# Énumération et numération

Victor Marsault

EDITE – ED 130

2016-ENST-0017

**Thèse de doctorat**
**De Télécom-Paristech**

Spécialité Informatique

# Énumération et numération


## Victor Marsault


Soutenue le 2016-03-01 devant un jury composé de :

| | |
|---|---|
| **Marie-Pierre Béal**, Professeur à l'université Paris-Est, | Examinateur |
| **Olivier Carton**, Professeur à l'université Paris VII, | Président |
| **Christiane Frougny**, Professeur à l'université Paris VIII , | Examinateur |
| **Jérôme Leroux**, Directeur de recherche au CNRS/LaBRI, | Rapporteur |
| **Michel Rigo**, Professeur à l'université de Liège, | Rapporteur |
| **Jacques Sakarovitch**, Directeur de recherche au CNRS/LTCI, | Directeur |
| **Jeffrey Shallit**, Professeur à l'université de Waterloo, Canada, | Rapporteur |


Télécom ParisTech
École de l'Institut Mines Télécom – membre de ParisTech
46, rue Barrault – 75634 Paris Cedex 13 – Tél. + 33 (0)1 45 81 77 77 – www.telecom-paristech.fr

# Énumération et numération

**Victor Marsault**

La présente version 5 de ce mémoire est postérieure au manuscrit officiel, des corrections mineures ont été apportées.
Dernière modification : 2017–06–22.

# Sommaire

# Remerciements

Je tiens tout d'abord à remercier Jacques Sakarovitch, qui a dirigé mes travaux de recherche ces dernières années. Son amour de la langue française et son aversion pour la syntaxe approximative m'ont notamment permis d'améliorer mon expression, qui a toujours été mon point faible.

Je suis reconnaissant à Jérôme Leroux, Michel Rigo et Jeffrey Shallit pour avoir bien voulu être les rapporteurs de ce mémoire et dont les suggestions m'ont bien aidé à le rendre meilleur. Je suis également reconnaissant à Marie-Pierre Béal, Olivier Carton et Christiane Frougny pour avoir accepté de faire partie de mon jury.

Merci plus particulièrement à Michel Rigo et son équipe pour l'intérêt permanent qu'ils ont porté à mes recherches et les fructueuses discussions qui en ont découlé.

Pour les même raisons, je remercie Shigeki Akiyama, toute l'équipe de Télécom et plus généralement tous les chercheurs avec qui j'ai eu le plaisir de travailler ou de partager un café.

Je veux aussi exprimer ma reconnaissance à ma famille et à tous ceux qui ont contribué à faire de moi ce que je suis : ma mère, qui selon la légende m'aurait précocement appris à compter avec des carreaux de chocolat, dans le système unaire donc ; mon père, qui m'a transmis très tôt son goût des sciences et de la programmation ; Patrice, qui m'a initié aux mathématiques dès l'age de huit ans alors que l'école a préféré en attendre dix de plus ; mon frère, avec qui j'ai perfectionné l'art de la conversation musclée.

Je remercie bien sûr tous mes amis, pour avoir notamment égayé mes soirées pendant les gardes de Lucie : l'équipe de REC qui a permis l'élaboration d'un jeu aussi injouable qu'ingrat ; les différents *groupes* qui ont traversé les épreuves d'Alkyal, d'Anthéon, de Celwynvian, de Lepidstadt, d'Oblivion et d'ailleurs ; mon génial acolyte de pâtisserie qui a *amélioré* avec moi des recettes de grands chefs, ainsi que les malheureux cobayes qui en ont subi les résultats ; les as de la coopération individualiste, alertes dans l'espace et suicidaires ailleurs, qui ont assisté à l'émergence et à la chute de l'Alliance Éternelle ; la ménagerie hétéroclite au sein de laquelle j'ai considérablement accru ma dextérité et ma vivacité, des qualités qui sont, vous en conviendrez, d'une rare inutilité pour l'écriture d'un mémoire.

Enfin, les mots me manquent pour exprimer ma gratitude envers ma femme, mon roc dans la tempête, pour son soutien indéfectible depuis tant d'années.

# Résumé

Ce mémoire aborde et résout des problèmes assez différents, ayant tous trait à la numération, avec une certaine unité conceptuelle quant aux moyens mis en œuvre pour les résoudre : la théorie des automates.

Nous considérons d'abord les bases entières et présentons un algorithme quasi-linéaire et structurel permettant de décider si le langage accepté par un automate donné est la représentation d'un ensemble ultimement périodique d'entiers.

Ensuite, nous étudions la base rationnelle $\frac{p}{q}$ et particulièrement le langage $L_{\frac{p}{q}}$ des représentations des entiers dans cette base. Il s'agit d'un langage relativement complexe selon la théorie classique des langages formels : il ne satisfait aucune forme de lemme d'itération. Nous montrons que chaque monoïde finiment engendré est représenté par un langage aussi complexe que $L_{\frac{p}{q}}$. Nous prenons ensuite une perspective différente pour étudier $L_{\frac{p}{q}}$ : à chaque entier est associé un mot infini, dit minimal, et l'on étudie la fonction qui associe le mot minimal d'un entier $n$ à celui de son successeur $(n+1)$ ; nous montrons en particulier que cette fonction est réalisée par un transducteur infini dont la structure est très proche de celle du langage $L_{\frac{p}{q}}$.

Enfin, nous décrivons une manière de sérialiser les arbres infinis et les langages en des mots, appelés signatures, par le moyen d'un parcours en largeur. On remarque d'abord que les langages réguliers sont associés aux mots morphiques, ce qui rejoint le lien entre les systèmes de numération abstraits réguliers et les systèmes de numération morphiques (aussi dit de Dumont-Thomas). On traite ensuite le cas des signatures périodiques et l'on montre qu'elles sont liées aux bases rationnelles ; ceci donne également une procédure pour construire $L_{\frac{p}{q}}$ de façon très simple. Enfin, nous définissons une transformation d'automate, la surminimisation, qui réduit le nombre d'états d'un automate au delà de ce que permet la minimisation classique ; en contrepartie, un automate et sa surminimisation n'acceptent pas le même langage, mais seulement des langages avec le même arbre ordonné sous-jacent.

# Abstract

This memoir involves several domains of discrete mathematics and theoretical computer science, such as formal languages, numeration, combinatorics on words, algorithmic, complexity, *etc.* In summary, various problems, all from the general area of numeration, are addressed by means of automata and transducers theory.

We first consider integer base numeration systems. Given as a parameter an integer base $b$, we give a quasi-linear and structural algorithm to decide whether the language accepted by a given automaton is the set of the representations (in base b) of an ultimately periodic set of integers.

Secondly, we consider the rational base $\frac{p}{q}$ and particularly the language $L_{\frac{p}{q}}$ of the representations of integers in this base. It is a quite complex language according to the usual criteria: in particular, it has a property called FLIP (for Finite Left Iteration Property) which implies that $L_{\frac{p}{q}}$ does not satisfies any kind of pumping lemma. We prove that, if a monoid $M$ is finitely generated and contains only numbers that are representable in base $\frac{p}{q}$, then the language of all the representations of the numbers of $M$ possesses the FLIP property. We then study $L_{\frac{p}{q}}$ from a different perspective: with every integer is associated an infinite word called minimal and we consider the function that maps the minimal word associated with $n$ to the minimal word associated with $(n+1)$; we show that this function is realised by an infinite transducer whose structure is virtually the same as the one of $L_{\frac{p}{q}}$.

We finally describe a way to serialise a infinite tree and language into an infinite word, called signatures, by means of a breadth-first traversal. We first note that the signatures of regular languages form a subclass of morphic words, a result linked to the classical transformation automaton/word morphism. We then treat the case of periodic signatures and show their intrinsic relationship with rational base numeration systems: for every base $\frac{p}{q}$ the language $L_{\frac{p}{q}}$ has a periodic signature; given a finite sequence $\mathbf{r}$ of integer (that we call rhythm) the signature $\mathbf{r}^\omega$ generates a language that is a non-canonical way to represent the set of all integers in base $\frac{p}{q}$, where $\frac{p}{q}$ is the average of components of $\mathbf{r}$. The notion of signature allows us to define an automaton transformation, called surminimisation, that reduces the number of states of the input automaton, more so than a classical minimisation. However, whereas an automaton and its minimisation accept the same language, it is not necessarily the case for an automaton and its surminimisation: the surminimisation process indeed preserves only the underlying ordered tree.

# Introduction

Un nombre est une quantité abstraite que l'on ne peut écrire, énoncer ou communiquer en tant que tel ; il faut pour cela utiliser des conventions partagées par tous les interlocuteurs. Par exemple, la quantité 81 s'écrit usuellement comme il vient de l'être, c'est-à-dire le mot formé des chiffres 8 puis 1. Mais c'est loin d'être la seule façon de l'exprimer ; on l'énonce *quatre-vingt-un* ou *huitante-et-un* ; on le représente par la suite de bits *1010001* dans la mémoire d'un ordinateur ; on peut l'écrire  en braille ; et même quelqu'un ne sachant pas compter peut la représenter par un sac contenant 81 pierres ou par 81 traits de craie sur un mur. En revanche, il ne viendrait à l'idée de personne de l'écrire 41 (en base 20) ou 100 (en base 3), et de l'énoncer *huit-un* ou *huit-dix-un* (en suivant le modèle de *huit-cent-un*).

Un système de numération est l'intermédiaire entre un nombre abstrait et son écriture ; il explique comment *représenter* les nombres par des suites de chiffres (ou mots) et inversement comment calculer la *valeur* (ou *évaluer*) un mot. Tous les mots ne sont pas nécessairement des représentations, comme par exemple le mot 081 auquel on donne volontiers la valeur 81 mais qui n'en est pas la représentation. Dans ce mémoire, on s'intéressera principalement à la représentation des nombres entiers dans plusieurs classes de systèmes de numération.

Formellement, un *alphabet* est un ensemble de lettres (ou de chiffres) ; un *mot* est une suite de lettres appartenant à un certain alphabet ; et un *langage* est un ensemble de mots. A chaque système de numération est associé un alphabet, de telle sorte que chaque (nombre) entier soit représenté par un mot sur cet alphabet et donc que chaque ensemble d'entiers soit représenté par un langage (sur ce même alphabet).

La théorie des langages formels fournit une hiérarchie précise permettant de classer les langages selon leur complexité (voir par exemple [19]). En particulier, un langage est dit *régulier* (ou reconnaissable, ou rationnel) s'il est accepté par un automate fini ; il est alors considéré comme *simple* suivant cette hiérarchie. Quand on s'intéressera aux ensembles d'entiers dont la représentation est un langage régulier, on utilisera fréquemment la théorie des automates et transducteurs (décrite par exemple dans [72]).

Un même ensemble d'entiers est représenté par des langages différents dans des systèmes de numération différents ; certains de ces langages sont simples et d'autres complexes. En général, la complexité du langage des représentations de tous les entiers est un bon indicateur de la complexité du système de numération considéré.

[19] Olivier Carton, 2008, *Langages formels, calculabilité et complexité*.
[72] Jacques Sakarovitch, 2003, *Éléments de théorie des automates*.

Ce mémoire se divise en trois parties largement indépendantes. La première étudie en détail un problème particulier Ultime Periodicité en *base entière* et la deuxième aborde des questions diverses liées à *la base rationnelle*. La troisième partie présente une perspective différente sur les arbres et langages qui s'apparente à la notion de *système de numération abstrait*.

# I Sur la base entière

Les bases entières, dont nous rappelons la définition au chapitre 2, sont les manières les plus anciennes et naturelles de représenter les nombres. C'est le cas de la base 10, utilisée aujourd'hui par la quasi-totalité des êtres humains ou du binaire (base 2) et de l'hexadécimal (base 16) utilisés en informatique. D'autres bases entières (et souvent des superpositions de bases entières) ont été utilisées dans l'Histoire à plus ou moins grande échelle.

En base entière $p$, on attribue à chaque chiffre le poids $p^i$, où $i$ est la position du chiffre considéré ; par exemple, dans le mot 805 le poids du chiffre 5 est $10^0 = 1$, celui du chiffre 0 est $10^1 = 10$ et celui du chiffre 8 est $10^2 = 100$. L'évaluation d'un mot est alors la somme de ses chiffres ainsi pondérée ; l'évaluation du mot 805 est l'entier $(8 \times 10^2) + (0 \times 10^1) + (5 \times 10^0)$. Inversement, la représentation d'un entier peut se calculer de deux manières différentes :

- du chiffre le plus significatif au chiffre le moins significatif, au moyen de l'*algorithme glouton* ;
- du chiffre le moins significatif au chiffre le plus significatif, au moyen de l'*algorithme d'Euclide*.

Le mot obtenu est formé de lettres prises dans l'*alphabet canonique (de taille $p$)*, noté $[\![p]\!]$ et composé des lettres (ou chiffres) allant de 0 à $(p-1)$. Un mot (sur $[\![p]\!]$) n'est la représentation d'aucun entier s'il commence par un 0 ; le langage des représentations d'entiers est donc très simple, celui des mots qui ne commencent pas par un 0.

Certains ensembles d'entiers sont représentés par des langages simples dans certaines bases et des langages complexes dans d'autres ; c'est le cas par exemple de l'ensemble $\{1, 2, 4, 8, 16, \ldots\}$ des puissances de 2 qui est représenté par le langage régulier $10^*$ en base 2 et par un langage complexe (non-algébrique) en base 3.

On peut alors se poser la question de quels ensembles d'entiers sont représentés par des langages simples dans toutes les bases. Il est connu depuis longtemps (certains disent depuis Pascal, voir prologue de [72]) que c'est le cas des ensembles ultimement périodiques. En revanche, ce ne sont que les travaux de Cobham [25] qui démontrent réciproquement qu'un ensemble d'entiers représenté par des langages réguliers dans toutes les bases est nécessairement ultimement périodique.

Soit un langage régulier $L$ représentant en base $p$ un ensemble $E$. Cet ensemble $E$ est donc simple pour la base $p$ ; pour déterminer s'il est simple pour toutes les

[72] Jacques Sakarovitch, 2003, *Éléments de théorie des automates*.

[25] Alan Cobham, 1969, *On the Base-Dependence of Sets of Numbers Recognizable by Finite Automata*.

autres bases, il faut et il suffit de déterminer si $E$ est ultimement périodique. Ce problème, appelé ULTIME PÉRIODICITÉ, est formalisé plus simplement par : peut on décider si un langage régulier sur $[\![p]\!]$ est la représentation d'un ensemble ultimement périodique.

Plusieurs réponses successives ont été données à ce problème. D'abord, par Honkala [40], puis, dans un contexte plus général, par Muchnik [62], jusqu'à arriver à l'algorithme polynomial de Leroux [46], à ce jour le plus efficace. Nous prenons, pour la base entière seulement, la convention déjà utilisée par Leroux que les automates lisent les représentations d'entiers avec le chiffre le moins significatif en premier, ce qui est contraire à l'usage courant et aux exemples donnés précédemment dans cette introduction. Dans le chapitre 3, nous donnons un algorithme quasi-linéaire pour résoudre ce problème dans sa forme originelle.

**THÉORÈME I** – *Soient une base entière $p$ et un automate déterministe $\mathcal{A}$ sur l'alphabet $[\![p]\!]$. On note $n$ le nombre d'états de $\mathcal{A}$ et $m$ son nombre de transitions. Le problème ULTIME PÉRIODICITÉ peut être décidé en temps $O(n \log(n) + m)$.*

La démonstration repose sur la définition de conditions structurelles, rassemblées sous le nom de critère (UP), que l'automate $\mathcal{A}$ ou que ses composantes fortement connexes doivent satisfaire. Le critère (UP) répond donc également à la question moins formelle : à quoi ressemble un automate dont le langage représente un ensemble ultimement périodique ?

# II Sur la base rationnelle

Les bases rationnelles, décrites précisément dans le chapitre 4, sont des généralisations des systèmes à base entière introduite dans [2] et dans lesquels la base est un nombre rationnel $\frac{p}{q}$ supérieur à 1.

La valeur d'un mot se calcule comme en base entière, si ce n'est que le poids du chiffre à la position $i$ n'est pas la $i$-ème puissance de la base mais $\frac{1}{q}\left(\frac{p}{q}\right)^i$ ; en base $\frac{3}{2}$, le mot 21 s'évalue donc à $(2 \times \frac{3}{4}) + (1 \times \frac{1}{2}) = 2$. Contrairement à la base entière, certains mots ont des valeurs non-entières, comme par exemple les mots 1 et 220 qui s'évaluent respectivement à $\frac{1}{2}$ et $\frac{15}{4}$ (toujours en base $\frac{3}{2}$). Néanmoins, chaque entier possède une représentation ; c'est un mot sur l'alphabet $[\![p]\!]$ qui est nécessairement calculé du chiffre le moins significatif au chiffre le plus significatif à l'aide d'une adaptation de l'algorithme d'Euclide utilisé pour les bases entières.

Le langage, noté $L_{\frac{p}{q}}$, des représentations des entiers en base $\frac{p}{q}$ est assez complexe selon la hiérarchie des langages formels (il est non-algébrique). De plus, il satisfait

[40] Juha HONKALA, 1986, *A Decision Method for The Recognizability of Sets Defined by Number Systems.*

[62] Andrei A. MUCHNIK, 2003, *The definable criterion for definability in Presburger arithmetic and its applications.*

[46] Jérôme LEROUX, 2005, *A polynomial time Presburger criterion and synthesis for number decision diagrams.*

[2] Shigeki AKIYAMA, Christiane FROUGNY et Jacques SAKAROVITCH, 2008, *Powers of rationals modulo 1 and rational base number systems.*

une propriété de forte irrégularité, que nous avons formalisée sous le nom de FLIP (acronyme de *Finite Left Iteration Property*), qui implique qu'il ne satisfait aucune forme de lemme d'itération. Cela n'empêche pas $L_{\frac{p}{q}}$ de posséder des propriétés qui suggèrent une forte régularité. En effet, l'algorithme d'Euclide modifié permet très facilement de calculer un des mots qui le compose. De plus, nous verrons dans la troisième partie qu'il est possible d'engendrer $L_{\frac{p}{q}}$ de façon périodique par un rythme (*cf.* théorème VI, plus loin).

L'étude de la base $\frac{p}{q}$ se confond largement avec l'étude de ce langage paradoxal à laquelle nous apportons notre contribution.

