \neq \varepsilon$, $|uv| \leq n$ et $uv^k w$ dans le langage pour tout k. On aurait donc $v = (^d$ avec $d > 0$, ce qui impliquerait que $uw = (^{n-d})^n$ est dans le langage, sachant que $n - d < n$.

2 Non équivalence des lemmes de pompage

Soit les trois versions qui suivent du lemme de pompage.

1. Soit L un langage rationnel sur un alphabet Σ. Alors

$$\exists n \in \mathbb{N}_{>0}, \forall u \in L : |u| \geq n \Rightarrow \exists v, t, w \in \Sigma^* \quad u = vtw \quad |t| > 0 \quad \forall m \in \mathbb{N} \quad vt^m w \in L \ .$$

2. Soit L un langage rationnel sur un alphabet Σ. Alors

$$\exists n \in \mathbb{N}_{>0}, \forall rus \in L : |u| \geq n \Rightarrow \exists v, t, w \in \Sigma^* \quad u = vtw \quad |t| > 0 \quad \forall m \in \mathbb{N} \quad rvt^m ws \in L \ .$$

3. Soit L un langage rationnel sur un alphabet Σ. Alors

$$\exists n \in \mathbb{N}_{>0}, \forall ru_1 \cdots u_n s \in L : \\ (\forall i, |u_i| \geq 1) \Rightarrow \exists 1 \leq i < j \leq n \quad \forall m \in \mathbb{N} \quad ru_1 \cdots u_i (u_{i+1} \cdots u_j)^m u_{j+1} \cdots u_n s \in L \ .$$

Exercice 2. Montrer que $L = \{w \in \{a, b\}^* \mid |w|_a = |w|_b\}$ vérifie le lemme 1 mais pas le lemme 2.

Solution 2. Pour tout mot $u \in L$ tel que $|u| \geq 1$, on peut nécessairement décomposer u comme $u = vtw$ avec $t \in \{ab, ba\}$, qui est une décomposition vérifiant que $vt^m w \in L$ pour tout $m \in \mathbb{N}$. Donc L vérifie le lemme 1.

Supposons maintenant que L vérifie le lemme 2 pour un certain $n \in \mathbb{N}_{>0}$. On aurait donc que $rus = a^n b^n \in L$ pour $r = a^n$, $u = b^n$ et $s = \varepsilon$ vérifierait que u se décomposerait comme $u = vtw$ avec $|t| > 0$ et $vt^k w \in L$ pour tout k. On aurait donc $t = b^d$ avec $d > 0$, ce qui impliquerait que $a^n b^{n-d} = rvws \in L$, sachant que $n - d < n$. Donc L ne vérifie pas le lemme 2.

Exercice 3. Montrer que $L = \{(ab)^n (cd)^n \mid n \in \mathbb{N}\} \cup \mathcal{L}(\Sigma^*(aa + bb + cc + dd + ac + bd)\Sigma^*)$ vérifie le lemme 2 mais pas le lemme 3.

Solution 3. Pour tout mot $rus \in L$ tel que $|u| \geq 3$, deux cas peuvent survenir.

Si $rus \in \mathcal{L}(\Sigma^*(aa + bb + cc + dd + ac + bd)\Sigma^*)$, alors il existe $x \in \mathcal{L}(aa + bb + cc + dd + ac + bd)$ apparaissant en tant que facteur dans rus : puisque $|u| \geq 3$, on peut toujours décomposer u comme $u = vtw$ avec t contenant une seule lettre de manière à ce que, pour tout $m \in \mathbb{N}$, $rvt^m ws$ contienne toujours x comme facteur et, par voie de conséquence, appartienne à L.

Si $rus = (ab)^n (cd)^n$ pour un certain $n \in \mathbb{N}_{>0}$, alors, puisque $|u| \geq 3$, on peut nécessairement décomposer u comme $u = vtw$ avec t contenant une seule lettre, $|v| \geq 1$ et $|w| \geq 1$. Ceci est une décomposition vérifiant que $vt^m w \in L$ pour tout $m \in \mathbb{N}$, puisque comme $|rv| \geq 1$ et $|ws| \geq 1$, si $m = 0$, alors $rvws$ contient nécessairement aa, bb, ac, bd, cc ou dd comme facteur, et si $m \geq 2$, alors $rvt^m ws$ contient nécessairement aa, bb, cc ou dd comme facteur.

Donc L vérifie le lemme 2.

