Faculté des Sciences

Conservation du caractère reconnaissable par opérations arithmétiques dans un système de numération abstrait

Mémoire présenté par
Émilie CHARLIER
en vue de l'obtention du Diplôme
d'Études Approfondies en Sciences

Année académique 2005-2006

Remerciements

Je tiens à remercier vivement Michel Rigo pour son attention à mon travail et sa toujours grande disponibilité.

J'adresse également de sincères remerciements à Pierre Lecomte pour les discussions intéressantes que nous avons eues relatives au sujet de ce travail.

Enfin, je remercie chaleureusement Georges Hansoul, Rémi Lambert et Ludovic Theate pour leur aide précieuse.

Table des matières

Introduction **v**

1 Notions de base et notations **1**
1.1 Langages et automates 1
1.2 Systèmes de numération abstraits 3

2 Langages réguliers polynomiaux **7**
2.1 Caractérisation 7
2.2 Théorèmes de séparation 12
2.3 Constructions 13
2.3.1 Langages de complexité n^k 13
2.3.2 Polynômes à coefficients dans $\mathbb{N}$ 15
2.3.3 Polynômes à coefficients dans $\mathbb{Z}$ 15
2.3.4 Polynômes à coefficients dans $\mathbb{Q}$ 16
2.3.5 Fonctions exponentielles polynômes 17
2.4 Propriété de la suite $v_n(L)$ 19

3 Multiplication par une constante **25**
3.1 Cas du langage a^*b^* 25
3.2 Cas des langages polynomiaux 32
3.2.1 Langages exactement polynomiaux 32
3.2.2 Langages polynomiaux quelconques 37
3.3 Cas des langages à complémentaire polynomial 40
3.4 Quelques remarques 42
3.4.1 Changement de l'ordre sur l'alphabet 42
3.4.2 Cas des langages *slender* 45

4 Généralisation **47**
4.1 $\mathcal{B}_\ell$-représentation d'un entier 48
4.2 Multiplication par $\lambda = \beta^\ell$ 50
4.3 Sous-ensembles réguliers de $\mathcal{B}_\ell$ 57
4.4 Conclusions 59

Bibliographie **61**

Introduction

Pour pouvoir manipuler et utiliser les nombres (qu'ils soient entiers ou réels), il faut pouvoir les représenter. Un système de numération n'est rien d'autre qu'un ensemble de règles à appliquer pour écrire tout nombre comme une suite de symboles appartenant à un alphabet déterminé de chiffres. On pourra par exemple considérer le système décimal bien connu de tous dans lequel tout nombre entier s'écrit de manière unique comme un mot fini sur l'alphabet $\{0, 1, ..., 9\}$.

Les systèmes de numération à base entière ont été généralisés de diverses façons, par exemple en considérant des systèmes construits sur une suite strictement croissante d'entiers satisfaisant une relation de récurrence linéaire. Dans ce contexte, une question naturelle est de déterminer quels ensembles d'entiers donnent lieu à des représentations particulièrement simples, c'est-à-dire, donnant lieu à des ensembles de représentations réguliers.

En particulier, on peut se poser la question de déterminer quelles suites d'entiers donnent lieu à des systèmes de numération pour lesquels l'ensemble des représentations de $\mathbb{N}$ tout entier donne un langage régulier. En effet, dans un tel cas de figure, déterminer si un mot est ou non une représentation valide peut être testé très simplement au moyen d'un automate fini. Ce problème a été étudié par J. Shallit dans [10].

Dans [6], P. Lecomte et M. Rigo ont choisi de contourner le problème en imposant que le langage de la numération soit régulier. Ils ont ainsi introduit la notion de système de numération abstrait. Un tel système est défini par un langage régulier infini L sur un alphabet totalement ordonné $(\Sigma, <)$. Ordonner les mots de L par ordre généalogique croissant fournit une bijection entre $\mathbb{N}$ et L, l'entier n étant alors représenté par le $(n+1)$-ième mot de L. Dans ce travail, nous considérons de tels systèmes de numération et nous nous intéressons aux parties de $\mathbb{N}$ correspondant à un ensemble de représentations régulier. On parlera alors de parties reconnaissables.

On peut noter que les cas classiques comme les numérations en base entière ou encore le système de Fibonacci sont des cas particuliers de systèmes abstraits. D'une manière plus générale, tout système de numération construit sur une suite linéaire récurrente dont le polynôme caractéristique est le polynôme minimum d'un nombre de Pisot ([1]) est un cas particulier

des systèmes de numération abstraits.

Pour les numérations basées sur un nombre de Pisot, les opérations arithmétiques élémentaires comme l'addition ou la multiplication par une constante préservent le caractère reconnaissable. Dans le cadre général présenté ici, la situation est plus complexe. La multiplication par une constante ne préserve pas nécessairement le caractère reconnaissable. Dans ce travail, on étudie principalement les systèmes de numération abstraits construits sur un langage régulier polynomial (i.e., dont la fonction de complexité comptant le nombre de mots de longueur n est bornée par un polynôme). Ainsi, un chapitre entier de ce mémoire est dédié à la structure des langages polynomiaux. Nous y proposons notamment une caractérisation de ceux-ci en termes d'automates et une construction effective d'un système de numération abstrait reconnaissant l'image de $\mathbb{N}$ par un polynôme à coefficients rationnels donné, par lequel l'image de tout entier est entière.

L'un des résultats principaux présentés dans ce travail concerne la numération construite sur le langage a^*b^*. On montre que, dans ce système, le caractère reconnaissable est préservé après multiplication par une constante λ si et seulement si λ est un carré impair. En particulier, nous donnons dans le dernier chapitre des indications sur une possible extension de ce résultat à un alphabet de plus de deux lettres.

Le cas de a^*b^* peut sembler anecdotique. En fait, pour tout langage polynomial dont la fonction de complexité est d'ordre n^k, en utilisant les propriétés des langages polynomiaux obtenues dans le chapitre consacré à ceux-ci, on peut montrer que les seuls multiplicateurs susceptibles de préserver le caractère reconnaissable sont des puissances $(k+1)$-ièmes d'entiers.

Chapitre 1

Notions de base et notations

Nous rappelons pour commencer les définitions et notions de base de la théorie des langages formels et des automates. Pour plus de détails à ce sujet, on peut par exemple consulter [3] ou [9]. Dans un deuxième temps, nous introduisons la notion de système de numération abstrait et nous en donnons quelques-unes des premières propriétés.

1.1 Langages et automates

Définition 1.1.1. Un *alphabet* Σ est un ensemble fini de symboles, généralement appelés *lettres* de cet alphabet. Un *mot* sur Σ est la concaténation d'un nombre fini de lettres de Σ. Le *mot vide* ε est le mot qui n'a aucune lettre. La longueur d'un mot w, notée $|w|$, est le nombre de lettres de w. L'ensemble des mots sur Σ est noté Σ^*. Muni de l'opération de concaténation, Σ^* a une structure de monoïde avec ε pour neutre. Un *langage* sur Σ est une partie de Σ^*.

Lorsque $\Sigma = \{\sigma\}$ est un alphabet unaire, on note σ^* plutôt que $\{\sigma\}^*$.

Définition 1.1.2. Si u et v sont deux mots sur Σ, le *shuffle* de u et v est le langage $u \sqcup\!\sqcup v = \{u_1 v_1 \cdots u_n v_n \mid u = u_1 \cdots u_n,\ v = v_1 \cdots v_n\}$. Si L et M sont deux langages sur Σ, le *shuffle* de L et M est l'union $L \sqcup\!\sqcup M$ des langages $u \sqcup\!\sqcup v$, avec $u \in L$ et $v \in M$.

Définition 1.1.3. Si w est un mot sur Σ, le *miroir* de w est le mot w^R défini récursivement par $\varepsilon^R = \varepsilon$ et $\sigma v = v^R \sigma$, pour tout $\sigma \in \Sigma$ et tout $v \in \Sigma^*$. Si L est un langage sur Σ, le *miroir* de L est le langage $L^R = \{w^R \mid w \in L\}$.

Définition 1.1.4. Si L est un langage sur Σ et w est un mot sur Σ, on note $w^{-1}.L = \{u \in \Sigma^* \mid wu \in L\}$ le langage des mots sur Σ qui, concaténés avec w, appartiennent à L.

Définition 1.1.5. Pour tout langage L et pour tout entier $n \geq 0$, on désigne

par $u_n(L)$ le nombre de mots de longueur n de L et par $v_n(L)$ le nombre de mots de longueur inférieure ou égale à n de L.

Définition 1.1.6. Un *automate fini déterministe* (AFD) est la donnée

$$\mathcal{A} = (Q, \Sigma, \delta, s, F)$$

d'un ensemble fini d'*états* Q, d'un alphabet Σ, d'une *fonction de transition* $\delta : Q \times \Sigma \to Q$, d'un état privilégié s de Q appelé *état initial* et d'un ensemble $F \subseteq Q$ d'*états finaux.* Si la fonction de transition est totale, l'automate $\mathcal{A}$ est dit *complet.*

Dans la suite de ce texte, sauf mention explicite du contraire, nous supposerons toujours travailler avec un AFD complet. La définition de la fonction de transition δ s'étend naturellement à $Q \times \Sigma^*$ par récurrence de la manière suivante : $\delta(q, \varepsilon) = q$ et $\delta(q, \sigma w) = \delta(\delta(q, \sigma), w)$, pour tous $q \in Q$, $\sigma \in \Sigma$ et $w \in \Sigma^*$.

Définition 1.1.7. Soit un AFD $\mathcal{A} = (Q, \Sigma, \delta, s, F)$. Un mot w sur Σ^* est *accepté* par $\mathcal{A}$ si $\delta(s, w) \in F$. Le langage accepté par $\mathcal{A}$ est l'ensemble des mots acceptés par $\mathcal{A}$ et est noté $L(\mathcal{A})$. Le *langage accepté à partir de l'état* q *de* Q est l'ensemble des mots acceptés $L_q(\mathcal{A}) = L(\mathcal{A}_q)$ par l'automate $\mathcal{A}_q = (Q, \Sigma, \delta, q, F)$, où on a remplacé l'état initial de $\mathcal{A}$ par q.

Si l'automate $\mathcal{A}$ est clairement défini par le contexte, on écrira simplement L_q au lieu de $L_q(\mathcal{A})$. De même, on notera $u_n(q)$ le nombre de mots de longueur n de L_q et par $v_n(q)$ le nombre de mots de longueur inférieure ou égale à n de L_q.

Définition 1.1.8. Soit un AFD $\mathcal{A} = (Q, \Sigma, \delta, s, F)$. Un état q de $\mathcal{A}$ est *accessible* si il peut être atteint à partir de l'état initial, en respectant la fonction de transition, i.e., si il existe un mot w sur Σ tel que $\delta(s, w) = q$. L'automate $\mathcal{A}$ est *accessible* si tous ses états sont accessibles. Il est *réduit* si les langages acceptés à partir d'états distincts sont distincts.

Parmi les automates acceptant un langage, on distingue souvent l'automate minimal de ce langage.

Définition 1.1.9. L'*automate minimal* d'un langage L sur un alphabet Σ est l'AFD $\mathcal{A}_L = (Q_L, \Sigma, \delta_L, s_L, F_L)$, où $Q_L = \{w^{-1}.\, L \mid w \in \Sigma^*\}$, l'état initial est $s_L = \varepsilon^{-1}.\, L = L$, $F_L = \{w^{-1}.\, L \mid w \in L\}$ et la fonction de transition δ_L est définie par $\delta_L(q, \sigma) = \sigma^{-1}.\, q$, pour tout $q \in Q_L$ et tout $\sigma \in \Sigma$.

Proposition 1.1.10. *L'automate minimal d'un langage accepte ce langage. Il est accessible et réduit.*

Définition 1.1.11. Un *automate fini non déterministe* (AFND) est un quin-

tuple

$$\mathcal{A} = (Q, \Sigma, \Delta, I, F),$$

où Q, Σ et F sont définis comme dans un AFD et où $\Delta \subseteq Q \times \Sigma^* \times Q$ est un ensemble fini, appelé la *relation de transition* et $I \subseteq Q$ est l'ensemble des *états initiaux.* Un mot w est *accepté par* $\mathcal{A}$ s'il existe $q \in I$ et $f \in F$ tels que $(q, w, f) \in \Delta$. Le *langage accepté par* $\mathcal{A}$ est l'ensemble des mots acceptés par $\mathcal{A}$, encore noté $L(\mathcal{A})$.

Proposition 1.1.12. *Un langage est accepté par un AFD si et seulement si il l'est aussi par un AFND.*

Définition 1.1.13. Un langage est *régulier* s'il est accepté par un automate fini.

Proposition 1.1.14. *Si L est un langage régulier sur un alphabet Σ, les suites $(u_n(L))_{n\geq 0}$ et $(v_n(L))_{n\geq 0}$ vérifient une relation de récurrence linéaire à coefficients constants.*

Proposition 1.1.15. *Si L et M sont deux langages réguliers sur un alphabet Σ, les langages L^R et $L \sqcup\!\sqcup M$ sont aussi réguliers.*

Les deux résultats suivants donnent des moyens de rejeter la régularité d'un langage. Le second est communément appelé *le lemme de la pompe.*

Proposition 1.1.16. *Si L est un langage régulier, alors l'ensemble des longueurs $|L| = \{|w| \mid w \in L\}$ des mots de L est une union finie de progressions arithmétiques.*

Proposition 1.1.17 (Lemme de la pompe)**.** *Si L est un langage régulier sur un alphabet Σ, alors il existe un entier $k > 0$ tel que tout mot w de L de longueur $|w| \geq k$ se décompose en $w = xyz$, avec $x, y, z \in \Sigma^*$, $y \neq \varepsilon$, $|xy| \leq k$ et $xy^*z \subseteq L$.*

1.2 Systèmes de numération abstraits

Définition 1.2.1. Un *système de numération* est une bijection entre l'ensemble des naturels $\mathbb{N}$ et un langage, appelé langage de la numération. En d'autres termes, c'est un moyen de représenter les nombres par des mots.

Définition 1.2.2. Soit un langage L sur un alphabet Σ. Si $<$ est un ordre total sur Σ, celui-ci induit un ordre total sur L, appelé *ordre généalogique,* de la manière suivante. Si u et v sont deux mots de L, $u < v$ si $|u| < |v|$ ou si $|u| = |v|$ et la première lettre par laquelle u et v diffèrent est plus petite dans u que dans v.

La définition suivante a été introduite par P. Lecomte et M. Rigo dans [6].

Définition 1.2.3. Un *système de numération abstrait* est la donnée

$$S = (L, \Sigma, <)$$

d'un langage L sur un alphabet totalement ordonné $(\Sigma, <)$. Pour tout entier n, $\mathrm{rep}_S(n)$ désigne le (n+1)-ième mot de L selon l'ordre généalogique induit par l'ordre $<$ sur Σ et est appelé la *S-représentation* de n. Ainsi, l'application $\mathrm{rep}_S : \mathbb{N} \to L$ est bien une bijection de $\mathbb{N}$ sur L.

L'application rep_S est même une bijection strictement croissante de $\mathbb{N}$ sur L et pour tout mot w de L, on note $\mathrm{val}_S(w) = \mathrm{rep}_S^{-1}(w)$. On appelle $\mathrm{val}_S(w)$ la *valeur numérique* de w selon S.

Définition 1.2.4. Soit $S = (L, \Sigma, <)$ un système de numération abstrait. Une partie X de $\mathbb{N}$ est *S-reconnaissable* si $\mathrm{rep}_S(X)$ est régulier.

Il est souvent agréable de travailler avec un système de numération pour lequel l'ensemble $\mathbb{N}$ tout entier est reconnaissable. On peut alors facilement détecter si un mot est une représentation valide ou non, c'est à dire s'il représente un entier ou non. Cela revient à supposer que le langage de la numération est régulier. C'est pourquoi, par la suite, on considérera presque toujours être en présence d'un système de numération abstrait construit sur un langage régulier.

Exemple 1.2.5. Plus loin, nous nous intéresserons plus particulièrement au langage a^*b^* et au système de numération abstrait $S = (a^*b^*, \{a, b\}, a < b)$ construit sur celui-ci. Pour ce système, on vérifie facilement que

$$\mathrm{val}_S(a^p b^q) = \frac{1}{2}(p+q)(p+q+1) + q,$$

pour tous entiers positifs p et q.

Définition 1.2.6. Soit $S = (L, \Sigma, <)$ un système de numération abstrait construit sur un langage régulier. Si $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ est un AFD acceptant L, pour tout état q de Q, on note val_{S_q} la fonction valeur relative au système abstrait $S_q = (L_q, \Sigma, <)$.

Lorsque l'ordre $<$ sur Σ est clairement déterminé par le contexte, on écrira simplement val_q.

Le résultat suivant donne une méthode de calcul effectif de la fonction val_S. Une démonstration de celui-ci se trouve dans [6].

Proposition 1.2.7. *Soit $S = (L, \Sigma, <)$ un système de numération abstrait construit sur un langage régulier et soit $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ un AFD acceptant ce langage. Si $q \in Q$ et si $w = xy \in L_q$, avec $y \neq \varepsilon$, alors*

$$\mathrm{val}_q(w) = val_{q.x}(y) + v_{|w|-1}(q) - v_{|y|-1}(q.x) + \sum_{x' < x, |x'| = |x|} u_{|y|}(q.x').$$

Une particularité remarquable de ces systèmes de numération abstraits est que les progressions arithmétiques y sont toujours reconnaissables. Une preuve de ce résultat se trouve dans [6].

Proposition 1.2.8. *Les progressions arithmétiques sont S-reconnaissables pour tout système de numération abstrait $S = (L, \Sigma, <)$ construit sur un langage régulier.*

Définition 1.2.9. Soit $S = (L, \Sigma, <)$ un système de numération abstrait. On désigne par $\mathrm{Min}(S)$ l'ensemble des plus petits mots de chaque longueur de L et par $\mathrm{Max}(S)$ l'ensemble des plus grands mots de chaque longueur de L.

Par la suite, lorsqu'il n'y a aucune ambiguïté sur l'ordre $<$ choisi sur Σ, nous nous autoriserons à noter ces langages $\mathrm{Min}(L)$ et $\mathrm{Max}(L)$.

La proposition suivante sera utile à plusieurs reprises dans la suite de ce travail. Une démonstration de celle-ci se trouve par exemple dans [10].

Proposition 1.2.10. *Soit $S = (L, \Sigma, <)$ un système de numération abstrait construit sur un langage régulier. Les langages $Min(S)$ et $Max(S)$ sont réguliers.*

Une autre particularité intéressante des systèmes de numération abstraits est la stabilité de la S-reconnaissablité par translation. Une preuve de ce résultat se trouve dans [6].

Théorème 1.2.11. *La S-reconnaissablité est stable par translation, pour tout système de numération abstrait $S = (L, \Sigma, <)$ construit sur un langage régulier.*

Dans le cadre de l'étude des systèmes de numération, on peut se poser plusieurs types de questions.

Premièrement, étant donné un système de numération abstrait S, quels sont les ensembles d'entiers S-reconnaissables ? Deuxièmement, étant donné un ensemble d'entiers X, peut-on construire un système de numération abstrait reconnaissant X ? Troisièmement, on peut vouloir étudier la stabilité des ensembles reconnaissables par opérations arithmétiques.

Dans le présent travail, nous nous efforcerons de commenter les résultats obtenus en terme de réponse ou de réponse partielle à chacune de ces questions.

Chapitre 2

Langages réguliers polynomiaux

Dans le cadre de ce travail, nous nous attachons particulièrement à étudier les systèmes de numération abstraits construits sur des langages réguliers polynomiaux. Ce chapitre est donc consacré à l'étude de ceux-ci. En premier lieu, nous donnons une caractérisation des langages réguliers polynomiaux en termes d'automates. Nous montrons comment en déduire que tout langage régulier est soit polynomial, soit exponentiel. Ensuite, nous proposons une construction effective d'un système de numération abstrait reconnaissant un ensemble du type $P(\mathbb{N})$, où P est un polynôme à coefficients rationnels dont l'image d'un entier est encore un entier. Enfin, nous donnons un résultat de convergence de la suite $\left(v_n(L)/n^{l+1}\right)_{n\in\mathbb{N}}$, où L est un langage dont la fonction de complexité est d'ordre n^l. Ce résultat sera particulièrement utile dans le chapitre 3, lorsque nous étudierons l'effet de la multiplication par une constante, dans le cas d'un langage polynomial quelconque.

2.1 Caractérisation

Définition 2.1.1. Si f et g sont deux fonctions définies sur $\mathbb{N}$, on dit que f est $O(g)$ si il existe un entier N et une constante réelle $C > 0$ tels que $f(n) \leq Cg(n)$, pour tout $n \geq N$; que f est $\Omega(g)$ si il existe une suite strictement croissante d'entiers $(n_i)_{i\in\mathbb{N}_0}$ et une constante réelle $C > 0$ telles que $f(n_i) \geq Cg(n_i)$, pour tout $i \in \mathbb{N}_0$; et que f est $\Theta(g)$ si f est à la fois $O(g)$ et $\Omega(g)$.

Définition 2.1.2. La *fonction de complexité* d'un langage L sur un alphabet Σ est la fonction $u_L : \mathbb{N} \to \mathbb{N}$ définie par $u_L(n) = \#(L \cap \Sigma^n)$. C'est donc la fonction qui à un entier n associe le nombre de mots de longueur n du langage. Un langage est dit *polynomial* ou *de densité polynomiale* si sa fonction de complexité est $O(n^l)$ pour un certain entier positif l et est dit

exponentiel si sa fonction de complexité est $2^{\Omega(n)}$.

Remarque 2.1.3. Afin de lever une éventuelle ambiguïté dans la définition 2.1.2, nous précisons qu'une fonction f est dite $2^{\Omega(n)}$ si on peut trouver une constante réelle $C > 0$ et une suite strictement croissante d'entiers $(n_i)_{i\in\mathbb{N}_0}$ telles que $f(n_i) \geq 2^{Cn_i}$, pour tout $i \in \mathbb{N}_0$. Remarquons qu'une fonction $2^{\Omega(n)}$ n'est pas nécessairement $\Omega(2^n)$.

Remarque 2.1.4. Pour tout $n \in \mathbb{N}$, nous avons bien sûr $u_n(L) = u_L(n)$. Par commodité, nous employerons indifféremment ces deux notations, privilégiant l'une ou l'autre selon que l'on manipule des suites ou des fonctions.

Dans cette partie, nous donnons une caractérisation des langages réguliers polynomiaux. Ensuite nous montrons que les langages réguliers sont sans exception, soit polynomiaux, soit exponentiels. Les résultats présentés ici proviennent essentiellement de [11].

Définition 2.1.5. Soit $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ un AFD. Pour tout mot sur Σ $w = \sigma_1 \cdots \sigma_n$, la *suite de transition d'états de w relativement à $\mathcal{A}$* est la suite d'états $STS_{\mathcal{A}}(w) = q_{i_0} q_{i_1} \cdots q_{i_n}$ où $q_{i_0} = s$ et $\delta(q_{i_k}, \sigma_{k+1}) = q_{i_{k+1}}$ pour tout $0 \leq k < n$. (Les lettres STS rappellent l'appellation anglophone *state transition sequence.*) On peut bien sûr étendre cette définition de manière à commencer la lecture de w à partir de n'importe quel état de l'automate. Pour tout mot $w = \sigma_1 \cdots \sigma_n$ sur Σ, on dira que la *suite de transition d'états de w à partir de q relativement à $\mathcal{A}$* est la suite d'états $STS^{(q)}_{\mathcal{A}}(w) = q_{i_0} q_{i_1} \cdots q_{i_n}$ où $q_{i_0} = q$ et $\delta(q_{i_k}, \sigma_{k+1}) = q_{i_{k+1}}$ pour tout $0 \leq k < n$.