Il est démontré dans [2] qu'en base rationnelle, l'addition peut s'effectuer de façon simple directement sur les représentations : un automate, appelé *additionneur* (en réalité, un transducteur droit, lettre-à-lettre et séquentiel) permet de transformer les représentations de deux entiers donnés $n$ et $m$ en la représentation de l'entier $(n + m)$. Il s'agit encore d'un paradoxe : on ne peut pas facilement énumérer les membres de $L_{\frac{p}{q}}$ alors que l'on peut aisément additionner deux de ses membres pour en obtenir un troisième.

L'existence de l'additionneur implique que l'ensemble $V_{\frac{p}{q}}$, formé des évaluations de tous les mots sur l'alphabet canonique $[\![p]\!]$, est stable par addition donc est un monoïde additif. Plus généralement, l'addition étant l'opération la mieux comprise en base rationnelle, il semble que les sous-monoïdes additifs de $V_{\frac{p}{q}}$ ont une importance particulière.

L'ensemble des entiers $\mathbb{N}$ est bien sûr un monoïde additif monogène, puisque tout entier $n$ peut s'écrire comme une somme $n = 1 + 1 + \cdots + 1$ de 1, et nous avons vu qu'il est représenté par le langage $L_{\frac{p}{q}}$ qui est FLIP. On montre dans le chapitre 5 que cela s'étend à tous les sous-monoïdes finiment engendrés de $V_{\frac{p}{q}}$.

**Théorème II** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q > 1$. Tous les sous-monoïdes additifs et finiment engendrés de $V_{\frac{p}{q}}$ sont représentés en base $\frac{p}{q}$ par des langages FLIP.*

La démonstration repose sur la définition de l'*incrémenteur*, un transducteur qui réalise l'addition par une constante (ou translation) et qui préserve la propriété FLIP ; il s'agit d'une spécialisation de l'additionneur. On montre ensuite que tout monoïde finiment engendré est inclus dans une union finie de translations de $\mathbb{N}$. Puisque la classe des langages FLIP est stable par toutes ces opérations (inclusion, union finie, translation), le théorème s'ensuit.

Dans tous les systèmes de numération, l'étude des représentations des ensembles périodiques est un passage obligé. Cette notion nécessite cependant une adaptation en base $\frac{p}{q}$ : un ensemble périodique d'entiers est inclus dans $\mathbb{N}$ donc sa représentation en base $\frac{p}{q}$ est nécessairement FLIP car incluse dans le langage FLIP $L_{\frac{p}{q}}$.

En généralisant la notion de congruence aux nombres rationnels, on montre que les ensembles périodiques de nombres dont la période est première avec $q$ sont tous représentés par des langages réguliers. Au contraire, si $E$ est un ensemble d'entiers périodique de période $q$ (dans le sens classique) on montre que $E$ est inséparable de son complémentaire dans $\mathbb{N}$ par un automate fini. Nous conjecturons que ce deuxième cas peut s'étendre à toutes les périodes qui ne sont pas premières avec $q$.

Dans le chapitre 6, nous prenons une perspective différente à l'étude de $L_{\frac{p}{q}}$. Une valeur réelle est donnée à chaque mot de $L_{\frac{p}{q}}$ en l'évaluant *après la virgule*. Par exemple en base $\frac{3}{2}$, le mot 21 est en fait évalué comme $0,21$ donc à $\frac{8}{9}$ ; en effet le poids du chiffre 2 est $\frac{1}{3} = (p^{-1}/q^{-1+1})$ et celui du chiffre 1 est $\frac{2}{9} = (p^{-2}/q^{-2+1})$. Le langage $L_{\frac{p}{q}}$ (en fait la clôture topologique de $0^*L_{\frac{p}{q}}$) peut alors être vu comme un arbre fractal infini, noté $\mathcal{T}_{\frac{p}{q}}$, dont les branches sont étiquetées par des mots infinis qui s'évaluent (après la virgule) à un nombre réel.

A chaque entier $n$ est associé un nœud de cet arbre fractal, le nœud atteint par la représentation de $n$. De ce nœud part un plus petit mot infini, appelé le *mot minimal* de $n$. On s'intéresse à la fonction $\xi$ qui associe le mot minimal de $n$ à celui $(n+1)$.

On construit pour cela un transducteur, noté $\mathcal{D}_{\frac{p}{q}}$, qui a une forme très particulière en base $\frac{3}{2}$, ou plus généralement dans les bases $\frac{p}{q}$ vérifiant $p = 2q-1$. Dans ce cas, $\mathcal{D}_{\frac{p}{q}}$ est construit en remplaçant chaque étiquette de $\mathcal{T}_{\frac{p}{q}}$ par un ensemble de couples de lettres (qui ne dépend que de cette étiquette). Le cas $p \neq (2q-1)$ nécessite une étape supplémentaire qui consiste à modifier préalablement l'alphabet de $\mathcal{T}_{\frac{p}{q}}$.

**THÉORÈME III** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q > 1$. Le transducteur $\mathcal{D}_{\frac{p}{q}}$ réalise la fonction $\xi$.*

Itérer la fonction $\xi$ est en quelque sorte équivalent à parcourir les branches (minimales) de $\mathcal{T}_{\frac{p}{q}}$. Si cette fonction avait été simple, cela aurait voulu dire que l'on peut facilement engendrer une branche de $\mathcal{T}_{\frac{p}{q}}$ à partir d'une autre, et donc trouver un peu d'ordre dans le chaos du langage $L_{\frac{p}{q}}$. Il est donc remarquable que $\mathcal{D}_{\frac{p}{q}}$ et $\mathcal{T}_{\frac{p}{q}}$ soient si proches : la structure nécessaire pour transformer une branche de $\mathcal{T}_{\frac{p}{q}}$ en la suivante est essentiellement l'arbre $\mathcal{T}_{\frac{p}{q}}$ lui-même.

À chaque entier $n$ est également associé *un mot maximal*, défini de façon similaire au mot minimal ; il est également l'image, par un morphisme strictement alphabétique, du mot minimal de $(n+1)$. Les évaluations après la virgule des mots partant de $n$ dans $\mathcal{T}_{\frac{p}{q}}$ forment un intervalle (réel) dont les bornes sont les valeurs respectives des mots minimal et maximal de $n$. On appelle *envergure* de $n$ la longueur de cet intervalle et on note $\mathsf{S}_{\frac{p}{q}}$ l'ensemble de toutes les envergures. Il s'avère que l'adhérence de cet ensemble, c'est-à-dire le plus petit ensemble fermé le contenant, possède des propriétés topologiques très différentes suivant le signe de $(p-(2\,q-1))$.

**THÉORÈME IV** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q > 1$.*

**a)** *Si $p \leqslant (2\,q-1)$, alors l'adhérence de $\mathsf{S}_{\frac{p}{q}}$ est un intervalle.*

**b)** *Si $p > (2\,q-1)$, alors l'adhérence de $\mathsf{S}_{\frac{p}{q}}$ est un ensemble de Cantor.*

La démonstration repose sur la relation triviale entre le mot minimal de $(n+1)$ et le mot maximal de $n$. La fonction $\xi$ transforme le mot minimal de $n$ en le mot minimal de $(n+1)$ ; elle est donc proche de celle transformant le mot minimal de $n$ en le mot maximal de $n$. La différence de comportements dans les deux cas du théorème précédent provient de l'étape «préalable» (à la construction de $\mathcal{D}_{\frac{p}{q}}$) qui modifie l'alphabet de $\mathcal{T}_{\frac{p}{q}}$ : quand $p > (2\,q-1)$ des arêtes de $\mathcal{T}_{\frac{p}{q}}$ sont supprimées, ce

qui produit la structure lacunaire de $\mathsf{S}_{\frac{p}{q}}$ ; dans l'autre cas, des arêtes sont ajoutées, ce qui laisse ses propriétés topologiques essentiellement inchangées.

# III Signature et étiquetage

Jusqu'à maintenant, nous n'avons utilisé que des alphabets de chiffres, donc naturellement ordonnés par $0 < 1 < 2 < \cdots$ Dans la troisième partie, on considère des alphabets de lettres mais néanmoins muni d'un ordre total.

L'ordre de l'alphabet induit sur les mots d'une part l'ordre lexicographique (utilisé par exemple dans les dictionnaires) et d'autre part l'ordre radiciel, une variante du précédent. Un mot $u$ est plus petit qu'un autre mot $v$ dans l'*ordre radiciel* si $u$ est plus court que $v$, ou s'ils sont de la même longueur et que $u$ est plus petit que $v$ dans l'ordre lexicographique.

Ce deuxième ordre est respecté par tous les systèmes de numération : étant donné deux entiers $n$ et $m$ tels que $n < m$, la représentation de $n$ est plus petite, dans l'ordre radiciel, que la représentation de $m$. En particulier, les bases entières et rationnelles respectent l'ordre radiciel. Les systèmes de numération abstraits (SNA) ont été introduits dans [44] en utilisant cette propriété comme une définition. Étant donné un langage $L$ (sur un alphabet ordonné), les mots de $L$ sont ordonnés selon l'ordre radiciel, ce langage $L$ est alors considéré comme un système de numération dit *abstrait* dans lequel chaque entier $n$ a pour représentation le $(n+1)$-ème mot de $L$ : le plus petit mot est la représentation de 0, le deuxième plus petit est la représentation de 1, etc.

Dans le chapitre 7 est introduite la notion de signature d'un arbre (enraciné, ordonné et de degré fini). Un tel arbre possède un parcours en largeur canonique et sa *signature* est définie comme la suite des degrés de ses nœuds pris dans l'ordre de ce parcours en largeur. La signature est caractéristique de l'arbre : deux arbres différents ont toujours des signatures différentes. Suivant la direction considérée, un arbre est *serialisé* en sa signature ou une signature *génère* l'arbre associé.

Si de plus les arcs de l'arbre sont étiquetés, on appelle *étiquetage* la suite des étiquettes de ses arcs visités par ce même parcours en largeur canonique. Le couple signature/étiquetage est similairement caractéristique de l'arbre étiqueté considéré. Puisqu'un langage clos par préfixe est essentiellement un arbre étiqueté, on parle également d'étiquetage et de signature d'un tel langage.

Calculer la signature d'un langage (clos par préfixe) consiste à parcourir le langage dans l'ordre radiciel et donc à le considérer comme un SNA. C'est pourquoi cette notion est particulièrement adaptée à l'étude des langages des représentations des entiers dans les systèmes de numération, comme nous verrons dans la suite.

On s'intéresse d'abord aux signatures des langages réguliers (et clos par préfixe), qui se révèlent être des mots morphiques. Un mot *purement morphique* est la limite $\sigma^\omega(\alpha)$ de l'itération d'un morphisme de mots $\sigma$ prolongeable en $\alpha$ (c'est-à-dire tel que le mot $\sigma(\alpha)$ commence par un $\alpha$). Un mot *morphique* est l'image $f(\sigma^\omega(\alpha))$

[44] Pierre Lecomte et Michel Rigo, 2001, *Numeration systems on a regular language.*

d'un mot purement morphique par un morphisme lettre-à-lettre $f$. On appelle s-morphique un mot morphique $f_\sigma(\sigma^\omega(\alpha))$ où $f_\sigma$ est l'indicateur de la longueur de $\sigma$, c'est-à-dire que pour toute lettre $\beta$, $f_\sigma(\beta)$ est la longueur de $\sigma(\beta)$.

**THÉORÈME V** – *Un langage clos par préfixe $L$ est régulier si et seulement si la signature étiquetée de $L$ est s-morphique.*

Ce théorème est une conséquence d'un résultat du à Rigo et Maes [71] et dont le principe remonte au travaux de Cobham dans [26]. La méthode utilisée est également proche de la définition des *systèmes de numération morphiques* introduits dans [30] (aussi appelés *de Dumont-Thomas* en référence à leurs auteurs). Ceux-ci n'étaient pas destinés à être utilisés en tant que systèmes de numération mais comme intermédiaires pour calculer des fonctions relatives aux morphismes de mots. Nous proposons une définition de ces systèmes en tant que SNA.

Dans le chapitre 8, on s'intéresse aux signatures (et étiquetages) purement périodiques dont on appelle la période *rythme*. Un rythme $\mathbf{r}$ est donc une suite finie d'entiers dont on appelle *paramètre directeur* le couple $(q, p)$, où $q$ est la longueur de $\mathbf{r}$ et $p$ la somme de ses composantes. A chaque rythme est associé un chemin dans le plan discret $\mathbb{Z}\times\mathbb{Z}$ qui va de l'origine $(0, 0)$ au point $(q, p)$. On montre alors que les signatures et étiquetages des langage $L_{\frac{p}{q}}$ sont périodiques et que les rythmes associés sont liés aux mots de Christoffel (étudiés en combinatoire des mots, *cf.* [13]).

**THÉORÈME VI** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q \geqslant 1$.*

**a)** *La signature du langage $L_{\frac{p}{q}}$ est $\mathbf{r}_{\frac{p}{q}}^{\omega}$, où $\mathbf{r}_{\frac{p}{q}}$ est le rythme dont le chemin dans $\mathbb{Z}\times\mathbb{Z}$ est le mot de Christoffel de pente $\frac{p}{q}$.*

**b)** *L'étiquetage du langage $L_{\frac{p}{q}}$ est $\boldsymbol{\gamma}_{\frac{p}{q}}^{\omega}$, où $\boldsymbol{\gamma}_{\frac{p}{q}}$ est la suite résultant de la génération de $\mathbb{Z}/p\mathbb{Z}$ par $q$ : $\boldsymbol{\gamma}_{\frac{p}{q}} = 0\,q\,(2\,q\,\%\,p)\cdots((p-1)\,q\,\%\,p)$.*

Ce théorème donne en particulier une manière très simple de construire $L_{\frac{p}{q}}$ (en tant qu'arbre infini étiqueté) ; on peut l'engendrer périodiquement par son rythme.

On prend ensuite un rythme quelconque de paramètre directeur $(q, p)$ dont on appelle *taux de croissance* la fraction irréductible $\frac{p'}{q'}$ égale à $\frac{p}{q}$. En s'inspirant de l'étiquetage de $L_{\frac{p'}{q'}}$, on construit un étiquetage spécial, noté $\boldsymbol{\gamma}_{\mathbf{r}}$. Le langage engendré par $\mathbf{r}$ et $\boldsymbol{\gamma}_{\mathbf{r}}$, s'il n'est pas égal à $L_{\frac{p'}{q'}}$, est néanmoins fortement lié à la base $\frac{p'}{q'}$.

**THÉORÈME VII** – *Soit un rythme $\mathbf{r}$ de taux de croissance $\frac{p'}{q'}$. Le langage engendré par $\mathbf{r}$ et l'étiquetage spécial associé $\boldsymbol{\gamma}_{\mathbf{r}}$ est une représentation non-canonique des entiers en base $\frac{p'}{q'}$.*

[71] Michel RIGO et Arnaud MAES, 2002, *More on Generalized Automatic Sequences.*

[26] Alan COBHAM, 1972, *Uniform Tag Sequences.*

[30] Jean-Marie DUMONT et Alain THOMAS, 1989, *Systèmes de Numération et Fonctions Fractales Relatifs aux Substitutions.*

[13] Jean BERSTEL, Aaron LAUVE, Christophe REUTENAUER et Franco SALIOLA, 2008, *Combinatorics on Words : Christoffel Words and Repetition in Words.*

Une *représentation non-canonique des entiers* en base $\frac{p'}{q'}$ est un langage $L$ sur un alphabet non-canonique (c'est-à-dire qui n'est pas $[\![p']\!]$) qui, lorsqu'il est considéré comme un SNA, admet la même fonction d'évaluation que la base $\frac{p'}{q'}$. Autrement dit, $L$ est une représentation non-canonique des entiers en base $\frac{p'}{q'}$ si, pour tout entier $n$, le $(n+1)$-ème mot de $L$ (dans l'ordre radiciel) est évalué, en base $\frac{p'}{q'}$, à $n$.

Le chapitre 9 définit une transformation d'automates appelée *surminimisation* qui permet de réduire le nombre d'états d'un automate au-delà de ce que permet la minimisation classique. En contrepartie, l'automate obtenu n'accepte plus le même langage mais seulement un langage dont l'arbre ordonné sous-jacent est identique ; en quelque sorte, elle ne conserve que la *silhouette* du langage.

Soit $\mathcal{A}$ un automate sur un alphabet ordonné $A$. L'ordre de $A$ peut être relevé aux transitions sortantes de chaque état : une transition étiquetée par une lettre $a$ est plus petite que les transitions étiquetées par des lettres plus grandes que $a$. Surminimiser $\mathcal{A}$ consiste en deux étapes. Les transitions de $\mathcal{A}$ sont d'abord réétiquetées en ne conservant que leur ordre : la plus petite transition est réétiquetée par 0, la deuxième plus petite par 1, etc. L'automate ainsi obtenu est ensuite minimisé.

Étant donné un langage régulier $L$, donc accepté par un automate $\mathcal{A}$, on appelle *réduction d'étiquetage* de $L$ le langage accepté par la surminimisation de $\mathcal{A}$. Un SNA régulier (ou SNAR) n'est rien d'autre qu'un langage régulier sur un alphabet ordonné, sur lequel la réduction d'étiquetage est donc défini. Deux SNAR sont dit T-équivalents si leurs réductions d'étiquetage respectives sont égales. Dans ce cas, la fonction de conversion de l'un en l'autre (c'est-à-dire qui envoie la représentation de chaque entier $n$ dans le premier sur la représentation de $n$ dans le second) est très simple.

**Théorème VIII** – *La fonction de conversion entre deux SNAR T-équivalents est réalisée par un transducteur fini, lettre-à-lettre et séquentiel pur.*

Un système de numération positionnel est défini par une suite croissante d'entiers $U = (U_i)_{i\in\mathbb{N}}$. Le $i$-ème élément de cette suite est le poids associé au $i$-ème chiffre ; la donnée de la suite $U$ définit ainsi une fonction d'évaluation sur les mots (dont les lettres sont des chiffres) : l'évaluation d'un mot $a_k \cdots a_1\, a_0$ est $a_k\, U_k + \cdots + a_1\, U_1 + a_0\, U_0$. La base entière $p$ est donc par exemple le système positionnel où, pour tout entier $i$, $U_i = p^i$.