Supposons maintenant que L vérifie le lemme 3 pour un certain $n \in \mathbb{N}_{>0}$. On aurait donc que $ru_1 \cdots u_n s = (ab)^n (cd)^n$ pour $r = (ab)^n$, $u_1 = \cdots = u_n = cd$ et $s = \varepsilon$ vérifierait qu'il existe $1 \leq i < j \leq n$ tels que $ru_1 \cdots u_i (u_{i+1} \cdots u_j)^m u_{j+1} \cdots u_n s \in L$. On aurait donc que $(ab)^n (cd)^{n-j+i} = ru_1 \cdots u_i u_{j+1} \cdots u_n s \in L$, sachant que $n - j + i < n$.

Donc L ne vérifie pas le lemme 3.

3 Quotients, théorème de Myhill-Nerode

Soit Σ un alphabet. Étant donné un langage L sur Σ et $u \in \Sigma^*$ on définit :

— le *quotient à gauche de L par u*, noté $u^{-1}L$, comme étant le langage sur Σ tel que $u^{-1}L = \{w \in \Sigma^* \mid uw \in L\}$;

— le *quotient à droite de L par u*, noté Lu^{-1}, comme étant le langage sur Σ tel que $Lu^{-1} = \{w \in \Sigma^* \mid wu \in L\}$.

Exercice 4. Montrer qu'un langage L sur un alphabet Σ est rationnel si et seulement s'il a un nombre fini de quotients à gauche, c'est-à-dire que $\{u^{-1}L \mid u \in \Sigma^*\}$ est fini.

Solution 4. Soit un langage rationnel L sur un alphabet Σ. Il existe donc un AFD $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ tel que $\mathcal{L}(\mathcal{A}) = L$. Étant donné $u \in \Sigma^*$, on a que

$$u^{-1}L = \{w \in \Sigma^* \mid uw \in L\} = \{w \in \Sigma^* \mid \hat{\delta}(q_0, uw) \in F\} = \{w \in \Sigma^* \mid \hat{\delta}(\hat{\delta}(q_0, u), w) \in F\} \ .$$

On peut donc associer tout quotient gauche $u^{-1}L$ pour $u \in \Sigma^*$ à un état $q_u \in Q$ tel que pour tout $w \in \Sigma^*$, $w \in u^{-1}L \Leftrightarrow \hat{\delta}(q_u, w) \in F$. Par conséquent, si $q_u = q_v$ pour $u, v \in \Sigma^*$, on a nécessairement que $u^{-1}L = v^{-1}L$, ce qui implique que le nombre de quotients à gauche de L est fini par finitude de Q.

Soit maintenant un langage L sur un alphabet Σ ayant un nombre fini de quotients à gauche. Observons que pour tous $u, v \in \Sigma^*$, si $u^{-1}L = v^{-1}L$, alors $(ua)^{-1}L = (va)^{-1}L$ pour tout $a \in \Sigma$, puisque, pour chaque $w \in \Sigma^*$,

$$w \in (ua)^{-1}L \Leftrightarrow uaw \in L \Leftrightarrow (aw) \in u^{-1}L \Leftrightarrow (aw) \in v^{-1}L \Leftrightarrow vaw \in L \Leftrightarrow w \in (va)^{-1}L \ .$$

Construisons l'AFD $\mathcal{A}_L = (\{u^{-1}L \mid u \in \Sigma^*\}, \Sigma, \delta_L, L, \{u^{-1}L \mid u \in \Sigma^*, \varepsilon \in u^{-1}L\})$ où $\delta_L \colon \{u^{-1}L \mid u \in \Sigma^*\} \times \Sigma \to \{u^{-1}L \mid u \in \Sigma^*\}$ est telle que pour tout $u \in \Sigma^*$ et $a \in \Sigma$, $\delta_L(u^{-1}L, a) = (ua)^{-1}L$ (qui est bien définie en vertu de l'observation que nous venons de faire). On peut montrer, par récurrence sur la longueur de w, que $\hat{\delta_L}(L, w) = w^{-1}L$ pour tout $w \in \Sigma^*$, ce qui implique que quel que soit $w \in \Sigma^*$,

$$\hat{\delta_L}(L, w) \in \{u^{-1}L \mid u \in \Sigma^*, \varepsilon \in u^{-1}L\} \Leftrightarrow \varepsilon \in w^{-1}L \Leftrightarrow w \in L \ .$$

Par conséquent, $\mathcal{L}(\mathcal{A}) = L$, et donc L est rationnel.