Définition 2.1.6. Soit $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ un AFD. Un mot w sur Σ est *t-tiered* (t-étagé), $t \geq 0$, relativement à $\mathcal{A}$ si sa suite de transition d'états est de la forme

$$STS_{\mathcal{A}}(w) = \alpha \beta_1^{d_1} \gamma_1 \cdots \beta_t^{d_t} \gamma_t$$

où

1) $0 \leq |\alpha| \leq \#Q$

et pour tout i, $1 \leq i \leq t$,

2) $1 \leq |\beta_i|, |\gamma_i| \leq \#Q$;
3) β_i et γ_i commencent par le même état q_i, apparaissant uniquement à ces deux endroits ;
4) $d_i > 0$.

Cette définition impose que $t \leq \#Q$, puisque les q_i sont tous distincts. Remarquons aussi qu'un mot *t-tiered* relativement à $\mathcal{A}$ a une suite de transition d'états de au moins $2t$ éléments, à savoir les q_i, apparaissant deux fois chacun.

Avec les notations de la définition 2.1.6, tout mot w accepté par l'auto-

mate $\mathcal{A}$ qui est *t-tiered* relativement à $\mathcal{A}$, $t > 0$, se décompose en

$$w = xy_1^{d_1} z_1 \cdots y_t^{d_t} z_t$$

avec $x, y_1, z_1, \ldots, y_t, z_t \in \Sigma^*$ et
$STS_{\mathcal{A}}(x) = \alpha q_1$;
$STS_{\mathcal{A}}^{(q_i)}(y_i) = \beta_i q_i$, $1 \leq i \leq t$;
$STS_{\mathcal{A}}^{(q_i)}(z_i) = \gamma_i q_{i+1}$, $1 \leq i < t$;
$STS_{\mathcal{A}}^{(q_t)}(z_t) = \gamma_t$.

Lemme 2.1.7. *Soient un AFD $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ et L son langage accepté. Si on peut trouver un mot w de L qui soit k-tiered relativement à $\mathcal{A}$, $k \in \mathbb{N}$, alors la fonction de complexité de L est $\Omega(n^{k-1})$.*

Démonstration. En gardant les mêmes notations que précédemment, on a la décomposition $w = xy_1^{d_1} z_1 \cdots y_k^{d_k} z_k$. On a bien sûr $xy_1^* z_1 \cdots y_k^* z_k \subseteq L$. Posons

$$C = \prod_{i=1}^{k} |y_i| \quad \text{et} \quad C_i = \frac{C}{|y_i|}, \ 1 \leq i \leq k.$$

Tout mot de la forme $xy_1^{t_1 C_1} z_1 \cdots y_k^{t_k C_k} z_k$, où $t_1, \ldots, t_k$ sont k entiers positifs de même somme $t > 0$, est dans L et de même longueur $n = |xz_1 \cdots z_k| + tC$. Vu le point 3) de la définition 2.1.6 et le fait que l'automate $\mathcal{A}$ est déterministe, deux tels k-uples distincts $(t_1, \ldots, t_k)$ et $(t'_1, \ldots, t'_k)$ donnent lieu à deux mots $xy_1^{t_1 C_1} z_1 \cdots y_k^{t_k C_k} z_k$ et $xy_1^{t'_1 C_1} z_1 \cdots y_k^{t'_k C_k} z_k$ distincts de L. On se ramène donc à l'évaluation de la cardinalité de l'ensemble

$$N_k(t) = \left\{ (t_1, \ldots, t_k) \in \mathbb{N}^k : \sum_{i=1}^{k} t_i = t \right\}$$

des k-uples d'entiers positifs ou nuls de même somme t. On peut montrer (voir par exemple [2] aux pages 80 et 102) que $\#N_k(t) = \binom{t+k-1}{t}$, qui est une fonction de t en $\Omega(t^{k-1})$. Comme n est linéaire en t, on a bien que la fonction de complexité du langage L est une fonction $\Omega(n^{k-1})$. □

Remarque 2.1.8. Soit L un langage vérifiant les hypothèses du lemme 2.1.7. Ce dernier affirme qu'on peut trouver une suite strictement croissante d'entiers $(n_i)_{i \in \mathbb{N}_0}$ et une constante $C > 0$ telles que $u_L(n_i) \geq C n_i^{k-1}$, pour tout $i \in \mathbb{N}_0$. En fait, au vu de la preuve de celui-ci, on peut même être plus précis en exigeant de la suite $(n_i)_{i \in \mathbb{N}_0}$ qu'elle soit périodique.

Lemme 2.1.9. *Un AFD $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ qui accepte un langage polynomial ne contient que des boucles simples, au sens suivant. On ne peut pas trouver d'état accessible q de Q tel que*

(i) $\delta(q, x) = q$ et $\delta(q, y) = q$, $x \neq y$, $x \neq \varepsilon$ et $y \neq \varepsilon$;

(ii) aucun préfixe non trivial x' de x ou y' de y ne vérifie $\delta(q, x') = q$ ou $\delta(q, y') = q$ et
(iii) le langage accepté à partir de q est non vide.

Démonstration. Procédons par l'absurde et supposons qu'on puisse trouver un état q de Q vérifiant (i), (ii) et (iii). En particulier, les deux premières conditions imposent que x ne peut pas être préfixe de y ni y de x. Comme q est un état accessible par hypothèse et vu la troisième condition, il existe des mots u et v sur Σ tels que $\delta(s, u) = q$ et $\delta(q, v) \in F$. Cette situation est illustrée par la figure 2.1. Pour tout $g \geq 0$, le langage $u(xy + yx)^g v$ contient

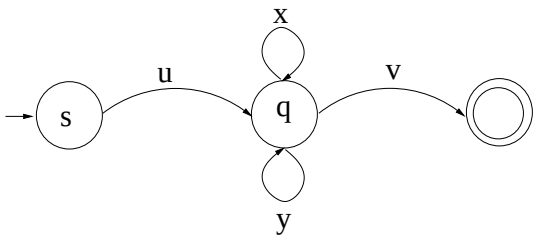

FIG. 2.1 – Langage accepté exponentiel.

alors exactement 2^g mots, tous acceptés et de même longueur $|uv| + g|xy|$. Ceci montre que le langage accepté est exponentiel, ce qui est contradictoire. □

Lemme 2.1.10. *Soient un AFD $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ et L son langage accepté. Si la fonction de complexité de L est $O(n^k)$ pour un certain entier $k \geq 0$, alors pour tout mot w de L, on peut trouver un entier $t \leq k + 1$ tel que w soit t-tiered relativement à $\mathcal{A}$.*

Démonstration. Considérons un mot w de L. Comme L est un langage polynomial, aucun état q de $STS_{\mathcal{A}}(w)$ ne vérifie simultanément les conditions (i) et (ii) du lemme 2.1.9. Il y a donc au plus une suite d'états non vide permettant de passer de q à q en respectant la fonction de transition de $\mathcal{A}$ et sans repasser par q. Ceci montre que w est *t-tiered* relativement à $\mathcal{A}$, pour un entier $t \leq \#Q$. En outre, on a nécessairement $t \leq k + 1$. En effet, sinon, vu le lemme 2.1.7, la complexité de L serait au moins en $\Omega(n^{k+1})$, ce qui n'est pas possible, vu l'hypothèse. □

Lemme 2.1.11. *Soient un AFD $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ et L son langage accepté. Si il existe $k \geq 0$ tel que tout mot de L soit t-tiered relativement à $\mathcal{A}$ pour un $t \leq k$, alors L peut être représenté par une union finie d'expressions régulières de la forme $xy_1^* z_1 \cdots y_t^* z_t$ avec $t \leq k$, x, y_i, $z_i \in \Sigma^*$ tels que $|x|, |z_i| < \#Q$ et $0 < |y_i| \leq \#Q$, pour tout $1 \leq i \leq t$.*

Démonstration. Soit w un mot de L. Par hypothèse, w est *t-tiered* relativement à $\mathcal{A}$ pour un entier $t \leq k$ et se décompose alors en $w = xy_1^{d_1}z_1 \cdots y_t^{d_t}z_t$ avec $|x|, |z_i| < \#Q$ et $0 < |y_i| \leq \#Q$, pour tout $1 \leq i \leq t$. Bien sûr, on a l'inclusion $xy_1^*z_1 \cdots y_t^*z_t \subseteq L$.
A chaque mot w de L correspond donc une et une seule expression régulière de ce type. On la note $E(w)$. Ainsi,

$$L = \bigcup_{w \in L} E(w).$$

Cette union est nécessairement une union finie puisque Q et Σ sont des ensembles finis. □

Lemme 2.1.12. *Si L est un langage représenté par une expression régulière de la forme $xy_1^*z_1 \cdots y_t^*z_t$ avec $t > 0$ et x, y_i, $z_i \in \Sigma^*$ pour tout $1 \leq i \leq t$, alors la fonction de complexité de L est $O(n^{t-1})$.*

Démonstration. On peut supposer que tous les mots y_i diffèrent du mot vide. Fixons un entier positif n et considérons les mots de L de longueur n. Ils sont de la forme $w = xy_1^{n_1}z_1 \cdots y_t^{n_t}z_t$, où $n_1, \ldots, n_t$ sont des entiers positifs ou nuls tels que $\sum_{i=1}^{t} n_i|y_i| + |xz_1 \cdots z_t| = n$. Notons $C = |xz_1 \cdots z_t|$ et $m = n - C$. On se ramène à nouveau (cf. lemme 2.1.7) à l'évaluation de la cardinalité de l'ensemble $N^t(m)$. En effet, l'application τ définie sur $\mathbb{N}^t$ par $\tau(n_1, \ldots, n_t) = (n_1|y_1|, \ldots, n_t|y_t|)$ est injective et pour tout $t-$uple $(n_1, \ldots, n_t)$ tel que le mot $xy_1^{n_1}z_1 \cdots y_t^{n_t}z_t$ est de longueur n, on a $\tau(n_1, \ldots, n_t) \in N^t(m)$. Le nombre de mots de longueur n ne peut donc pas pas dépasser $\#N^t(m)$. Pour conclure, on remarque que $\#N^t(m) = \binom{m+t-1}{t-1} = \binom{n-C+t-1}{t-1}$ est une fonction en $O(n^{t-1})$, puisque C et t sont des constantes. □

Le théorème suivant découle directement des lemmes qui précèdent. Il donne une caractérisation des langages réguliers polynomiaux.

Théorème 2.1.13. *Soient L un langage régulier sur un alphabet Σ et $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ un AFD acceptant L. Soit un entier $k \geq 0$.*
Les assertions suivantes sont équivalentes.

i) La fonction de complexité de L est une fonction $O(n^k)$.
ii) Tout mot de L est $t-$tiered relativement à $\mathcal{A}$ pour un entier $t \leq k+1$.
iii) Le langage L est une union finie d'expressions régulières du type $xy_1^*z_1 \cdots y_t^*z_t$, avec $0 \leq t \leq k+1$, x, y_i, $z_i \in \Sigma^*$ et $|x|, |y_i|, |z_i| \leq \#Q$ pour tout $1 \leq i \leq t$.

De ce théorème, on tire quelques conséquences directes. Tout d'abord, énonçons une forme plus faible de celui-ci, mais souvent suffisante.

Théorème 2.1.14. *Soit un entier $k \geq 0$. Un langage régulier L a une complexité en $O(n^k)$ si et seulement si L est une une union finie d'expressions*

régulières du type $xy_1^*z_1 \cdots y_t^*z_t$, *avec* $0 \leq t \leq k+1$ *et* $x, y_i, z_i \in \Sigma^*$, *pour tout* $1 \leq i \leq t$.

Ensuite, le corollaire suivant est simplement un cas particulier du théorème 2.1.14. Il fait apparaître les langages *slender,* i.e. dont la fonction de complexité est bornée par une constante, qui seront étudiés dans le chapitre 3.

Corollaire 2.1.15. *Soient L un langage régulier. La fonction de complexité de L est bornée par une constante si et seulement si L est une union finie d'expressions régulières de la forme xy^*z avec x, y, $z \in \Sigma^*$.*

Le résultat suivant reprend une remarque déjà formulée plus haut.

Corollaire 2.1.16. *Soient L un langage régulier et $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ un AFD acceptant L. Si L est polynomial et si C est le nombre d'états de $\mathcal{A}$, alors la fonction de complexité de L est une fonction $O(n^{C-1})$.*

Démonstration. Vu la définition 2.1.6, les états q_i sont tous distincts et par conséquent, tout mot de L est au plus $C - tiered$ relativement à $\mathcal{A}$. Par le théorème 2.1.13, la fonction de complexité est alors au plus en $O(n^{C-1})$. □

Le théorème 2.1.13 permet de montrer la réciproque du lemme 2.1.9.

Corollaire 2.1.17. *Soient un AFD $\mathcal{A} = (Q, \Sigma, \delta, s, F)$ et L son langage accepté. Si on ne peut trouver aucun état accessible q de Q vérifiant simultanément les conditions (i), (ii) et (iii) du lemme 2.1.9, alors L est un langage polynomial.*

Démonstration. Supposons qu'aucun état accessible q de Q ne vérifie simultanément les conditions (i), (ii) et (iii) du lemme 2.1.9. Par le même argument que dans la preuve du lemme 2.1.10, on peut montrer que tout mot w de L est t-*tiered* pour un entier $t \leq \#Q$. De là, le théorème 2.1.13 permet de conclure. □

2.2 Théorèmes de séparation

Le théorème suivant est un premier théorème de séparation des langages réguliers. C'est une conséquence directe des résultats de la section précédente.

Théorème 2.2.1. *Pour tout entier $k \geq 0$, on ne peut pas trouver de langage régulier L de fonction de complexité qui ne soit ni $O(n^k)$ ni $\Omega(n^{k+1})$.*

Démonstration. Procédons par l'absurde et supposons qu'il existe un langage régulier L dont la fonction de complexité $u_L(n)$ ne soit ni $O(n^k)$ ni $\Omega(n^{k+1})$. Alors $u_L(n)$ serait $O(n^{k+1})$. Si $\mathcal{A}$ est un AFD acceptant ce langage L, alors, par le lemme 2.1.10, tout mot de L est t-*tiered* relativement à $\mathcal{A}$, pour un entier $t \leq k+2$. Si L contient un mot $(k+2)$-*tiered* relativement

à $\mathcal{A}$, par le lemme 2.1.7, $u_L(n)$ serait $\Omega(n^{k+1})$, ce qui n'est pas possible. Sinon, tout mot de L est *t-tiered* relativement à $\mathcal{A}$, pour un entier $t \leq k+1$ et par le théorème 2.1.13, $u_L(n)$ devrait être $O(n^k)$, ce qui est également contradictoire. □

Voici le second théorème de séparation. Il montre, comme annoncé plus haut, que seules les fonctions de classe polynomiale ou exponentielle ont une chance d'être la fonction de complexité d'un langage régulier.

Théorème 2.2.2. *On ne peut pas trouver de langage régulier L de fonction de complexité qui ne soit ni $O(n^k)$, pour un certain entier $k \geq 0$, ni $2^{\Omega(n)}$.*

Démonstration. Procédons par l'absurde et supposons qu'il existe un langage régulier L dont la fonction de complexité $u_L(n)$ ne soit ni $O(n^k)$, pour aucun entier $k \geq 0$, ni $2^{\Omega(n)}$. Comme il s'agit d'un langage régulier, L est accepté par un AFD $\mathcal{A} = (Q, \Sigma, \delta, s, F)$. Deux cas se présentent. Soit on peut trouver un état accessible q de Q qui vérifie les conditions (i), (ii) et (iii) du lemme 2.1.9, soit on ne peut pas. Dans le premier cas, L serait exponentiel, par le même argument que dans la preuve du lemme 2.1.9. Et dans le second cas, par le corollaire 2.1.17, L serait polynomial et nous avons supposé le contraire. □

Corollaire 2.2.3. *Soit un langage régulier $L \subseteq \Sigma^*$. Soit il existe $k \geq 0$ tel que u_L est $\Theta(n^k)$, soit $u_L(n)$ est $2^{\Theta(n)}$.*

Démonstration. Vu le théorème 2.2.2, L est soit polynomial, soit exponentiel. Dans le deuxième cas, comme on a toujours $u_L(n) \leq (\#\Sigma)^n$, quelque soit le langage $L \subseteq \Sigma^*$, on a bien que $u_L(n)$ est $2^{\Theta(n)}$. Dans le premier cas, $u_L(n)$ est $O(n^k)$, pour un entier $k \geq 0$. On choisit k minimum pour cette propriété. Montrons qu'alors $u_L(n)$ est aussi $\Omega(n^k)$. Vu le lemme 2.1.10, si $\mathcal{A}$ est un AFD acceptant L, tout mot de L est *t-tiered* relativement à $\mathcal{A}$, pour un $t \leq k+1$. Supposons qu'il n'existe pas de mot de L qui soit $(k+1)$-*tiered* relativement à $\mathcal{A}$. Alors, par le théorème 2.1.13, $u_L(n)$ serait $O(n^{k-1})$, ce qui est impossible, vu le choix de k. D'où on peut trouver un mot w de L qui est $(k+1)$-*tiered* relativement à $\mathcal{A}$ et par le lemme 2.1.7, $u_L(n)$ est alors une fonction $\Omega(n^k)$. □

2.3 Constructions

2.3.1 Langages de complexité n^k

Dans la section précédente, on a montré que les seules fonctions candidates pour être fonction de complexité d'un langage régulier étaient les fonctions de classe polynomiale et les fonctions de classe exponentielle. En fait, on peut même aller plus loin et montrer que pour chaque fonction de

classe polynomiale f, il existe un langage régulier de fonction de complexité en $\Theta(f)$.

Plus précisément, on se propose de donner ici une construction effective d'un langage régulier de complexité exactement n^k, $k \geq 0$. Les résultats qui suivent proviennent essentiellement de [8].

Définition 2.3.1. Soit un langage L sur un alphabet Σ. L'*alpabet minimal* de L est la plus petite partie de Σ qui contient les symboles nécessaires pour écrire les mots de L.

On veut construire des langages réguliers L_k tels que $u_{L_k}(n) = n^k$, pour tout $k \in \mathbb{N}$. La construction proposée se base sur le résultat suivant, facile à vérifier.

Proposition 2.3.2. *Soient un langage L sur un alphabet Σ et un symbole σ n'appartenant pas à l'alphabet minimum de L. Si M est le langage $\{\sigma\} \sqcup\!\sqcup L$, alors $u_M(n) = n\, u_L(n-1)$, $n \geq 1$.*

Tout d'abord, comme cas de base, il suffit de prendre $L_0 = a^*$ sur l'alphabet $\{a\}$. Considérons maintenant un entier $k > 0$ et supposons avoir à notre disposition des langages réguliers L_j tels que $u_{L_j}(n) = n^j$, $n \geq 0$, pour tout $j < k$. Vu la proposition 2.3.2, pour obtenir un langage L_k de complexité n^k, il suffit de construire un langage M_k de complexité

$$u_{M_k}(n) = (n+1)^{k-1} = \sum_{j=0}^{k-1} \binom{k-1}{j} n^j,\ n \geq 0.$$

Un tel langage s'obtient comme union finie des langages L_j sur des alphabets disjoints, $j < k$. Plus précisément, on écrit M_k sous la forme

$$M_k = \bigcup_{j=0}^{k-1} \bigcup_{i=1}^{\binom{k-1}{j}} L_{j,i},$$

où les langages $L_{j,i}$ sont des langages de complexité n^j et d'alphabets $\Sigma_{j,i}$ minimaux disjoints. Ainsi, si σ_k n'appartient pas à l'alphabet minimal de M_k, le langage $L_k = \{\sigma_k\} \sqcup\!\sqcup M_k$ est bien de complexité exactement n^k.

Dans la suite de ce texte, L_k et M_k désigneront encore les langages ainsi construits. Si on se place dans le contexte des systèmes de numération, contexte qui nous intéresse précisément dans ce travail, avoir à disposition de tels langages est très utile. En effet, cela va nous permettre de montrer que toute partie de $\mathbb{N}$ de la forme $P(\mathbb{N})$, où P est un polynôme à coefficients rationnels par lequel l'image de tout entier est entière, est reconnaissable par un système de numération abstrait. La suite de ce chapitre est précisément consacrée à la construction effective d'un système de numération abstrait reconnaissant l'ensemble des image des entiers par un tel

polynôme donné. Nous considérerons successivement des polynômes à coefficients dans $\mathbb{N}$, dans $\mathbb{Z}$ et finalement dans $\mathbb{Q}$.

2.3.2 Polynômes à coefficients dans $\mathbb{N}$

L'idée principale est de construire un langage régulier L tel que les positions des premiers mots de chaque longueur pour l'ordre généalogique sont exactement les valeurs prises par le polynôme. Ceci est réalisé, éventuellement à un nombre fini d'éléments près, lorsque la fonction de complexité du langage L est de la forme $u_L(n) = P(n+1) - P(n)$, $n \geq n_0$, et que le nombre de mots de longueur strictement inférieure à n_0 est $P(n_0)$. Il s'ensuit alors que

$$\mathrm{rep}_S(\{P(\mathbb{N}) \mid n \geq n_0\}) = \mathrm{Min}(L) \cap \Sigma^{\geq n_0}.$$

Proposition 2.3.3. *Soit $P \in \mathbb{N}[X]$. Il existe un système de numération construit sur un langage régulier $S = (L, \Sigma, <)$ tel que $P(\mathbb{N})$ est une partie S-reconnaissable.*

Démonstration. Si P est un polynôme constant, le résultat est évident. Supposons donc P non constant. Comme la reconnaissabilité est stable par translation (cf. le théorème 1.2.11), on peut supposer que $P(0) = 0$. Le polynôme $P(X+1) - P(X)$ étant aussi à coefficients dans $\mathbb{N}$, on obtient un langage L de complexité $u_L(n) = P(n+1) - P(n)$, $n \geq 1$, en prenant une union disjointe de langages L_k de complexité n^k introduits précédemment. On munit ensuite l'alphabet minimal de ce langage d'un ordre total et on considère le système de numération abstrait S associé. Quitte à ajouter des symboles nouveaux à l'alphabet minimal de L, on peut supposer que le premier mot w de longueur 2 de L vérifie $\mathrm{val}_S(w) = P(2)$. A un nombre fini d'éléments près, $P(\mathbb{N})$ coïncide alors avec la partie $\mathrm{val}_S(\mathrm{Min}(L))$ de $\mathbb{N}$, dont on sait qu'elle est S-reconnaisable (cf. la proposition 1.2.10). □

Remarquons que cette construction est indépendante de l'ordre choisi sur l'aphabet du système. En effet, seule intervient la position du premier mot de chaque longueur, et non le mot lui-même.