Sous certaines conditions, tous les entiers peuvent être représentés par des mot sur un certain alphabet $[\![p_U]\!]$ ; ces représentations sont calculées par *l'algorithme glouton* et forment le langage $L(U)$ des représentations des entiers dans le système $U$.

Les systèmes positionnels sont tous des SNA mais ils ne sont pas toujours toujours des SNAR. Après une généralisation de la réduction d'étiquetage aux langages quelconques (c'est-à-dire pas nécessairement réguliers) et donc aux SNA, on vérifie que tous les systèmes positionnels ont un étiquetage irréductible.

**Théorème IX** – *Pour tout système positionnel $U$, le langage calable $0^*L(U)$ est irréductible.* [1]

1. L'adjectif *calable* signifie que des zéros de tête sont autorisés. Voir définition 1.17, page 28.

En définitive, ce mémoire aborde et résout des problèmes assez différents, ayant tous trait à la numération avec une certaine unité conceptuelle quant aux moyens mis en œuvre pour les résoudre : la théorie des automates.

Chapitre 1

# Préliminaires

Ce chapitre rassemble les notions usuelles sur les mots, les langages, les automates et les systèmes de numération abstraits.

## Chiffres, lettres, mots, langages

Étant donné deux entiers strictement positifs $n$ et $m$, on note $\frac{n}{m}$ leur division dans $\mathbb{Q}$ et respectivement par $n \% m$ et $n \div m$ le reste et le quotient de la division Euclidienne de $n$ par $m$ c'est-à-dire qui vérifient $n = (n \div m)\, m + (n \% m)$ et $0 \leqslant (n \% m) < m$. L'entier $n \div m$ est aussi égal à $\left\lfloor \frac{n}{m} \right\rfloor$.

Un *alphabet* est un ensemble fini de symboles appelés *lettres* ; quand ces lettres sont des entiers, ils sont appelés *chiffres*. On utilisera principalement des alphabets *ordonnés*, c'est-à-dire dont les lettres sont totalement ordonnées ; pour les alphabets de chiffres, cet ordre coïncide avec l'ordre usuel des entiers. En particulier, on note $[\![n]\!] = \{0, 1, 2, \ldots, (n-1)\}$ l'alphabet des $n$ plus petites lettres. L'ensemble des *mots* (fini) sur l'alphabet $A$ est le monoïde libre engendré par $A$, noté $A^*$ ; un mot (fini) $u \in A^*$ est donc une suite finie $(a_0, a_1, \ldots, a_{n-1})$ de lettres de $A$ ; on note ce mot plus simplement $u = a_0\, a_1 \cdots a_{n-1}$ et on appelle *longueur* l'entier $|u| = |a_0\, a_1 \cdots a_{n-1}| = n$. Le mot vide est noté $\varepsilon$.

Le monoïde $A^*$ est naturellement muni de la *concaténation*, une loi de composition interne qui à deux mots $u = a_0\, a_1 \cdots a_{n-1} \in A^*$ et $v = b_0\, b_1 \cdots b_{m-1} \in A^*$ associe le mot $u\, v = a_0\, a_1 \cdots a_{n-1}\, b_0\, b_1 \cdots b_{m-1}$. Si un mot $w$ se factorise comme $w = u\, v$, alors $u$ est appelé un préfixe de $w$, ce que l'on note $u \sqsubseteq w$, et $v$ un suffixe de $w$.

Un langage $L$ sur l'alphabet $A$ est un sous-ensemble de $A^*$. On note $\mathsf{Pre}\,(L)$ l'ensemble de tous les préfixes des mots de $L$ :

$$\mathsf{Pre}\,(L) = \{\, u \mid u \sqsubseteq w \ \text{ et } \ w \in L \,\}\,;$$

On appelle $\mathsf{Pre}\,(L)$ *le langage des préfixes de* $L$. On dit qu'un langage $L$ est *clos par préfixe* s'il est égal au langage de ses préfixes : $\mathsf{Pre}\,(L) = L$.

## Topologie des mots infinis

Nous donnons ici une description succincte de l'ensemble des mots infinis sur un alphabet fini, voir par exemple [67] pour plus de détails. Un mot infini sur un alphabet $A$ est une suite de lettres $a_0\,a_1\cdots a_k\cdots$ appartenant à $A$. L'ensemble des mots infinis sur $A$ est noté $A^\omega$.

Un *préfixe* d'un mot infini $a_0\,a_1\cdots a_k\cdots$ est un mot fini de la forme $a_0\,a_1\cdots a_i$, pour un certain entier $i$. Pour tout langage $L \subseteq A^\omega$, le langage $\mathsf{Pre}\,(L)$ est donc un sous-ensemble de $A^*$ entièrement défini.

La topologie sur les mots infinis suit intuitivement la règle selon laquelle deux mots infinis sont d'autant plus proches que leur plus grand préfixe commun est long. Formellement, la *distance* entre deux mots infinis $v$ et $w$ est $d(v,w) = 2^{-|u|}$ où $u$ est le plus grand préfixe commun de $v$ et $w$. La topologie sur $A^\omega$ est induite par cette distance. Elle possède entre autres la propriété suivante.

**Lemme 1.1** – *Toute suite à valeurs dans $A^\omega$ possède une sous-suite convergente.*

Démonstration. Soit une suite $(w_n)_{n\in\mathbb{N}}$ à valeurs dans $A^\omega$. Pour tout mot fini $u \in A^*$, on note $I_u \subseteq \mathbb{N}$ l'ensemble des indices tels que $w_i$ admet $u$ comme préfixe. On note $u_0 = \varepsilon$ et donc $I_{u_0} = \mathbb{N}$ est infini. Soit $j \in N$, supposons que $I_{u_j}$ est infini, alors il existe une lettre $a_j \in A$ tel que $I_{u_j a_j}$ est infini (car $A$ est fini). On note donc $u_{(j+1)} = u_j\,a_j$, si bien que pour tout $j \in \mathbb{N}$, $I_{u_j}$ est infini.

On définit l'extraction $\varphi : \mathbb{N} \to \mathbb{N}$ par $\varphi(0) = 0$ et, pour tout entier $j > 0$, $\varphi(j)$ est le plus petit indice de $I_{u_j}$ strictement plus grand que $\varphi(j-1)$. La sous-suite $(w_{\varphi(n)})_{n\in\mathbb{N}}$ converge alors vers $a_0\,a_1\cdots a_k\cdots$. □

Puisque $A^\omega$ est un espace métrique, une partie de $A^\omega$ est compacte si et seulement si elle est séquentiellement compacte ; le lemme suivant découle donc directement du lemme précédent.

**Lemme 1.2** – *Tout ensemble fermé de $A^\omega$ est compact.*

Par abus de langage on appelle *clôture topologique* d'un langage $L$ de mots **finis** l'ensemble de mots infinis suivant :

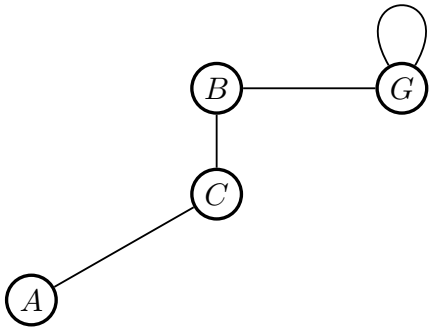
$$\mathsf{adh}(L) \;=\; \{\, a_0a_1\cdots a_k\cdots \mid \forall i \in \mathbb{N} \qquad a_0a_1\cdots a_i \in \mathsf{Pre}\,(L)\,\}\ .$$

ou, de façon équivalente, $\mathsf{adh}(L)$ est le plus grand langage $K$ de mot infinis tel que $\mathsf{Pre}\,(K) \subseteq \mathsf{Pre}\,(L)$.

## Graphes et graphes orientés

Un *graphe* est un couple $(V,E)$ formé d'un ensemble de *sommets* $V$ et d'un ensemble d'*arêtes* $E \subseteq V\times V$. L'ensemble $E$ doit être *symétrique*, c'est-à-dire qu'une arête $(s,t)$ appartient à $E$ si et seulement si $(t,s)$ lui appartient également. Par exemple, la figure 1 montre un graphe et sa représentation.

[67] Dominique Perrin et Jean-Éric Pin, 2004, *Infinite words : automata, semigroups, logic and games*.




$$V_1 = \{ \quad A, B, C, D \quad \}$$

$$E_1 = \{ \quad (A,C), (B,C), (B,G)\,(C,A) \quad (C,B), (G,B), (G,G) \quad \}$$

Figure 1 – Un graphe $G_1$

Un *chemin* dans un graphe $G = (V, E)$ est une suite $s_0\, s_1 \cdots s_k$ de sommets tels que pour tout entier $i$, $0 \leqslant i < k$, $(s_i, s_{i+1}) \in E$ ou $(s_{i+1}, s_i) \in E$. Un chemin est dit *élémentaire* s'il ne passe pas deux fois par le même sommet, et *simple* s'il n'utilise pas deux fois la même arête. Un *cycle* est un chemin dont les extrémités sont égales. Par exemple, sont donnés ci-dessous des chemins de $G_1$ :

$$A\,C\,B\,,\ B\,G\,G\,,\ B\,C\,B \text{ et } G\,G\,;$$

le premier est élémentaire ; le deuxième est simple ; le troisième est un cycle et le dernier est un cycle simple.

Un graphe est dit *connexe* si pour toute paire $(s, s')$ de ses sommets il existe un chemin dont $s$ et $s'$ sont les extrémités. Un *arbre* est un graphe qui ne contient aucun cycle simple. Le graphe $G_1$ n'est donc pas un arbre ; néanmoins retirer la boucle sur le sommet $G$ (l'arête $(G, G)$) produise un arbre.

Un *graphe orienté* est un graphe dont l'ensemble des arêtes n'est pas nécessairement symétrique ; elles sont alors appelées des *arcs*, et les cycles sont appelés des *circuits*. La figure 2 représente un graphe orienté.

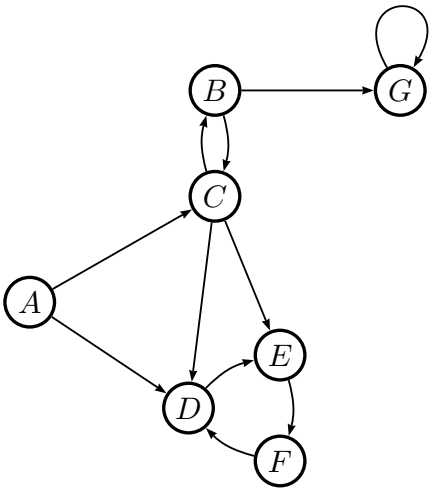



Figure 2 – Un graphe orienté $G_2$

Soit un graphe orienté $G$. Il est dit *fortement connexe* si pour tout couple $(s, s')$

de ses sommets, il existe un chemin de $s$ vers $s'$. Le *graphe non-orienté sous-jacent* de $G$ est une copie de $G$ dans lequel l'ensemble des arcs est symétrisé (et devient donc un ensemble d'arêtes). Le graphe orienté $G$ est dit *connexe* si son graphe non-orienté sous-jacent est connexe.

Un *DAG* (anagramme de Directed Acyclic Graph), parfois aussi appelé *hiérarchie*, est un graphe orienté connexe et sans circuits.

# Automates

Cette section présente quelques notions élémentaires de théorie des automates. Nous suivons essentiellement [72] pour les notations et la terminologie,

Un automate est un graphe orienté dont les arcs sont étiquetés par des lettres prises dans un alphabet fini et dont certains sommets sont distingués (final, initial ou les deux). Les sommets sont alors appelés des *états* et les arcs des *transitions*. Un automate $\mathcal{A}$ est formellement noté comme un 5-uplet $\langle Q, A, E, I, F \rangle$ où

- $Q$ est un ensemble fini d'états (dans certains cas, on considérera des automates *infinis*, auquel cas $Q$ sera infini) ;
- $A$ est un alphabet fini ;
- $E \subseteq (Q \times A \times Q)$ est l'ensemble de transitions ;
- $I \subseteq Q$ est l'ensemble des états initiaux ;
- $F \subseteq Q$ est l'ensemble des états finals.

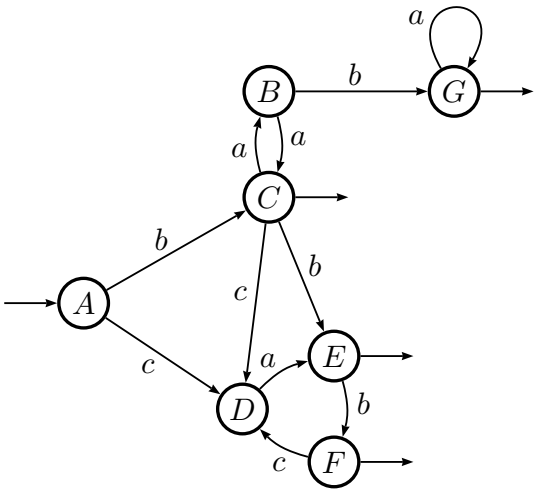



Figure 3 – Un automate $\mathcal{A}_3$

Soient $s, s'$ deux états de $\mathcal{A}$, et $a$ une lettre de $A$. On écrit $s \xrightarrow[\mathcal{A}]{a} s'$ (ou simplement $s \xrightarrow{a} s'$ quand il n'y a pas ambiguïté) si $(s, a, s')$ appartient à $E$ ; dans ce cas, on dit que $s$ est un *successeur par* $a$ de $s$. De manière analogue, si $u = a_0 a_1 \cdots a_k$ est un mot de $A^*$, on écrit $s \xrightarrow{u} s'$ si il existe des états $s_1, s_2 \cdots s_k$ dans $Q$ tels que

$$s \xrightarrow{a_0} s_1 \xrightarrow{a_1} s_2 \xrightarrow{a_2} \cdots \xrightarrow{a_{k-1}} s_k \xrightarrow{a_k} s' \ .$$

[72] Jacques Sakarovitch, 2003, *Éléments de théorie des automates*.

Le chemin ci-dessus est usuellement appelé un *calcul* de $u = a_0 a_1 \cdots a_k$ partant de $s$ ; nous ajoutons une condition supplémentaire : $s$ doit être un état initial. Si de plus, $s'$ est final, le calcul est dit *acceptant* et $u$ est dit accepté par $\mathcal{A}$. Le *langage accepté par* $\mathcal{A}$, noté $L(\mathcal{A})$ est l'ensemble des mots acceptés par $\mathcal{A}$ et deux automates sont dit *équivalents* s'ils acceptent le même langage. Inversement, tout langage accepté par un automate est dit régulier (ou rationnel, ou reconnaissable, voir remarque 1.13).

Un état $s$ de $\mathcal{A}$ est dit *accessible* s'il existe un calcul l'atteignant et il est dit *co-accessible* si, symétriquement, il existe un mot $u$ et un état final $f \in F$ tel que $s \xrightarrow{u} f$. Quand tous les états de $\mathcal{A}$ sont accessibles (resp. co-accessibles), $\mathcal{A}$ est dit lui-même *accessible* (resp. *co-accessible*). Il est dit *émondé* si il est à la fois accessible et co-accessible.

On dit qu'un automate $\mathcal{A} = \langle Q, A, E, I, F \rangle$ est *déterministe* s'il vérifie les deux conditions suivante :

- l'ensemble des états initiaux, $I$, est un singleton $\{i\}$ ;
- il n'existe pas deux transitions distinctes $(s, a, s')$ et $(s, a, s'')$ dans $E$.

Si $\mathcal{A}$ est déterministe, chaque mot admet au plus un calcul dans $\mathcal{A}$.

Dans ce cas, on note plutôt $\mathcal{A} = \langle Q, A, \delta, i, F \rangle$ où $i$ est *l'état initial* et $\delta : (Q \times A) \to Q$ est la *fonction (partielle) de transitions* que l'on note simplement $\delta(s, a) = (s \cdot a)$ quand ce n'est pas ambigu. En particulier, on dira que $(s \cdot a)$ *existe* si la fonction $\delta$ est définie sur $(s, a)$.

**Théorème 1.3** [69] – *Tout automate non-déterministe admet un automate déterministe qui lui est équivalent.*

Néanmoins, le processus de déterminisation provoque dans le pire des cas une *explosion exponentielle* du nombre d'états : il existe des automates non-déterministes à $n$ états dont le plus petit automate déterministe équivalent possède $2^n$ états.

Dans la suite, on considérera exclusivement des automates déterministes.

**Notation 1.4** – *Quand un automate $\mathcal{A}$ n'est pas explicitement défini à l'aide de chevrons $\langle\ \rangle$, on considère implicitement que son ensemble d'état est noté $Q_{\mathcal{A}}$, sa fonction de transition est noté $\delta_{\mathcal{A}}$, etc.*

**Définition 1.5** – *Soient deux automates $\mathcal{A}$ et $\mathcal{M}$. Un* morphisme d'automates $\mathcal{A} \to \mathcal{M}$ *est une fonction surjective $\varphi : Q_{\mathcal{A}} \to Q_{\mathcal{M}}$ qui vérifie les trois conditions suivantes :*

$$\varphi(i_{\mathcal{A}}) = \varphi(i_{\mathcal{B}}) \tag{1.1a}$$

$$\varphi^{-1}(F_{\mathcal{M}}) = F_{\mathcal{A}} \tag{1.1b}$$

$$\forall a \in A\ ,\quad \forall s \in Q_{\mathcal{A}} \qquad \varphi(s \cdot a) \;=\; \varphi(s) \cdot a \tag{1.1c}$$

*On dit alors que deux états $s$ et $s'$ de $\mathcal{A}$ sont* $\varphi$-equivalent *si $\varphi(s) = \varphi(s')$, que $\mathcal{M}$ est* un quotient *de $\mathcal{A}$ et que $\mathcal{A}$ est un* revêtement *de $\mathcal{M}$. Si de plus un troisième automate $\mathcal{B}$ est également un revêtement de $\mathcal{M}$ il est dit* en bisimulation avec *(ou bisimilaire à) $\mathcal{A}$.*

[69] Michael O. Rabin et Dana Scott, 1959, *Finite automata and their decision problems.*

**Propriété 1.6** – *Deux automates en bisimulation acceptent le même langage.*

Un automate est dit *complet* si la fonction de transition est une fonction totale ; chaque état possède alors au plus un successeur par chaque lettre.