Exercice 5. Montrer que pour tout langage rationnel L sur un alphabet Σ, tout AFD le reconnaissant contient au moins $\left|\{u^{-1}L \mid u \in \Sigma^*\}\right|$ états, que ce minorant est atteint et que tous les AFD avec ce nombre minimal d'états sont équivalents à « renommage » des états près (ce qui nous autorise à parler de « l'automate minimal de L »).

Solution 5. Soit L un langage rationnel sur un alphabet Σ.

Pour tout AFD $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ émondé (c'est-à-dire, dont chacun des états est accessible depuis l'état initial) reconnaissant L, on peut définir la fonction $f_{\mathcal{A}} \colon Q \to \{u^{-1}L \mid u \in \Sigma^*\}$ telle que $f_{\mathcal{A}}(q) = \{w \in \Sigma^* \mid \hat{\delta}(q, w) \in F\}$, qui est égal, pour tout $u \in \Sigma^*$ tel que $\hat{\delta}(q_0, u) = q$, à $u^{-1}L = \{w \in \Sigma^* \mid \hat{\delta}(q, w) \in F\}$. Quel que soit $q \in Q$, il existe au moins un tel $u \in \Sigma^*$, puisque $\mathcal{A}$ est émondé, donc $f_{\mathcal{A}}$ est bien définie. En outre, par construction, on a que $u^{-1}L = f_{\mathcal{A}}(\hat{\delta}(q_0, u))$ pour tout $u \in \Sigma^*$, ce qui montre que $f_{\mathcal{A}}$ est surjective. On en conclut donc, par finitude de Q, que $\left|\{u^{-1}L \mid u \in \Sigma^*\}\right| \leq |Q|$. Notons que, bien évidemment, cette inégalité est aussi vraie pour tout AFD non émondé reconnaissant L.

L'AFD $\mathcal{A}_L = (\{u^{-1}L \mid u \in \Sigma^*\}, \Sigma, \delta_L, L, \{u^{-1}L \mid u \in \Sigma^*, \varepsilon \in u^{-1}L\})$ donné par la construction de l'exercice précédent nous permet de montrer que ce minorant est atteint.

Soit maintenant un AFD $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ avec le nombre minimal d'états $\left|\{u^{-1}L \mid u \in \Sigma^*\}\right|$. Sachant qu'il est nécessairement émondé, on a que la fonction $f_{\mathcal{A}} \colon Q \to \{u^{-1}L \mid u \in \Sigma^*\}$ définie plus haut est donc bijective, en plus de vérifier que :

— $f_{\mathcal{A}}(q_0) = L$;
— $f_{\mathcal{A}}(\delta(q, a)) = \delta_L(f_{\mathcal{A}}(q), a)$ pour tout $a \in \Sigma$;
— $f_{\mathcal{A}}(F) = \{u^{-1}L \mid u \in \Sigma^*, \varepsilon \in u^{-1}L\}$.

$\mathcal{A}$ est donc équivalent à $\mathcal{A}_L$ à « renommage » des états près (donné par la fonction $f_{\mathcal{A}}$).

4 Algorithme de Brzozowski

Soit Σ un alphabet.

Étant donné un mot $w \in \Sigma^*$, on définit le *mot renversé de* w, noté $w^{\mathcal{R}}$, comme étant $w^{\mathcal{R}} = \varepsilon$ si $w = \varepsilon$ et $w^{\mathcal{R}} = a_n \cdots a_1$ si $w = a_1 \cdots a_n$ pour $n \in \mathbb{N}_{>0}$.

Étant donné un langage L sur Σ, le *renversé de* L, noté $L^{\mathcal{R}}$, est simplement le langage $L^{\mathcal{R}} = \{w^{\mathcal{R}} \mid w \in L\}$.

Pour tout AFD $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$, on définit l'AFN $\mathcal{A}^{\mathcal{R}} = (Q, \Sigma, \delta^{\mathcal{R}}, F, \{q_0\})$ où $\delta^{\mathcal{R}} \colon Q \times \Sigma \to \mathfrak{P}(Q)$ est telle que $\delta^{\mathcal{R}}(q, a) = \{p \in Q \mid \delta(p, a) = q\}$ pour tous $q \in Q$ et $a \in \Sigma$. (Précisons que, dans cette section, nous autoriserons les AFN à avoir un ensemble non vide d'états initiaux au lieu d'uniquement un seul et que nous les interdirons d'avoir des ε-transitions).

On dira enfin qu'un AFD est *émondé* lorsque chacun de ses états est accessible depuis l'état initial.