Corollaire 2.3.4. *Pour tout entier k positif ou nul, il existe un système de numération construit sur un langage régulier $S = (L, \Sigma, <)$ tel que l'ensemble $\{x^k \mid x \in \mathbb{N}\}$ est une partie S-reconnaissable.*

2.3.3 Polynômes à coefficients dans $\mathbb{Z}$

Lemme 2.3.5. *Soient k, $\alpha \in \mathbb{N}_0$. Il existe un langage régulier L tel que*

$$u_L(n) = \begin{cases} n^k - \alpha n^{k-1}, & n \geq \alpha \\ 0, & n < \alpha. \end{cases}$$

Démonstration. Si $k = 1$, le langage $L = a^{\alpha+1}a^*b^*$ convient. Supposons maintenant que $k \geq 2$. En reprenant les notations de notre construction des langages de complexité n^k, pour tout entier $i = 1, \ldots, n$, le langage L_k possède exactement n^{k-1} mots de longueur n avec σ_k en i-ième position. Il s'ensuit que le langage

$$L = L_k \setminus \bigcup_{j=0}^{\alpha-1} \Sigma^* \sigma_k \Sigma^j$$

convient. □

Proposition 2.3.6. *Soit $P \in \mathbb{Z}[X]$. Si $P(\mathbb{N}) \subseteq \mathbb{N}$, alors il existe un système de numération construit sur un langage régulier $S = (L, \Sigma, <)$ tel que $P(\mathbb{N})$ est une partie S-reconnaissable.*

Démonstration. On considère encore le polynôme $P(X+1) - P(X)$. Comme on a supposé que $P(\mathbb{N}) \subseteq \mathbb{N}$, celui-ci est de coefficient dominant strictement positif et si l est son degré, en ajoutant des termes de la forme $x^j - x^j$, on peut le réécrire sous la forme

$$\sum_{k=0}^{l} b_k x^k + (x^{i_1+1} - a_{i_1} x^{i_1}) + \cdots + (x^{i_r+1} - a_{i_r} x^{i_r})$$

où $i_j \in \{0, \ldots, l-1\}$, $a_{i_j} \in \mathbb{N}_0$ pour $1 \leq j \leq r$ et $b_k \in \mathbb{N}$ pour $0 \leq k \leq l$. Notons $\alpha = \sup\{a_{i_j} \mid 1 \leq j \leq r\}$. Par le lemme 2.3.5, on construit des langages réguliers R_j tels que $u_{R_j}(n) = n^{i_j+1} - a_{i_j} n^{i_j}$, $n \geq \alpha$. Ensuite, en prenant des unions des langages R_j et L_k, on construit un langage régulier L tel que $u_L(n) = P(n+1) - P(n)$, $n \geq \alpha$. On munit l'aphabet minimal de ce langage d'un ordre total et on considère le système de numération abtrait S associé. On peut supposer que le premier mot w de L vérifie $\mathrm{val}_S(w) = P(\alpha)$. Ceci nécessite éventuellement un ajout ou un retrait d'un nombre fini d'éléments aux mots du langage, opérations qui ne modifient en rien sa régularité. On a donc obtenu que $P(\mathbb{N})$ coïncide, à un nombre fini d'éléments près, avec la partie $\mathrm{val}_S(\mathrm{Min}(L))$ de $\mathbb{N}$, qui est S-reconnaissable (cf. le lemme 1.2.10). □

2.3.4 Polynômes à coefficients dans $\mathbb{Q}$

Théorème 2.3.7. *Soit $P \in \mathbb{Q}[X]$. Si $P(\mathbb{N}) \subseteq \mathbb{N}$, alors il existe un système de numération construit sur un langage régulier $S = (L, \Sigma, <)$ tel que $P(\mathbb{N})$ est une partie S-reconnaissable.*

Démonstration. Le polynôme P peut s'écrire sous la forme

$$P = \frac{a_k}{b_k} X^k + \cdots + \frac{a_1}{b_1} X + \frac{a_0}{b_0},$$

avec $b_j, a_k \in \mathbb{N}_0$ pour $0 \leq j \leq k$ et $a_j \in \mathbb{Z}$ pour $0 \leq j \leq k-1$. Si on note $m = \text{ppcm}\{b_j \mid 0 \leq j \leq k\}$, alors $P = \frac{P'}{m}$, où $P' \in \mathbb{Z}[X]$. Comme par hypothèse $P(\mathbb{N}) \subseteq \mathbb{N}$, on a $P'(\mathbb{N}) \subseteq m\mathbb{N}$. Vu la preuve de la proposition 2.3.6, on peut trouver un langage régulier L' et une constante $\alpha \in \mathbb{N}$ tels que

$$u_{L'}(n) = P'(n+1) - P'(n) = m\left(P(n+1) - P(n)\right),\ n \geq \alpha$$

et tels que L' contienne exactement $P'(\alpha)$ mots de longueur strictement inférieure à α. Formellement,

$$\sum_{i=0}^{\alpha-1} u_{L'}(i) = P'(\alpha) = mP(\alpha).$$

On choisit un ordre total $<$ sur l'alphabet minimal Σ de L' et on considère $S' = (L', \Sigma, <)$ le système de numération abstrait associé. Les progressions arithmétiques étant toujours reconnaissables (cf. le théorème 1.2.8), $m\mathbb{N}$ est une partie S'-reconnaissable de $\mathbb{N}$. Par conséquent, le langage $L = \text{rep}_{S'}(m\mathbb{N}) \subseteq \Sigma^*$ est régulier et tel que

$$\sum_{i=0}^{\alpha-1} u_L(i) = P(\alpha) \text{ et } u_L(n) = P(n+1) - P(n),\ n \geq \alpha.$$

Le premier mot de longueur α de L est le premier de longueur α de L' et sa position pour l'ordre généalogique induit par $<$ sur L est $P(\alpha)$. On conclut en utilisant encore la proposition 1.2.10 avec le système abstrait $S = (L, \Sigma, <)$. □

2.3.5 Fonctions exponentielles polynômes

Définition 2.3.8. Une *fonction exponentielle polynôme* est une fonction de la forme

$$f(n) = \sum_{i=1}^{k} P_i(n)\ \alpha_i^n,$$

où $P_i \in \mathbb{Q}[X]$ et $\alpha_i \in \mathbb{N}$, pour tout $1 \leq i \leq k$.

Les constructions obtenues précédemment nous amènent naturellement à considérer ce type de fonction. On obtient comme conséquence presque directe une construction d'un système de numération abstrait reconnaissant $f(\mathbb{N})$, où f est une fonction exponentielle polynôme donnée.

Proposition 2.3.9. *Pour tout entier α positif ou nul, il existe un système de numération construit sur un langage régulier $S = (L, \Sigma, <)$ tel que l'ensemble $\{\alpha^n \mid n \in \mathbb{N}\}$ est une partie S-reconnaissable.*

Démonstration. Si α vaut 0 ou 1, c'est évident. Supposons $\alpha \geq 2$. Une manière de procéder est d'utiliser encore une fois la proposition 1.2.10 et de construire un langage régulier L de complexité $u_L(n) = \alpha^{n+1} - \alpha^n = (\alpha - 1)\,\alpha^n$, $n \geq 1$. Ceci peut être réalisé en prenant $\alpha - 1$ copies disjointes de Σ^*, où Σ est un alphabet à α éléments. □

Proposition 2.3.10. *Soient $\alpha \in \mathbb{N}$ et $P \in \mathbb{Q}[X]$ tel que $P(\mathbb{N}) \subseteq \mathbb{N}$. Il existe un système de numération construit sur un langage régulier $S = (L, \Sigma, <)$ tel que l'ensemble $\{P(n)\,\alpha^n \mid n \in \mathbb{N}\}$ est une partie S-reconnaissable.*

Démonstration. Si $\alpha = 0$, c'est évident. Si $\alpha = 1$, c'est le théorème 2.3.7. Supposons maintenant $\alpha \geq 2$. Dans un premier temps, nous supposons que $P \in \mathbb{N}[X]$. L'idée est toujours de construire un langage régulier L de complexité $u_L(n) = P(n+1)\,\alpha^{n+1} - P(n)\,\alpha^n = (\alpha\,P(n+1) - P(n))\,\alpha^n$. Bien sûr, $\alpha\,P(n+1) - P(n) \in \mathbb{N}[X]$. Il suffit alors de montrer comment construire des langages $L_{k,\alpha}$ contenant exactement $n^k\,\alpha^n$ mots de longueur $n > k \geq 1$.

Tout d'abord, si $k = 1$, en prenant α copies disjointes de Σ^*, avec $\#\Sigma = \alpha$, on obtient un langage $M_{1,1}$ tel que $u_{M_{1,1}}(n-1) = \alpha^n$, $n \geq 2$. Vu la proposition 2.3.2, si σ n'apparaît pas dans l'alphabet minimal de $M_{1,1}$, le langage $L_{1,\alpha} = \{\sigma\} \sqcup\!\sqcup M_{1,1}$ vérifie alors $u_{L_{1,\alpha}}(n) = n\alpha^n$, $n \geq 2$. Ensuite, si $k = 2$, avec α^2 copies disjointes de Σ^*, on obtient un langage $M_{2,1}$ tel que $u_{M_{2,1}}(n-2) = \alpha^n$, $n \geq 3$. Si σ_1 n'apparaît pas dans l'alphabet minimal de $M_{2,1}$, alors $u_{\{\sigma_1\} \sqcup\!\sqcup M_{2,1}}(n-1) = (n-1)\alpha^n$, $n \geq 3$. Le langage $M_{2,2} = (\{\sigma_1\} \sqcup\!\sqcup M_{2,1}) \cup M_{1,1}$ vérifie donc $u_{M_{2,2}}(n-1) = n\alpha^n$, $n \geq 3$, les unions étant réalisées à chaque fois sur des alphabets disjoints. Si σ_2 est un nouveau symbole, alors le langage $L_{2,\alpha} = \{\sigma_2\} \sqcup\!\sqcup M_{2,2}$ convient. En continuant de cette manière, on trouve le langage $L_{3,\alpha}$ et tous les suivants. La construction de $L_{3,\alpha}$ est synthétisée dans le tableau 2.1.

	description	complexité
$M_{3,1}$	α^3 copies de Σ^*	$u(n-3) = \alpha^n$
$M_{3,2}$	$M_{3,1} \sqcup\!\sqcup \{\sigma_1\} \bigcup$ 2 copies de $M_{2,1}$	$u(n-2) = n\,\alpha^n$
$M_{3,3}$	$M_{3,2} \sqcup\!\sqcup \{\sigma_2\} \bigcup$ 1 copie de $M_{2,2}$	$u(n-1) = n^2\,\alpha^n$
$L_{3,\alpha}$	$M_{3,3} \sqcup\!\sqcup \{\sigma_3\}$	$u(n) = n^3\,\alpha^n$

TAB. 2.1 – Construction de $L_{3,\alpha}$, $(n \geq 4)$.

En procédant comme dans le lemme 2.3.5 et la proposition 2.3.6, puis comme dans le théorème 2.3.7, on peut supposer que les coefficients de P sont rationnels. □

Le théorème suivant est une conséquence directe de la proposition 2.3.10.

Théorème 2.3.11. *Soient $P_i \in \mathbb{Q}[X]$ tels que $P(\mathbb{N}) \subseteq \mathbb{N}$ et $\alpha_i \in \mathbb{N}$, pour $i = 1, \ldots, k$, $k \geq 1$. Soit*

$$f(n) = \sum_{i=1}^{k} P_i(n)\ \alpha_i^n.$$

Il existe un système de numération $S = (L, \Sigma, <)$ construit sur un langage régulier tel que $f(\mathbb{N})$ est une partie S-reconnaissable.

2.4 Propriété de la suite $v_n(L)$

Soit L un langage régulier de fonction de complexité en $\Theta(n^l)$. En général, la suite $\left(u_n(L)/n^l\right)_{n\in\mathbb{N}}$ ne converge pas. A cet égard, l'exemple suivant est démonstratif.

Exemple 2.4.1. Considérons le langage $L = a^*b^* \cap \left(\{a,b\}^2\right)^*$ des mots de a^*b^* de longueurs paires. On a

$$\begin{cases} u_{2n+1}(L) = 0 \\ u_{2n}(L) = 2n+1 \end{cases} \quad \text{et} \quad v_{2n}(L) = v_{2n+1}(L) = (n+1)^2.$$

En particulier, $u_L(n)$ est $\Theta(n)$. La suite $(u_n(L)/n)_{n\in\mathbb{N}}$ ne converge pas parce qu'elle possède deux sous-suites convergeant vers des limites différentes :

$$\lim_{n\to+\infty} \frac{u_{2n}(L)}{n} = 2 \text{ et } \lim_{n\to+\infty} \frac{u_{2n+1}(L)}{n} = 0.$$

Par contre, on observe que $\lim_{n\to+\infty} \frac{v_n(L)}{n^2} = 1 > 0$.

Cette dernière observation de l'exemple 2.4.1 est en fait générale. Si L est un langage régulier de complexité en $\Theta(n^l)$, alors la suite $\left(v_n(L)/n^{l+1}\right)_{n\in\mathbb{N}}$ converge, et ce, vers une limite strictement positive. Avant de donner une preuve de ce résultat, nous montrons les deux lemmes suivants.

Lemme 2.4.2. *Soient ρ_j, θ_j, ϕ_j, $1 \leq j \leq k$, des nombres réels tels que pour tous i, j distincts, $\theta_i \not\equiv \theta_j \pmod{2\pi}$ et pour tout i, $\rho_i \neq 0$. Alors il existe $\varepsilon > 0$ tel que $M_n = |\sum_{j=1}^{k} \rho_j e^{i(n\theta_j+\phi_j)}| > \varepsilon$, pour une suite infinie d'entiers n.*

Démonstration. Par le théorème de Bolzano-Weierstrass, on peut trouver des complexes $z_1, \ldots, z_k$ et une sous-suite $t(n)$ tels que $\rho_j e^{i(t(n)\theta_j+\phi_j)} \to z_j$ si $n \to +\infty$ et $|z_j| = |\rho_j| \neq 0$. Procédons par l'absurde et supposons que pour tout $\varepsilon > 0$, on n'a $M_n > \varepsilon$ seulement pour un nombre fini d'entiers n. Autrement dit, on suppose que la suite $(M_n)_{n\in\mathbb{N}}$ converge vers 0. Mais alors, on aurait $\sum_{j=1}^{k} z_j = 0$ et pour tout $0 \leq l \leq k-1$,

$$\sum_{j=1}^{k} \rho_j e^{i(\,(t(n)+l)\,\theta_j+\phi_j\,)} \to \sum_{j=1}^{k} z_j e^{il\theta_j} = 0.$$

De là, les nombres $z_1, \ldots, z_k$ devraient vérifier

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ e^{i\theta_1} & e^{i\theta_2} & \cdots & e^{i\theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ e^{i(k-1)\theta_1} & e^{i(k-1)\theta_2} & \cdots & e^{i(k-1)\theta_k} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

ce qui est impossible puisque le déterminant de Vandermonde ne s'annule pas, vu les hypothèses. □

Rappelons (voir par exemple [4]) que la somme des puissances p-ièmes des $n+1$ premiers entiers est donnée par

$$\sum_{i=0}^{n} i^p = \frac{(n+B+1)^{p+1} - B^{p+1}}{p+1}, \tag{2.1}$$

où les termes de la forme B^m sont remplacés par les *nombres de Bernoulli* B_m définis par l'identité

$$\frac{x}{e^x - 1} = \sum_{m=0}^{+\infty} \frac{B_m x^m}{m!}.$$

Cette formule sera utile pour montrer le lemme suivant ainsi que dans la démonstration du théorème 2.4.5.

Lemme 2.4.3. *Si L est un langage régulier tel que $u_L(n)$ est $\Theta(n^l)$, pour $l \in \mathbb{N}$, alors $v_L(n) = \sum_{i=0}^{n} u_L(i)$ est une fonction $\Omega(n^{l+1})$.*

Démonstration. Par hypothèse, on peut trouver une suite strictement croissante d'entiers $(n_i)_{i\in\mathbb{N}_0}$ et une constante $C > 0$ telles que $u_L(n_i) \geq C n_i^l$, pour tout $i \in \mathbb{N}_0$. Vu le théorème 2.1.13 et la remarque 2.1.8, on peut supposer que $n_i = n_1 + (i-1)N$, pour tout $i \in \mathbb{N}_0$, avec $N \in \mathbb{N}_0$. Dès lors, vu la formule 2.1,

$$\begin{aligned} v_L(n_i) = \sum_{j=0}^{n_i} u_L(j) &\geq \sum_{j=1}^{i} u_L(n_j) \geq C \sum_{j=1}^{i} (n_1 + (j-1)N)^l \\ &\geq CN^l \sum_{j=0}^{i-1} j^l = CN^l \, \frac{(i+B)^{l+1} - B^{l+1}}{l+1}. \end{aligned}$$

En remplaçant ensuite i par $\frac{1}{N}(n_i - n_1 + N)$ dans cette inégalité, on a la conclusion. □

Remarque 2.4.4. Avec les mêmes hypothèses que dans le lemme précédent, il est clair qu'on a en fait que $v_L(n)$ est une fonction $\Theta(n^{l+1})$.

Théorème 2.4.5. *Si L est un langage régulier tel que $u_L(n)$ est $\Theta(n^l)$, pour $l \in \mathbb{N}$, alors la suite*

$$\left(\frac{v_n(L)}{n^{l+1}}\right)_{n\in\mathbb{N}}$$

converge vers une limite strictement positive. De plus, 1 est racine du polynôme caractéristique de la relation linéaire de récurrence satisfaite par $(u_n(L))_{n\in\mathbb{N}}$ avec une multiplicité supérieure ou égale à $l+1$.

Démonstration. Nous savons de la théorie générale des automates que la suite $(u_L(n))_{n\in\mathbb{N}}$ satisfait une relation linéaire de récurrence à coefficients constants (cf. la proposition 1.1.14). Le terme général de celle-ci peut s'écrire comme une somme finie $u_L(n) = \sum_i P_i(n) z_i^n$, où les P_i sont des polynômes et les z_i des nombres complexes distincts. Comme L est polynomial, on a nécessairement $|z_i| \leq 1$, pour tout i. De plus, au moins un des nombres z_i est de module 1. Soit d le plus haut degré des polynômes P_k correspondant aux nombres z_k de module 1. Fixons quelques notations. Soient $z_1 = e^{i\theta_1}, \ldots, z_r = e^{i\theta_r}$ les nombres de module 1 dont le polynôme correspondant P_k est de degré d, avec $\theta_i \not\equiv \theta_j (\mathrm{mod}\ 2\pi)$ si $i \neq j$. Le coefficient de n^d dans $P_k(n)$ est noté c_k. On note $z_{r+1}, \ldots, z_s$ les autres nombres. Ainsi, il vient

$$|u_L(n)| = n^d \left| \sum_{j=1}^{r} c_j\, e^{in\theta_j} + R_n \right|,$$

où

$$R_n = \frac{1}{n^d}\left(\sum_{j=1}^{r}(P_j(n) - c_j n^d) z_j^n + \sum_{j=r+1}^{s} P_j(n) z_j^n\right) \to 0 \text{ si } n \to +\infty.$$

Vu le lemme 2.4.2, il existe $\varepsilon > 0$ et une suite infinie d'entiers n tels que $|u_L(n)| \geq n^d\,(\varepsilon - |R_n|)$. Pour n suffisamment grand, on a $|R_n| \leq \frac{\varepsilon}{2}$ et donc

$$|u_L(n)| \geq \frac{\varepsilon}{2}\, n^d,$$

pour une infinité d'entiers n. Comme par hypothèse, $u_L(n)$ est $O(n^l)$, ceci implique que $d \leq l$. On peut donc réécrire $u_L(n)$ sous la forme

$$u_L(n) = \sum_{j=0}^{k} Q_j(n) e^{in\theta_j} + T(n),$$

où les Q_j sont des polynômes de degré inférieur ou égal à l, $\theta_i \not\equiv \theta_j (\mathrm{mod}\ 2\pi)$ si $i \neq j$, $\theta_0 = 0$ et $T(n) = \sum_{i:|z_i|<1} P_i(n)\, z_i^n$. Notons q_j le coefficient de n^l dans $Q_j(n)$, celui-ci pouvant éventuellement être nul. Ceci donne

$$u_L(n) = q_0 n^l + \sum_{j=1}^{k} q_j e^{in\theta_j} n^l + \sum_{j=0}^{k} (Q_j(n) - q_j n^l)\, e^{in\theta_j} + T(n).$$

Par conséquent,

$$\begin{aligned}\frac{v_n(L)}{n^{l+1}} &= \frac{1}{n^{l+1}}\sum_{p=0}^{n} u_L(p)\\ &= q_0\,\frac{\sum_{p=0}^{n} p^l}{n^{l+1}} + \sum_{j=1}^{k} q_j\,\frac{\sum_{p=0}^{n} e^{ip\theta_j}p^l}{n^{l+1}}\\ &\quad + \sum_{j=0}^{k}\frac{\sum_{p=0}^{n}\left(Q_j(p)-q_jp^l\right)\,e^{ip\theta_j}}{n^{l+1}} + \frac{\sum_{p=0}^{n}T(p)}{n^{l+1}}.\end{aligned}$$

Vu la formule 2.1, on a que $\lim_{n\to+\infty}\frac{\sum_{p=0}^{n}p^l}{n^{l+1}} = \frac{1}{l+1}$. Montrons que

$$\lim_{n\to+\infty}\frac{\sum_{p=0}^{n} e^{ip\theta_j}p^l}{n^{l+1}} = 0. \tag{2.2}$$

D'une part,

$$\left(z\frac{\partial}{\partial z}\right)^l \sum_{p=0}^{n} z^p = \sum_{p=0}^{n} p^l z^p.$$

D'autre part, on peut vérifier que

$$\left(z\frac{\partial}{\partial z}\right)^l \frac{z^{n+1}-1}{z-1} = \sum_{k=0}^{l} n^k\,\frac{z^n\,R_k(z)}{(z-1)^{l+1-k}} + \frac{R(z)}{(z-1)^{l+1}},$$

où les R_k et R sont des polynômes de degré inférieur ou égal à $l+1$. Si $z = e^{i\theta_j}$, alors les modules des fractions dans le membre de droite sont bornés, ce qui suffit. Comme les polynômes $Q_j(p) - q_jp^l$ sont de degré strictement inférieur à l, on a également

$$\lim_{n\to+\infty}\frac{\sum_{p=0}^{n}\left(Q_j(n)-q_jp^l\right)\,e^{ip\theta_j}p^l}{n^{l+1}} = 0.$$

Enfin, on montre que $\lim_{n\to+\infty}\frac{\sum_{p=0}^{n}T(p)}{n^{l+1}} = 0$. On considère un terme quelconque de $T(n)$, qu'on note simplement $P(n)z^n$. On a $|z| = \xi < 1$ et $P(n) = \sum_{i=0}^{\delta} a_in^i$, avec $\delta \in \mathbb{N}$. Alors

$$\left|\frac{1}{n^{l+1}}\sum_{p=0}^{n}P(p)z^p\right| \le \frac{1}{n^{l+1}}\sum_{p=0}^{n}\xi^p\sum_{i=0}^{\delta}|a_i|p^i \le \sum_{i=0}^{\delta}\frac{|a_i|}{n^{l+1}}\sum_{p=0}^{n}\xi^p p^i.$$

En procédant de la manière que pour montrer 2.2, on obtient bien encore une limite nulle. Finalement, $\lim_{n\to+\infty}\frac{v_n(L)}{n^{l+1}} = \frac{q_0}{l+1}$ et par le lemme 2.4.3, cette limite est strictement positive.

Pour la deuxième partie de la thèse, il suffit de remarquer que puisque $q_0 > 0$, 1 est racine du polynôme caractéristique de la relation de récurrence satisfaite par $(u_n(L))_{n\in\mathbb{N}}$ et comme il s'agit du coefficient de n^l dans le polynôme qui lui correspond, sa multiplicité est au moins $l+1$. □

Remarque 2.4.6. Le dernier point de la preuve du théorème 2.4.5 est un fait connu par ailleurs. Par exemple, la caractérisation des langages polynomiaux 2.1.13 permet de travailler directement sur un automate acceptant le langage polynomial considéré. Il est alors facile de vérifier que le théorème de Perron-Frobenius permet de tirer la même conclusion.

Chapitre 3

Multiplication par une constante

Une des questions relatives à l'étude des systèmes de numération abstraits est celle de la stabilité du caractère reconnaissable par opérations arithmétiques. Dans ce cadre, il semble naturel de commencer par étudier l'effet de la multiplication par une constante. En effet, si la multiplication par 2 ne conserve pas le caractère reconnaissable pour un système donné, on peut montrer qu'alors le graphe de l'addition n'est pas régulier. Nous détaillerons ceci plus loin.

Ce chapitre est divisé en quatre parties, organisées comme suit. Tout d'abord, nous regardons en détails le cas du système de numération abstrait construit sur le langage a^*b^*. Le résultat principal pour ce système est que la multiplication par une constante préserve le caractère reconnaissable si et seulement si cette constante est un carré parfait impair. Ensuite, nous élargissons notre étude au cas des systèmes de numération abstraits construits sur des langages polynomiaux quelconques. Nous donnons, dans le cas de ces systèmes, une condition nécessaire de la conservation du caractère reconnaissable par multiplication par une constante. Nous poursuivons avec quelques considérations pour le cas des systèmes construits sur des langages exponentiels, en séparant notre étude selon que le langage est de complémentaire polynomial ou exponentiel. Enfin, nous terminons par plusieurs remarques, notamment concernant le problème du changement de l'ordre sur l'alphabet du système.