**Théorème 1.7** [63, 64] – *Soit un langage régulier $L$. Il existe un automate, appelé* l'automate minimal complet de $L$, *qui est le quotient de tous les automates complets qui acceptent $L$.*

On appelle *automate minimal de $L$*, noté $\mathcal{M}_L$, la partie émondé de l'automate complet minimal ; c'est le quotient de tous les automates émondés qui acceptent $L$. Il existe plusieurs algorithmes pour calculer l'automate minimal : l'algorithme de Brzozowski [17], l'algorithme de Moore [59] ou l'algorithme de Hopcroft [41] qui est, à ce jour, le plus efficace.

**Théorème 1.8** [41] – *L'automate $\mathcal{N}_L$ (resp. $\mathcal{M}_L$) peut être calculé en temps $O(n \log(n))$ à partir d'un automate $\mathcal{A}$ complet (resp. émondé) acceptant $L$ ayant $n$ états.*

Soit un automate $\mathcal{A} = \langle Q_\mathcal{A}, A, \delta_\mathcal{A}, i_\mathcal{A}, F_\mathcal{A} \rangle$ émondé acceptant un langage $L$. Il existe d'après le théorème 1.7 un morphisme d'automates $\varphi : \mathcal{A} \to \mathcal{M}_L$. Deux états $s, s'$ de $\mathcal{A}$ sont *Nérode-équivalents* s'ils sont $\varphi$-équivalents pour ce *morphisme de minimisation*, ou, de façon équivalente si

$$\forall u \in A^* \qquad (s \cdot u) \in F_\mathcal{A} \iff (s' \cdot u) \in F_\mathcal{A}$$

**Définition 1.9** – *Soit un automate $\mathcal{A}$. La fonction* induite *par un mot $u \in A^*$ est la fonction partielle $f_u : Q_\mathcal{A} \to Q_\mathcal{A}$ qui associe à chaque état $s$ l'état $(s \cdot u)$ s'il existe. L'ensemble des fonctions induites par des mots de $A^*$ muni de la composition est un monoïde fini, appelé* monoïde de transitions de $\mathcal{A}$.

*On dit que $\mathcal{A}$ est un* automate à groupe *si ce monoïde est un groupe.*

**Propriété 1.10** – *Un automate est à groupe si et seulement si la fonction induite par chaque lettre est une permutation des états de $\mathcal{A}$.*

En particulier la classe des automates à groupe est stable par morphisme d'automates, comme énoncé par la proposition suivante.

**Proposition 1.11** – *Tout quotient d'un automate à groupe est un automate à groupe.*

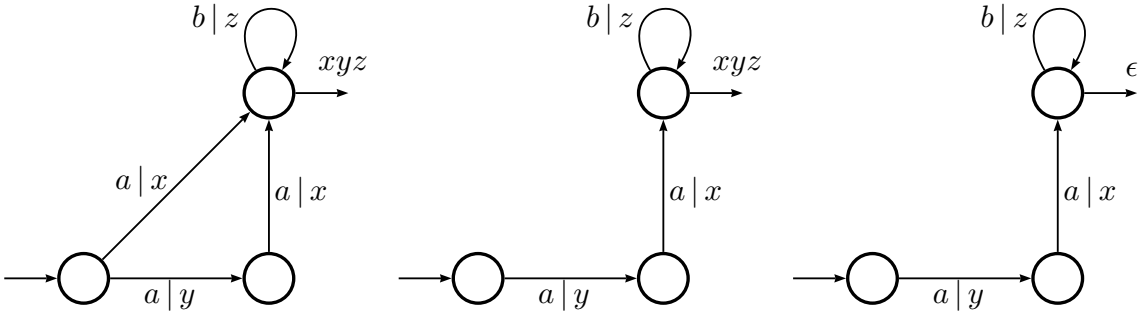
[17] John A. Brzozowski, 1963, *Canonical regular expressions and minimal state graphs for definite events.*
[63] J. Myhill, 1957, *Finite automata and the representation of events.*
[64] Anil Nerode, 1958, *Linear Automaton Transformations.*
[17] John A. Brzozowski, 1963, *Canonical regular expressions and minimal state graphs for definite events.*
[59] Edward F. Moore, 1956, *Gedanken experiments on sequential machines.*
[41] John E. Hopcroft, 1971, *An $n \log n$ algorithm for minimizing states in a finite automaton.*

# Transducteurs

Nous n'utilisons et n'allons définir qu'une sous-classe restreinte des transducteurs ; pour plus de détails, nous référons une nouvelle fois le lecteur à [72]. Un transducteur lettre-à-lettre de $A^*$ dans $B^*$ est essentiellement un automate dont l'alphabet est le produit cartésien $A \times B$ ; il est noté comme un 6-uplet $\langle Q, A, B, E, I, \omega \rangle$ où

- $Q$ est un ensemble fini d'états (on utilisera parfois des transducteurs *infinis*, auquel cas l'ensemble des états sera infini) ;
- $A$ est un alphabet fini dit *d'entrée* ;
- $B$ est un alphabet fini dit *de sorte* ;
- $E \subseteq (Q \times A \times B \times Q)$ est l'ensemble de transitions ;
- $I \subseteq Q$ est l'ensemble des états initiaux ;
- $\omega$ est la *fonction finale*, une fonction partielle $Q \to B^*$ ; un état est dit *final* si $\omega$ est défini sur lui.

On note $s \xrightarrow[\mathcal{D}]{a \mid b} s'$ (au lieu de $(s, a, b, s') \in E$) une transition de $\mathcal{D}$, et $s \xrightarrow[\mathcal{D}]{u \mid v} s'$ un chemin de $\mathcal{D}$. Si $u$ est un mot de l'alphabet d'entrée $A^*$, on appelle *calcul de* $u$, tout chemin $s \xrightarrow[\mathcal{D}]{u \mid v} s'$ pour certains $v \in B^*$, $s \in I$ et $s' \in Q$ ; si de plus $s'$ est un état final, on dit que ce calcul est *acceptant* et que le couple $u \mid (v\, \omega(q))$ est accepté par $\mathcal{D}$. La *relation réalisée* par $\mathcal{D}$ est l'ensemble des couples $u \mid v$ acceptés par $\mathcal{D}$.




(a) Un transducteur $\mathcal{D}_1$ lettre-à-lettre non-séquentiel

(b) Un transducteur $\mathcal{D}_2$ lettre-à-lettre et séquentiel

(c) Un transducteur $\mathcal{D}_2$ lettre-à-lettre et séquentiel pur

Figure 4 – Quelques transducteurs

Un transducteur $\mathcal{D}$ est dit *séquentiel* si chaque mot de $A^*$ admet au plus un calcul dans $\mathcal{D}$ ; il est dit *séquentiel pur* si sa fonction finale $\omega$ envoie chaque état (pour lequel elle est définie) sur le mot vide $\varepsilon$. [1]

1. La définition usuelle (*cf.* [72]) de transducteur séquentiel pur impose de plus que la fonction finale $\omega$ est définie sur tous les états ; enlever cette restriction permet, notamment dans le chapitre 9, des démonstrations bien plus simples.

[72] Jacques Sakarovitch, 2003, *Éléments de théorie des automates*.

La relation réalisée par un transducteur (lettre-à-lettre et) séquentiel est donc une fonction et s'il est de plus séquentiel pur cette fonction conserve la longueur. Dans les deux cas on appelle, par abus de langage, *image d'un mot* par un tel transducteur l'image de ce mot (si elle existe) par la fonction réalisée par ce transducteur : l'image d'un mot $u$ par un transducteur $\mathcal{D}$ est notée $\mathcal{D}(u)$.

On ne considérera dans la suite que des transducteurs lettre-à-lettre et séquentiel (parfois pur).

On appelle *automate d'entrée* d'un transducteur $\mathcal{D} = \langle Q, A, B, E, I, \omega \rangle$ l'automate $\mathcal{A} = \langle Q, A, E', I, F \rangle$, où $F$ est l'ensemble des états finals de $\mathcal{D}$ ($F = \{s \in \mathbb{Q} \mid \omega \text{ est définie sur } s\}$) ; et où $E'$ contient la transition $s \xrightarrow{a} s'$ s'il existe une transition $s \xrightarrow{a \mid b} s'$ dans $E$ pour une certaine lettre $b \in B$. L'*automate de sortie* est défini de façon duale.

Le *langage d'entrée de* $\mathcal{D}$ est le domaine de définition de la fonction réalisée par $\mathcal{D}$ ; il est accepté par l'automate d'entrée de $\mathcal{D}$. Le *langage de sortie de* $\mathcal{D}$ est l'image de la fonction réalisée par $\mathcal{D}$ ; si $\mathcal{D}$ est séquentiel pur, ce langage est accepté par l'automate de sortie de $\mathcal{D}$.

# Langages formels

Nous ne présentons ici que des notions élémentaires de théorie des langages formels ; voir par exemple [19] pour plus de détails.

L'ensemble des *langages rationnels* sur un alphabet $A$, noté $\mathsf{Rat}\ A^*$ est le plus petit ensemble de parties de $A^*$ tel que

- pour toute lettre $a \in A$, $\{a\} \in \mathsf{Rat}\ A^*$ ;
- $\mathsf{Rat}\ A^*$ est fermé par union, intersection et étoile.

L'ensemble des *langages reconnaissables* sur un alphabet $A$, noté $\mathsf{Rec}\ A^*$, est l'ensemble des langages qui sont acceptés par des automates finis (voir section 1.4).

**Théorème 1.12** [43] – *Pour tout alphabet fini $A$, $\mathsf{Rec}\ A^* = \mathsf{Rat}\ A^*$.*

**Remarque 1.13** – *Les notions de langage reconnaissable et langage rationnel coïncident dans le cas où l'on considère un monoïde libre ; ils sont aussi appelés* réguliers *dans ce cas. Nous utiliserons préférentiellement le terme* régulier*, car les deux autres pourront prêter à confusion pour les raisons suivantes :*

- *les ensembles d'entiers ultimement périodiques (qui est le sujet du chapitre 3) sont souvent appelés reconnaissables dans la littérature ;*
- *les systèmes de numération à base rationnelle (c'est-à-dire dont la base est un nombre de $\mathbb{Q}$) constitue l'objet d'étude central de cette thèse.*

Le lemme d'itération (pour les langages réguliers), énoncé ci-dessous, donne une condition nécéssaire pour qu'un langage soit régulier.

[19] Olivier Carton, 2008, *Langages formels, calculabilité et complexité*.

[43] Stephen C. Kleene, 1956, *Representation of events in nerve nets and finite automata*.

[9] Yehoshua Bar-Hillel, Micha A. Perles et Eli Shamir, 1961, *On formal properties of simple phrase structure grammars*.

**Lemme 1.14** [9] – *Soit $L$ un langage régulier. Il existe un entier $N \in \mathbb{N}$ tel que tout mot $w \in L$ de longueur $|w| \geqslant N$ possède une factorisation $w = x\,u\,y$ qui vérifie*

**a)** $0 < |u|$
**b)** $|xu| \leqslant N$
**c)** $x\,u^n\,y \in L$ *pour tout entier* $n$.

Une *grammaire algébrique* $G$ est un quadruplet $(N, A, S, P)$ où

- $N$ est un alphabet fini de symboles non-terminaux ;
- $A$ est un alphabet fini de symboles terminaux disjoint de $N$ ;
- $S \in N$ est un symbole non-terminal particulier, appelé axiome ;
- $P \subseteq N \times (A \cup N)^*$ est un ensemble de règles de production.

Soit un mot $u\,X\,v \in (A \cup N)^*$ dans lequel le symbole non-terminal $X \in N$ apparaît. L'application d'une règle $(X, w) \in P$ sur ce mot est la substitution du symbole $X$ par le mot $w$, ce qui produit donc le mot $u\,w\,v$. On note $u \to v$ si le mot $v$ peut être obtenue par l'application d'une règle de $P$ sur le mot $u$. On appelle *dérivation*, noté $\to^*$, la fermeture transitive de $\to$ : on dit donc que le mot $v$ est *dérivé* du mot $u$, noté $u \to^* v$, s'il existe une suite de mots $u_1, u_2, \ldots, u_k$ tels que

$$u \to u_1 \to u_2 \to \cdots \to u_k \to v$$

Le *langage engendré par la grammaire* $G$ est alors l'ensemble des mots (de symboles terminaux) dérivés depuis l'axiome $S$ :

$$L(G) \;=\; \{\, v \in A^* \mid S \to^* v \,\} \;.$$

Un langage est dit *algébrique* s'il est engendré par une grammaire algébrique. Il existe un lemme d'itération pour les langages algébriques, donné ci-dessous.

**Lemme 1.15** [9] – *Soit $L$ un langage algébrique. Il existe un entier $N \in \mathbb{N}$ tel que tout mot $w \in L$ de longueur $|w| \geqslant N$ possède une factorisation $w = x\,u\,y\,v\,z$ qui vérifie*

**a)** $0 < |u\,v|$
**b)** $|u\,y\,v| \leqslant N$
**c)** $x\,u^n\,y\,v^n\,z \in L$ *pour tout entier* $n$.

# Systèmes de numération abstraits

Un *système de numération* $S$, que l'on appellera parfois simplement *un système*, est une façon d'associer les nombres entiers et les mots sur un certain alphabet (fini) $A$. Un système $S$ est composé des deux fonctions suivantes :

- *l'évaluation*, une fonction partielle $\pi_S : A^* \to \mathbb{N}$ ;
- *la représentation*, une fonction totale $\mathbb{N} \to A^*$ qui envoie chaque entier $n$ sur un mot appelé, *$S$-représentation de $n$*, et noté $\langle n \rangle_S$ ;

de telle sorte que $\langle \mathbb{N} \rangle_S \subseteq \mathsf{Dom}(\pi_S)$ et pour tout entier $n \in \mathbb{N}$, $\pi_S(\langle n \rangle_S) = n$.

Les systèmes de numération abstraits ont été introduit par Lecomte et Rigo [44](voir aussi [45]). Il s'agit de considérer comme une définition une propriété commune à tous les systèmes de numération : les représentations des entiers suivent toujours l'*ordre radiciel* (parfois aussi appelé *généalogique*, défini ci-dessous), c'est-à-dire que la représentation de l'entier $(n + p)$ est toujours plus grande selon cet ordre que la représentation d'un entier $n$.

**Définition 1.16** – *Soit un alphabet $A$ totalement ordonné. Le monoïde libre $A^*$ est alors totalement ordonné par l'ordre radiciel défini ci-dessous.*

*Soient deux mots $u, v \in A^*$. On dit que $u$ est strictement inférieur à $v$ dans l'ordre radiciel, noté $u <_{rad} v$, si l'une des deux conditions suivantes est satisfaite.*

- $|u| < |v|$
- $|u| = |v|$ *et il existe trois mots $w, u', v' \in A^*$ et deux lettres $a, b \in A$ tels que $a < b$, $u = w\,a\,u'$ et $v = w\,b\,v'$.*

Un système de numération abstrait (SNA) est entièrement défini par, et dans un sens équivalent à, un langage $L$ sur un alphabet ordonné $A$, si bien que l'on note le système simplement $L$. Les mots de $L$ sont ordonnés par l'*ordre radiciel* et la représentation d'un entier $n$ dans le SNA $L$, noté $\langle n \rangle_L$, est défini comme le $(n+1)$-ème mot de $L$ dans l'ordre radiciel. La fonction d'évaluation est simplement l'inverse de la fonction de représentation, si bien que son domaine de définition est $L$ ; elle n'est généralement pas utilisée.

Les adjectifs employés pour un langage $L$ sont utilisées pour le SNA correspondant. En particulier, si $L$ est un langage régulier, alors il définit un système de numération abstrait régulier (SNAR).

Il est parfois utile d'avoir une lettre qui permet de modifier arbitrairement la longueur des mots. Dans ce cas, il est usuel d'ajouter une lettre spéciale # (*cf.* [45], parfois aussi notée \$, *cf.* [7]) qui n'est pas dans l'alphabet $A$ (et donc n'apparaît pas dans $L$) et de considérer l'alphabet $A \cup \{\#\}$ ; l'ordre de cet alphabet reprend celui de $A$, et le complète en définissant que # est plus petite que toutes les lettres de $A$. Cette solution ne correspond pas à ce que l'on observe dans les systèmes de numération classiques, dans lesquels une lettre de calage (généralement 0) existe naturellement. Par exemple, en base 10, la lettre 0 est dans le mot 080, utilisée à gauche comme lettre de calage et à droite comme lettre signifiante.

**Définition 1.17** – *Soit un alphabet ordonné $A$. On dit qu'un langage de $A^*$ est calable s'il peut s'écrire sous la forme $z^*L$ et que tous les mots de $L$ commencent par des lettres plus grandes que $z$. On appelle alors $z$ la* lettre de calage *de $z^*L$.*

*Un* SNA calable *est construit à partir d'un langage calable $z^*L$ et donne les mêmes représentations que le SNA $L$ ; la représentation de $n$ est toujours noté $\langle n \rangle_L$. En revanche, il est fourni avec sa lettre de calage $z$ de telle sorte qu'on puisse évaluer $z^k\langle n \rangle_L$ à $n$ pour tout entier $k$.*

[44] Pierre Lecomte et Michel Rigo, 2001, *Numeration systems on a regular language.*

[45] Pierre Lecomte et Michel Rigo, 2010, *Abstract Numeration Systems.*

[7] Pierre-Yves Angrand et Jacques Sakarovitch, 2010, *Radix enumeration of rational languages.*

Par exemple, le système de numération en base entière dont on donne une définition dans le chapitre suivant est le SNA calable $[\![p]\!]^*$. En général, étudier le SNA calable $z^*L$ plutôt que le SNA $L$ permet de retirer une singularité à l'origine.

# Première partie

# Sur la base entière

Chapitre 2

# Ensembles p-reconnaissables de nombres

Ce chapitre présente les *systèmes de numération à base entière*, ou plus simplement *les bases entières*. Ce sont les manières les plus naturelles de représenter les entiers ; elles généralisent le système en base 10 utilisé dans la vie de tous les jours.