Exercice 6. Soit L un langage sur Σ et $\mathcal{A}$ un AFD émondé le reconnaissant. Montrer que $\mathcal{A}^{\mathcal{R}}$ est un AFN reconnaissant $L^{\mathcal{R}}$.

Solution 6. Posons $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$.

Il est direct de voir que pour tous $n \in \mathbb{N}_{>0}$, $p_0, p_1, \ldots, p_n \in Q$ et $a_1, \ldots, a_n \in \Sigma$, on a que $\delta(p_{i-1}, a_i) = p_i$ pour tous $i \in [\![1, n]\!]$ si et seulement si $p_{i-1} \in \delta^{\mathcal{R}}(p_i, a_i)$ pour tous $i \in [\![1, n]\!]$.

Par conséquent, pour tous $q, p \in Q$ et $w \in \Sigma^*$, on a que $\hat{\delta}(q, w) = p$ si et seulement si $q \in \hat{\delta^{\mathcal{R}}}(p, w^{\mathcal{R}})$. Cela permet de conclure que $\mathcal{A}^{\mathcal{R}}$ reconnaît $L^{\mathcal{R}}$.

Pour tout AFN $\mathcal{B} = (Q, \Sigma, \delta, S, F)$, on notera $\mathcal{B}^{\mathcal{D}}$ l'AFD obtenu de $\mathcal{B}$ par la construction par sous-ensembles, c'est-à-dire la version émondée de l'AFD $(\mathfrak{P}(Q), \Sigma, \delta', S, \{P \subseteq Q \mid P \cap F \neq \emptyset\})$, où δ' est telle que $\delta'(P, a) = \bigcup_{q \in P} \delta(q, a)$ pour tous $P \subseteq Q$ et $a \in \Sigma$.

Exercice 7. Soit L un langage sur Σ et $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ un AFD émondé le reconnaissant. Considérons l'AFD $(\mathcal{A}^{\mathcal{R}})^{\mathcal{D}} = (Q', \Sigma, \delta', q_0', F')$, qui reconnaît $L^{\mathcal{R}}$. Montrer que pour tous $u, v \in \Sigma^*$, $Lu^{-1} = Lv^{-1}$ implique que $\hat{\delta'}(q_0', u^{\mathcal{R}}) = \hat{\delta'}(q_0', v^{\mathcal{R}})$.

Solution 7. Posons $A^{\mathcal{R}} = (Q, \Sigma, \delta^{\mathcal{R}}, F, \{q_0\})$. On a alors que $q_0' = F$.

Soit $u, v \in \Sigma^*$ tels que $Lu^{-1} = Lv^{-1}$. Puisque $\hat{\delta'}(q_0', u^{\mathcal{R}})$ et $\hat{\delta'}(q_0', v^{\mathcal{R}})$ sont des sous-ensembles de Q, on va montrer que tout élément dans l'un est aussi dans l'autre et vice versa.

Soit donc $q \in \hat{\delta'}(q_0', u^{\mathcal{R}})$. Puisque $q_0' = F$, cela veut donc dire qu'il existe $p \in F$ tel que $q \in \hat{\delta^{\mathcal{R}}}(p, u^{\mathcal{R}})$. Par ce qui a été montré dans l'exercice précédent, il s'ensuit que $\hat{\delta}(q, u) = p$. Or, $\mathcal{A}$ étant émondé, il existe nécessairement $w \in \Sigma^*$ tel que $\hat{\delta}(q_0, w) = q$, qui vérifie donc que $\hat{\delta}(q_0, wu) = p \in F$. Cela implique que $wu \in L$, autrement dit que $w \in Lu^{-1}$. Mais puisque $Lu^{-1} = Lv^{-1}$, on a $wv \in L$ et il doit donc exister $p' \in F$ tel que $p' = \hat{\delta}(q_0, wv) = \hat{\delta}(\hat{\delta}(q_0, w), v) = \hat{\delta}(q, v)$. Par ce qui a été montré dans l'exercice précédent, il s'ensuit que $q \in \hat{\delta^{\mathcal{R}}}(p', v^{\mathcal{R}})$, et on peut donc conclure que $q \in \hat{\delta'}(q_0, v^{\mathcal{R}})$, puisque $p' \in F = q_0'$.

On montre de manière symétrique que tout q dans $\hat{\delta'}(q_0', v^{\mathcal{R}})$ appartient aussi à $\hat{\delta'}(q_0', u^{\mathcal{R}})$.