3.1 Cas du langage a^*b^*

On place ici l'étude de la conservation du caractère reconnaissable dans le cas du système de numération abstrait $S = (a^*b^*, \{a, b\}, a < b)$. Pour ce système, il a été montré ([6]) que la multiplication par une constante préserve la S-reconnaissabilité si et seulement si cette constante est un carré parfait

impair. Nous en donnons ici la preuve originale des auteurs.

Le lemme suivant reste valable dans le cadre plus général des systèmes de numération abtraits construits sur un langage du type $a_1^* \cdots a_\ell^*$. Nous aborderons ce point plus en détails dans la chapitre 4. Nous y expliquerons notamment comment obtenir une nouvelle preuve du théorème 3.1.2.

Lemme 3.1.1. *Soit une matrice $A \in \mathbb{N}_\ell^\ell$ régulière. Pour $1 \leq i \leq \ell$, posons $h_i(\vec{n}) = A_{i1}n_1 + \cdots + A_{i\ell}n_\ell - b_i$, où $\vec{n} = (n_1, \ldots, n_\ell) \in \mathbb{N}^\ell$ et $b_i \in \mathbb{Z}$. Si les éléments de $\det(A).A^{-1}$ sont tous positifs ou nuls, alors le langage $L = \{a_1^{h_1} \cdots a_\ell^{h_\ell} \mid h_i := h_i(\vec{n}) \geq 0,\ 1 \leq i \leq \ell,\ \vec{n} \in \mathbb{N}^\ell\}$ est un sous-ensemble régulier de $a_1^* \cdots a_\ell^*$.*

Démonstration. Si $\vec{n} \in \mathbb{N}^\ell$ satisfait $h_i(\vec{n}) \geq 0$, alors

$$(A\vec{n})_i = A_{i1}n_1 + \cdots + A_{i\ell}n_\ell = u_i + b_i,$$

pour un certain $u_i \geq 0$. D'où

$$n_i = (A^{-1}A\,\vec{n})_i = \sum_{j=1}^{\ell} (A^{-1})_{ij}\,(u_j + b_j),\ u_j \geq 0.$$

Déterminons donc quels sont les vecteurs $\vec{u} = (u_1, \ldots, u_\ell) \in \mathbb{N}^\ell$ définissant ainsi des entiers n_i positifs ou nuls. Si $\det(A) < 0$, par hypothèse, les éléments de A^{-1} sont négatifs ou nuls et il existe seulement un nombre fini de tels vecteurs $\vec{u}$. On a alors que L est un langage fini et donc régulier. Supposons maintenant que $\det(A) > 0$. Alors les éléments de A^{-1} sont positifs ou nuls, et pour des u_j suffisamment grands, on a $n_i \geq 0$. Il reste encore à montrer que les n_i ainsi définis sont des entiers. Si $\mathcal{A}$ est la matrice des cofacteurs de A, on a que n_i est un entier si et seulement si $\sum_{j=1}^{\ell} \mathcal{A}_{ji}\,(u_j + b_j)$ est multiple de $\det(A)$. En posant $r_j \equiv u_j\,(\mod \det(A)) \in \{0, \ldots, \det(A) - 1\}$, pour tout $1 \leq j \leq \ell$, la condition devient que $\sum_{j=1}^{\ell} \mathcal{A}_{ji}\,(r_j + b_j)$ est multiple de $\det(A)$. Or il y a seulement un nombre fini de tels vecteurs $(r_1, \ldots, r_\ell)$. Ceci montre que le langage L est en fait une union finie de langages réguliers de la forme

$$a_1^{r_1 + s_1 \det(A)} \left(a_1^{\det(A)}\right)^* \cdots\ a_\ell^{r_\ell + s_\ell \det(A)} \left(a_\ell^{\det(A)}\right)^*,$$

où les s_i sont des entiers choisis suffisamment grands pour assurer que les n_i soient bien tous positifs ou nuls. □

Théorème 3.1.2. *Soit $S = (a^*b^*, \{a, b\}, a < b)$. Soit $\lambda \in \mathbb{N}$. La multiplication par λ conserve la S-reconnaissabilité si et seulement si λ est un carré parfait impair.*

Démonstration. Si λ n'est pas un carré parfait, alors la multiplication par λ ne conserve pas la S-reconnaissabilité. Ceci est une conséquence d'un résultat plus général (cf. théorème 3.2.4) qui sera exposé plus bas. Supposons

donc que $\lambda = \beta^2$, pour un entier β. Nous allons montrer comment diviser l'espace $\mathbb{N}^2$ des couples d'entiers positifs ou nuls en $\beta + 1$ régions R_i dans chacune desquelles on peut donner une formule explicite pour la fonction $M : \mathbb{N}^2 \to N^2$ définie par $M(p, q) = (r, t)$ si $\lambda \operatorname{val}_S(a^p b^q) = \operatorname{val}_S(a^r b^t)$.

Nous illustrons ce phénomène pour le cas $\beta = 5$. La figure 3.1 montre l'effet de la multiplication par 25 sur l'ensemble des entiers représentés par $a^{3i}b^{4j}$, $0 \leq i, j \leq 30$. Chaque point de coordonnée (p, q) correspond au mot $a^p b^q$. Les figures 3.2 et 3.3 montrent l'effet de la multiplication par 25 sur chacune des régions. La colonne de gauche (resp. de droite) contient les différentes régions avant (resp. après) multiplication. On voit alors plus clairement apparaître la régularité des images.

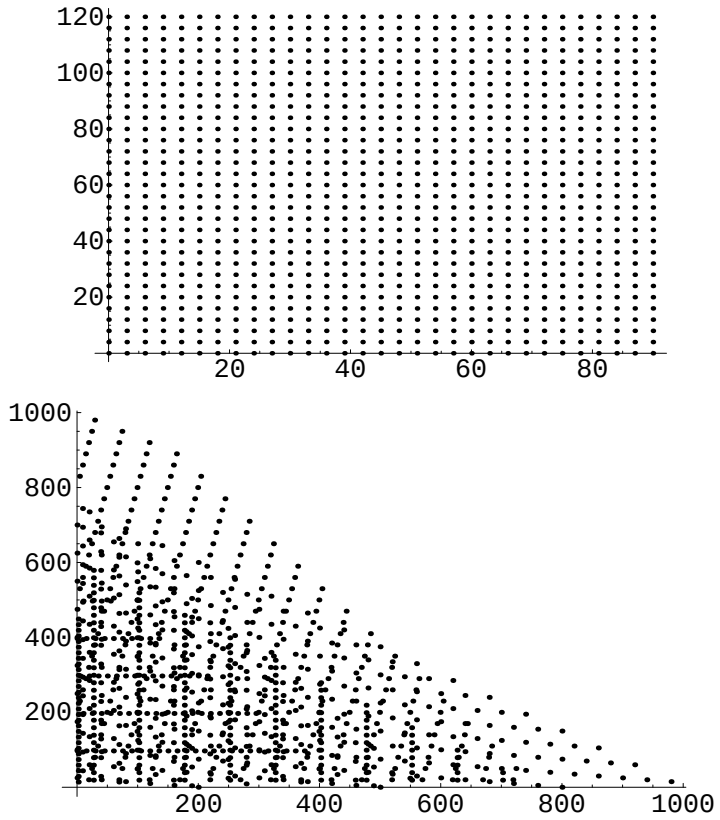

FIG. 3.1 – La multiplication par 25 dans a^*b^*.

En posant $k = p+q$ et $k' = r+t$, vu la forme de val_S (voir exemple 1.2.5), on obtient que le couple (r, t) vérifie $\lambda \operatorname{val}_S(a^p b^q) = \operatorname{val}_S(a^r b^t)$ si et seulement si

$$2r = 2p\lambda + k'^2 + 3k' - k^2\lambda - 3k\lambda. \tag{3.1}$$

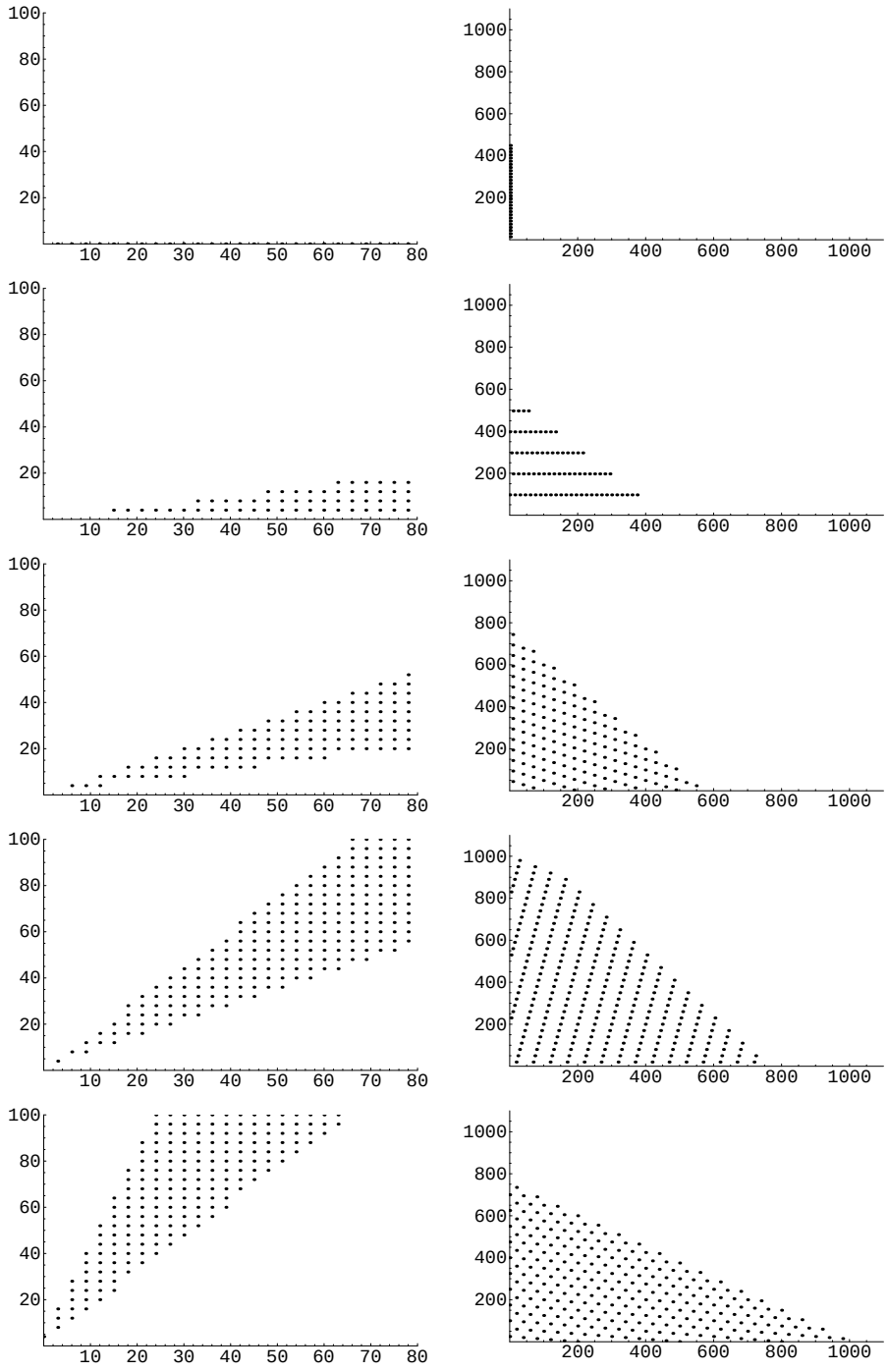

FIG. 3.2 – Les régions R_i avant et après multiplication par 25.

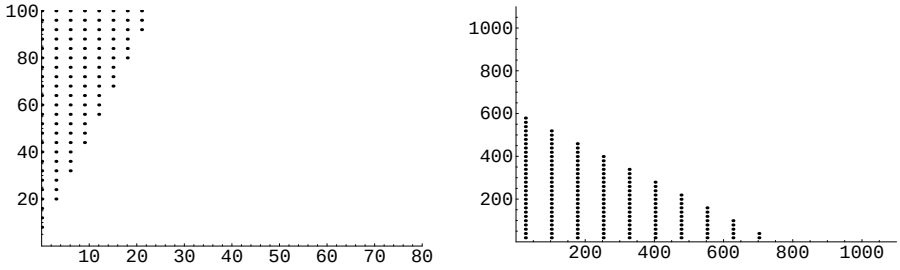

FIG. 3.3 – La dernière région R_i.

Montrons que nécessairement $k' \geq \beta k$. En effet, tout entier positif x a une représentation de longueur k si et seulement si $\sum_{i=1}^{k} i \leq x < \sum_{i=1}^{k+1} i$ et il est clair que $\sum_{i=1}^{\beta k} i \leq \beta^2 \sum_{i=1}^{k} i$. On peut donc écrire $k' = \beta k + u$, pour un certain entier $u \geq 0$. Montrons maintenant que, quelque soit la longueur de départ k donnée, on n'a en fait qu'un choix fini de tels entiers u possible. Les mots $a^p b^q$ d'une même longueur k prennent comme valeurs

$$\mathrm{val}_S(a^k) \leq \mathrm{val}_S(a^p b^q) \leq \mathrm{val}_S(b^k)$$

et vu la forme de val_S, on trouve que

$$\lambda k(k+1) \leq 2\lambda \mathrm{val}_S(a^p b^q) \leq \lambda k(k+3)$$
$$k'(k'+1) \leq 2\mathrm{val}_S(a^r b^t) \leq k'(k'+3)$$

et donc que

$$k'(k'+1) \leq \beta^2 k(k+3) \tag{3.2}$$
$$\beta^2 k(k+1) \leq k'(k'+3). \tag{3.3}$$

En remplaçant k' par $\beta k + u$ d'abord dans l'inégalité 3.2, on obtient après calcul que $u < \frac{3\beta - 1}{2}$. En particulier, u est une fonction bornée. L'inégalité 3.3 donne alors ensuite $u \geq \frac{\beta - 3}{2}$.

1) Supposons que β est impair. Les inégalités obtenues donnent alors

$$\left\lfloor \frac{\beta}{2} \right\rfloor - 1 \leq u < \left\lfloor \frac{\beta}{2} \right\rfloor + \beta - 1$$

et on peut finalement écrire $u = \lfloor \beta/2 \rfloor + i$, avec $-1 \leq i \leq \beta - 1$ et, pour chaque i, vu la condition 3.1, on obtient

$$\begin{cases} r = r_i(p,q) := \beta(i+1)p - \beta(\beta - i - 1)q + \frac{1}{8}\left((\beta + 2i + 2)^2 - 9\right) \\ t = t_i(p,q) := -\beta i p + \beta(\beta - i)q - \frac{1}{8}\left((\beta + 2i)^2 - 1\right). \end{cases}$$

Ces équations définissent, avec les conditions r, $t \geq 0$, $\beta + 1$ régions

$$R_i = \left\{ (p,q) \in \mathbb{N}^2 \mid \left| \mathrm{rep}_S\left(\beta^2 \mathrm{val}_S(a^p b^q)\right) \right| = \beta(p+q) + \left\lfloor \frac{\beta}{2} \right\rfloor + i \right\}$$

qui partitionnent $\mathbb{N}^2$. On note R'_i les parties de a^*b^* correspondantes, i.e. les ensembles des mots a^pb^q tels que $(p,q) \in R_i$. Les sous-ensembles réguliers de a^*b^* sont les unions finies d'ensembles de la forme

$$D = \left\{ a^{y+fz} b^{w+gx} \mid f,\, g \geq 0 \right\},\ w,\, x,\, y,\, z \geq 0.$$

Les ensembles

$$D_i = \mathrm{rep}_S\left(\lambda\, \mathrm{val}_S(D \cap R'_i)\right) = \left\{ a^{r_i(y+fz,w+gx)} b^{t_i(y+fz,w+gx)} \mid f,\, g \geq 0 \right\}$$

sont de la forme de L dans le lemme 3.1.1, avec

$$A = \begin{pmatrix} z\beta(i+1) & -x\beta(\beta-i-1) \\ -z\beta i & x\beta(\beta-i). \end{pmatrix}$$

et

$$h_1(f,\, g) = z\beta(i+1)f - x\beta(\beta-i-1)\, g + b_1$$
$$h_2(f,\, g) = -z\beta i f + x\beta(\beta-i)\, g + b_2.$$

Puisque $\det(A) = xz\beta^3$, les hypothèses du lemme 3.1.1 sont satisfaites lorsque $i \geq 0$ et $xz \neq 0$ et dans ces conditions, D_i est régulier. Prouvons maintenant la régularité de D_i dans les autres cas.

Si $x = z = 0$, alors $D_i = \{a^{b_1}b^{b_2}\}$, qui est bien sûr régulier. On peut donc supposer avoir $x \neq 0$ ou $z \neq 0$.

Supposons $i \geq 0$ avec $xz = 0$. Supposons que $x = 0$. Si $i \neq 0$, la double condition $r,\, t \geq 0$ impose une borne supérieure pour f. Ainsi D_i est fini, donc régulier. Si $i = 0$, alors D_i est de la forme $\{a^{z\beta f+b_1}b^{b_2} \mid f \geq -b_1/\beta z\}$, qui est bien régulier. Le cas $z = 0$ se traite de manière analogue.

Il reste encore à vérifier la régularité de D_{-1}. Si $x = 0$, alors $D_{-1} = \{a^{b_1}b^{\beta f z+b_2} \mid f \geq -b_2/\beta z\}$, qui est régulier. Sinon, la condition $r \geq 0$ donne une borne supérieure pour g, celui-ci ne pouvant ainsi prendre qu'un nombre fini de valeurs $g_1, \ldots, g_n$. Si $z = 0$, D_{-1} est fini, donc régulier. Sinon, pour tout $1 \leq j \leq n$, la condition $t \geq 0$ donne une borne inférieure k_j pour f et

$$D_{-1} = \bigcup_{j=1}^{n} \left\{ a^{-\beta^2 g_j x + b_1} b^{\beta f z + \beta(\beta+1) g_j x + b_2} \mid f \geq k_j \right\},$$

qui est bien régulier.

2) Considérons maintenant le cas où β est pair. Ecrivons $\beta = 2\gamma$. On a alors

$$\gamma - 1 \leq u < 3\gamma - 1,$$

ce qui nous permet d'écrire $u = \gamma + i$, avec $-1 \leq i \leq \beta - 1$ et comme pour le cas impair, on obtient des équations

$$\begin{cases} r = r_i(p,q) := \beta(i+\frac{3}{2})p - \beta(\beta-i-\frac{3}{2})q + \frac{1}{8}\left((\beta+2i+3)^2 - 9\right) \\ t = t_i(p,q) := -\beta(i+\frac{1}{2})p + \beta(\beta-i-\frac{1}{2})q - \frac{1}{8}\left((\beta+2i)^2 + 4i + 2\beta\right), \end{cases}$$

définissant, avec les conditions $r,\ t \geq 0$, $\beta + 1$ régions

$$R_i = \left\{ (p,q) \in \mathbb{N}^2 \mid \left| \mathrm{rep}_S \left(\beta^2 \mathrm{val}_S(a^p b^q) \right) \right| = \beta(p+q) + \frac{\beta}{2} + i \right\}$$

qui partitionnent $\mathbb{N}^2$. On peut vérifier que pour tout p suffisamment grand, $(p,0) \in R_{-1}$. Dès lors, $\mathrm{rep}_S\left(\lambda\, \mathrm{val}_S(a^p)\right) = a^r b^t$, avec

$$\begin{cases} r = \gamma p + \frac{1}{2}\gamma(\gamma+1) - 1 \\ t = \gamma p - \frac{1}{2}\gamma(\gamma-1). \end{cases}$$

Ceci montre que le langage $\mathrm{rep}_S\left(\mathrm{val}_S(a^*)\right)$ n'est pas régulier et par conséquent, la multiplication par β^2 ne conserve pas la S-reconnaissabilité. □

Signalons tout de même qu'à l'origine (voir [6]), le cas où le multiplicateur n'est pas un carré parfait avait été directement montré comme réfutable, sur base de la théorie sur les équations de Pell et du fait que si un langage est régulier, l'ensemble des longueurs de ses mots est ultimement périodique (cf. la proposition 1.1.16).

On peut considérer $L \subseteq (\Sigma \times \Delta)^*$ comme un langage avec comme convention $(x,y)(x',y') = (xx', yy')$, les lettres étant les couples (σ, σ'), $\sigma \in \Sigma$, $\sigma' \in \Delta$. Muni de cette opération de concaténation, $(\Sigma \times \Delta)^*$ a bien une structure de monoïde avec $(\varepsilon, \varepsilon)$ pour neutre.

Définition 3.1.3. Soient Σ et Δ deux alphabets et des mots $x \in \Sigma^*$ et $y \in \Delta^*$. Soit $\$ \notin \Sigma \cup \Delta$. Si $|x| = |y| + i$, $i \in \mathbb{N}$, alors $(x,y)^{\$} = (x, \$^i y)$ et si $|y| = |x| + i$, $i \in \mathbb{N}$, alors $(x,y)^{\$} = (\$^i x, y)$.

Définition 3.1.4. Une relation R sur $\Sigma^* \times \Delta^*$ est *régulière* si $R^{\$}$ est un langage régulier sur l'alphabet $\Sigma \times \Delta$.

Définition 3.1.5. Une application $f : \Sigma^* \to \Delta^*$ est *régulière* si son graphe $\widehat{f} = \{(x, f(x))^{\$} \mid x \in \Sigma^*\}$ est un langage régulier sur l'alphabet $\Sigma \times \Delta$.

Ces définitions s'étendent naturellement à des n-uples de mots.

Corollaire 3.1.6. *Dans le cas du système abstrait $S = (a^*b^*, \{a,b\}, a < b)$, l'addition n'est pas une application régulière.*

Démonstration. Par le théorème 3.1.2, on peut trouver une partie X de $\mathbb{N}$ qui est S-reconnaissable mais telle que $2X$ n'est pas S-reconnaissable. Procédons par l'absurde et supposons que le graphe de l'addition

$$\widehat{\mathcal{G}} = \{(\mathrm{rep}_S(x), \mathrm{rep}_S(y), \mathrm{rep}_S(x+y))^{\$} \mid x,\ y \in \mathbb{N}\}$$

soit un langage régulier. Comme bien sûr,

$$A = \{(\mathrm{rep}_S(x), \mathrm{rep}_S(x), w)^{\$} \mid x \in X,\ w \in \{a,b\}^*\}$$

est régulier, le langage

$$A \cap \widehat{\mathcal{G}} = \{(\mathrm{rep}_S(x), \mathrm{rep}_S(x), \mathrm{rep}_S(2x))^{\$} \mid x \in X\}$$

serait alors aussi régulier. Si p_3 est l'homomorphisme canonique de projection sur la troisième composante d'un vecteur, on aurait alors aussi que le langage $p_3(A \cap \widehat{\mathcal{G}}) = \mathrm{rep}_S(2X)$ serait régulier, ce qui est une contradiction. □

Remarque 3.1.7. Nous voudrions attirer l'attention sur le fait que le corollaire 3.1.6 ne prouve pas que l'addition ne conserve pas la S-reconnaissabilité. En effet, si le fait qu'une application soit régulière implique bien qu'elle préserve la régularité, le contraire n'est pas vrai en général.

Nous donnons ici un exemple d'une application de graphe non régulier mais qui préserve la régularité.