Néanmoins, contrairement à l'usage, nous allons considérer que les nombres sont représentés avec le *chiffre de poids faible en premier*. Par exemple, les mots "015" "0150" et "015000000" ont tous trois la valeur 510. Cette convention a été prise pour des raisons de clarté. En effet, nous allons considérer dans le chapitre suivant exclusivement des automates qui lisent les représentations avec le chiffre de poids faible en premier. L'alternative aurait été de ne considérer que des automates droits, c'est-à-dire qui lisent les mots de droite à gauche ce qui aurait rendu les constructions habituelles sur les automates beaucoup moins lisibles.

## $p$-représentation des entiers

Soit un entier $p \geqslant 2$ dorénavant appelé la *base*. On lui associe l'alphabet *canonique* $[\![p]\!] = \{0, 1, \ldots, p-1\}$ composé des $p$ plus petits entiers ; ce sont les *chiffres* de la base $p$.

A tout mot $w$ sur l'alphabet $[\![p]\!]$ est associé une *valeur* entière par la *fonction d'évaluation* :

$$\begin{array}{rccl} \pi_p \ : & [\![p]\!]^* & \longrightarrow & \mathbb{N} \\ & w = a_0 a_1 \cdots a_k & \longmapsto & \pi_p(w) \ = \ \displaystyle\sum_{i=0}^{k} a_i p^i \end{array} \tag{2.1}$$

Notez que la définition de $\pi_p$ implique déjà que que l'on représente les entiers avec le chiffre de poids faible en premier. Les trois équations suivantes donnent des formules pour calculer récursivement la valeur d'un mot ; elles découlent immédiatement de la définition de $\pi_p$.

$$\forall u \in [\![p]\!]^*, \ \forall a \in [\![p]\!] \qquad \pi_p(u\,a) \ = \ \pi_p(u) + a\,p^{|u|} \tag{2.2a}$$

$$\forall a \in [\![p]\!] \ , \ \forall v \in [\![p]\!]^* \qquad \pi_p(a\,v) \ = \ a + \pi_p(v)\,p \tag{2.2b}$$

$$\forall u, v \in [\![p]\!]^* \qquad \pi_p(u\,v) \ = \ \pi_p(u) + \pi_p(v)\,p^{|u|} \tag{2.2c}$$

Pour tout entier $n$, il existe des mots dont la valeur est égale à $n$ ; un en particulier est appelé *la représentation de $n$ en base $p$* (ou plus simplement *$p$-représentation de $n$*), est noté $\langle n \rangle_p$, et se calcule par l'*algorithme d'Euclide* suivant. On pose $N_0 = n$ et

$$\forall i \qquad N_i \;=\; p\,N_{(i+1)} + a_i \;, \tag{2.3}$$

où $a_i$ est le reste de la division euclidienne de $N_i$ par $p$, donc appartient à $[\![p]\!]$. L'entier $N_{(i+1)}$ est le quotient de cette même division euclidienne, ce qui implique que la suite $(N_i)_{i\in\mathbb{N}}$ strictement décroissante jusqu'à devenir stationnaire à $N_{(k+1)} = 0$. La $p$-représentation de $n$ est alors le mot $\langle n \rangle_p = a_0\, a_1 \cdots a_k$ et une simple vérification montre que

$$\pi_p\Big(\langle n \rangle_p\Big) \;=\; n \quad \text{ et que } \quad \langle n \rangle_p \text{ ne commence pas par un 0.} \tag{2.4}$$

On peut alternativement calculer $\langle n \rangle_p$ par l'algorithme glouton décrit ci-dessous. Soit $k$ l'entier tel que $p^k \leqslant n < p^{(k+1)}$. On pose $M_k = n$ et

$$\forall i,\ \ 0 \leqslant i \leqslant k, \qquad \begin{array}{l} b_i \text{ est le plus grand entier tel que } (M_i - b_i\, p^i) \geqslant 0 \\ M_{(i-1)} = (M_i - b_i\, p^i) \end{array} \tag{2.5}$$

Une simple vérification montre encore une fois que

$$\pi_p(b_0\, b_1 \cdots b_k) \;=\; n \quad \text{ et que } \quad b_0 \neq 0 \;. \tag{2.6}$$

Il découle alors du théorème 2.1, plus loin que $\langle n \rangle_p = b_0\, b_1 \cdots b_k$.

L'algorithme d'Euclide calcule les chiffres de poids faible en premier (donc, avec notre convention, de gauche à droite) alors que l'algorithme glouton calcule d'abord les chiffres de poids fort (donc de droite à gauche). En effet, les deux équations suivantes sont vérifiées pour tout entier $i$, $0 \leqslant i \leqslant (k+1)$ :

$$\langle n \rangle_p \;=\; a_0\, a_1 \cdots a_{i-1} \langle N_i \rangle_p \quad \text{ et } \quad \langle n \rangle_p \;=\; \langle M_{k-i} \rangle_p b_{(k-i+1)}\, b_{(k-i+2)} \cdots b_k$$

**Théorème 2.1** – *Soient un mot $u \in [\![p]\!]^*$ et un entier $n$ tels que $\pi_p(u) = n$. Alors $u$ est de la forme $\langle n \rangle_p 0^j$ pour un certain entier $j$.*

Démonstration. Le théorème 2.1 découle de l'assertion suivante qui exprime l'injectivité de $\pi_p$ sur les mots d'une longueur donnée.

**Assertion 2.1.1** – *Soient deux mots $u, v \in [\![p]\!]^*$. S'ils vérifient $\pi_p(u) = \pi_p(v)$ et $|u| = |v|$ alors ils sont égaux.*

*Démonstration de l'assertion.* On note $u = a_0\, a_1 \cdots a_k$ et $v = b_0\, b_1 \cdots b_k$. Prouvons par récurrence sur $i$ que les préfixes de longueurs $i$ respectifs de $u$ et $v$ sont égaux ; ceci est évidemment vrai pour le cas $i = 0$.

Supposons donc que $u$ et $v$ ont un préfixe commun de longueur $i \leqslant k$ noté $w = a_0\, a_1 \cdots a_{(i-1)} = b_0\, b_1 \cdots b_{(i-1)}$ et prouvons que $a_i = b_i$. On note $u'$ et $v'$ les mots tels que $u = w\, a_i\, u'$ et $v = w\, b_i\, v'$. L'équation (2.2c), appliquée à ces deux mots, donne les deux équations suivantes :

$$\pi_p(u) \;=\; \pi_p(w\, a_i\, u') \;=\; \pi_p(w) + a_i\, p^i + \pi_p(u') p^{(i+1)} \;\equiv\; \pi_p(w) + a_i\, p^i \ [p^{(i+1)}] \,;$$

$$\pi_p(v) \;=\; \pi_p(w\, b_i\, v') \;=\; \pi_p(w) + b_i\, p^i + \pi_p(v') p^{(i+1)} \;\equiv\; \pi_p(w) + b_i\, p^i \ [p^{(i+1)}] \;.$$

Puisque $\pi_p(u) = \pi_p(v)$, alors en particulier $\pi_p(u) \equiv \pi_p(v)\ [p^{(i+1)}]$. Il découle des deux équations précédentes que $(\pi_p(w) + a_i\, p^i) \equiv (\pi_p(w) + b_i\, p^i)\ [p^{(i+1)}]$, donc que $a_i = b_i$, ce qui conclut la récurrence. □

Il découle du théorème 2.1 que deux mots qui ne finissent pas par un zéro ont nécessairement des valeurs distinctes, ce qui implique la proposition suivante.

**Proposition 2.2** – *Les représentations des entiers en base $p$ forment le langage régulier* $\{\varepsilon\} \cup [\![p]\!]^*([\![p]\!] \setminus 0) = [\![p]\!]^* \setminus ([\![p]\!]^* 0)$.

# Additionneur, normalisateur

L'addition est une opération de $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$, donc induit (à base $p$ fixé) une fonction $[\![p]\!]^* \times [\![p]\!]^* \to [\![p]\!]^*$ telle que la somme des valeurs des deux entrées soit égale à celle de la sortie. Cette fonction est réalisée en base entière par un transducteur fini, lettre-à-lettre et séquentiel appelé *additionneur*. Il s'agit d'un cas particulier du *normalisateur* qui prend en entrée un mot sur un alphabet fini arbitraire et renvoie en sortie un mot de même valeur sur l'alphabet canonique $[\![p]\!]$.

Soit $D \subseteq \mathbb{Z}$ un alphabet fini de chiffres. La fonction $\pi_p$ est naturellement étendue sur $D^*$ et plus généralement, elle pourra dans la suite s'appliquer à n'importe quel mot sur un alphabet de chiffres. La *fonction de normalisation* $\chi_D$ est la fonction partielle définie par :

$$\begin{array}{rcl} \mathsf{Dom}(\chi_D) & = & \{\, u \in D^* \mid \exists v \in [\![p]\!]^* \quad \pi_p(u) \;=\; \pi_p(v) \,\} \\ \chi_D : \; D^* & \longrightarrow & [\![p]\!]^* \\ u & \longmapsto & v, \text{ le mot le plus court} \\ & & \quad \text{tel que } |v| \;\geqslant\; |u| \;\text{ et }\; \pi_p(u) \;=\; \pi_p(v). \end{array} \tag{2.7}$$

Le domaine de définition de $\chi_D$ est *le plus grand possible*, c'est l'ensemble des mots de $D^*$ qui admettent un mot de même valeur sur l'alphabet canonique. Puisque $D$ est un alphabet d'entiers relatifs, les mots de $D^*$ ont tous une valeur dans $\mathbb{Z}$ ; le domaine de définition de $\chi_D$ s'exprime donc de manière équivalente par :

$$\mathsf{Dom}(\chi_D) \;=\; \left\{\, v \in D^* \;\middle|\; \pi_p(v) \;\geqslant\; 0 \,\right\} .$$

**Théorème 2.3** – *Soit $D \subseteq \mathbb{Z}$ un alphabet fini de chiffres. Il existe un transducteur fini, lettre-à-lettre et séquentiel $\mathcal{C}_D$ de $D^*$ vers $[\![p]\!]^*$ qui réalise $\chi_D$ la fonction $\chi_D$.*

La définition 2.4, ci dessous, donne la définition d'un transducteur infini ; il sera démontré par la suite que sa partie accessible est finie et qu'il réalise la fonction $\chi_D$. Il est noté également $\mathcal{C}_D$ par abus de langage.

**Définition 2.4** – *Le normalisateur $\mathcal{C}_D$ est le transducteur infini*

$$\mathcal{C}_D \;=\; \langle\, \mathbb{Z},\, D,\, [\![p]\!],\, 0,\, \delta,\, \eta,\, \omega \,\rangle$$

*dont les états sont les entiers relatifs, l'alphabet d'entrée est $D$, l'alphabet de sortie est $[\![p]\!]$, les transitions sont définies par*

$$\forall s, s' \in \mathbb{Z},\;\; d \in D,\;\; a \in [\![p]\!] \qquad s \xrightarrow{\;d \mid a\;} s' \quad\Longleftrightarrow\quad s + d = p\, s' + a \tag{2.8a}$$

*et la fonction finale est définie par*

$$\forall s \in \mathbb{Z} \quad \begin{cases} \omega(s) \;=\; \langle s \rangle_p & \text{si} \;\; s \geqslant 0 \\ \omega(s) \text{ n'est pas défini} & \text{si} \;\; s < 0 \end{cases} \tag{2.8b}$$

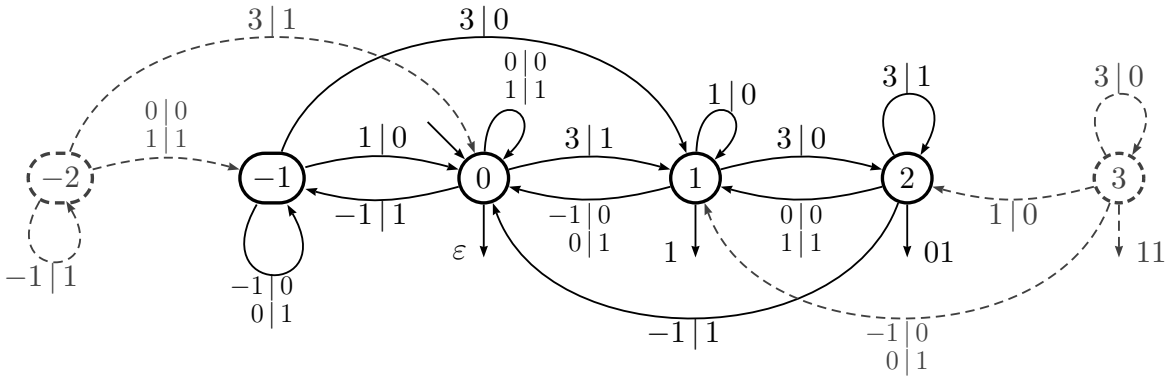



Figure 1 – Le normalisateur $\mathcal{C}_{D_1}$ en base 2, où $D_1 = \{-1, 0, 1, 3\}$

La figure 5 représente le normalisateur de $\mathcal{C}_{D_1}$, où $D_1 = \{-1, 0, 1, 3\}$ ; les états inaccessibles (et leurs transitions sortantes) sont tiretés et grisés.

Il découle de la définition 2.4 que $\mathcal{C}_D$ est lettre-à-lettre ; de plus, l'équation (2.8a) implique que pour tout état $s$ et toute lettre $d \in D$, il existe un unique état $s \in \mathbb{Z}$ et une unique lettre $a \in [\![p]\!]^*$ tel que $s \xrightarrow{d \mid a} s'$. Le normalisateur $\mathcal{C}_D$ est donc séquentiel et complet en entrée.

Bien que $\mathcal{C}_D$ ne soit pas fini, sa partie accessible l'est comme l'exprime la proposition 2.6. Il est ensuite démontré à a proposition 2.7 que $\mathcal{C}_D$ réalise la fonction $\chi_D$. Les démonstrations de ces deux propositions nécessitent le lemme préliminaire suivant, qui détaille le fonctionnement interne du normalisateur.

**Lemme 2.5** – *Soient deux mots $u \in D^*$, $v \in [\![p]\!]^*$ et deux états $s, t$ de $\mathcal{C}_D$. Le chemin $s \xrightarrow{u \mid v} t$ existe dans $\mathcal{C}_D$ si et seulement si $(\pi_p(u) + s) = (\pi_p(v) + tp^{|u|})$ et $|u| = |v|$.*

Démonstration. Notez que puisque $\mathcal{C}_D$ est lettre-à-lettre, l'hypothèse $s \xrightarrow{u \mid v} t$ implique que $|u| = |v|$. On peut supposer que $|u| = |v|$ dans les deux sens de l'équivalence.

Par récurrence sur la longueur de $u$ et $v$ ; si $u = v = \varepsilon$, alors $\pi_p(u) = \pi_p(v) = 0$ et les deux côtés de l'équivalence sont simplement $s = t$.

Soient deux mots $d\,u \in D^+$ et $a\,v \in [\![p]\!]^+$. Les valeurs de ces deux mots peuvent se calculer par les formules suivantes (*cf.* équation (2.2b)) :

$$\pi_p(d\,u) \;=\; d + p\,\pi_p(u) \quad \text{et} \quad \pi_p(a\,v) \;=\; a + p\pi_p(v)\;. \tag{$*$}$$

L'existence de $s, t$ tel que $s \xrightarrow{d \mid a} s' \xrightarrow{u \mid v} t$ est équivalente à (d'après l'hypothèse de récurrence et l'équation (2.8a)) :

$$(\pi_p(u) + s') \;=\; (\pi_p(v) + tp^{|u|}) \quad \text{et} \quad s + d \;=\; p\,s' + a\;.$$

Utiliser les deux équations $(*)$ montrent que l'équation précédente est équivalente à

$$\pi_p(d\,u) + s \;=\; \pi_p(a\,v) + t\,p^{|d\,u|} \quad \text{et} \quad \begin{array}{l} s' \text{ est le quotient de la division} \\ \text{euclidienne de } (s+d) \text{ par } p. \end{array} \qquad \square$$

**Proposition 2.6** – *La partie accessible de $\mathcal{C}_D$ est finie.*

Démonstration. Soit $e$ le plus grand chiffre de $D$. Prouvons que tout état $s' > \frac{e}{p-1}$ n'est pas accessible depuis un état plus petit que lui, donc pas accessible depuis 0.

Par l'absurde. Supposons qu'il existe un état $s < s'$ et deux lettres $a \in [\![p]\!]$, $d \in D$ tels que $s \xrightarrow{d \mid a} s'$ ; ce qui est équivalent à $s + d = p\,s' + a$ (équation (2.8a)). Puisque $s' > s$ et $a \geqslant 0$, alors $s' + d > p\,s'$ donc $d > (p-1)s'$ ce qui contredit la conjonction d'hypothèses : $e \geqslant d$ et $s' > \frac{e}{p-1}$.

On démontre de manière analogue que tout état $s' < (\frac{f}{p-1} - 1)$ ne peut pas être atteint depuis un état plus grand que lui, si $f$ est la plus petite lettre de $D$. $\square$

**Proposition 2.7** – *Le normalisateur $\mathcal{C}_D$ réalise la fonction $\chi_D$.*

Démonstration. Soit un mot $u \in D^*$. Puisque le normalisateur est complet en entrée et séquentiel il existe un unique mot $v$ et un unique état $t$ tel que

$$0 \xrightarrow{u \mid v} t\ .$$

Il découle du lemme 2.5 que

$$\pi_p(u) \;=\; \pi_p(v) + t\,p^{|u|} \quad \text{et} \quad |u| \;=\; |v|\ . \qquad (*)$$

Puisque $v$ est un mot de $[\![p]\!]^*$ (de la même longueur que $u$), $0 \leqslant \pi_p(v) < p^{|v|} = p^{|u|}$. Il découle donc de l'équation précédente que $t \geqslant 0$ si et seulement si $\pi_p(u) \geqslant 0$.

Les assertions suivantes sont alors équivalentes : l'image $\mathcal{C}_D(u)$ existe ; $t$ est un état final ; $t$ est positif ou nul ; $\pi_p(u)$ est positif ou nul ; $u$ est dans le domaine de définition de $\chi_D$. Dans ce cas, $\mathcal{C}_D(u) = v\langle t\rangle_p$, un mot dont la valeur est égale à $\big(\pi_p(v) + tp^{|u|}\big)$, donc à $\pi_p(u)$ d'après $(*)$.