Exercice 8. En déduire que pour tout langage L sur Σ et AFD émondé $\mathcal{A}$ le reconnaissant, $(\mathcal{A}^{\mathcal{R}})^{\mathcal{D}}$ est l'automate minimal reconnaissant $L^{\mathcal{R}}$.

Solution 8. Posons $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ et $(\mathcal{A}^{\mathcal{R}})^{\mathcal{D}} = (Q', \Sigma, \delta', q_0', F')$.

Soit $u, v \in \Sigma^*$ tels que $u^{-1}L^{\mathcal{R}} = v^{-1}L^{\mathcal{R}}$. On observe que pour tout $w \in \Sigma^*$, on a

$$wu^{\mathcal{R}} \in L \Leftrightarrow uw^{\mathcal{R}} \in L^{\mathcal{R}} \Leftrightarrow w^{\mathcal{R}} \in u^{-1}L^{\mathcal{R}} \Leftrightarrow w^{\mathcal{R}} \in v^{-1}L^{\mathcal{R}} \Leftrightarrow vw^{\mathcal{R}} \in L^{\mathcal{R}} \Leftrightarrow wv^{\mathcal{R}} \in L \ ,$$

d'où il s'ensuit que $L(u^{\mathcal{R}})^{-1} = L(v^{\mathcal{R}})^{-1}$. Par ce qui a été démontré dans l'exercice précédent, cela implique donc que $\hat{\delta'}(q_0', u) = \hat{\delta'}(q_0', v)$.

Il s'ensuit que $|Q| \leq \big|\{u^{-1}L^{\mathcal{R}} \mid u \in \Sigma^*\}\big|$, et que donc, par le résultat de la section précédente, l'AFD $(\mathcal{A}^{\mathcal{R}})^{\mathcal{D}}$ est minimal.

Exercice 9. Proposer un algorithme de minimisation se basant sur le résultat précédent et en discuter la complexité.

Solution 9. Étant donné un AFD $\mathcal{A}$, il suffit de calculer $(((\mathcal{A}^{\mathcal{R}})^{\mathcal{D}})^{\mathcal{R}})^{\mathcal{D}}$.

Considérons pour tout $n \in \mathbb{N}_{>0}$ le langage $L_n = \mathcal{L}\big((a+b)^{n-1}a(a+b)^*\big)$: bien que l'AFD minimal de L_n ait $n+1$ états, on sait que l'AFD minimal de $L_n^{\mathcal{R}} = \mathcal{L}\big((a+b)^*a(a+b)^{n-1}\big)$ a 2^n états. Il peut donc falloir un temps exponentiel en le nombre d'états de l'AFD donné en entrée à cet algorithme pour effectuer la minimisation.

5 Reconnaissance par monoïdes

Un *monoïde* est la donnée $(M, \star)$ d'un ensemble M et d'une loi de composition interne $\star\colon M \times M \to M$ tels que

— $\star$ est associative : $\forall x, y, z \in M, (x \star y) \star z = x \star (y \star z)$;

— $\star$ a un élément neutre : $\exists e \in M, \forall m \in M, e \star m = m \star e = m$.

Il est classique de montrer que cet élément neutre est unique, et on le notera donc $1_{(M,\star)}$.

Si Σ est un alphabet, l'ensemble de tous les mots Σ^* muni de la concaténation $\cdot$ forme un monoïde $(\Sigma^*, \cdot)$ dont l'élément neutre est le mot vide, appelé le *monoïde libre engendré par* Σ.

Étant donné deux monoïdes $(M, \star)$ et $(N, \perp)$, un *morphisme de* $(M, \star)$ *dans* $(N, \perp)$ est une application $\varphi\colon M \to N$ telle que :

— $\forall m_1, m_2 \in M, \varphi(m_1) \perp \varphi(m_2) = \varphi(m_1 \star m_2)$;

— $\varphi(1_{(M,\star)}) = 1_{(N,\perp)}$.

Si Σ est un alphabet, on dira qu'un monoïde $(M, \star)$ *reconnaît* un langage L sur Σ si et seulement s'il existe un morphisme $\varphi\colon \Sigma^* \to M$ de $(\Sigma^*, \cdot)$ dans $(M, \star)$ tel que $L = \varphi^{-1}(P)$ pour un certain $P \subseteq M$.

Exercice 10. Montrer que si un langage est reconnu par un monoïde fini, alors il est rationnel.