Exemple 3.1.8. Considérons le morphisme f sur l'alphabet unaire $\Sigma = \{\sigma\}$ défini par $f(\sigma) = \sigma^2$. Bien sûr, f préserve la régularité. Son graphe est donné par $\widehat{f} = \{(\sigma^n, \sigma^{2n})^{\$} \mid n \in \mathbb{N}\} = \{(\$^n\sigma^n, \sigma^{2n}) \mid n \in \mathbb{N}\}$. Le lemme de la pompe 1.1.17 montre alors que $\widehat{f}$ ne peut être régulier.

3.2 Cas des langages polynomiaux

Dans le cas d'un système de numération abstrait construit sur un langage polynomial, on peut donner une condition nécessaire pour avoir la conservation du caractère reconnaissable par multiplication par une constante. Nous séparons la démonstration de celle-ci en deux parties. Nous considérons en premier lieu le cas des langages de complexité exactement polynomiale et dans un deuxième temps seulement, nous étudierons le cas d'un langage polynomial quelconque. Les résultats exposés dans cette partie sont issus de [7].

3.2.1 Langages exactement polynomiaux

Définition 3.2.1. Un langage est *exactement polynomial* si sa fonction de complexité est un polynôme.

Lemme 3.2.2. *Soit $f : \mathbb{N} \to \mathbb{N}$ une fonction strictement croissante telle que $f(\mathbb{N})$ est une union finie de progressions arithmétiques et soient*

$$y_0 \in f(\mathbb{N}) \text{ et } \Gamma \in \mathbb{N}_0 \text{ tels que } \forall y \geq y_0,\ y \in f(\mathbb{N}) \Leftrightarrow y + \Gamma \in f(\mathbb{N}).$$

Soit $k = f^{-1}(y_0 + \Gamma) - f^{-1}(y_0)$. Alors pour tout $x \geq f^{-1}(y_0)$ et pour tout $n \in \mathbb{N}$, on a $f(x + nk) = f(x) + n\Gamma$.

Démonstration. C'est une simple vérification par double récurrence. □

Lemme 3.2.3. *Si H est un polynôme sur $\mathbb{Q}$ tel que $H(\mathbb{N}_0) \subseteq \mathbb{Z}$, alors $H(\mathbb{Z}) \subseteq \mathbb{Z}$.*

Démonstration. On procède par induction sur le degré de H. Si deg(H)=1, alors $H = aX + b$, avec a, $b \in \mathbb{Z}$ et $H(\mathbb{Z}) \subseteq \mathbb{Z}$. Supposons que le résultat est vrai pour les polynômes de degré $k \geq 1$. Soit H un polynôme de degré $k+1$. Alors $R(X) = H(X+1) - H(X)$ est un polynôme de degré k et $R(\mathbb{N}_0) \subseteq \mathbb{Z}$. Par hypothèse de récurrence, $R(\mathbb{Z}) \subseteq \mathbb{Z}$. Ainsi, $H(0) = H(1) - R(0) \in \mathbb{Z}$ et on conclut par induction sur x pour $x < 0$, puisque $H(x) = H(x+1) - R(x)$, pour tout x. □

Théorème 3.2.4. *Soit $L \subseteq \Sigma^*$ un langage régulier tel que*

$$u_L(n) = \begin{cases} a_l n^l + \cdots + a_1 n + a_0, & \text{si } n > 0 \\ 1, & \text{sinon} \end{cases}$$

avec $a_i \in \mathbb{Q}$, $0 \leq i \leq l$, et $a_l > 0$. Soit $S = (L, \Sigma, <)$ un système de numération abstrait construit sur ce langage. Si $\lambda \in \mathbb{N} \setminus \{n^{l+1} \mid n \in \mathbb{N}\}$, alors on peut trouver une partie X de $\mathbb{N}$ telle que $\mathrm{rep}_S(X)$ est régulier et $\mathrm{rep}_X(\lambda X)$ n'est pas régulier.

Démonstration. Montrons tout d'abord comment construire un polynôme $P \in \mathbb{Q}[X]$ de degré $l+1$ tel que $P(0) = 0$ et $P(n+1) = P(n) + u_L(n)$, pour tout $n \geq 1$. On note $P = b_{l+1}X^{l+1} + \cdots + b_1 X$. Considérons le polynôme $Q = a_l X^l + \cdots + a_1 X + a_0$. En identifiant les coefficients dans l'identité polynomiale $Q(X) = P(X+1) - P(X)$, on obtient que les coefficients de P doivent satisfaire les $l+1$ équations $a_i = \sum_{j=i+1}^{l+1} \binom{j}{i} b_j$, $0 \leq i \leq l$. Celles-ci forment un système triangulaire à $l+1$ inconnues b_j, $1 \leq j \leq l+1$. Le polynôme P est donc ainsi complètement déterminé et est bien à coefficients rationnels.

Comme pour tout $n \in \mathbb{N}_0$, $Q(n) = u_L(n) \in \mathbb{N}$, le lemme 3.2.3 donne $Q(\mathbb{Z}) \subseteq \mathbb{Z}$. Vu l'identité polynomiale $Q(X) = P(X+1) - P(X)$, on a alors aussi $P(\mathbb{Z}) \subseteq \mathbb{Z}$. Comme on a aussi $P(n+1) - P(n) = u_L(n) > 0$, pour n suffisamment grand, il existe $N \in \mathbb{N}_0$ tel que pour tout $n \geq N$, $P(n) > 0$.

Soit $x \in \mathbb{N}_0$. Remarquons que

$$\begin{aligned} |\mathrm{rep}_S(x)| = n &\Leftrightarrow v_{n-1}(L) \leq x \leq v_n(L) - 1 \\ &\Leftrightarrow P(n) - a_0 + 1 \leq x \leq P(n+1) - a_0. \end{aligned} \tag{3.4}$$

En effet, $P(1) = a_0$ et

$$v_n(L) = \sum_{i=0}^{n} u_L(i) = 1 + \sum_{i=1}^{n} (P(i+1) - P(i)) = 1 + P(n+1) - P(1).$$

Le polynôme P ainsi construit est tel que $\{P(n) \mid n \geq N\}$ est, à un nombre fini d'éléments près, un translaté de $\mathrm{val}_S(\mathrm{Min}(S))$. Comme $\mathrm{Min}(S)$

est S-reconnaissable (cf. la proposition 1.2.10) et comme le caractère reconnaissable est stable par translation (cf. la proposition 1.2.11), on obtient que $\{P(n) \mid n \geq N\}$ est une partie S-reconnaissable de $\mathbb{N}$.

Soit $\lambda \in \mathbb{N} \setminus \{n^{l+1} \mid n \in \mathbb{N}\}$. Notre but est de montrer que l'ensemble $\lambda\{P(n) \mid n \geq N\}$ n'est pas S-reconnaissable. Nous procédons par étapes.

Nous montrons d'abord que, pour n suffisamment grand,

$$n \leq |\mathrm{rep}_S(\lambda P(n))| < \left\lfloor \lambda^{\frac{1}{l}} n \right\rfloor \leq \lambda^{\frac{1}{l}} n. \tag{3.5}$$

La première inégalité est évidente vu la condition 3.4. Montrons la seconde inégalité. Ecrivons $P = b_{l+1}X^{l+1} + P'$, avec $b_{l+1} > 0$ et Q un polynôme de degré inférieur ou égal à l. Ainsi

$$P\left(\left\lfloor \lambda^{\frac{1}{l}} n \right\rfloor\right) - \lambda P(n) - a_0 + 1 > b_{l+1}\left(\left(\lambda^{\frac{1}{l}} n - 1\right)^{l+1} - \lambda n^{l+1}\right) + O(n^l).$$

Pour n suffisamment grand, l'expression de droite est strictement positive et on conclut encore grâce à la condition 3.4.

Ensuite, nous montrons que, pour n suffisamment grand,

$$|\mathrm{rep}_S(\lambda P(n+1))| > |\mathrm{rep}_S(\lambda P(n))|.$$

Vu la condition 3.4, cela revient à montrer que $\lambda P(n+1) > P(i+1) - a_0$, pour n suffisamment grand, où $i = |\mathrm{rep}_S(\lambda P(n))|$. Par construction du polynôme P et vu la définition de i,

$$\lambda P(n+1) = \lambda P(n) + \lambda u_L(n) > P(i) - a_0 + \lambda u_L(n).$$

Par conséquent, il suffit de montrer que $P(i)+\lambda u_L(n) > P(i+1)$. En utilisant les inégalités 3.5, on obtient

$$\begin{aligned}
& \lambda u_L(n) - (P(i+1) - P(i)) \;=\; \lambda u_L(n) - u_L(i) \\
= \;& a_l(\lambda n^l - i^l) + \cdots + a_k(\lambda n^k - i^k) + \cdots + a_0(\lambda - 1) \\
= \;& n^l \left(\underbrace{a_l\left(\lambda - \left(\frac{i}{n}\right)^l\right)}_{>0} + \cdots + \underbrace{\frac{a_k}{n^{l-k}}}_{\to 0} \underbrace{\left(\lambda - \left(\frac{i}{n}\right)^k\right)}_{\text{borné}} + \cdots + \underbrace{\frac{a_0}{n^l}}_{\to 0}(\lambda - 1) \right)
\end{aligned}$$

et donc on a bien que $\lim_{n\to+\infty} \frac{\lambda u_L(n) - u_L(i)}{n^l} > 0$.

Procédons par l'absurde et supposons que $\mathrm{rep}_S(\lambda\{P(n) \mid n \geq N\})$ soit régulier. Alors, vu la proposition 1.1.16, $|\mathrm{rep}_S(\lambda\{P(n) \mid n \geq N\})|$ devrait être une union finie de progressions arithmétiques. On pourrait donc trouver $l_0 \in |\mathrm{rep}_S(\lambda\{P(n) \mid n \geq N\})|$ et $\Gamma \in \mathbb{N}_0$ tels que pour tout $l \geq l_0$, on ait

$$l \in |\mathrm{rep}_S(\lambda\{P(n) \mid n \geq N\})| \;\Leftrightarrow\; l + \Gamma \in |\mathrm{rep}_S(\lambda\{P(n) \mid n \geq N\})|.$$

Vu ce qui précède, la fonction $|\mathrm{rep}_S(\lambda P(\cdot))|$ devient strictement croissante à partir d'un certain cran, et donc, pour n suffisamment grand, on aurait que $|\mathrm{rep}_S(\lambda P(n))| > l_0$. Par le lemme 3.2.2, on pourrait trouver $k \in \mathbb{N}_0$ tel que pour n suffisamment grand et pour tout $\alpha \in \mathbb{N}$, on ait

$$|\mathrm{rep}_S(\lambda P(n+\alpha k))| = |\mathrm{rep}_S(\lambda P(n))| + \alpha\Gamma = i(n) + \alpha\Gamma.$$

Vu les inégalités 3.4, ceci donnerait

$$P(i+\alpha\Gamma) - a_0 + 1 \leq \lambda P(n+\alpha k) \leq P(i+\alpha\Gamma+1) - a_0.$$

D'une part, la première de ces inégalités donne

$$\lambda P(n+\alpha k) - P(i+\alpha\Gamma) + a_0 - 1 \geq 0,$$

pour tout $\alpha \in \mathbb{N}$. D'où le coefficient de α^{l+1} dans le membre de gauche doit être positif ou nul, c'est à dire $b_{l+1}(\lambda k^{l+1} - \Gamma^{l+1}) \geq 0$. Remarquons que ce coefficient s'annule si et seulement si $\lambda = (\Gamma/k)^{l+1}$. Montrons que ceci n'arrive jamais. En effet, Γ/k ne peut pas être entier par hypothèse. Mais alors on aurait $\Gamma/k \in \mathbb{Q} \setminus \mathbb{N}$ et $\lambda = (\Gamma/k)^{l+1}$ ne serait pas un entier, ce qui ne se peut pas non plus. Finalement, on a obtenu que $b_{l+1}(\lambda k^{l+1} - \Gamma^{l+1}) > 0$, et donc que

$$\lambda k^{l+1} > \Gamma^{l+1}.$$

D'autre part, la seconde inégalité donne

$$\lambda P(n+\alpha k) - P(i+\alpha\Gamma+1) + a_0 \leq 0,$$

pour tout $\alpha \in \mathbb{N}$. Le coefficient de α^{l+1} dans le membre de gauche, qui est encore $b_{l+1}(\lambda k^{l+1} - \Gamma^{l+1})$, doit alors être strictement négatif, et donc

$$\lambda k^{l+1} < \Gamma^{l+1},$$

d'où une contradiction. □

Ce résultat montre en particulier (voir théorème 3.1.2) que seuls les carrés parfaits sont candidats pour préserver la S-reconnaissabilité après multiplication, dans le cas du système abstrait $S = (a^*b^*, \{a, b\}, a < b)$.

La proposition suivante montre que l'argument utilisé dans la démonstration du théorème 3.2.4 ne peut malheureusement pas être étendu au cas des constantes $\lambda \in \{n^{l+1} \mid n \in \mathbb{N}\}$. En effet, il montre en particulier que si un langage L a pour fonction de complexité un polynôme de degré l et si P est le polynôme construit dans la preuve du théorème 3.2.4, alors l'ensemble $|\mathrm{rep}_S\left(\{\beta^{l+1}P(n) \mid n \geq N\}\right)|$ est une union finie de progressions arithmétiques, pour tout entier β.

Proposition 3.2.5. *Dans les mêmes conditions que le théorème 3.2.4, on peut trouver une constante $C \in \mathbb{Z}$ telle que pour n suffisamment grand, on a $|\mathrm{rep}_S\left(\beta^{l+1}P(n)\right)| = \beta n + C$, pour tout $\beta \in \mathbb{N}$.*

Démonstration. Soit $\beta \in \mathbb{N}$. Considérons le même polynôme P que dans la preuve du théorème 3.2.4. Vu la condition 3.4, nous devons trouver $C \in \mathbb{Z}$ tel que pour tout n suffisamment grand, on a

$$P(\beta n + C + 1) - a_0 - \beta^{l+1} P(n) \geq 0 \tag{3.6}$$

$$\beta^{l+1} P(n) - P(\beta n + C) + a_0 - 1 \geq 0. \tag{3.7}$$

Dans chacune des expressions de gauche de 3.6 et de 3.7, les termes en n^{l+1} se suppriment. Il s'agit donc dans les deux cas de polynômes de degré l. Si on pose $a_l = b_{l+1}(l+1) > 0$, le coefficient de n^l dans 3.6 est

$$\beta^l \left(a_l (C+1) + b_l (1-\beta) \right), \tag{3.8}$$

qui est une fonction linéaire strictement croissante de C s'annulant uniquement en $C = C_1 := \frac{b_l(\beta-1)-a_l}{a_l}$. Parallèlement, le coefficient de n^l dans 3.7 est

$$-\beta^l \left(a_l \, C + b_l (1-\beta) \right), \tag{3.9}$$

qui est une fonction linéaire strictement décroissante de C et s'annulant uniquement en $C = C_2 := \frac{b_l(\beta-1)}{a_l} = C_1 + 1$. Cette situation est illustrée par la figure 3.4.

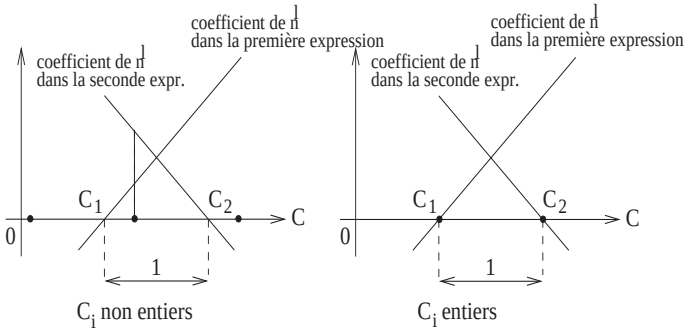

FIG. 3.4 – Les coefficients de n^l dans 3.6 et 3.7.

Si C_1 et C_2 ne sont pas des entiers, alors on peut trouver $C \in]C_1, C_2[\cap \mathbb{Z}$ tel que les coefficients 3.8 et 3.9 soient tous les deux strictement positifs. Dans ce cas, les inégalités 3.6 et 3.7 sont bien satisfaites, pour n suffisamment grand. Sinon, les seuls choix possibles pour la constante C sont $C = C_1$ et $C = C_2$. En effet, pour tout autre entier C, un des deux coefficients est forcément strictement négatif. Si $C = C_1$, alors le coefficient 3.9 est strictement positif et l'inégalité 3.7 est satisfaite pour n suffisamment grand. De même, si $C = C_2$, alors le coefficient 3.8 est strictement positif et l'inégalité 3.6 est satisfaite pour n suffisamment grand. Pour $1 \leq i \leq l-1$, le coefficient de n^i dans 3.6 avec $C = C_1$ est l'opposé du coefficient de n^i dans 3.7 avec $C = C_2$. Le terme

indépendant de 3.6 avec $C = C_1$ est $P(C_2) - a_0$ et celui de 3.7 avec $C = C_2$ est $-P(C_2) + a_0 - 1$. D'où, si l'expression 3.6 avec $C = C_1$ s'écrit sous la forme

$$A_{l-1}n^{l-1} + \cdots + A_1 n + P(C_2) - a_0,$$

alors l'expression 3.7 avec $C = C_2$ s'écrit

$$-A_{l-1}n^{l-1} - \cdots - A_1 n - P(C_2) + a_0 - 1.$$

Si les coefficients A_i ne sont pas tous nuls, on pose $j = \max\{i \mid A_i \neq 0\}$. Si $A_j > 0$, alors on prend $C = C_1$ et si $A_j < 0$, on choisit $C = C_2$. Sinon, on a $A_i = 0$, pour tout $1 \leq i \leq l-1$. Si $P(C_2) - a_0 \geq 0$, alors on prend $C = C_1$. Sinon, on a $-P(C_2) + a_0 > 0$ et comme, lors de la construction du polynôme P, on avait montré que $P(\mathbb{Z}) \subseteq \mathbb{Z}$, il s'agit d'un entier. Ceci implique que $-P(C_2) + a_0 - 1 \geq 0$ et on choisit donc $C = C_2$. □

3.2.2 Langages polynomiaux quelconques

Nous démontrons à présent la condition nécessaire annoncée dans le cas d'un système de numération abstrait construit sur un langage polynomial quelconque.

Théorème 3.2.6. *Soit $L \subseteq \Sigma^*$ un langage régulier tel que $u_L(n)$ est $\Theta(n^l)$, pour $l \in \mathbb{N}$. Soit $S = (L, \Sigma, <)$ un système de numération abstrait construit sur ce langage. Si $\lambda \in \mathbb{N} \setminus \{n^{l+1} \mid n \in \mathbb{N}\}$, alors on peut trouver une partie X de $\mathbb{N}$ telle que $\mathrm{rep}_S(X)$ est régulier et $\mathrm{rep}_S(\lambda X)$ n'est pas régulier.*

Démonstration. Comme $u_L(n)$ est $\Theta(n^l)$ par hypothèse, on peut trouver une suite $(n_i)_{i\in\mathbb{N}}$ et une constante $b_0 > 0$ telles que $u_L(n_i) \geq b_0 n_i^l$, pour tout $i \in \mathbb{N}$. Vu le théorème 2.1.13 de caractérisation des langages réguliers polynomiaux et la remarque 2.1.8, on peut choisir la suite $(n_i)_{i\in\mathbb{N}}$ telle que $n_i = n_0 + iC$, pour tout $i \in \mathbb{N}$, avec $C \in \mathbb{N}_0$. Remarquons qu'alors on a $n + 1 \leq |\mathrm{rep}_S(v_n(L))| \leq n + C + 1$, pour n suffisamment grand. En effet, parmi C valeurs de $u_n(L)$ consécutives, au moins une est strictement positive. Vu le théorème 2.4.5, la suite $\left(v_n(L)/n^{l+1}\right)_{n\in\mathbb{N}}$ converge vers une limite strictement positive. Si a est cette limite, alors pour tout $K > a$ et pour tout $J < a$, on peut trouver des entiers n_K et n_J tels que

$$v_n(L) \leq Kn^{l+1},\ n \geq n_K, \tag{3.10}$$

$$v_n(L) \geq Jn^{l+1},\ n \geq n_J. \tag{3.11}$$

Fixons de tels K et J.

Supposons dans un premier temps que $\lambda > \left(\frac{K}{J}\right)^l$. Montrons que, pour n suffisamment grand,

$$n + 1 \leq |\mathrm{rep}_S(\lambda v_n(L))| \leq \left\lceil \lambda^{\frac{1}{l}} n \right\rceil + C - 1 < \lambda^{\frac{1}{l}} n + C. \tag{3.12}$$

Il suffit de montrer que $\lambda v_n(L) < v_{\lceil \lambda^{1/l} n \rceil + C - 1}(L)$. Pour n suffisamment grand, vu l'inégalité 3.11, on a

$$v_{\lceil \lambda^{1/l} n \rceil}(L) \geq J\left(\left\lceil \lambda^{\frac{1}{l}} n \right\rceil\right)^{l+1} \geq J\left(\lambda^{\frac{1}{l}} n\right)^{l+1}.$$

Ensuite l'inégalité 3.10 donne $\lambda v_n(L) \leq \lambda K n^{l+1} < J\lambda^{(l+1)/l} n^{l+1}$, ce qui suffit puisque la fonction $n \mapsto v_n$ est croissante.

Par hypothèse, $u_L(n)$ est $O(n^l)$ et on peut donc trouver une constante $b_1 \geq b_0$ telle que $u_L(n) \leq b_1 n^l$, pour n suffisamment grand. Soit s un entier tel que $sb_0 > b_1$. Nous montrons maintenant que la fonction $i \mapsto |\mathrm{rep}_S(\lambda v_{n_{si}-1})|$ est strictement croissante à partir d'un certain cran. On doit montrer que, pour i suffisamment grand, on a

$$|\mathrm{rep}_S(\lambda v_{n_{s(i+1)}-1})| = |\mathrm{rep}_S(\lambda v_{n_{si}+sC-1})| > |\mathrm{rep}_S(\lambda v_{n_{si}-1})|.$$

Si $g = |\mathrm{rep}_S(\lambda v_{n_{si}-1})|$, alors $v_{g-1}(L) \leq \lambda v_{n_{si}-1} < v_g(L)$ et la thèse devient

$$\lambda v_{n_{si}+sC-1} = \lambda v_{n_{si}-1} + \lambda \sum_{j=0}^{sC-1} u_L(n_{si}+j) \geq v_g(L) = v_{g-1}(L) + u_L(g).$$

Il suffit alors de montrer que $\lambda \sum_{j=0}^{sC-1} u_L(n_{si}+j) \geq u_L(g)$. Premièrement, vu l'inégalité 3.12, on a $g < \lambda^{1/l}(n_{si}-1) + C$ et donc

$$u_L(g) \leq b_1 g^l < b_1 \left(\lambda^{\frac{1}{l}}(n_{si}-1) + C\right)^l,$$

pour i suffisamment grand. Deuxièmement, on a

$$\lambda \sum_{j=0}^{sC-1} u_L(n_{si}+j) \geq \lambda \sum_{j=0}^{s-1} u_L(n_{si}+jC) \geq \lambda\, b_0\, s\, n_{si}^l > \lambda\, b_1\, n_{si}^l,$$

ce qui suffit.