Ce mot $\mathcal{C}_D(u) = v\langle t\rangle_p$ est de longueur $|u| + |\langle t\rangle_p|$, donc au moins aussi long que $u$. S'il n'est pas le mot le plus court dont la valeur est $\pi_p(u)$, alors il termine par un 0 (théorème 2.1), ce qui implique que $\langle t\rangle_p = \varepsilon$, donc que $|\mathcal{C}_D(u)| = |u|$. Qu'il termine ou non par un 0, $\mathcal{C}_D(u)$ est donc le mot $w$ le plus court tel que $\pi_p(w) = \pi_p(u)$ et $|w| \geqslant |u|$, donc $\mathcal{C}_D(u) = \chi_D(u)$. $\square$

Ceci conclut la preuve du théorème 2.3 : la partie accessible de $\mathcal{C}_D$ est un transducteur fini, lettre-à-lettre et séquentiel qui réalise la fonction $\chi_D$.

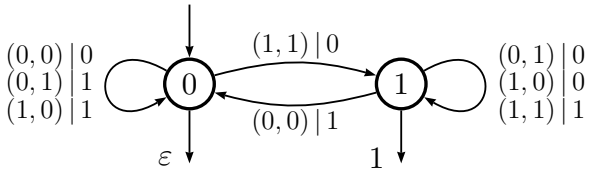



Figure 2 – L'additionneur en base 2

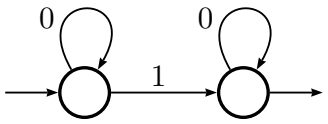
L'additionneur est le cas particulier du normalisateur où $D = [\![2p-1]\!]$. En effet, l'*addition chiffre-à-chiffre*, définie ci-après, transforme deux mots $u, v \in [\![p]\!]^*$ en un troisième mot $w \in D^*$ qui satisfait : $\pi_p(u) + \pi_p(v) = \pi_p(w)$. Soient deux mots sur $[\![p]\!]$, $u = a_0\, a_1 \cdots a_k$ et $u = b_0\, b_1 \cdots b_k$ ; on définit le mot $w = d_0\, d_1 \cdots d_k$ dans $D^*$ tel que pour tout entier $i$, $d_i = a_i + b_i$. Une simple vérification montre que $\pi_p(d_k \cdots d_1 d_0) = \pi_p(a_k \cdots a_1 a_0) + \pi_p(b_k \cdots b_1 b_0)$. Si bien que $\mathcal{C}_D(d_k \cdots d_1 d_0)$ est un mot de $[\![p]\!]^*$ dont la valeur est la somme $\pi_p(a_k \cdots a_1 a_0) + \pi_p(b_k \cdots b_1 b_0)$.

## Ensembles $p$-reconnaissables

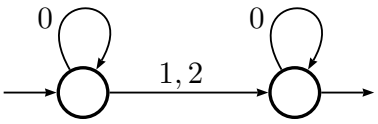
Dans le chapitre suivant, nous utilisons des automates pour reconnaître des ensembles d'entiers représentés en base $p$. Il sera toujours supposé que les automates acceptent *par valeur*, c'est-à-dire que si un mot de valeur $n$ est accepté par un automate $\mathcal{A}$, alors tous les autres mots de valeur $n$ sont également acceptés par $\mathcal{A}$. Dans le cas des bases entières, un automate $\mathcal{A}$ accepte donc par valeur si, pour tout état $s$ de $\mathcal{A}$, $s$ et $(s \cdot 0)$ ont le même statut final/non-final [1].

Dans ce cas, on appelle *ensemble d'entiers accepté (en base p)* par un automate $\mathcal{A}$, noté $N(\mathcal{A})$, l'ensemble $N(\mathcal{A}) = \pi_p(L(\mathcal{A}))$ ; un mot $u$ est accepté par $\mathcal{A}$ si et seulement si sa valeur appartient à $N(\mathcal{A})$.

Un ensemble d'entier est dit *p-reconnaissable* s'il est accepté (en base $p$) par un automate. Par exemple l'ensemble $S_1 = \{2^i \mid i \in \mathbb{N}\}$ des puissances de 2 est 2-reconnaissable (*cf.* figure 3a) et 4-reconnaissable (*cf.* figure 3b) mais n'est pas 3-reconnaissable.




(a) Automate acceptant $S_1$ en base 2




(b) Automate acceptant $S_1$ en base 4

Figure 3 – L'ensemble $S_1$ des puissances de 2 est à la fois 2- et 4-reconnaissable

On dit que deux entiers $p$ et $q$ sont *multiplicativement dépendants* si il existe deux entiers $k$ et $i$ tels que $p^k = q^i$. Dans ce cas, les notions de $p$-reconnaissable et de $q$-reconnaissable sont équivalentes, comme énoncé ci-dessous.

**Lemme 2.8** – *Soient deux entiers $p, q > 1$. Si $p$ et $q$ sont multiplicativement dépendants, alors un ensemble est $q$-reconnaissable si et seulement si il est $p$-reconnaissable.*

Démonstration. Il suffit de traiter le cas où $q^k = p$.

Pour toute lettre $a \in [\![p]\!]$ il existe un unique mot de $[\![q]\!]^*$ qui est de longueur $k$ et de valeur $a$. Ce mot existe car $a$ est un entier plus petit que $p = q^k$ donc d'après l'algorithme glouton, $|\langle a \rangle_q| \leqslant k$ ; d'autre part, ce mot est unique d'après le théorème 2.1.

1. Et tout état qui n'a pas de successeur par 0 ne peut être final.

On définit le morphisme de mots $\varphi$ de $[\![p]\!]^*$ vers $[\![q]\!]^*$ par : l'image $\varphi(a)$ d'une lettre $a \in [\![p]\!]$ est l'unique mot de $[\![q]\!]^*$ de valeur $a$ et de longueur $k$. Une simple vérification donne alors la formule suivante

$$\forall u \in [\![p]\!]^* \qquad \pi_p(u) \;=\; \pi_q(\varphi(u)) \;.$$

Sens direct. Soit $\mathcal{A} = \langle\, Q, [\![q]\!], \delta, i, F \,\rangle$ un automate acceptant un ensemble d'entiers $N(\mathcal{A})$ en base $q$. On définit $\mathcal{B} = \langle\, Q, [\![p]\!], \delta', i, F \,\rangle$ un automate qui ne diffère de $\mathcal{A}$ que par 1) l'alphabet qui est $[\![p]\!]$ au lieu de $[\![q]\!]$ et 2) les transitions qui sont définies comme suit :

$$\forall a \in [\![p]\!] \qquad s \xrightarrow[\mathcal{B}]{a} s' \quad\Longleftrightarrow\quad s \xrightarrow[\mathcal{A}]{\varphi(a)} s' \;.$$

Soit un mot $u \in [\![q]\!]^*$. Ce mot est accepté par $\mathcal{B}$ si et seulement si $\varphi(u)$ est accepté par $\mathcal{A}$, c'est-à-dire si et seulement si $\pi_p(u) = \pi_q(\varphi(u))$ appartient à $N(\mathcal{A})$. Il s'ensuit que $N(\mathcal{B}) = N(\mathcal{A})$.

Sens réciproque. Soit $\mathcal{A} = \langle\, Q_{\mathcal{A}}, [\![p]\!], \delta_{\mathcal{A}}, i_{\mathcal{A}}, F_{\mathcal{A}} \,\rangle$ un automate acceptant un ensemble d'entiers $N(\mathcal{A})$ en base $p$. On définit $\mathcal{B} = \langle\, Q_{\mathcal{A}} \times [\![q]\!]^{<k}, [\![q]\!], \delta_{\mathcal{B}}, (i_{\mathcal{A}}, \varepsilon), F_{\mathcal{B}} \,\rangle$ où les transitions sont définies par

$$\forall u \in [\![q]\!]^{<k},\ \forall a \in [\![q]\!],\ \forall s \in Q_{\mathcal{A}} \qquad \begin{cases} (s,u) \xrightarrow[\mathcal{B}]{a} (s, ua) & \text{si } |ua| < k \\ (s,u) \xrightarrow[\mathcal{B}]{a} (s', \varepsilon) & \text{sinon, où } s' \text{ est l'état tel que } s \xrightarrow[\mathcal{A}]{\varphi^{-1}(ua)} s' \end{cases}$$

et $F_{\mathcal{B}}$ est formés de tous les état $(s,u)$ tels que $(s,u) \xrightarrow{0^i} (s', \varepsilon)$ et $s' \in F_{\mathcal{A}}$, pour un certain entier $i$. Pour tout mot $v \in [\![p]\!]^*$, le calcul de $\varphi(v)$ est acceptant dans $\mathcal{B}$ si et seulement si celui de $v$ est acceptant dans $\mathcal{A}$ ; ceci montre que $\mathcal{B}$ accepte un mot de longueur multiple de $k$ si et seulement si sa valeur est dans $N(\mathcal{A})$. Soit un mot $u$ de longueur quelconque ; il existe un entier $i$ tel que $u0^i$ ai pour longueur un multiple de $k$, et la définition de $F_{\mathcal{B}}$ implique que $u$ est accepté si et seulement si $u\,0^i$ est accepté. Si bien que $\mathcal{B}$ accepte $u$ si et seulement si $\pi_p(u) = \pi_p(u0^i)$ est dans $N(\mathcal{A})$, donc $N(\mathcal{B}) = N(\mathcal{A})$. □

On note $E_n^R$ l'ensemble d'entiers *purement périodique* dont la période est $n \in \mathbb{N}$ et l'ensemble des restes est $R \subseteq \mathbb{Z}/n\mathbb{Z}$, c'est-à-dire :

$$E_n^R \;=\; \{\, i \in \mathbb{N} \mid (i \,\%\, n) \in R \,\} \;. \tag{2.9}$$

On dit que le couple $(n, R)$ est *canonique* si $n$ est la plus petite période de $E_n^R$, c'est-à-dire s'il n'existe pas $n' < n$ et $R'$ tel que $E_n^R = E_{n'}^{R'}$. Par exemple, le couple $(n, R) = (10, \{0, 5\})$ n'est pas canonique car le reste d'un entier par division par 10 est égale à 0 ou 5 si et seulement s'il est divisible par 5 : $E_{10}^{\{0,5\}} = E_5^{\{0\}}$

De plus, on note $E_{n,m}^R$ l'ensemble des entiers de $E_n^R$ plus grands que la *pré-période* $m \in \mathbb{N}$, c'est-à-dire :

$$E_{n,m}^R \;=\; \{\, i \in \mathbb{N} \;\mid\; (i \,\%\, n) \in R \quad \text{et} \quad i \;\geqslant\; m \,\} \;=\; E_n^R \cap (m + \mathbb{N}) \;. \tag{2.10}$$

Un ensemble est dit *ultimement périodique* s'il peut s'écrire comme $(I \cup E_{n,m}^R)$ pour certains $n \in \mathbb{N}$, $R \subseteq \mathbb{Z}/n\mathbb{Z}$, $m \in \mathbb{N}$ et $I \subseteq \{0, 1, \ldots, m-1\}$.

**Remarque 2.9** – *La période d'un ensemble d'entiers est usuellement notée $p$ et la base entière $b$. Dans notre cas, la base entière est le cas particulier de la base rationnelle, usuellement notée $\frac{p}{q}$, où $q = 1$. Nous avons choisi de noter la base $p$ et la période $n$.*

Les ensembles ultimement périodiques ont une place particulière dans l'étude des ensembles $p$-reconnaissables, car ils sont représentés par des langages réguliers dans toutes les bases.

**Lemme 2.10** – *Tout ensemble ultimement périodique est $p$-reconnaissable, pour toute base $p$.*

Dans le chapitre suivant (sous-section 3.3), nous construisons un automate acceptant un ensemble ultimement périodique arbitraire.

Le théorème de Cobham [25], énoncé ci-dessous, est la réciproque des lemmes 2.8 et 2.10. C'est un résultat fondamental dont la démonstration est difficile (voir aussi [16]).

**Théorème 2.11** [25] – *Soient deux entiers $p, q > 1$. Si $p$ et $q$ sont multiplicativement indépendants, alors tout ensemble d'entiers qui est à la fois $p$- et $q$-reconnaissable est ultimement périodique.*

**Corollaire 2.12** – *Un ensemble d'entiers est $p$-reconnaissable pour toute base $p$ si et seulement si il est ultimement périodique.*

Après le théorème de Cobham se pose la question de décider si un ensemble $p$-reconnaissable donné est ultimement périodique ou non. Ce problème, énoncé plus formellement ci-dessous, a été résolu par Honkala [40].

| Ultime périodicité | |
|---|---|
| Données: | • une base $p$ ;<br>• un automate fini déterministe $\mathcal{A}$ sur l'alphabet $[\![p]\!]$. |
| Question: | Le langage accepté par $\mathcal{A}$ est-il la représentation en base $p$ d'un ensemble ultimement périodique ? |

**Théorème 2.13** [40] – *Le problème Ultime périodicité est décidable.*

La démonstration du théorème 2.13 donnée dans [40] consiste à borner les période et pré-période éventuelles par une fonction du nombre d'états de l'automate donné en entrée ; puis à énumérer tous les automates dont les paramètres sont plus petits que ces bornes en testant pour chacun s'il est équivalent à l'automate donné

[25] Alan Cobham, 1969, *On the Base-Dependence of Sets of Numbers Recognizable by Finite Automata.*

[16] Véronique Bruyère, Georges Hansel, Christian Michaux et Roger Villemaire, 1994, *Logic and p-recognizable sets of integers.*

[40] Juha Honkala, 1986, *A Decision Method for The Recognizability of Sets Defined by Number Systems.*

en entrée. La question de la complexité d'un algorithme qui mettrait en œuvre une telle procédure n'était pas envisagée.

Une approche en dimension $d$ supérieure a permis d'apporter un éclairage nouveau sur ce problème. Soit $\mathbb{N}^d$ le monoïde additif des $d$-uplets d'entiers. Les sous-ensembles reconnaissables ou rationnels de $\mathbb{N}^d$ sont définis de façon usuelle (*cf.* [72]). Un sous-ensemble de $\mathbb{N}^d$ est dit *reconnaissable* s'il est saturé par une congruence d'index fini et il est dit *rationnel* s'il peut être exprimé par une expression rationnelle. La famille des ensembles rationnels est noté $\mathsf{Rat}(\mathbb{N}^d)$ et celle des ensembles reconnaissables est notée $\mathsf{Rec}(\mathbb{N}^d)$. Si $d > 1$, le monoïde $\mathbb{N}^d$ n'est pas libre, ce qui implique que $\mathsf{Rat}(\mathbb{N}^d) \neq \mathsf{Rec}(\mathbb{N}^d)$ ; puisque par ailleurs $\mathbb{N}^d$ est un monoïde finiment engendré, $\mathsf{Rat}(\mathbb{N}^d) \subseteq \mathsf{Rec}(\mathbb{N}^d)$. [2]

Une caractérisation de $\mathsf{Rat}(\mathbb{N}^d)$ est donnée par le théorème suivant. C'est de ce théorème que vient l'appellation *ensemble Presburger-définissable* qui est souvent employée dans la littérature.

**Théorème 2.14** [37] – *Les ensembles rationnels de $\mathbb{N}^d$ sont exactement les ensembles qui peuvent être définis par une formule de l'arithmétique de Presburger, c'est-à-dire une formule de la logique du premier ordre dont les prédicats atomiques sont l'égalité et l'addition.*

Un $d$-uplet d'entiers peut être représenté en base $p$ par un $d$-tuple de mots *de la même longueur* (des zéros de queue peuvent être ajoutés à des représentations plus courtes). De tels $d$-uplets peuvent être lus par des automates finis sur l'alphabet $(\llbracket p \rrbracket^d)^*$ (c'est-à-dire des automates qui lisent de façon synchronisée sur $d$ bandes). Par analogie avec la dimension $d = 1$, un sous-ensemble de $\mathbb{N}^d$ est dit $p$-reconnaissable [3] si ses éléments sont acceptés par un tel automate.

Les deux lemmes suivants généralisent donc les lemmes 2.8 et 2.10 à la dimension $d$.

**Lemme 2.15** – *Soient deux bases $p, q > 1$ et une dimension $d$. Si $p$ et $q$ sont multiplicativement dépendants, alors toute partie de $\mathbb{N}^d$ est $p$-reconnaissable si et seulement si elle est $q$-reconnaissable.*

**Lemme 2.16** – *Tout ensemble de $\mathsf{Rat}(\mathbb{N}^d)$ est $p$-reconnaissable dans toute base $p \in \mathbb{N}$, et toute dimension $d \in \mathbb{N}$.*

Une généralisation du théorème de Cobham (théorème 2.11) en dimension $d$ est due à Semenov (voir aussi [16]).

2. Au contraire, dans le cas de la dimension $d = 1$, $\mathbb{N}$ est isomorphe au monoïde libre $\{a\}^*$ et $\mathsf{Rat}(\mathbb{N}) = \mathsf{Rec}(\mathbb{N})$ est formé des ensembles d'entiers ultimement périodiques.

3. On pourrait aussi dire $p$-rationnel, voire même $p$-régulier.

[72] Jacques Sakarovitch, 2003, *Éléments de théorie des automates*.

[37] Seymour Ginsburg et Edwin H. Spanier, 1966, *Semigroups, Presburger formulas and languages*.

[16] Véronique Bruyère, Georges Hansel, Christian Michaux et Roger Villemaire, 1994, *Logic and p-recognizable sets of integers*.

**Théorème 2.17** [75] – *Soient deux bases $p$, $q > 1$ et une dimension $d$. Si $p$ et $q$ sont multiplicativement indépendants, alors toute partie de $\mathbb{N}^d$ qui est à la fois $p$- et $q$-reconnaissable est rationnelle.*

La généralisation du problème Ultime périodicité en dimension $d$ est alors exprimée par le problème ci-dessous.

| Sous-ensemble rationnel | |
|---|---|
| Données: | • une base $p$<br>• une dimension $d$<br>• un automate fini déterministe $\mathcal{A}$ sur l'alphabet $[\![p]\!]^d$. |
| Question: | Le langage accepté par $\mathcal{A}$ est-il dans $\mathsf{Rat}(\mathbb{N}^d)$ ? |

Une première réponse à ce problème est d'abord proposé par Muchnik, puis une seconde, plus efficace, par Leroux.