Solution 10. Soit L un langage sur un alphabet Σ et $(M, \star)$ un monoïde fini le reconnaissant. Il existe donc un morphisme $\varphi\colon \Sigma^* \to M$ de $(\Sigma^*, \cdot)$ dans $(M, \star)$ tel que $\varphi^{-1}(P) = L$ pour un certain $P \subseteq M$.

Construisons l'AFD $\mathcal{A} = (M, \Sigma, \delta, 1_{(M,\star)}, P)$ où $\delta\colon M \times \Sigma \to M$ est tel que $\delta(m, a) = m \star \varphi(a)$ pour tous $m \in M$ et $a \in \Sigma$. On peut montrer que pour tout $w \in \Sigma^*$, $\hat{\delta}(1_{(M,\star)}, w) = \varphi(w)$, et que donc, pour tout $w \in \Sigma^*$,

$$w \in \mathcal{L}(\mathcal{A}) \Leftrightarrow \hat{\delta}(1_{(M,\star)}, w) \in P \Leftrightarrow \varphi(w) \in P \Leftrightarrow w \in \varphi^{-1}(P) = L \ .$$

Par conséquent, $\mathcal{A}$ reconnaît L, et on en conclut donc que L est rationnel.

Exercice 11. Montrer que si un langage est rationnel, alors il est reconnu par un monoïde fini.

Solution 11. Soit L un langage rationnel sur un alphabet Σ. Il existe donc un AFD $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ reconnaissant L.

Considérons le monoïde fini $(Q^Q, \circ)$, où $\circ$ dénote la composition des fonctions de Q dans Q (en les appliquant cela dit de la gauche vers la droite). Comme $(\Sigma^*, \cdot)$ est précisément le monoïde libre engendré par Σ, on peut définir un morphisme $\varphi\colon \Sigma^* \to Q^Q$ comme étant l'unique morphisme de $(\Sigma^*, \cdot)$ dans $(Q^Q, \circ)$ tel que, pour tout $a \in \Sigma$, $\varphi(a) = f_a$ où $f_a\colon Q \to Q$ vérifie que $f_a(q) = \delta(q, a)$ pour tout $q \in Q$. Posons également $P = \{f \in Q^Q \mid f(q_0) \in F\}$. On peut montrer que pour tout $w \in \Sigma^*$, $\varphi(w) = f$ où $f\colon Q \to Q$ vérifie que $f(q) = \hat{\delta}(q, w)$ pour tout $q \in Q$ et que donc, pour tout $w \in \Sigma^*$,

$$w \in L \Leftrightarrow \hat{\delta}(q_0, w) \in F \Leftrightarrow (\varphi(w))(q_0) \in F \Leftrightarrow \varphi(w) \in P \Leftrightarrow w \in \varphi^{-1}(P) \ .$$

On en conclut donc que le monoïde fini $(Q^Q, \circ)$ reconnaît L.

Exercice 12. Montrer, en utilisant ce résultat, que le langage $\{w \in \{a, b\}^* \mid |w|_a = |w|_b\}$ n'est pas rationnel. Donner un monoïde permettant de le reconnaître.

Solution 12. Posons $L = \{w \in \{a, b\}^* \mid |w|_a = |w|_b\}$.

Soit $(M, \star)$ un monoïde reconnaissant L. Il existe donc un morphisme $\varphi\colon \Sigma^* \to M$ de $(\Sigma^*, \cdot)$ dans $(M, \star)$ tel que $\varphi^{-1}(P) = L$ pour un certain $P \subseteq M$.

Pour tous $n, m \in \mathbb{N}, n \neq m$, on a que $a^n b^n \in L$ tandis que $a^m b^n \notin L$, ce qui implique que $\varphi(a^n b^n) \in P$ tandis que $\varphi(a^m b^n) \notin P$, et donc, par voie de conséquence, que nécessairement $\varphi(a^n) \neq \varphi(a^m)$. Il s'ensuit que l'ensemble $\{\varphi(a^n) \mid n \in \mathbb{N}\} \subseteq M$ est infini, et que donc le monoïde $(M, \star)$ ne peut être fini.

Par ce qui a été montré plus haut, L n'est donc pas rationnel.

Un monoïde le reconnaissant est $(\mathbb{Z}, +)$, en utilisant l'unique morphisme $\varphi\colon \{a, b\}^* \to \mathbb{Z}$ de $(\{a, b\}^*, \cdot)$ dans $(\mathbb{Z}, +)$ tel que $\varphi(a) = 1$ et $\varphi(b) = -1$, qui vérifie que $L = \varphi^{-1}(\{0\})$.