Considérons l'ensemble d'entiers $X = \{v_{n_{si}-1} \mid i \in \mathbb{N}\}$. Comme pour tout $i \in \mathbb{N}$, on a $u_L(n_{si}) > 0$, $\mathrm{rep}_S(v_{n_{si}-1})$ est le premier mot de longueur $n_{si} = n_0 + siC$ de L. D'où

$$\mathrm{rep}_S(X) = \mathrm{Min}(S) \cap \Sigma^{n_0} \left(\Sigma^{sC}\right)^*$$

et, vu la proposition 1.2.10, X est S-reconnaissable. Notre but est de montrer que l'ensemble λX n'est pas S-reconnaissable. Procédons par l'absurde et supposons qu'il le soit. Alors, vu la proposition 1.1.16, $|\mathrm{rep}_S(\lambda X)|$ serait une union finie de progressions arithmétiques. Vu ce qui précède, on peut appliquer le lemme 3.2.2 à la fonction $i \mapsto |\mathrm{rep}_S(\lambda v_{n_{si}-1})|$. Celui-ci nous donne des constantes strictement positives k et Γ telles que pour tout entier α positif ou nul et pour i suffisamment grand,

$$|\mathrm{rep}_S(\lambda v_{n_0+sC(i+\alpha k)-1})| = |\mathrm{rep}_S(\lambda v_{n_0+sCi-1})| + \alpha\Gamma$$

ou de manière équivalente, si $z = |\mathrm{rep}_S(\lambda v_{n_0+sCi-1})|$,

$$v_{z+\alpha\Gamma-1} \leq \lambda v_{n_0+sC(i+\alpha k)-1} < v_{z+\alpha\Gamma}.$$

D'une part, vu les inégalités 3.11 et 3.10, la première de ces inégalités donne

$$J(z+\alpha\Gamma-1)^{l+1} \leq v_{z+\alpha\Gamma-1} \leq \lambda K(n_0+sCi+sCk\alpha-1)^{l+1},$$

pour tout α et pour i suffisamment grand. Comme cette inégalité doit être vraie pour tout entier α, les coefficients des termes dominants en α doivent vérifier $J\Gamma^{l+1} \leq \lambda K(sCk)^{l+1}$. D'autre part, la seconde inégalité donne

$$K(z+\alpha\Gamma)^{l+1} \geq v_{z+\alpha\Gamma} > \lambda J(n_0+sCi+sC\alpha k-1)^{l+1},$$

pour tout $\alpha \in \mathbb{N}$ et pour i suffisamment grand. Par le même argument que précédemment, on doit avoir aussi $J\Gamma^{l+1} \geq \lambda K(sCk)^{l+1}$, donc

$$\lambda = \frac{J}{K}\left(\frac{\Gamma}{sCk}\right)^{l+1}.$$

Si $(K_m)_{m\in\mathbb{N}_0}$ et $(J_m)_{m\in\mathbb{N}_0}$ sont les suites définies par

$$K_m = a + \frac{1}{m} \text{ et } J_m = a - \frac{1}{m},$$

alors pour tout $m \in \mathbb{N}_0$, on a $J_m \leq \frac{v_n(L)}{n^{l+1}} \leq K_m$, si n est suffisamment grand. Ce qu'on a montré jusqu'à présent reste donc vrai si on remplace J par J_m et K par K_m, quelque soit $m \in \mathbb{N}_0$. En faisant tendre m vers l'infini, poser la condition $\lambda > (K_m/J_m)^{l+1}$ revient à supposer $\lambda \geq 2$. Ceci impliquerait

$$\lambda = \left(\frac{\Gamma}{sCk}\right)^{l+1},$$

ce qui est impossible vu l'hypothèse sur λ. □

Par le même raisonnement que dans le cas du langage a^*b^*, le théorème 3.2.6 admet le corollaire suivant.

Corollaire 3.2.7. *Dans le cas d'un système de numération abstrait construit sur un langage régulier polynomial, l'addition n'est pas une application régulière.*

Pour un système de numération construit sur un langage dont la fonction de complexité est bornée par une constante, au vu du théorème 3.2.6, tous les entiers sont des multiplicateurs susceptibles de conserver le caractère reconnaissable. Un tel langage est dit *slender* et les sytèmes de numération construits sur ces langages seront discutés pour leurs propriétés propres à la fin de ce chapitre.

3.3 Cas des langages à complémentaire polynomial

Considérons à présent un système de numération abstrait construit sur un langage régulier quelconque. Vu le théorème de séparation 2.2.2, le complémentaire de celui-ci est lui-même un langage soit polynomial, soit exponentiel. Nous montrons dans cette partie que la multiplication par une constante ne préserve en général pas le caractère reconnaissable pour un système de numération abstrait construit sur un langage régulier à complémentaire polynomial.

Définition 3.3.1. Si L est un langage sur Σ et x est un mot sur Σ, on note $L_x = \{w \in L \mid \exists y \in \Sigma^*, w = xy\}$ l'ensemble des mots de L dont x est préfixe.

Bien sûr, si $\sigma \in \Sigma$, alors $L_{\sigma^{i+1}} \subseteq L_{\sigma^i}$, pour tout $i \in \mathbb{N}$. Comme $L_x \subseteq L$, on a aussi $u_{L_x}(n) \leq u_L(n)$, pour tout $n \in \mathbb{N}$. Si L est de complexité polynomiale, il en est donc de même pour L_x. Remarquons aussi que si L s'écrit $L = \Sigma^* \backslash M$, avec $M \subseteq \Sigma^*$, alors $L_x = \Sigma_x^* \setminus M_x$, quelque soit le mot x sur Σ.

Exemple 3.3.2. Considérons le langage $M = \Sigma^* \setminus L$, avec $\Sigma = \{a, b\}$ et $L = a^*b^*$. Notons $S = (M, \Sigma, a < b)$. On a

$$u_L(n) = n + 1 \text{ et } u_M(n) = 2^n - n - 1;$$
$$v_n(M) = \sum_{i=0}^{n} u_M(i) = 2^{n+1} - \frac{n(n+3)}{2} - 2,$$

pour tout $n \in \mathbb{N}$. L'ensemble $\{v_n(M) \,|\, n \in \mathbb{N}\}$ est bien sûr S-reconnaissable. Mais on peut montrer que, par contre, l'ensemble $\{2v_n(M) \,|\, n \in \mathbb{N}\}$, n'est pas S-reconnaissable.

Ceci est en fait une conséquence d'un résultat plus général.

Théorème 3.3.3. *Si $S = (\Sigma^* \setminus L, \Sigma, <)$ est un système de numération abstrait construit sur un langage régulier à complémentaire L de fonction de complexité en $\Theta(n^l)$, $l \in \mathbb{N}$, et d'alphabet Σ de cardinalité $t \geq 2$, alors on peut trouver une partie S-reconnaissable X de $\mathbb{N}$ telle que, pour tout entier $j \geq 1$, t^jX n'est pas S-reconnaissable.*

Démonstration. Notons $\Sigma = \{\sigma_1 < \cdots < \sigma_{t-1} < \tau\}$ et $M = \Sigma^* \setminus L$. Pour $0 \leq k < n$, on a

$$u_n\left(M_{\tau^{n-k}}\right) = u_n(\Sigma^*_{\tau^{n-k}}) - u_n(L_{\tau^{n-k}}) = t^k - u_n(L_{\tau^{n-k}}),$$

où $u_n(L_{\tau^{n-k}})$ est une fonction $O(n^l)$ et

$$v_n(M) = \sum_{i=0}^{n} \left(u_i(\Sigma^*) - u_i(L)\right) = \frac{t^{n+1} - 1}{t - 1} - v_n(L).$$

Si $X = \{v_n(M) \mid n \in \mathbb{N}\}$, alors $\mathrm{rep}_S(X) = \mathrm{Min}(S)$ et vu la proposition 1.2.10, X est une partie S-reconnaissable de $\mathbb{N}$. Nous allons montrer que, pour tout entier $j \geq 1$, $t^j X$ n'est pas S-reconnaissable. Fixons un entier $j \geq 1$. Pour n suffisamment grand, on a

$$v_{n+j-1}(M) \leq t^j v_n(M) < v_{n+j}(M).$$

En effet, d'une part, on a

$$v_{n+j}(M) - t^j v_n(M) = t^j v_n(L) - v_{n+j}(L) + \frac{t^j - 1}{t-1}.$$

Vu le théorème 2.4.5, la suite $\left(v_n(L)/n^{l+1}\right)_{n\in\mathbb{N}}$ converge vers une limite strictement positive. Si a est cette limite, alors

$$\lim_{n\to+\infty} \frac{v_{n+j}(M) - t^j v_n(M)}{n^{l+1}} = (t^j - 1)a > 0.$$

D'autre part, vu le lemme 2.4.3, $v_n(L)$ est une fonction $\Theta(n^{l+1})$ et

$$t^j v_n(M) - v_{n+j-1}(M) = t^{n+j} - \frac{t^j - 1}{t-1} + v_{n+j-1}(L) - t^j v_n(L)$$

a un terme dominant exponentiel, ce qui suffit. Il s'ensuit que

$$|\mathrm{rep}_S(t^j v_n(M))| = n + j.$$

Ensuite, pour tout n suffisamment grand, il existe un unique entier $i = i(n)$ tel que

$$\underbrace{u_{n+j}\left(M_{\tau^{n+j-i+1}}\right)}_{=t^{i-1}-u_{n+j}(L_{\tau^{n+j-i+1}})} < v_{n+j}(M) - t^j v_n(M) \leq \underbrace{u_{n+j}\left(M_{\tau^{n+j-i}}\right)}_{=t^{i}-u_{n+j}(L_{\tau^{n+j-i}})}$$

et on peut écrire $\mathrm{rep}_S(t^j v_n(M)) = \tau^{n+j-i}\sigma z$, avec $|z| = i - 1$ et $\sigma \neq \tau$. Montrons que $\lim_{n\to+\infty}(n - i(n)) = +\infty$. Si $n - i(n)$ était borné, en divisant chaque membre des inégalités ci-dessus par t^n et en passant à la limite sur n, on aurait que les extrémités tendraient vers des limites strictement positives tandis que la limite centrale serait nulle, ce qui serait absurde.

Nous procédons par l'absurde en supposant que $\mathrm{rep}_S(t^j X)$ est accepté par un AFD à q états. Alors vu ce qui précède, on a $n - i(n) \geq q$ pour n suffisamment grand et on peut donc trouver des entiers n_0 et $s \geq 0$ tels que $\mathrm{rep}_S(t^j v_{n_0}(M)) = \tau^{q+s}\sigma z_0$, avec $\sigma \neq \tau$. Dès lors, par le lemme de la pompe 1.1.17, il existe une constante strictement positive α telle que les mots $\tau^{q+s+m\alpha}\sigma z$ sont dans $\mathrm{rep}_S(t^j X)$, pour tout entier positif m. Ceci n'est pas possible, puisque le suffixe z_0 est de longueur fixe. □

Les seuls candidats restants pour conserver le caractère reconnaissable d'un système de numération abstrait par multiplication par une constante sont donc les systèmes construits sur un langage exponentiel à complémentaire exponentiel. Nous ne développerons pas dans ce texte le cas de ces langages. Signalons tout de même qu'on peut obtenir pour ces systèmes des conditions suffisantes garantissant la conservation de la reconnaissabilité par multiplication par une constante et par addition ([7]). Nous donnons ici un exemple d'un tel langage.

Exemple 3.3.4. Soit le langage $L = a\{a,b\}^*$ sur l'alphabet $\{a,b\}$. Le complémentaire de L est le langage $M = b\{a,b\}^*$ et leurs fonctions de complexité sont données par

$$u_L(n) = u_M(n) = 2^{n-1},\ n \in \mathbb{N}.$$

3.4 Quelques remarques

3.4.1 Changement de l'ordre sur l'alphabet

Jusqu'à présent, nous avons étudié le lien entre le caractère reconnaissable d'un ensemble d'entiers dans un système de numération abstrait et différentes opérations arithmétiques effectuées sur cet ensemble. Dans la même veine de questionnement, on peut aussi se demander quelle est l'influence d'un changement de l'ordre de l'aphabet du système.

Si $S = (L, \Sigma, <)$ et $T = (L', \Sigma', \prec)$ sont deux systèmes de numération abstraits, on pose

$$\xi_{S,T} = \mathrm{rep}_T \circ \mathrm{val}_S : L \to L' \text{ et } \xi'_{S,T} = \mathrm{val}_T \circ \mathrm{rep}_S : \mathbb{N} \to \mathbb{N}.$$

Lorsque S et T sont clairement identifiés par le contexte, on écrit simplement ξ et ξ'.

Nous commençons par donner un exemple positif.

Exemple 3.4.1. Considérons les systèmes abstraits $S = (a^*b^*, \{a,b\}, a < b)$ et $T = (a^*b^*, \{a,b\}, b \prec a)$. On observe que $\xi_{S,T}(a^pb^q) = a^qb^p,\ p,q \in \mathbb{N}$. D'où, pour tout mot w de a^*b^*, on a $\xi_{S,T}(w) = h(w^R)$, où $h : \{a,b\} \to \{a,b\}$ est le morphisme défini par $h(a) = b$ et $h(b) = a$. Ceci montre que si une partie de $\mathbb{N}$ est S-reconnaissable, alors elle est aussi T-reconnaissable.

Malgré cet exemple positif, on n'a en général pas la conservation du caractère reconnaissable après changement de l'ordre de l'alphabet du système. En effet, le résultat suivant montre que dans le cas du système de numération abstrait construit sur le complémentaire de a^*b^*, changer l'ordre altère le caractère reconnaissable des ensembles d'entiers.

Lemme 3.4.2. *Soit $n \in \mathbb{N}$ et soient les systèmes $U = (\{a,b\}^*, \{a,b\}, a < b)$ et $V = (\{a,b\}^*, \{a,b\}, b \prec a)$. On a $\xi'_{U,V}(n) = 3.2^l - n - 3$, où $l = |rep_U(n)|$.*

Démonstration. C'est une simple vérification. □

Proposition 3.4.3. *Soient les systèmes de numération abstraits*

$$S = (\{a,b\}^* \setminus a^*b^*, \{a,b\}, a < b) \;\; et \;\; T = (\{a,b\}^* \setminus a^*b^*, \{a,b\}, b \prec a);$$
$$U = (\{a,b\}^*, \{a,b\}, a < b) \;\; et \;\; V = (\{a,b\}^*, \{a,b\}, b \prec a).$$

Pour tout entier $n \geq 2$, si $l = |rep_U(n-1)|$, alors

$$\xi_{S,T}(bab^n) = aba^{n-l-1}brep_U(n-1).$$

Démonstration. Pour faciliter la lecture, notons $L = \{a,b\}^* \setminus a^*b^*$. D'une part,

$$\begin{aligned} \mathrm{val}_S(bab^n) &= \underbrace{v_{n+1}(L)}_{|w|<n+2} + \underbrace{u_{n+1}(L)}_{w=au,|u|=n+1} + \underbrace{u_n(\{a,b\}^*)}_{w=abu,|u|=n} - 1 \\ &= v_{n+1}(L) + 2^{n+1} + 2^n - n - 3. \end{aligned}$$

D'autre part, on construit $\mathrm{val}_T\left(aba^{n-l-1}b\,\mathrm{rep}_U(n-1)\right)$ en utilisant la proposition 1.2.7 avec l'automate minimal de L. Celui-ci est représenté par la figure 3.5.

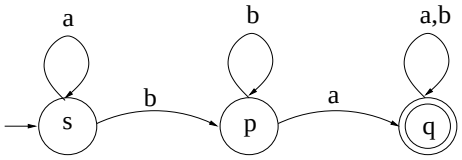

FIG. 3.5 – Automate minimal de a^*b^*.

Ainsi, on a

$$\begin{aligned} & \mathrm{val}_p(a^{n-l-1}b\,\mathrm{rep}_U(n-1)) + v_{n+1}(L) - v_{n-1}(p) + u_n(p) + u_n(q) \\ = \;& \mathrm{val}_q(a^{n-l-2}b\,\mathrm{rep}_U(n-1)) - v_{n-2}(q) + u_{n-1}(p) + v_{n+1}(L) + 2^{n+1} - 1 \\ \vdots \;& \\ = \;& \mathrm{val}_q(b\,\mathrm{rep}_U(n-1)) - v_l(q) + \sum_{i=l+1}^{n-2} u_i(q) + v_{n+1}(L) + 2^{n+1} + 2^{n-1} - 2 \\ = \;& \xi'_{U,V}(n-1) - v_{l-1}(q) + v_{n+1}(L) + 2^{n+1} + 2^n - 2.2^l - 2 \\ = \;& v_{n+1}(L) + 2^{n+1} + 2^n - n - 3. \end{aligned}$$

où on a utilisé le lemme 3.4.2 à la dernière étape. □

Nous déduisons de la proposition 3.4.3 un exemple négatif.

Exemple 3.4.4. Considérons cette fois les systèmes de numération abstraits $S = (\{a,b\}^* \setminus a^*b^*, \{a,b\}, a < b)$ et $T = (\{a,b\}^* \setminus a^*b^*, \{a,b\}, b \prec a)$. Vu la proposition 3.4.3 et le lemme de la pompe 1.1.17, $\mathrm{val}_S(bab^2b^*)$ n'est pas un ensemble T-reconnaissable d'entiers. On n'a donc pas conservation du caractère reconnaissable si on change l'ordre sur $\{a,b\}$.

Le caractère reconnaissable n'est donc en général pas conservé si on change l'ordre sur l'aphabet du système. Malgré cela, on peut tout de même mettre en évidence des classes de langages pour lesquels le changement de l'ordre de l'aphabet n'a pas d'effet sur la reconnaissabilité. La première d'entre elles est celle des langages vérifiant les hypothèses de la proposition suivante.

Proposition 3.4.5. *Soient $S = (L, \Sigma, <)$ et $T = (L, \Sigma, \prec)$ deux systèmes de numération abstrait construit sur un même langage L et d'ordres $<$ et $\prec$ distincts. Soit $n_0 \in \mathbb{N}$. Si $u_n(p) = u_n(q)$, $n \geq n_0$, pour tous états p et q de l'automate minimal $\mathcal{A}_L = (Q, \Sigma, \delta, s, F)$ de L, alors toute partie X de $\mathbb{N}$ est S-reconnaissable si et seulement si elle est aussi T-reconnaissable.*

Démonstration. Les éléments de Σ s'ordonnent en $\Sigma = \{\sigma_1 < \cdots < \sigma_p\} = \{\sigma_{\nu_1} \prec \cdots \prec \sigma_{\nu_p}\}$, où ν est une permutation de $\{1, \ldots, p\}$. Nous allons montrer que le graphe $\widehat{\xi} = \{(x,y) \in L \times L \mid \mathrm{val}_S(x) = \mathrm{val}_T(y)\}$ de $\xi_{S,T}$ est un langage régulier sur l'alphabet $\Sigma \times \Sigma$. On aura alors que

$$\mathrm{rep}_T(X) = p_2\left(\widehat{\xi} \cap p_1^{-1}\left(\mathrm{rep}_S(X)\right)\right)$$

est régulier si et seulement si $\mathrm{rep}_S(X)$ l'est aussi, où p_1, $p_2 : (\Sigma \times \Sigma)^* \to \Sigma^*$ sont les homomorphismes de projections canoniques.

Soit $(x,y) \in \widehat{\xi}$. Comme les systèmes abtraits S et T possèdent la même suite $(v_n)_{n\in\mathbb{N}}$, on a bien $|x| = |y|$. Si $|x| \geq n_0$, alors, par construction de la fonction rep dans un système de numération abstrait et vu l'hypothèse sur les états de l'automate minimal du langage, x et y se décomposent en

$$x = \underbrace{\sigma_{i_1} \cdots \sigma_{i_l}}_{\alpha} \beta \text{ et } y = \underbrace{\sigma_{\nu_{i_1}} \cdots \sigma_{\nu_{i_l}}}_{\alpha'} \beta',$$

avec $|\beta| = |\beta'| = n_0$, $\beta \in L_{s.\alpha}$, $\beta' \in L_{s.\alpha'}$ et $\mathrm{val}_{S_{s.\alpha}}(\beta) = \mathrm{val}_{T_{s.\alpha'}}(\beta')$. Pour conclure, il suffit de remarquer que les mots de $\widehat{\xi}$ de longueur plus grande à égale à n_0 sont exactement les mots acceptés par l'AFND qui a comme ensemble d'états $(Q \times Q) \cup \{f\}$, où (s,s) est l'unique état initial et f est l'unique état final et dont les relations de transition sont de deux types : premièrement, celles de la forme $((q,q'),(\sigma_i,\sigma_{\nu_i}),(q.\sigma_i,q'.\sigma_{\nu_i}))$ et deuxièmement, celles de la forme $((q,q'),(\beta,\beta'),f)$, avec $|\beta| = |\beta'| = n_0$, $\beta \in L_q$, $\beta' \in L_{q'}$ et $\mathrm{val}_q(\beta) = \mathrm{val}_{q'}(\beta')$. □

Une autre classe de langages pour lesquels le changement de l'ordre de l'aphabet n'a pas d'effet sur le caractère reconnaissable est celle des langages dits *slender*. Ces langages possèdent plusieurs propriétés propres intéressantes, que nous détaillons dans la section suivante.

3.4.2 Cas des langages *slender*

La classe des langages *slender* préserve non seulement le caractère reconnaissable sous changement d'ordre de l'alphabet du système, mais elle préserve également celui-ci par multiplication par une constante et par addition.

Définition 3.4.6. Soit un entier $d > 0$. Un langage L est *d-slender* si $u_L(n) \leq d$, pour tout $n \geq 0$. Un langage *slender* est un langage qui est *d-slender* pour un certain entier $d > 0$.

Vu le corollaire 2.1.15, les langages réguliers *slender* admettent la caractérisation suivante. Un langage $L \subseteq \Sigma^*$ régulier est *slender* si et seulement si il est de la forme

$$L = \bigcup_{i=1}^{k} x_i y_i^* z_i,\ k \geq 1,\ x_i, y_i, z_i \in \Sigma^*,$$

où l'union est une union disjointe.

La preuve de la proposition suivante est une preuve directe. En fait, on peut aussi déduire ce résultat du théorème 3.4.8 donné ci-dessous.

Proposition 3.4.7. *Si L est un langage régulier slender et si $S = (L, \Sigma, <)$ et $T = (L, \Sigma, \prec)$ sont deux systèmes de numération abstrait construit sur L d'ordres $<$ et $\prec$ distincts, alors toute partie X de $\mathbb{N}$ est S-reconnaissable si et seulement si elle est aussi T-reconnaissable.*

Démonstration. Par le même argument que dans la proposition 3.4.5, il suffit de montrer que le graphe $\widehat{\xi} = \{(x, y) \in L \times L \mid \mathrm{val}_S(x) = \mathrm{val}_T(y)\}$ de $\xi_{S,T}$ est régulier. Soit $d > 0$ tel que L est *d-slender*. Posons $I_{1,<} = \mathrm{Min}(S)$, $I_{1,\prec} = \mathrm{Min}(T)$ et pour $2 \leq i \leq d$,

$$I_{i,<} = \mathrm{Min}\left(L \setminus \left(\bigcup_{j=1}^{i-1} I_{j,<}\right), <\right) \text{ et } I_{i,\prec} = \mathrm{Min}\left(L \setminus \left(\bigcup_{j=1}^{i-1} I_{j,\prec}\right), \prec\right).$$

Bien sûr, pour tout $1 \leq i \leq d$, les langages $I_{i,<}$ et $I_{i,\prec}$ sont réguliers et

$$L = \bigcup_{j=1}^{d} I_{j,<} = \bigcup_{j=1}^{d} I_{j,\prec},$$

où les unions sont disjointes. Pour tout $x \in L$, on a $|x| = |\xi(x)|$. D'où

$$\widehat{\xi} = \bigcup_{j=1}^{d} \left((I_{j,<} \times I_{j,\prec}) \cap (\Sigma \times \Sigma)^* \right),$$

qui est bien régulier. □

Théorème 3.4.8. *Si S est un système de numération abstrait construit sur langage régulier slender, alors une partie X de $\mathbb{N}$ est S-reconnaissable si et seulement si elle est une union finie de progressions arithmétiques.*

Démonstration. Comme les progressions arithmétiques sont reconnaissables par tout système de numération abstrait (voir proposition 1.2.8), la condition est évidemment suffisante. Montrons qu'elle est aussi nécessaire. Par la caractérisation des langages *slender,* on a

$$L = \left(\bigcup_{i=1}^{k} x_i y_i^* z_i \right) \cup F_0,\ k \geq 1,\ x_i, y_i, z_i \in \Sigma^*,\ y_i \neq \varepsilon,$$

où les unions sont disjointes et F_0 est un ensemble fini. La suite $(u_n(L))_{n\in\mathbb{N}}$ est ultimement périodique de période $C = \text{ppcm}\{|y_i| \mid 1 \leq i \leq k\}$. De plus, on observe que si $x_i y_i^n z_i$ est le l-ième mot de L de longueur $|x_i z_i| + n|y_i|$, alors

$$x_i\, y_i^{\,n+\frac{C}{|y_i|}}\, z_i$$

est le l-ième mot de L de longueur $|x_i z_i| + n|y_i| + C$. Soit X une partie de $\mathbb{N}$ qui est S-reconnaissable. Le langage $\text{rep}_S(X)$ est un sous-ensemble régulier de L et s'écrit donc sous la forme

$$\text{rep}_S(X) = \left(\bigcup_{j\in J} x_j (y_j^{\alpha_j})^* z_j \right) \cup F_0',$$

avec $J \subseteq \{1, \ldots, k\}$, $\alpha_j \in \mathbb{N}$ pour tout $j \in J$ et F_0' un sous-ensemble fini de L. Vu ce qui précède, X est bien ultimement périodique. □

Corollaire 3.4.9. *Si S est un système de numération abstrait construit sur un langage régulier slender, alors la multiplication par une constante préserve la S-reconnaissabilité.*

Remarquons que le corollaire 3.4.9 est en accord avec le théorème 3.2.6. En effet, un langage slender est de complexité $O(n^0)$.