**Théorème 2.18** [61, 62] – *Le problème Sous-ensemble rationnel est décidable.*

**Théorème 2.19** [46] – *Le problème Sous-ensemble rationnel est décidable en temps polynomial.*

Les algorithmes sous-jacents aux preuves des théorèmes 2.18 et 2.19 ont des complexités temporelles respectivement triplement exponentielle et quadratique. Le second est donc considérablement plus rapide ; il repose sur des constructions géométriques sophistiquées.

Une autre façon de démontrer la décidabilité de Ultime périodicité est donnée dans [4]. Ses auteurs remarquent qu'il existe une formule du premier ordre (utilisant comme prédicats atomiques l'ordre et l'égalité) qui est satisfaite si et seulement si un ensemble $E$ d'entiers est ultimement périodique :

$$\Phi(E) \;=\; \exists n, m \in \mathbb{N},\ \forall i \in E \qquad i > m \implies [i \in E \iff (i+n) \in E]\ .$$

Puisque l'addition et l'ordre sont réalisés en bases $p$ par des transducteurs finis, lettre-à-lettre et séquentiel, si $E$ est accepté par un automate fini, alors on peut construire un automate acceptant $\Phi(E)$. L'ensemble $E$ est ultimement périodique si et seulement si cet automate accepte toutes les entrées, ce qui est décidable en temps linéaire. Cette procédure résout donc le problème Ultime périodicité en complexité exponentielle (car la formule $\Phi$ a une alternance de quantificateurs).

[75] Alexei L. Semenov, 1977, *Presburgerness of predicates regular in two number systems (in Russian).*

[61] Andrei A. Muchnik, 1991, *The definable criterion for definability in Presburger arithmetic and its applications (in Russian).*

[46] Jérôme Leroux, 2005, *A polynomial time Presburger criterion and synthesis for number decision diagrams.*

[4] Jean-Paul Allouche, Narad Rampersad et Jeffrey Shallit, 2009, *Periodicity, repetitions, and orbits of an automatic sequence.*

Cette solution est à première vue moins efficace que celle de Leroux et traite un cas plus restreint : la dimension $d = 1$. Néanmoins, elle peut être généralisée dans une direction orthogonale à celle de la dimension si l'on considère le problème suivant.

| Ultime Périodicité Généralisé | |
|---|---|
| Données: | • un SNAR $S$ sur un alphabet $A$<br>• un automate fini déterministe $\mathcal{A}$ sur $A$. |
| Question: | Le langage accepté par $\mathcal{A}$ est-il la représentation dans le système $S$ d'un ensemble ultimement périodique ? |

L'idée de [4] donne alors immédiatement une réponse partielle à ce problème, comme d'ailleurs noté dans un article ultérieur [22]. Le problème Ultime Périodicité Généralisé est décidable en temps exponentiel pour les systèmes de numération abstraits réguliers dans lesquels l'addition est réalisée par un transducteur lettre à-lettre et séquentiel.

Dans [12] est donnée une autre solution partielle à ce problème. Il est décidable pour une large classes de système de numération positionnels (*cf.* [36] ou sous-section 9.3.1) ; cette classe est incomparable avec celle traitée par l'idée précédente, et ne contient notamment pas les bases entières. La démonstration consiste, comme dans l'article original de Honkala [40], à borner les période et prépériode par des fonctions du nombre d'états ; la question de la complexité de l'algorithme proposé n'est pas abordée.

Enfin, Rigo et Maes montrent dans [71] que le problème Ultimem Périodicité Généralisé est équivalent [4] à un problème sur les morphismes de mot. Durand [33] et Mitrofanov [58] ont récemment démontré indépendamment que ce problème était décidable.

4. En disant que les problèmes sont *équivalents*, on entend que chaque instance d'un des problèmes peut être transformé en une instance de l'autre de telle sorte que les réponses aux deux problèmes soient identiques. La transformation sous-jacente est utilisée dans la section 7.2.2 (page 185 et suivantes) à d'autres fins.

[22] Emilie Charlier, Narad Rampersad et Jeffrey Shallit, 2012, *Enumeration and Decidable Properties of Automatic Sequences.*

[12] Jason Bell, Emilie Charlier, Aviezri S. Fraenkel et Michel Rigo, 2009, *A Decision Problem for Ultimately Periodic Sets in Nonstandard Numeration Systems.*

[36] Christiane Frougny et Jacques Sakarovitch, 2010, *Number representation and finite automata.*

[71] Michel Rigo et Arnaud Maes, 2002, *More on Generalized Automatic Sequences.*

[33] Fabien Durand, 2013, *Decidability of the HD0L ultimate periodicity problem.*

[58] Ivan Mitrofanov, 2011, *A proof for the decidability of HD0L ultimate periodicity (in Russian).*

## Chapitre 3

# Des ensembles ultimement périodiques

Ce chapitre présente un algorithme quasi-linéaire pour résoudre le problème Ultime Périodicité présenté dans le chapitre précédent. La construction sous-jacente a été partiellement inspirée de celle de Leroux et comporte donc des similarités avec celle-ci. Néanmoins l'amélioration de complexité apportée n'est pas due à une simplification de l'algorithme à la dimension 1.

**Théorème I** – *Soient une base entière $p$ et un automate déterministe $\mathcal{A}$ sur l'alphabet $[\![p]\!]$. On note $n$ le nombre d'états de $\mathcal{A}$ et $m$ son nombre de transitions. Le problème ULTIME PÉRIODICITÉ peut être décidé en temps $O(n \log(n) + m)$.*

Puisque l'on ne s'intéresse qu'à des automates déterministes, $m \leqslant pn$. Donc, si l'on considère que $p$ comme un paramètre du problème, la solution proposée est plus simplement en complexité $O(n \log(n))$.

L'algorithme associé au théorème I résulte d'une caractérisation structurelle (appelé critère (UP)) des automates qui acceptent les ensembles ultimement périodiques. Cette façon de procéder répond donc également à la question informelle

*À quoi ressemble un automate acceptant un ensemble ultimement périodique ?*

Tout d'abord, la section 3.1 traite un cas particulier du problème Ultime Périodicité : le cas où l'automate donné en entrée est fortement connexe. Il y est décrit des automates canoniques appelés *automates de Pascal* qui acceptent les ensembles purement périodiques dont la période est première avec la base. Il est ensuite montré comment décider en temps linéaire si un automate donné est le quotient d'un tel automate.

Dans la section 3.2, nous donnons des propriétés structurelles d'automates concernant les formes et les positions relatives de ses composantes fortement connexes (CFC). Ces propriétés sont rassemblées sous le nom de *critère (UP)* que l'on peut décider en temps linéaire, comme l'énonce le théorème suivant.

**Théorème 3.37** – *Soit $\mathcal{A}$ un automate minimal dont le nombre de transitions est $m$. Il peut être décidé en temps $O(m)$ si $\mathcal{A}$ satisfait le critère (UP) ou non.*

---

Les résultats présentés dans ce chapitre ont été publiés dans les actes de DLT 2013, voir [54].

L'algorithme présenté dans la section 3.1 (pour décider si un automate donné est le quotient d'un automate de Pascal) est utilisé de façon cruciale par celui sous-jacent au théorème 3.37. En réalité, la complexité linéaire du second algorithme découle presque immédiatement de celle du premier ; la principale difficulté de la démonstration du théorème 3.37 est donc préalablement résolue dans la section 3.1.

La section 3.3 est consacrée à la construction, étant donné un ensemble d'entiers $E$ ultimement périodique arbitraire, d'un automate qui accepte $E$ et qui satisfait le critère (UP). En d'autres termes, il y est établi le théorème suivant.

**Théorème 3.38** – *Soient une période $n \in \mathbb{N}$, une prépériode $m \in \mathbb{N}$, un ensemble $R \subseteq \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$ et un ensemble $I \subseteq \{0, 1, \ldots, m-1\}$. Il existe un automate complet qui accepte l'ensemble d'entiers $(I \cup E^R_{n,m})$ et qui satisfait le critère (UP).*

Dans la section 3.4, nous montrons que le critère (UP) est stable par quotient ; il découle donc du théorème 3.38 que l'automate minimal acceptant un ensemble ultimement périodique donné satisfait le critère (UP). Inversement, nous y établissons que tout automate satisfaisant le critère (UP) accepte un ensemble ultimement périodique d'entiers. Le théorème suivant résume les deux sens.

**Théorème 3.52** – *Soit $\mathcal{A}$ un automate minimal et, complet ou émondé. L'automate $\mathcal{A}$ accepte par valeur un ensemble ultimement périodique d'entiers si et seulement si il satisfait le critère (UP).*

Or, la minimisation d'un automate à $n$ état peut être effectuée en temps $\mathsf{O}(n \log(n))$ grâce à l'algorithme classique d'Hopcroft (*cf.* [41, 28]). Le théorème I découle donc des théorèmes 3.37 et 3.52.

# Automates de Pascal

Cette première section traite le problème Quotient d'un Pascal, un cas particulier du problème Ultime Périodicité : la pré-période $m$ est nulle ($m = 0$) et la période $n$ est première avec la base ; aucune restriction n'est imposée à l'ensemble de restes $R$ qui est donc une partie de $\mathbb{Z}/n\mathbb{Z}$.

Nous définissons alors *l'automate de Pascal*[1] *de paramètre* $(n, R)$, noté $\mathcal{P}^R_n$, qui accepte l'ensemble d'entiers $E^R_n$ : $L(\mathcal{P}^R_n) = \{u \mid \pi_p(u) \,\%\, n \in R\}$. Si le paramètre $(n, R)$ est canonique, on dit par abus de langage que $\mathcal{P}^R_n$ est un automate de Pascal *canonique*.

1. En l'honneur du mathématicien et philosophe Blaise Pascal (1623–1662), qui a décrit dès 1654 une procédure pour déterminer si un entier $k$, écrit *dans n'importe quelle base $p$* est divisible par un autre entier $n$ (*cf.* [66] ou le prologue de [73]).

[41] John E. Hopcroft, 1971, *An nlogn algorithm for minimizing states in a finite automaton.*
[28] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest et Clifford Stein, 2002, *Introduction à l'algorithmique (2ème ed.).*
[66] Blaise Pascal, 1963, *Œuvres complètes.*
[73] Jacques Sakarovitch, 2009, *Elements of Automata Theory.*

Ainsi tout automate **minimal** qui accepte l'ensemble d'entiers $E_n^R$ est le quotient de $\mathcal{P}_n^R$ que l'on peut de plus supposer canonique ; on peut donc réduire ce cas particulier de Ultime Périodicité au problème suivant.

| Quotient d'un Pascal | |
|---|---|
| Données: | • une base $p$<br>• un automate fini déterministe $\mathcal{A}$ sur l'alphabet $[\![p]\!]$. |
| Question: | $\mathcal{A}$ est-il un quotient d'un automate de Pascal canonique ? |

Le but de cette section est de prouver que ce problème peut-être résolu en temps linéaire, comme énoncé ci-dessous.

**Théorème 3.1** – *Le problème Quotient d'un Pascal est décidable en temps $O(m)$ où $m$ est le nombre de transitions de l'automate en entrée.*

**Remarque 3.2** – *On peut assez facilement décider ce problème avec une complexité polynomiale, ou même quadratique, mais obtenir une complexité linéaire présente une réelle difficulté. En effet, le nombre d'états d'un automate de Pascal n'est généralement pas linéaire en le nombre d'états de sa minimisation. Dans le pire des cas, il faut donc vérifier l'existence d'un morphisme d'automates sans parcourir en entier le revêtement. Ceci n'est rendu possible que grâce à la remarquable régularité des monoïdes de transitions des automates de Pascal.*

## Définition

Soit un entier $n \geqslant 1$ premier avec la base $p$ et un ensemble $R \subseteq \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$. Puisque $n$ et $p$ sont premiers entre eux, $p$ est un élément inversible de l'anneau $\mathbb{Z}/n\mathbb{Z}$ ; il en est de même pour $p^2, p^3 \ldots$ Or les éléments inversibles de $\mathbb{Z}/n\mathbb{Z}$ sont en nombre fini, il existe donc $i$ et $j$, tels que $p^i = p^j$ et $i > j$ donc $p^{(i-j)} = 1$. Il existe donc un (plus petit) entier (strictement positif) $\psi$ tel que

$$p^{\psi} \equiv 1 \quad [n] \quad \text{de telle sorte que} \quad \forall k \in \mathbb{N} \quad p^k \equiv p^{(k \,\%\, \psi)} \quad [n] \,. \tag{3.1}$$

L'entier $\psi$ n'est autre que l'ordre de $p$ dans le groupe (multiplicatif) des éléments inversibles de $\mathbb{Z}/n\mathbb{Z}$.

Il découle des équations (2.2) et (3.1) que $(\pi_p(u\,a) \,\%\, n)$ peut être calculé à partir de $(\pi_p(u) \,\%\, n)$ et de $(|u| \,\%\, \psi)$ grâce à la formule

$$\forall u \in [\![p]\!]^*, \ \forall a \in [\![p]\!] \qquad \pi_p(u\,a) \,\%\, n \equiv (\pi_p(u) \,\%\, n) + a\, p^{|u| \,\%\, \psi} \quad [n] \,. \tag{3.2}$$

**Définition 3.3** – *Pour tout entier $n \geqslant 1$ premier avec la base $p$ et tout ensemble $R \subseteq \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$,* l'automate de Pascal de paramètre $(n, R)$ *est*

$$\mathcal{P}_n^R = \langle\, \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z},\ [\![p]\!],\ \delta,\ (0,0),\ R \times \mathbb{Z}/\psi\mathbb{Z} \,\rangle$$

*dont la fonction de transition $\delta$ est définie par*

$$\forall (s,t) \in \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z}, \ \forall a \in [\![p]\!] \qquad \delta((s,\,t),\,a) = (s,\,t) \cdot a = (s + a\,p^t,\ t+1) \,. \tag{3.3}$$

**Exemple 3.4** – *La figure 1 représente $\mathcal{P}_3^{\{2\}}$, l'automate de Pascal qui accepte les entiers écrits en binaire et dont la valeur est congrue à 2 modulo 3 ; c'est-à-dire que $p = 2$, $n = 3$, $R = \{2\}$ et un simple calcul montre que $\psi = 2$.*

*Pour alléger la figure, les étiquettes sont omises ; les transitions étiquetées par la lettre 1 sont dessinées avec des traits rouges gras et celles étiquetées par la lettre 0 sont dessinées avec des traits noirs fins.*

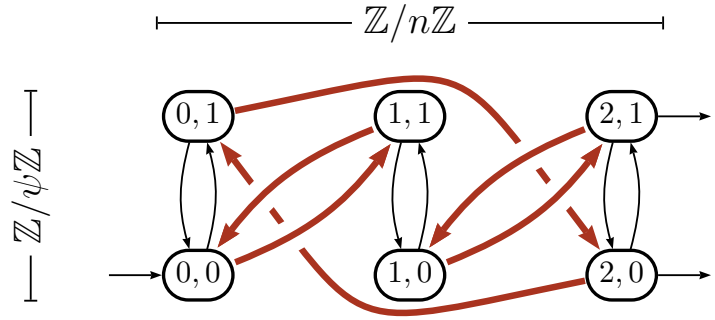



Figure 1 – L'automate de Pascal $\mathcal{P}_3^2$

L'automate de Pascal ainsi défini est correct, comme l'énonce la proposition suivante.

**Proposition 3.5** – *L'ensemble d'entiers $E_n^R$ est accepté par $\mathcal{P}_n^R$.*

Cette proposition est une conséquence du corollaire 3.7 du lemme suivant.

**Lemme 3.6** – *Soit un mot $u \in [\![p]\!]^*$. On note $h = \pi_p(u) \,\%\, n$ et $k = |u| \,\%\, \psi$. Alors, pour tout état $(s,t)$ de $\mathcal{P}_n^R$, $(s,t) \cdot u = (s + h\,p^t,\, t + k)$.*

Démonstration. Par induction sur la longueur de $u$ ; si $u$ est le mot vide, alors $\pi_p(u) = |u| = 0$ et le lemme est vérifié.

Soit un mot non-vide $u\,a \in [\![p]\!]^+$. On note $h = \pi_p(u) \,\%\, n$ et $k = |u| \,\%\, \psi$, si bien que l'hypothèse de récurrence est

$$(s,t) \xrightarrow{u} (s + h\,p^t,\, t + k)\ .$$

Il découle donc de l'équation (3.3) (appliquée à l'état $(s + h\,p^t,\, t + k)$ et à la lettre $a$) que

$$((s,t) \cdot u\,a) \;=\; \big(s + (h + a\,p^k)\,p^t,\, t + k + 1\big)\ .$$

Or, appliquer l'équation (3.2) donne :

$$\pi_p(u\,a) \;\equiv\; \pi_p(u) \,\%\, n + a\,p^{|u| \,\%\, \psi} \;=\; h + a\,p^k\ ,$$

ce qui implique que

$$((s,t) \cdot u\,a) \;=\; (s + h'\,p^t,\, t + k')$$

où $h' = \pi_p(u\,a) \,\%\, n$ et $k' = |u\,a| \,\%\, \psi$, concluant la récurrence. □

**Corollaire 3.7** – *Soit un mot $u \in [\![p]\!]^*$. Le calcul de $u$ dans $\mathcal{P}_n^R$ atteint l'état $(\pi_p(u),\, |u|)$.*

## Propriétés des automates de Pascal

Soit un automate de Pascal $\mathcal{P}_n^R$ fixé dans la suite. Nous étudions dans cette sous-partie les propriétés de $\mathcal{P}_n^R$, en particulier celles de son monoïde de transitions.

**Proposition 3.8** – *L'automate $\mathcal{P}_n^R$ est émondé.*

Démonstration. Soit un état $(s,t)$ de $\mathcal{P}_n^R$.

Montrons que $(s,t)$ est accessible. On définit l'entier

$$k \;=\; (\psi - |\langle s\rangle_p|) + t \quad \text{ et le mot } \quad u \;=\; \langle s\rangle_p 0^k\,;$$

une simple vérification montre que $|u|\,\%\,\psi \;=\; t$ et que $\pi_p(u) \;=\; s$. D'après le corollaire 3.7, le calcul du mot $u$ atteint l'état $(s,t)$.