Corollaire 3.4.10. *Si S est un système de numération abstrait construit sur un langage régulier slender, alors l'addition préserve la S-reconnaissabilité.*

Chapitre 4

Généralisation

Dans cette dernière partie, nous généralisons le problème de la conservation du caractère reconnaissable par mutiplication par une constante dans le cas d'un système abstrait construit sur le langage a^*b^* au cas d'un système construit sur un langage de la forme $a_1^* \cdots a_\ell^*$.

Précisons tout d'abord quelques notations. On désignera par $\mathcal{B}_\ell$ le langage $a_1^* \cdots a_\ell^*$ sur l'alphabet $\Sigma_\ell = \{a_1, \ldots, a_\ell\}$ de cardinalité $\ell \geq 1$. On supposera toujours que $(\Sigma_\ell, <)$ est totalement ordonné par $a_1 < \cdots < a_\ell$. De plus, on notera rep_ℓ et val_ℓ les fonctions de représention et de valeur numérique du système de numération abstrait associé ainsi que $u_\ell(n) := \mathbf{u}_{\mathcal{B}_\ell}(n)$ et $v_\ell(n) := \mathbf{v}_{\mathcal{B}_\ell}(n)$. Si w est un mot sur Σ_ℓ, $|w|$ désigne sa longueur et $|w|_{a_j}$ compte le nombre de lettres a_j apparaissant dans w. La *fonction de Parikh* Ψ est l'application qui à un mot $w \in \Sigma_\ell^*$ associe le vecteur $\Psi(w) := (|w|_{a_1}, \ldots, |w|_{a_\ell})$.

Remarque 4.0.11. Dans le cas des langages $\mathcal{B}_\ell$, la fonction de Parikh Ψ est une bijection. Dans ce qui suit, nous ne ferons donc pas de distinction entre un entier n, sa $\mathcal{B}_\ell$-représentation $\mathrm{rep}_\ell(n) = a_1^{i_1} \cdots a_\ell^{i_\ell} \in \mathcal{B}_\ell$ et le vecteur de Parikh correspondant $\Psi(\mathrm{rep}_\ell(n)) = (i_1, \ldots, i_\ell) \in \mathbb{N}^\ell$.

Dans les exemples, lorsque nous considérerons les cas $\ell = 2$ ou 3, nous utiliserons les alphabets $\{a < b\}$ ou $\{a < b < c\}$.

Ce chapitre est organisé de la manière suivante. En premier lieu, nous obtenons une méthode de calcul de $\mathrm{val}_\ell(a_1^{n_1} \cdots a_\ell^{n_\ell})$. Nous en déduisons une preuve du fait que tout entier positif ou nul s'écrit de manière unique comme somme de coefficients binomiaux :

$$n = \binom{z_\ell}{\ell} + \binom{z_{\ell-1}}{\ell-1} + \cdots + \binom{z_1}{1}$$

avec $z_\ell > z_{\ell-1} > \cdots > z_1 \geq 0$. Ensuite, nous énonçons la conjecture que les entiers λ préservant le caractère reconnaissable pour le système abstrait construit sur $\mathcal{B}_\ell$ sont exactement ceux de la forme $(\prod_{i=1}^k p_i^{\theta_i})^\ell$ où $p_1, \ldots, p_k$

sont des nombres premiers strictement supérieurs à ℓ. Nous proposons des résultats partiels dans cette direction. Finalement, nous donnons une forme explicite des sous-ensembles réguliers de $\mathcal{B}_\ell$ et nous terminons par une application à la $\mathcal{B}_\ell$-reconnaissabilité des progressions arithmétiques.

4.1 $\mathcal{B}_\ell$-représentation d'un entier

Rappelons que le coefficient binomial $\binom{i}{j}$ s'annule lorsque $i < j$.

Lemme 4.1.1. *Pour tous entiers $\ell \geq 1$ et $n \geq 0$, on a*

$$\mathbf{u}_{\ell+1}(n) = \mathbf{v}_\ell(n) \tag{4.1}$$

et

$$\mathbf{u}_\ell(n) = \binom{n+\ell-1}{\ell-1}. \tag{4.2}$$

Démonstration. La relation (4.1) provient du fait que l'ensemble des mots de longueur n de $\mathcal{B}_{\ell+1}$ se partitionne en

$$\bigcup_{i=0}^{n} \left(a_1^* \cdots a_\ell^* \cap \Sigma_\ell^i\right) a_{\ell+1}^{n-i}.$$

Pour obtenir (4.2), on procède par induction sur $\ell \geq 1$. Pour $\ell = 1$, il est clair que $\mathbf{u}_1(n) = 1$ pour tout $n \geq 0$. Supposons maintenant que l'égalité (4.2) est vérifiée en ℓ et montrons qu'elle l'est encore en $\ell + 1$. Vu l'égalité (4.1), on a

$$\mathbf{u}_{\ell+1}(n) = \sum_{i=0}^{n} \mathbf{u}_\ell(i) = \sum_{i=0}^{n} \binom{i+\ell-1}{\ell-1}.$$

Pour conclure, il suffit de vérifier par induction sur $n \geq 0$ que

$$\sum_{i=0}^{n} \binom{i+\ell-1}{\ell-1} = \binom{n+\ell}{\ell},$$

ce qui se fait de manière directe. □

Lemme 4.1.2. *Soit $S = (\mathcal{B}_\ell, \Sigma_\ell, <)$. On a*

$$\mathrm{val}_\ell(a_1^{n_1} \cdots a_\ell^{n_\ell}) = \sum_{i=1}^{\ell} \binom{n_i + \cdots + n_\ell + \ell - i}{\ell - i + 1}.$$

Par conséquent, pour tout $n \in \mathbb{N}$,

$$|\mathrm{rep}_\ell(n)| = k \Leftrightarrow \underbrace{\binom{k+\ell-1}{\ell}}_{\mathrm{val}_\ell(a_1^k)} \leq n \leq \underbrace{\sum_{i=1}^{\ell} \binom{k+i-1}{i}}_{\mathrm{val}_\ell(a_\ell^k)}.$$

Démonstration. De la structure même du système considéré, on observe que

$$\mathrm{val}_\ell(a_1^{n_1}\cdots a_\ell^{n_\ell}) = \mathrm{val}_\ell(a_1^{n_1+\cdots+n_\ell}) + \mathrm{val}_{\ell-1}(a_1^{n_2}\cdots a_{\ell-1}^{n_\ell}).$$

En itérant cette décomposition, on obtient

$$\mathrm{val}_\ell(a_1^{n_1}\cdots a_\ell^{n_\ell}) = \sum_{i=1}^{\ell} \mathrm{val}_{\ell-i+1}(a_1^{n_i+\cdots+n_\ell}).$$

Comme a_1^n est le premier mot de longueur n, on a $\mathrm{val}_\ell(a_1^n) = \mathbf{v}_\ell(n-1)$ et la conclusion s'ensuit, en utilisant les relations (4.1) et (4.2). □

L'exemple suivant illustre la décomposition de $\mathrm{val}_\ell(a_1^{n_1}\cdots a_\ell^{n_\ell})$ utilisée dans cette dernière preuve pour le cas $\ell = 3$.

Exemple 4.1.3. Considérons les mots de longueur 3 du langage $a^*b^*c^*$:

$$aaa < aab < aac < abb < abc < acc < bbb < bbc < bcc < ccc.$$

On a $\mathrm{val}_3(aaa) = \binom{5}{3} = 10$ et $\mathrm{val}_3(acc) = 15$. Si on applique le morphisme effaçant $\varphi : \{a,b,c\}^* \to \{a,b,c\}^*$ défini par $\varphi(a) = \varepsilon$, $\varphi(b) = b$ et $\varphi(c) = c$ sur les mots de longueur 3, on obtient

$$\varepsilon < b < c < bb < bc < cc < bbb < bbc < bcc < ccc.$$

Ainsi la liste ordonnée des mots de longueur 3 de $a^*b^*c^*$ contient une copie ordonnée des mots de longueur au plus 3 du langage b^*c^* et pour obtenir $\mathrm{val}_3(acc)$, on ajoute à $\mathrm{val}_3(aaa)$ la position du mot cc dans le langage ordonné b^*c^*. En d'autres termes, $\mathrm{val}_3(acc) = \mathrm{val}_3(aaa) + \mathrm{val}_2(cc)$ où val_2 est considéré comme une application définie sur le langage b^*c^*.

Le résultat suivant apparaît dans [5]. Nous en obtenons ici une preuve alternative, utilisant uniquement les propriétés des systèmes de numération abstraits construits sur un langage du type $a_1^*\cdots a_\ell^*$.

Corollaire 4.1.4. *Soit $\ell \in \mathbb{N}_0$. Tout entier positif ou nul n s'écrit de manière unique sous la forme*

$$n = \binom{z_\ell}{\ell} + \binom{z_{\ell-1}}{\ell-1} + \cdots + \binom{z_1}{1} \tag{4.3}$$

avec $z_\ell > z_{\ell-1} > \cdots > z_1 \geq 0$.

Démonstration. L'application $\mathrm{rep}_\ell : \mathbb{N} \to a_1^*\cdots a_\ell^*$ étant bijective, tout entier positif ou nul n a une représentation unique de la forme $a_1^{n_1}\cdots a_\ell^{n_\ell}$ et la conclusion provient directement du lemme 4.1.2. □

La méthode générale de construction de la représentation d'un entier dans un système de numération abstrait donnée dans [6, Algorithm 1] prend une forme spéciale dans le cas du langage $\mathcal{B}_\ell$. Nous en déduisons un algorithme calculant la décomposition (4.3) ou, de manière équivalente, la B_ℓ-représentation de tout entier.

Algorithme 1. Soit `n` un entier et `l` un entier positif ou nul. L'algorithme suivant produit les entiers `z(l)`,...,`z(1)` correspondant aux z_i apparaissant dans la décomposition (4.3) de `n`.

```
Pour i=l, l-1, ..., 1 effectuer
   si n>0, répéter
      trouver t tel que C(t,i) ≤ n < C(t+1,i)
      z(i)←t
      n←n-C(t,i)
   sinon, z(i)←i-1
```

Considérons maintenant le système triangulaire ayant $n_1, \ldots, n_\ell$ comme inconnues :

$$n_i + \cdots + n_\ell = \mathtt{z}(\ell - i + 1) - \ell + i, \quad i = 1, \ldots, \ell.$$

On a $\mathrm{rep}_\ell(\mathtt{n}) = a_1^{n_1} \cdots a_\ell^{n_\ell}$.

Remarque 4.1.5. Pour accélérer le calcul de `t` dans l'algorithme ci-dessus, on peut utiliser des méthodes d'analyse numérique. En effet, pour des entiers `i` et `n` donnés, $\binom{\mathtt{t}}{\mathtt{i}} - \mathtt{n}$ est un polynôme en `t` de degré `i` et nous recherchons en fait la plus grande racine réelle x de ce polynôme. D'où $\mathtt{t} = \lfloor x \rfloor$.

Exemple 4.1.6. Pour $\ell = 3$, on obtient par exemple

$$12345678901234567890 = \binom{4199737}{3} + \binom{3803913}{2} + \binom{1580642}{1}$$

et en résolvant le système

$$\begin{cases} n_1 + n_2 + n_3 &= 4199737 - 2 \\ n_2 + n_3 &= 3803913 - 1 \\ n_3 &= 1580642 \end{cases}$$

on trouve $(n_1, n_2, n_3) = (395823, 2223270, 1580642)$. Ainsi

$$\mathrm{rep}_3(12345678901234567890) = a^{395823} b^{2223270} c^{1580642}.$$

4.2 Multiplication par $\lambda = \beta^\ell$

Dans le cas du langage $\mathcal{B}_3$, si la multiplication par une constante préserve le caractère reconnaissable, alors, par le théorème 3.2.6 et le lemme

4.1.1, cette constante doit être un cube. Dans le cas de $\mathcal{B}_2$, uniquement les carrés impairs possèdent la propriété de conservation. On peut de manière analogue se demander quels cubes préservent le caractère reconnaissable pour le système abstrait construit sur $\mathcal{B}_3$.

Théorème 4.2.1. *Soit $S = (a^*b^*c^*, \{a, b, c\}, a < b < c)$. Si $\beta \in \mathbb{N}_0$ est tel que $\beta \not\equiv \pm 1 \pmod 6$, alors la multiplication par β^3 ne préserve pas la S-reconnaissabilité.*

Démonstration. Supposons d'abord $\beta \equiv 2 \pmod 6$. Pour n suffisamment grand, on a

$$\mathrm{rep}_3\left[(6k+2)^3 \,\mathrm{val}_3(a^n)\right] = a^r\, b^{s+(3k+1)n}\, c^{t+(3k+1)n}$$

où les constantes r, s, t sont données par

$$r = 4k + 6k^2,\ s = 5k + 11k^2 + 24k^3 + 18k^4,\ t = -3k - 17k^2 - 24k^3 - 18k^4.$$

On vérifie cette identité en appliquant val_3 des deux côtés et en utilisant le lemme 4.1.2. Si $\beta \equiv 3 \pmod 6$, alors pour n suffisamment grand,

$$\mathrm{rep}_3\left[(6k+3)^3 \,\mathrm{val}_3(a^n)\right] = a^r\, b^{s+(2k+1)n}\, c^{t+(4k+2)n}$$

où les constantes r, s, t sont données par

$$r = 1 + 6k + 6k^2,\ s = 1 + 11k + 27k^2 + 36k^3 + 18k^4,$$

$$t = -1 - 11k - 33k^2 - 36k^3 - 18k^4.$$

Si $\beta \equiv 4 \pmod 6$, alors n suffisamment grand,

$$\mathrm{rep}_3\left[(6k+4)^3 \,\mathrm{val}_3(a^n)\right] = a^r\, b^{s+(3k+2)n}\, c^{t+(3k+2)n}$$

où les constantes r, s, t sont données par

$$r = 2 + 8k + 6k^2,\ s = 4 + 23k + 47k^2 + 48k^3 + 18k^4,$$

$$t = -4 - 25k - 53k^2 - 48k^3 - 18k^4.$$

Finalement, si $\beta \equiv 0 \pmod 6$, alors pour n suffisamment grand,

$$\mathrm{rep}_3\left[(6k)^3 \,\mathrm{val}_3(a^n)\right] = a^r\, b^{s+5k\,n}\, c^{t+k\,n}$$

où les constantes r, s, t sont données par

$$r = -1 + 6k^2, \quad s = -1 + 5k - 3k^2, \quad t = k - 3k^2 - 18k^4.$$

Notons $X = \mathrm{val}_3(a^*)$. Vu les formules qui précèdent, il est évident que X est S-reconnaissable mais que $\mathrm{rep}_3(\beta^3 X)$ n'est pas régulier. □

Grâce à des expérimentations informatiques, nous conjecturons le résultat suivant.

Conjecture 1. *La multiplication par β^ℓ préserve la S-reconnaissabilité pour le système de numération abstrait $S = (\mathcal{B}_\ell, \Sigma_\ell, \{a_1 < \cdots < a_\ell\})$ si et seulement si*

$$\beta = \prod_{i=1}^{k} p_i^{\theta_i}$$

où $p_1, \ldots, p_k$ ($k \in \mathbb{N}_0$) sont des nombres premiers strictement plus grands que ℓ. En d'autres termes, la multiplication par β^ℓ ne préserve pas la S-reconnaissabilité si et seulement

$$\exists M \in \{2, \ldots, \ell\} : \beta \equiv 0 \pmod{M}.$$

Considérons des entiers $\ell \geq 2$ et $\beta \in 2\mathbb{N} + 3$ fixés une fois pour toutes. Le résultat suivant relie les longueurs des $\mathcal{B}_\ell$-représentations de n et $\beta^\ell n$ grossièrement par un facteur β.

Lemme 4.2.2. *Pour $n \in \mathbb{N}$ suffisamment grand, on a*

$$|\mathrm{rep}_\ell(\beta^\ell n)| = \beta\,|\mathrm{rep}_\ell(n)| + \frac{(\beta-1)(\ell-1)}{2} + i$$

avec $i \in \{-1, 0, \ldots, \beta - 1\}$.

Démonstration. Un entier n admet une $\mathcal{B}_\ell$-représentation $\mathrm{rep}_\ell(n)$ de longueur k si et seulement si $\mathrm{val}_\ell(a_1^k) \leq n < \mathrm{val}_\ell(a_1^{k+1})$. Par le lemme 4.1.2, pour un tel entier n, on a

$$\beta^\ell \binom{\ell + k - 1}{\ell} \leq \beta^\ell n < \beta^\ell \binom{\ell + k}{\ell}. \tag{4.4}$$

Tout d'abord, on observe que

$$\mathrm{val}_\ell(a_1^{\beta k}) = \binom{\ell + \beta k - 1}{\ell} \leq \beta^\ell \binom{\ell + k - 1}{\ell} \leq \beta^\ell n.$$

Ceci signifie que $k' := |\mathrm{rep}_\ell(\beta^\ell n)| \geq \beta k$. On peut donc écrire k' sous la forme $\beta k + u$, pour un certain entier $u \geq 0$. Nous devons montrer que

$$\frac{(\beta-1)(\ell-1)}{2} - 1 \leq u \leq \frac{(\beta-1)(\ell-1)}{2} + \beta - 1.$$

Par définition de k', on a

$$\binom{\ell + k' - 1}{\ell} \leq \beta^\ell n < \binom{\ell + k'}{\ell}. \tag{4.5}$$

De (4.4) et (4.5), on déduit que

$$\binom{\ell + k' - 1}{\ell} < \beta^\ell \binom{\ell + k}{\ell} \text{ et } \beta^\ell \binom{\ell + k - 1}{\ell} < \binom{\ell + k'}{\ell}.$$

En substituant k' par $\beta k + u$ dans la première inégalité, on obtient

$$\prod_{v=0}^{\ell-1} (\beta k + u + v) < \beta^\ell \prod_{v=1}^{\ell} (k + v). \tag{4.6}$$

L'expression dans le membre de gauche peut s'écrire sous la forme

$$\sum_{i=0}^{\ell} \alpha_i (\beta k + u)^i = \sum_{j=0}^{\ell} \sum_{i=j}^{\ell} \alpha_i \binom{i}{j} (\beta k)^{i-j} u^j \tag{4.7}$$

pour des coefficients entiers positifs α_i avec $\alpha_\ell = 1$, $\alpha_{\ell-1} = s(\ell - 1)$, ..., $\alpha_1 = (\ell - 1)!$ et $\alpha_0 = 0$, où $s(n) := 1 + 2 + \cdots + n$. De la même manière, le membre de droite s'écrit $\beta^\ell \sum_{i=0}^{\ell} \gamma_i k^i$ avec $\gamma_\ell = 1$, $\gamma_{\ell-1} = s(\ell)$, ..., $\gamma_0 = \ell!$. En soustrayant de chaque côté de (4.6) le terme correspondant à $j = 0$ dans (4.7), on obtient

$$u \underbrace{\sum_{j=1}^{\ell} \sum_{i=j}^{\ell} \alpha_i \binom{i}{j} (\beta k)^{i-j} u^{j-1}}_{=: T(k)} < \sum_{i=0}^{\ell-1} (\beta^{\ell-i} \gamma_i - \alpha_i)(\beta k)^i.$$

Remarquons que le terme en $(\beta k)^l$ disparait dans le membre de droite. Montrons que u est borné. Supposons que $u \geq 1$. On peut écrire T sous la forme $T(k) = \sum_{i=0}^{\ell-1} \delta_i (\beta k)^i$, avec $\delta_i \geq 0$ pour tout $0 \leq i \leq \ell - 1$. On a $\delta_{\ell-1} = \ell$ et

$$u\, \ell\, (\beta k)^{\ell-1} < \sum_{i=0}^{\ell-1} (\beta^{\ell-i} \gamma_i - \alpha_i)(\beta k)^i.$$

Ainsi, en divisant par $\ell\, (\beta k)^{\ell-1}$, on a

$$u < \frac{\beta s(\ell) - s(\ell - 1)}{\ell} + \sum_{i=0}^{\ell-2} \frac{\beta^{\ell-i} \gamma_i - \alpha_i}{\ell} (\beta k)^{i-\ell+1}.$$

En laissant k tendre vers l'infini, on obtient alors

$$u \leq \frac{\beta s(\ell) - s(\ell - 1)}{\ell} = \frac{(\beta - 1)(\ell - 1)}{2} + \beta.$$

Pour conclure, on doit avoir $u \leq \frac{(\beta-1)(\ell-1)}{2} + \beta - 1$. Ceci peut être obtenu en montrant (de la même manière que dans la remarque 4.2.7) que pour k suffisamment grand, on a

$$\beta^\ell n < \beta^\ell \binom{\ell + k}{\ell} \leq \binom{\beta k + \frac{(\beta-1)(\ell-1)}{2} + \beta + \ell - 1}{\ell}$$

signifiant que $\beta^\ell n$ est plus petit que la valeur numérique du premier mot de longueur $\beta k + \frac{(\beta-1)(\ell-1)}{2} + \beta$.

Pour obtenir la borne inférieure annoncée pour u, on procède de manière analogue en utilisant le fait que u est borné. □

En utilisant le lemme 4.2.2, on peut définir une partition de $\mathbb{N}$.

Définition 4.2.3. Pour tout $i \in \{-1, 0, \ldots, \beta - 1\}$ et $k \in \mathbb{N}$ suffisamment grand, on définit

$$\mathcal{R}_{i,k} := \left\{ n \in \mathbb{N} : |\mathrm{rep}_\ell(n)| = k \text{ et } |\mathrm{rep}_\ell(\beta^\ell n)| = \beta\, k + \frac{(\beta-1)(\ell-1)}{2} + i \right\}.$$

Le résultat suivant est technique mais il est intéressant de remarquer que ses hypothèses sont exactement les conditions introduites dans notre conjecture 1.

Lemme 4.2.4. *Si β satisfait aux conditions énoncées dans la conjecture 1, alors pour tout $u \geq \ell$, on a*

$$\binom{u}{\ell} \equiv \binom{u+\beta^\ell}{\ell} \pmod{\beta^\ell}.$$

Démonstration. Soit $u, v \geq \ell$. On a

$$\binom{v}{\ell} - \binom{u}{\ell} = \frac{v(v-1)\cdots(v-\ell+1) - u(u-1)\cdots(u-\ell+1)}{\ell!}.$$

Le numérateur du membre de droite est un entier divisible par $\ell!$. En outre, ce numérateur est aussi divisible par $v-u$. En effet, il est de la forme $P(v)-P(u)$ pour un certain polynôme P.

Remarquons que si $v = u + \beta^\ell$, le numérateur correspondant est divisible par $\ell!$ et aussi par β^ℓ. Mais comme tout facteur premier de β est plus grand que ℓ, $\ell!$ et β^ℓ sont premiers entre eux. Par conséquent, le numérateur correspondant est divisible par $\beta^\ell \ell!$. □

Supposons que β vérifie les conditions de la conjecture 1. Une inspection minutieuse de la multiplication par β^ℓ en utilisant la partition induite par le lemme 4.2.2 nous amène à l'observation suivante.

Proposition 4.2.5. *Soit $i \in \{0, \ldots, \beta - 1\}$. Il existe une constante $\mathbf{L} \geq 0$ (dépendant uniquement de ℓ et β) telle que pour tout $k \geq \mathbf{L}$, si $m = \min \mathcal{R}_{i,k}$ et $n = \min \mathcal{R}_{i,k+\beta^{\ell-1}}$, alors*

$$\forall t \in \{2, \ldots, \ell\} : |\mathrm{rep}_\ell(\beta^\ell m)|_{a_t} = |\mathrm{rep}_\ell(\beta^\ell n)|_{a_t}.$$

De plus, $|\mathrm{rep}_\ell(\beta^\ell m)|_{a_1} + \beta^\ell = |\mathrm{rep}_\ell(\beta^\ell n)|_{a_1}$

Démonstration. Comme $m \in \mathcal{R}_{i,k}$, par le lemme 4.1.2, on a

$$\underbrace{\binom{\beta\, k + \frac{(\beta-1)(\ell-1)}{2} + i + \ell - 1}{\ell}}_{=:\, A_i(k)} \le \beta^\ell m \le \underbrace{\sum_{j=1}^{\ell} \binom{\beta\, k + \frac{(\beta-1)(\ell-1)}{2} + i + j - 1}{j}}_{=:\, B_i(k)}.$$

Et comme $m - 1 \in \mathcal{R}_{i-1,k}$, on a aussi

$$\underbrace{\binom{\beta\, k + \frac{(\beta-1)(\ell-1)}{2} + i + \ell - 2}{\ell} + \beta^\ell}_{=:\, C_i(k)} \le \beta^\ell m \le \underbrace{\sum_{j=1}^{\ell} \binom{\beta\, k + \frac{(\beta-1)(\ell-1)}{2} + i + j - 2}{j} + \beta^\ell}_{=:\, D_i(k)}.$$

On vérifie directement que $D_i(k) = A_i(k) - 1 + \beta^\ell$,

$$A_i(k) - C_i(k) = \frac{\beta^{\ell-1}}{(\ell-1)!}\, k^{\ell-1} + o(k^{\ell-1})$$

et

$$B_i(k) - D_i(k) = \frac{\beta^{\ell-1}}{(\ell-1)!}\, k^{\ell-1} + o(k^{\ell-1}).$$

D'où il existe $\mathbf{L}$ tel que, pour tout $k \ge \mathbf{L}$, on a $A_i(k) > C_i(k)$, $B_i(k) > D_i(k)$ et

$$A_i(k) \le \beta^\ell m < A_i(k) + \beta^\ell.$$

Ainsi, si on fixe un tel $\mathbf{L}$, pour tout $k \ge \mathbf{L}$, on peut trouver un unique entier $\mu_i(k)$ tel que

$$\beta^\ell m = A_i(k) + \mu_i(k) \quad \text{et} \quad 0 \le \mu_i(k) \le \beta^\ell - 1.$$

En particulier, pour tout $k \ge \mathbf{L}$, on a également un unique entier $\mu_i(k+\beta^{\ell-1})$ tel que

$$\beta^\ell n = A_i(k + \beta^{\ell-1}) + \mu_i(k + \beta^{\ell-1}) \quad \text{et} \quad 0 \le \mu_i(k + \beta^{\ell-1}) \le \beta^\ell - 1.$$

On déduit du lemme 4.2.4 que $A_i(k) \equiv A_i(k + \beta^{\ell-1}) \pmod{\beta^\ell}$ et par conséquent, $\mu_i(k) = \mu_i(k + \beta^{\ell-1})$. Du lemme 4.1.2 on déduit

$$\mathrm{rep}_\ell(\beta^\ell m) = a_1^t\, \mathrm{rep}_{\{a_2,\ldots,a_\ell\}}(\mu_i(k))$$

où t est tel que $|\mathrm{rep}_\ell(\beta^\ell m)| = \beta\, k + \frac{(\beta-1)(\ell-1)}{2} + i$ et

$$\mathrm{rep}_\ell(\beta^\ell n) = a_1^{t+\beta^\ell}\, \mathrm{rep}_{\{a_2,\ldots,a_\ell\}}(\mu_i(k)).$$

□

Fixons un entier $\mathbf{L}$ vérifiant les conditions de la proposition 4.2.5.

Corollaire 4.2.6. *Soit $i \in \{0, \dots, \beta - 1\}$ et $k \geq \mathbf{L}$. Posons*

$$A_i(k) := \binom{\beta\, k + \frac{(\beta-1)(\ell-1)}{2} + i + \ell - 1}{\ell}.$$

Le premier élément appartenant à $\mathcal{R}_{i,k}$ est donné par $\left\lceil \frac{A_i(k)}{\beta^\ell} \right\rceil$.

Démonstration. C'est une conséquence directe de la preuve de la proposition 4.2.5. □

Remarque 4.2.7. La proposition 4.2.5 ne dit rien à propos des plus petits éléments de $\mathcal{R}_{-1,k}$. Pour k suffisamment grand, on peut montrer qu'il existe une constante $C > 0$ telle que

$$\underbrace{\binom{\beta k + \frac{(\beta-1)(\ell-1)}{2} + \ell - 1}{\ell}}_{\mathrm{val}_\ell(a_1^{\beta k + (\beta-1)(\ell-1)/2})} - \beta^\ell \underbrace{\binom{k+\ell-1}{\ell}}_{\mathrm{val}_\ell(a_1^k)} = C\, k^{\ell-2} + o(k^{\ell-2}) > 0,$$

ce qui implique que a_1^k est le premier mot de longueur k dont l'image par val_ℓ appartient à $\mathcal{R}_{-1,k}$.

Définition 4.2.8. Soit $i \in \{0, \dots, \beta - 1\}$. Au vu de la proposition 4.2.5, si m_k est le premier mot de $\mathcal{R}_{i,k}$ ($k \geq \mathbf{L}$) alors pour $t = 2, \dots, \ell$, le nombre $|\mathrm{rep}_\ell(\beta^\ell m_k)|_{a_t}$ dépend uniquement de $k \pmod{\beta^{\ell-1}}$ et est alors noté $n_t^{(j)}$ si $0 \leq j < \beta^{\ell-1}$ et $k \equiv j \pmod{\beta^{\ell-1}}$.
Pour $i \in \{0, \dots, \beta - 1\}$ et $j \in \{0, \dots, \beta^{\ell-1} - 1\}$, on note donc

$$M_{i,j} := a_2^{n_2^{(j)}} \cdots a_\ell^{n_\ell^{(j)}} \quad \text{et} \quad \mu_{i,j} := \mathrm{val}_{\{a_2,\dots,a_\ell\}}(M_{i,j}).$$

De la preuve de la proposition 4.2.5, il est clair que

$$\mu_{i,j} = -A_i(j + n\,\beta^{\ell-1}) \pmod{\beta^\ell}$$

pour tout n tel que $j + n\,\beta^{\ell-1} \geq \mathbf{L}$.

Exemple 4.2.9. Soient $\ell = 3$ et $\beta = 5$. Le nombre 171717 (resp. 172739) est le premier élément de $\mathcal{R}_{0,100}$ (resp. $\mathcal{R}_{1,100}$). On a

$$\mathrm{rep}_3(171717) = a^{95}b^3c^2 \text{ et } \mathrm{rep}_3(5^3\,171717) = a^{490}\mathbf{b^{14}c^0},$$
$$\mathrm{rep}_3(172739) = a^{55}b^{41}c^4 \text{ et } \mathrm{rep}_3(5^3\,172739) = a^{493}\mathbf{b^0c^{12}}.$$

Ainsi $M_{0,0} = b^{14}$ (resp. $M_{1,0} = c^{12}$) et $\mu_{0,0} = \mathrm{val}_{\{b,c\}}(b^{14}) = 105$ (resp. $\mu_{1,0} = \mathrm{val}_{\{b,c\}}(c^{12}) = 90$). Le nombre 333396 (resp. 334986) est le premier élément de $\mathcal{R}_{0,125}$ (resp. $\mathcal{R}_{1,125}$) et

$$\mathrm{rep}_3(333396) = a^{119}b^6c^0 \text{ and } \mathrm{rep}_3(5^3\,333396) = a^{615}\mathbf{b^{14}c^0},$$
$$\mathrm{rep}_3(333396) = a^{69}b^{41}c^{15} \text{ and } \mathrm{rep}_3(5^3\,333396) = a^{618}\mathbf{b^0c^{12}}.$$

Le résultat suivant décrit précisément comment la multiplication par β^ℓ affecte les représentations au sein d'une région $\mathcal{R}_{i,k}$.

Proposition 4.2.10. *Soient $k \geq \mathbf{L}$ tel que $k \equiv j \pmod{\beta^{\ell-1}}$ et m le premier élément de $\mathcal{R}_{i,k}$, $0 \leq i \leq \beta - 1$. Si $m + t$ appartient à $\mathcal{R}_{i,k}$, alors*

$$\mathrm{rep}_\ell(\beta^\ell(m+t)) = a_1^{s(t)}\, \mathrm{rep}_{\{a_2,\ldots,a_\ell\}}(\mu_{i,j} + t\ \beta^\ell)$$

où $s(t)$ est tel que $|\mathrm{rep}_\ell(\beta^\ell(m+t))| = \beta\, k + \frac{(\beta-1)(\ell-1)}{2} + i$.

Démonstration. C'est évident. □

Corollaire 4.2.11. *Soit $i \in \{0, \ldots, \beta-1\}$ and $j \in \{0, \ldots, \beta^{\ell-1} - 1\}$. On a*

$$\bigcup_{n : j + n\beta^{\ell-1} \geq \mathbf{L}} \{\mathrm{rep}_\ell(\beta^\ell\, m) \mid m \in \mathcal{R}_{i, j+n\beta^{\ell-1}}\}$$
$$= \left[a_1^*\, \mathrm{rep}_{\{a_2,\ldots,a_\ell\}}(\mu_{i,j} + \mathbb{N}\ \beta^\ell)\right] \bigcap \Sigma_\ell^C (\Sigma_\ell^{\beta^\ell})^*$$

où $C = \beta j + \frac{(\beta-1)(\ell-1)}{2} + i + n_0\beta^\ell$ et $n_0 = \min\{n \mid j + n\beta^{\ell-1} \geq \mathbf{L}\}$.

Exemple 4.2.12. Soient $\ell = 3$ et $\beta = 5$. Le premier élément de $\mathcal{R}_{0,100}$ (resp. $\mathcal{R}_{0,125}$) est $m = 171717$ (resp. $n = 333396$). On a $\#\mathcal{R}_{0,100} = 1022$ et $\#\mathcal{R}_{0,100} = 1590$. On obtient le tableau suivant illustrant la proposition 4.2.10.

t	$\Psi(\mathrm{rep}_3(5^3\,(m+t)))$	$\Psi(\mathrm{rep}_3(5^3\,(n+t)))$	$\Psi(\mathrm{rep}_{\{b,c\}}(\mu_{0,0} + 5^3\,t))$
0	$(490, 14, 0)$	$(615, 14, 0)$	$(14, 0)$
1	$(484, 0, 20)$	$(609, 0, 20)$	$(0, 20)$
2	$(478, 22, 4)$	$(603, 22, 4)$	$(22, 4)$
⋮	⋮	⋮	⋮
1021	$(0, 34, 470)$	$(125, 34, 470)$	$(34, 470)$
1022	×	$(124, 415, 90)$	$(415, 90)$
⋮	⋮	⋮	⋮
1589	×	$(0, 34, 595)$	$(34, 595)$

4.3 Sous-ensembles réguliers de $\mathcal{B}_\ell$

Pour étudier la conservation du caractère reconnaissable après multiplication par une constante, nous avons considéré une partie reconnaissable arbitraire X de $\mathbb{N}$ et nous avons cherché à montrer que $\beta^\ell X$ était encore reconnaissable. Nous rappelons donc ici la forme générale des sous-ensembles réguliers de $\mathcal{B}_\ell$.

Définition 4.3.1. Une partie X de $\mathbb{N}^k$ est dite *linéaire* si on peut trouver $p_0, p_1, \cdots, p_t \in \mathbb{N}^k$ tel que

$$X = p_0 + \mathbb{N}\, p_1 + \cdots + \mathbb{N}\, p_t = \{p_0 + \lambda_1 p_1 + \cdots + \lambda_t p_t \mid \lambda_1, \ldots, \lambda_t \in \mathbb{N}\}.$$

Les vecteurs $p_1, \ldots, p_t$ sont les *périodes* de X. Une partie X de $\mathbb{N}^k$ est *semi-linéaire* si elle est une union finie de parties linéaires. L'ensemble des périodes d'une partie semi-linéaire est l'union des ensembles des périodes des parties linéaires correspondantes. Si $x \in \mathbb{N}^k$, $[x]_i$ désigne sa i-ième composante.

Lemme 4.3.2. *Une partie X de $\mathbb{N}$ est $\mathcal{B}_\ell$-reconnaissable si et seulement si $\Psi(\mathrm{rep}_\ell(X))$ est une partie semi-linéaire dont les périodes sont des multiples entiers des vecteurs canoniques $\mathbf{e}_i$ où $[\mathbf{e}_i]_j = \delta_{i,j}$.*

Démonstration. C'est immédiat. □

Avec une telle caractérisation, nous obtenons le théorème 1.2.8 d'une manière différente.

Nous utilisons les notations suivantes pour l'automate minimal $\mathcal{A}_\ell$ de $\mathcal{B}_\ell$. On désigne par $\{q_1, \ldots, q_\ell\}$ l'ensemble des états de $\mathcal{A}_\ell$. Chacun des états est final, q_1 est initial et pour $1 \leq i \leq j \leq n$, on a une transition $q_i \xrightarrow{a_j} q_j$.

Proposition 4.3.3. *Soient p, $q \in \mathbb{N}$. L'ensemble $\Psi(\mathrm{rep}_\ell(q + \mathbb{N}\, p)) \subseteq \mathbb{N}^\ell$ est une union finie d'ensembles linéaires de la forme*

$$p_0 + \mathbb{N}\,\theta\,\mathbf{e}_1 + \cdots + \mathbb{N}\,\theta\,\mathbf{e}_\ell \quad \textit{pour un certain } \theta \in \mathbb{N}.$$

Démonstration. Pour tous $n_1, \ldots, n_\ell \in \mathbb{N}$, on a

$$\mathrm{val}_\ell(a_1^{n_1} \cdots a_\ell^{n_\ell}) = \sum_{j=1}^{\ell} \mathbf{v}_{q_j}(n_j + \cdots + n_\ell - 1).$$

En effet, nous devons compter les mots généalogiquement plus petits que $a_1^{n_1} \cdots a_\ell^{n_\ell}$ du langage. Nous avons d'abord les mots de longueur plus petite que $n_1 + \cdots + n_\ell$. On en compte $v_{q_1}(n_1 + \cdots + n_\ell - 1)$. Ensuite viennent les mots de longueur $n_1 + \cdots + n_\ell$. Parmi eux, il y a $v_{q_2}(n_2 + \cdots + n_\ell - 1)$ mots commençant par au moins $n_1{+}1$ lettres a_1. Après ça, il y a $v_{q_3}(n_3{+}\cdots{+}n_\ell{-}1)$ mots commençant par $a_1^{n_1}$ suivi d'au moins $n_2{+}1$ lettres a_2 et ainsi de suite.

Pour un entier $j \in \{1, \ldots, \ell\}$ donné, la suite $(v_{q_j}(n) \mod p)_{n\in\mathbb{N}}$ est ultimement périodique, de période π_j et de prépériode τ_j. En effet, on sait que la suite $(v_{q_j}(n))_{n\in\mathbb{N}}$ satisfait une relation linéaire de récurrence à coefficients constants (cf. la proposition 1.1.14). Soit $P = \mathrm{ppcm}_j\, \pi_j$ et $T = \max_j \tau_j$. Ainsi, pour tout $j \in \{1, \ldots, \ell\}$, si $n_1, \ldots, n_\ell > T$,

$$\mathrm{val}_\ell(a_1^{n_1} \cdots a_j^{n_j} \cdots a_\ell^{n_\ell}) \equiv \mathrm{val}_\ell(a_1^{n_1} \cdots a_j^{n_j+P} \cdots a_\ell^{n_\ell}) \pmod p.$$

Nous avons montré que pour tout $x \in \mathbb{N}^\ell$ tel que $T < \sup_i x_i \leq T + P$, x appartient à $\Psi(\mathrm{rep}_\ell(q + \mathbb{N}\, p))$ si et seulement $x + t_1\, P\mathbf{e}_1 + \cdots + t_\ell\, P\mathbf{e}_\ell$ appartient au même ensemble, pour tous $t_1, \ldots, t_\ell \in \mathbb{N}$. La conclusion s'ensuit facilement :

$$\Psi(\mathrm{rep}_\ell(q + \mathbb{N}\, p)) = F \cup \bigcup_{\substack{\mathrm{val}_\ell(a_1^{x_1} \cdots a_\ell^{x_\ell}) \in q + \mathbb{N}\, p \\ T < \sup x_i \leq T + P}} x + \mathbb{N}\, P\mathbf{e}_1 + \cdots + \mathbb{N}\, P\mathbf{e}_\ell$$

où $F = \{x \in \mathbb{N}^\ell \mid \mathrm{val}_\ell(a_1^{x_1} \cdots a_\ell^{x_\ell}) \in q + \mathbb{N}\, p \ \text{ et } \ \sup_i x_i \leq T\}$. □

Exemple 4.3.4. Dans la figure 4.1, l'axe des x (resp. l'axe des y) compte le nombre de a_1 (resp. a_2) dans un mot. Le mot vide correspond au coin inférieur gauche. Un point de $\mathbb{N}^2$ de coordonnée (i, j) a sa couleur déterminée par la valeur de $\mathrm{val}_2(a_1^i\, a_2^j)$ modulo p (avec $p = 3$, 5, 6 et 8 respectivement). Il y a donc p couleurs possibles. Sur cette figure, nous représentons les mots $a_1^i\, a_2^j$ pour $0 \leq i,\, j \leq 19$.

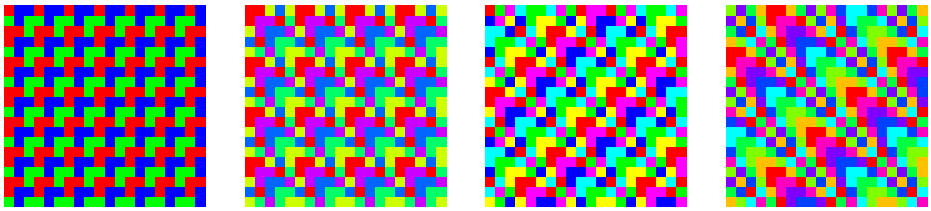

FIG. 4.1 – $\Psi(\mathrm{rep}_2(p\,\mathbb{N} + k))$ pour $p = 3$, 5, 6, 8.

4.4 Conclusions

Nous avons obtenu plusieurs résultats structurels sur les mots de $\mathcal{B}_\ell$ avant et après multiplication par β^ℓ. Malgré cela, une preuve de la conjecture 1 est toujours hors de notre portée.

Une façon de montrer la conjecture 1 pourrait être de considérer un ensemble linéaire de la forme $\mathcal{L} = p_0 + \mathbb{N}\,\gamma_1\, \mathbf{e_1} + \cdots + \mathbb{N}\,\gamma_\ell\, \mathbf{e}_\ell \subseteq \mathbb{N}^\ell$ avec $\gamma_i \in \mathbb{N}$ (en effet, les parties reconnaissables correspondent aux unions finies de tels ensembles) et d'étudier séparément la multiplication par β^ℓ pour chacun des

$$X_{i,j} := \mathrm{val}_\ell(\Psi^{-1}(\mathcal{L})) \bigcap \left(\bigcup_{n : j + n\beta^{\ell-1} \geq \mathbf{L}} \mathcal{R}_{i, j + n\beta^{\ell-1}} \right),$$

$i \in \{-1, \ldots, \beta - 1\}$, $j \in \{0, \ldots, \beta^{\ell-1} - 1\}$. Malheureusement, il n'est pas facile de voir que $\mathrm{rep}_\ell(\beta^\ell\, X_{i,j})$ est encore régulier même en tenant compte du lemme 4.2.2 et des propositions 4.2.5 et 4.2.10.

Jusqu'à présent, une telle approche semble être fructueuse seulement pour un alphabet de taille $\ell = 2$. En effet, pour un ensemble donné $\mathcal{L} = p_0 + \mathbb{N}\,\gamma_1\,\mathbf{e_1} + \mathbb{N}\,\gamma_2\,\mathbf{e_2}$, si on désigne par $m_{i,k}$ (resp. $\tau_{i,k}$) le premier élément de $\mathcal{R}_{i,k}$ (resp. le premier élément de $\mathcal{R}_{i,k} \cap \mathrm{val}_2(\Psi^{-1}(\mathcal{L}))$), alors $g_i(k) := \tau_{i,k} - m_{i,k}$ est ultimement périodique. Cet argument peut être utilisé en complément de la proposition 4.2.10 pour obtenir une preuve alternative du théorème 3.1.2. Mais nos essais pour prouver la conjecture 1 pour $\ell > 2$ avec les mêmes arguments restent sans succès. Une des difficultés rencontrées est que la différence entre deux éléments consécutifs de $\mathcal{R}_{i,k} \cap \mathrm{val}_\ell(\Psi^{-1}(\mathcal{L}))$ est constante pour $\ell = 2$ mais est non-constante pour $\ell \geq 3$. Par contre, des résultats partiels comme le lemme 4.2.4 semblent plaider en faveur de notre conjecture.

Bibliographie

[1] V. Bruyère, G. Hansel, Bertrand numeration systems and recognizability, *Theoret. Comput. Sci.* **181** (1997) 17–43.

[2] D.I.A. Cohen, *Basic Techniques of Combinatorial Theory,* John Wiley & Sons, New-York-Chichester-Brisbane, (1978).

[3] S. Eilenberg, *Automata, Languages, and Machines* vol. A, Academic Press, New York, (1974).

[4] R. L. Graham, D. E. Knuth and O. Patashnik, *Bernoulli Numbers*, Concrete Mathematics, 2nd edition, Addison-Wesley Publ. Co., Reading, MA, (1994).

[5] G. Katona, A theorem on finite sets, Theory of Graphs, Proc. Colloquium, Tihany, Hungary (1966), 187–207.

[6] P.B.A. Lecomte, M. Rigo, Numeration systems on a regular language, *Theory Comput. Syst.* **34** (2001), 27–44.

[7] M. Rigo, Numeration systems on a regular language : arithmetic operations, recognizability and formal power series, *Theoret. Comput. Sci.* **269** (2001), 469–498.

[8] M. Rigo, Construction of regular languages and recognizability of polynomials, *Discrete Math.* **254** (2002), 485–496.

[9] M. Rigo, *Théorie des automates et langages formels*, notes de cours, Université de Liège, (2003).

[10] J. Shallit, Numeration systems, linear recurrences, and regular sets, *Inform. and Comput.* **113** (1994), 331–347.

[11] A.Szilard, S. Yu, K. Zhang, J. Shallit, Characterizing regular languages with polynomial densities, *Proceedings of the 17th International Symposium on Mathematical Foundations of Computer Science, Lect. Notes in Comput. Sci.* **629** (1992), 494–503.