Montrons que $(s,t)$ est co-accessible. On garde les notation du paragraphe précédent et on définit de plus les entiers

$$i \;=\; (\psi - t) \;\;,\quad j \;=\; |\langle n-s\rangle_p| \quad \text{ et le mot } \quad v \;=\; 0^i\langle n-s\rangle_p 0^j \;.$$

Il s'ensuit que $|u\,v|\,\%\,\psi = 0$ et que

$$\begin{aligned}\pi_p(u\,v) \;=\; \pi_p(u\,0^i\langle n-s\rangle_p 0^j) \;&=\; \pi_p(u\,0^i) + \pi_p(\langle n-s\rangle_p 0^j)\,p^{|u|+i}\\ &=\; s + (n-s)\,p^{|u|+i}\;.\end{aligned}$$

Or, $(|u|+i) \equiv t + (\psi - t)\ [\psi] \equiv 0\ [\psi]$ donc $p^{(|u|+i)} \equiv 1\ [n]$ ce qui implique que $\pi_p(u\,v)\,\%\,n = 0$. Il découle donc du corollaire 3.7 que le calcul du mot $u\,v$ atteint l'état $(0,0)$ ; donc le chemin $(0,0) \xrightarrow{u} (s,t) \xrightarrow{v} (0,0)$ est dans $\mathcal{P}_n^R$. □

**Proposition 3.9** – *L'automate $\mathcal{P}_n^R$ est un automate à groupe.*

Démonstration. Soit un état $(s,t)$ de $\mathcal{P}_n^R$ et une lettre $a \in [\![p]\!]^*$. D'après l'équation (3.3), un état $(s',t')$ un est prédécesseur de $(s,t)$ par $a$ si et seulement si $s' = (s - a\,p^{(t-1)})$ et $t' = t-1$, donc $(s',t')$ existe et est unique. Toute lettre $a$ induit donc une permutation sur les états de $\mathcal{P}_n^R$ qui est donc un automate à groupe. □

À $n$ fixé, les automates de Pascal $\mathcal{P}_n^{R'}$ pour tout $R' \subseteq [\![n]\!]$ ne diffèrent que sur les états finals, ils ont donc tous le même monoïde de transitions, que l'on note $G_n$. Ce monoïde est un groupe d'après le lemme précédent et nous établissons dans la suite qu'il a une forme très particulière, comme l'énonce la proposition 3.11 après une définition préalable.

On rappelle qu'un sous-groupe $H$ d'un groupe $G$ est dit *distingué* s'il est stable par conjugaison, c'est-à-dire si pour tout élément $g \in G$, $g\,H\,g^{-1} = H$.

**Définition 3.10** – *Soient $H$ et $K$ deux groupes et $f : K \to \mathsf{Aut}(H)$ un morphisme de $K$ vers les automorphismes de $H$. Le produit semi-direct (externe) $H \rtimes_f K$ est le produit cartésien de $H$ et $K$ muni de la loi*

$$\forall h_1, h_2 \in K,\ \forall k_1, k_2 \in K \qquad (h_1,k_1)(h_2,k_2) \;=\; (h_1\,f(k_1)(h_2), k_1k_2)\;.$$

*On dit qu'un groupe $G$ est le produit semi-direct (interne) d'un sous-groupe distingué $H$ par un sous-groupe $K$ si*

$$H \cap K \;=\; \{e\} \quad \textit{et} \quad H\,K \;=\; G$$

*où $e$ est l'élément neutre de $G$.*

**Proposition 3.11** – *Le monoïde de transitions de $\mathcal{P}_n^R$ est isomorphe au produit semi-direct $\mathbb{Z}/n\mathbb{Z} \rtimes \mathbb{Z}/\psi\mathbb{Z}$.*

Le groupe $G_n$ est, par définition, formé des permutations $\tau_u$, pour tout mot $u$ de $[\![p]\!]^*$, définies par

$$\begin{array}{rccc} \tau_u : & \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z} & \longrightarrow & \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z} \\ & (s,t) & \longmapsto & (s,t)\cdot u \end{array}$$

Il découle du lemme 3.6 que la permutation $\tau_u$ ne dépend que de $\pi_p(u)\,\%\,n$ et de $|u|\,\%\,\psi$. Il s'ensuit que $G_n$ est isomorphe au groupe $(\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z},\, \diamond)$, dont la loi est définie par :

$$(s,t) \diamond (h,k) \;=\; (s + h\,p^t,\; t+k)\;, \tag{3.4}$$

par l'isomorphisme $g : G_n \to \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z}$ défini par

$$\forall u \in [\![p]\!]^* \qquad g(\tau_u) \;=\; (\pi_p(u), |u|) \;=\; \tau_u((0,0))\;. \tag{3.5}$$

Le lemme suivant découle immédiatement du lemme 3.6.

**Lemme 3.12** – *Pour tout mot $u \in [\![p]\!]^*$ et tout état $(s,t) \in \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z}$,*

$$((s,t)\cdot u) \;=\; (s,t) \diamond (\pi_p(u), |u|)\;.$$

Les propriétés suivantes montrent que le groupe $(\mathbb{Z}/n\mathbb{Z}\times\mathbb{Z}/\psi\mathbb{Z},\, \diamond)$ est le produit semi-direct de deux de ses sous-groupes.

**Propriété 3.13** –

**a)** *L'ensemble $H = \mathbb{Z}/n\mathbb{Z} \times \{0\}$ est un sous-groupe distingué de $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z}$.*

**b)** *L'ensemble $K = \{0\} \times \mathbb{Z}/\psi\mathbb{Z}$ est un sous-groupe de $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z}$.*

**c)** *Le groupe $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z}$ est le produit semi-direct interne $H \rtimes K$.*

Démonstration. **a)** Soient $(h,0)$ et $(h',0)$ deux éléments de $H$. D'après l'équation (3.4), $(h,0) \diamond (h',0) = (h+h',0) \in H$ ; $H$ est donc un sous-groupe de $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z}$.

Soit $(s,t)$ un élément de $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z}$ ; son inverse est nécessairement $(s',-t)$, pour un certain $s'$. Pour tout élément $(h,0)$ de $H$ le produit $(s,t) \diamond (h,0) \diamond (s',-t)$ est égal à $(h',0)$, pour un certain $h'$ ; $H$ est donc stable par conjugaison, il est donc un sous-groupe *distingué* de $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z}$.

Le point (**b**) est démontré de manière analogue.

**c)** D'après l'équation (3.4), $(h,0) \diamond (0,k) = (h,k)$, pour tout $(h,0) \in H$ et $(0,k) \in K$, donc $H\,K = G$. De plus $H \cap G = \{(0,0)\}$, donc $G$ est le produit semi-direct interne $H \rtimes K$. □

Pour résumer, $G_n$ est isomorphe au groupe $(\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z},\, \diamond)$ et le lemme précédent établit que celui-ci est égal au produit semi-direct interne $(\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z},\, \diamond) = ((\mathbb{Z}/n\mathbb{Z}\times\{0\}) \rtimes (\{0\}\times\mathbb{Z}/\psi\mathbb{Z}))$ ou plus simplement $(\mathbb{Z}/n\mathbb{Z}\times\mathbb{Z}/\psi\mathbb{Z},\, \diamond) = \mathbb{Z}/n\mathbb{Z} \rtimes \mathbb{Z}/\psi\mathbb{Z}$. Ceci conclut la preuve de la proposition 3.11.

Dans la suite, on identifie $G_n$ et $\mathbb{Z}/n\mathbb{Z} \rtimes \mathbb{Z}/\psi\mathbb{Z}$, si bien que l'on pourra écrire par exemple *la permutation* $(s,t) \in G_n$. Cette permutation est induite par tous les mots $u$ vérifiant $\pi_p(u) = s$ et $|u| = t$ et correspond donc à la multiplication à droite par $(s,t)$, c'est-à-dire $(h,k) \mapsto (h,k) \diamond (s,t)$, pour tout $(h,k)$ in $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/\psi\mathbb{Z}$.

Par définition des monoïdes (ici groupes) de transitions, $G_n$ est engendré par les permutations induites par les lettres de $[\![p]\!]$ ; les valeurs explicites de ces permutations sont données par le lemme suivant (qui est une conséquence directe du lemme 3.12).

**Lemme 3.14** – *La lettre $a \in [\![p]\!]$ induit la permutation $(a,1)$ de $G_n$.*

Le groupe $G_n$ est le produit (semi-direct) de deux groupes monogènes. Or il est engendré par la famille $\{\,(a,1) \mid a \in [\![p]\!]\,\}$ dont le cardinal est $p \geqslant 2$. Le lemme suivant montre qu'il peut être engendré par deux éléments seulement.

**Lemme 3.15** – *Le groupe $G_n$ est engendré par les actions des lettres 0 et 1.*

Démonstration. Il suffit de montrer que tout état $(s,t)$ de $\mathcal{P}_n^R$ peut être atteint depuis l'état initial en ne lisant que des 0 ou des 1.

Une simple application de l'équation (3.3) régissant les transition de $\mathcal{P}_n^R$ montre que

$$\forall k \in \mathbb{Z}/n\mathbb{Z} \qquad (k,0) \xrightarrow{10^{\psi-1}} (k+1,0)\ .$$

En itérant ce procédé, il s'ensuit que

$$(0,0) \xrightarrow{10^{\psi-1}} (1,0) \xrightarrow{10^{\psi-1}} \cdots \xrightarrow{10^{\psi-1}} (s,0)\ .$$

D'autre part, pour tout $t \in \mathbb{Z}/\psi\mathbb{Z}$, $(s,0) \xrightarrow{0^t} (s,t)$. □

Néanmoins, les éléments $\tau_0 = (0,1)$ et $\tau_1 = (1,1)$ ne correspondent pas aux générateurs naturels d'un produit (fut-il semi-direct) ; on l'a bien vu lors de la démonstration précédente, où toute utilisation de la lettre 1 nécessite d'être suivie par $(\psi - 1)$ fois la lettre 0. On définit donc une nouvelle lettre, notée $g$, de telle sorte que son action soit la permutation $\tau_g = (1,0)$. Il découle de l'équation (3.4) que l'égalité $(1,0) = (1,1) \diamond (0,\psi-1)$ est vérifiée si bien que l'action de la lettre $g$ est celle du mot $10^{(\psi-1)}$, c'est-à-dire :

$$\forall (s,t) \in G_n \qquad (s,t) \xrightarrow{g} (s + b^t,\ t) \ \ (= (s,t) \diamond (1,0)) \tag{3.6}$$

L'action de toute lettre $a \in [\![p]\!]$ est identique à celle du mot $g^a 0$, le lemme suivant découle donc du lemme 3.14 (ou 3.15).

**Lemme 3.16** – *Le groupe $G_n$ est engendré par les actions des lettres 0 et $g$.*

On peut donc construire un *automate de Pascal simplifié*, noté $\mathcal{P}'^R_n$, dont l'alphabet est $\{0,g\}$ ; il est obtenu en supprimant de $\mathcal{P}_n^R$ toutes les transitions étiquetées par des lettres différentes de 0 et en ajoutant des transitions étiquetées par $g$ dont l'action est celle du mot $10^{\psi-1}$. Le lemme précédent nous assure qu'aucune information n'a été perdue : $\mathcal{P}'^R_n$ et $\mathcal{P}_n^R$ ont le même monoïde de transition. Ainsi $\mathcal{P}'^R_n$ est le graphe (orienté) de Cayley du groupe $G_n$ engendré par $\tau_0$ et $\tau_g$.

**Exemple 3.17** – *La figure 2 montre l'automate $\mathcal{P}'^{\{2\}}_3$, à comparer avec la figure 1 représentant l'automate de Pascal $\mathcal{P}^{\{2\}}_3$. Une fois encore les étiquettes sont omises ; les transitions étiquetées par la lettre 0 sont dessinées avec des traits simples et les transitions étiquetées par la lettre g sont dessinées avec des traits verts doubles.*

*La structure de produit semi-direct y est très visible. En effet, la lettre* 0 *induit une permutation de chaque colonne et la lettre g induit une permutation de chaque ligne. De plus, l'action de* 0 *ne dépend pas de la colonne alors que l'action de g dépend de la ligne, ce qui brise la symétrie colonne/ligne et souligne le caractère semi-direct du produit.*

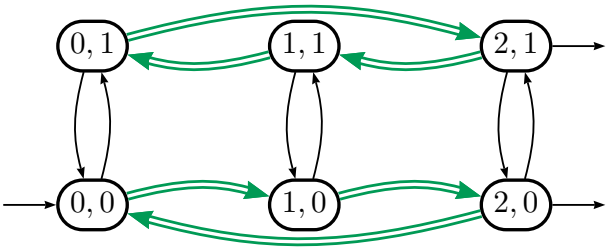



Figure 2 – L'automate de Pascal simplifié $\mathcal{P}'^{\{2\}}_3$

**Remarque 3.18** – *L'élément $(0, \psi - 1)$ est, dans $G_n$, l'inverse de $(0, 1)$. Dans la section 3.1.4, on s'autorisera à prendre les transitions des automates* ***à l'envers****, l'action de la lettre g est alors la même que celle du mot $10^{-1}$ ; ce mot a l'avantage d'être plus court et de ne pas dépendre de $\psi$ (ni donc de n).*

La proposition suivante utilise une propriété qui sera démontrée dans la section suivante sur les quotients d'un automate de Pascal canonique. Nous l'avons néanmoins placée car elle énonce un résultat sur les automates de Pascal.

**Proposition 3.19** – *Si $\mathcal{P}^R_n$ est canonique, alors le monoïde syntaxique de $L(\mathcal{P}^R_n)$ est isomorphe à $G_n$, le monoïde de transitions de $\mathcal{P}^R_n$.*

Démonstration. Le monoïde syntaxique de $L(\mathcal{P}^R_n)$ est isomorphe au monoïde de transitions de l'automate minimal acceptant $L(\mathcal{P}^R_n)$, noté $\mathcal{M}$ (voir par exemple [73]).

Soit $\varphi : \mathcal{P}^R_n \to \mathcal{M}$ le morphisme d'automates associé à cette minimisation. Il suffit donc de prouver que si deux mots $u$ et $v$ ont des actions différentes sur les états de $\mathcal{P}^R_n$, alors $u$ et $v$ ont également des actions différentes sur les états de $\mathcal{M}$.

Soient $u$ et $v$ de tels mots ; on note $(s, t) = \tau_u$ et $(s', t') = \tau_v$ les éléments de $G_n$ respectivement induits par $u$ et $v$ (donc en particulier $(s, t) \neq (s', t')$). La démonstration de toute la proposition consiste à prouver qu'il existe un état $(h, k)$ tel que $\varphi((h, k) \diamond (s, t)) \neq \varphi((h, k) \diamond (s', t'))$.

[73] Jacques Sakarovitch, 2009, *Elements of Automata Theory*.

D'après l'équation (3.4),

$$\forall (h,k) \in G_n \qquad \begin{array}{rcl} (h,k) \diamond (s,t) & = & (h + s\,p^k,\ k+t)\,; \\ (h,k) \diamond (s',t') & = & (h + s'\,p^k,\ k+t')\ . \end{array} \qquad (*)$$

Par l'absurde : on suppose que

$$\forall (h,k) \in G_n \qquad \varphi((h + s\,p^k,\ k+t)) \ = \ \varphi((h + s'\,p^k, k+t'))\ . \qquad (**)$$

Montrons que $s = s'$. Un morphisme d'automates respecte les états finals (*cf.* équation (1.1b)), d'où

$$\forall h \in \mathbb{Z}/n\mathbb{Z} \qquad (h + s\,p^k) \in R \iff (h + s'\,p^k) \in R\ .$$

Si $s \neq s'$, alors $p^k(s-s') \neq 0$ est une période plus petite que $n$ de $E_n^R$, ce qui contredit l'hypothèse comme quoi le paramètre $(n,R)$ est canonique, donc $s = s'$.

Montrons que $t = t'$. Il découle des équations $(*)$ que $((h,k) \diamond (s,t))$ et $((h,k) \diamond (s',t'))$ ont la même première composante (puisque $s = s'$) et des secondes composantes respectivement égale à $k+t$ et $k+t'$. Appliquer alors la contraposée du lemme 3.26 (qui sera démontré dans la sous-section suivante) montre que $k+t = k+t'$ donc $t = t'$

Nous avions pris comme hypothèse que $(s,t) \neq (s,t')$. Contradiction. □

**Remarque 3.20** – *On note $L_n^R$ le langage des mots de $[\![p]\!]^*$ dont la valeur appartient à $E_n^R$, si bien que $L_n^R = \langle E_n^R \rangle_p = L(\mathcal{P}_n^R)$. Il découle de la proposition 3.19 que l'automate de Pascal $\mathcal{P}_n^R$ est le graphe orienté de Cayley du monoïde syntaxique de $L_n^R$.*

Le lemme suivant qui conclut cette sous-section exprime le caractère isotrope des automates de Pascal. En effet, changer l'état initial d'un automate de Pascal produit un autre automate de Pascal de même période. Ce lemme permettra notamment dans les sections suivantes de parler d'automate de Pascal pour des composantes fortement connexes, sans état initial prédéfini.

**Lemme 3.21** (d'isotropisme) – *Changer l'état initial de l'automate de Pascal $\mathcal{P}_n^R$ produit l'automate de Pascal $\mathcal{P}_n^S$, pour un certain ensemble $S \subseteq \mathbb{Z}/n\mathbb{Z}$.*

Démonstration. Soient un état $(s,t)$ de $\mathcal{P}_n^R$ et $\mathcal{A}$ une copie de $\mathcal{P}_n^R$ où l'état initial est $(s,t)$ au lieu de $(0,0)$. Un mot $u \in [\![p]\!]^*$ est accepté par $\mathcal{A}$ si et seulement si $(s,t) \xrightarrow{u} (s',t')$ avec $s' \in R$ c'est-à-dire si et seulement si $(s + \pi_p(u)\,p^t) \,\%\, n \in R$ (lemme 3.6). Le langage accepté par $\mathcal{A}$ est donc

$$L(\mathcal{A}) \ = \ \{\, u \mid \pi(u) \,\%\, n \in S \,\} \quad \text{avec} \quad S \ = \ \{\, (r-s)\,p^{\psi - t} \mid r \in R \,\}\ .$$

Une simple vérification montre de plus que l'isomorphisme d'automates $A \to \mathcal{P}_n^S$ envoie l'état $(s',t')$ de $\mathcal{A}$ sur l'état $((s'-s)\,p^{\psi-t},\ t'-t)$ de $\mathcal{P}_n^S$. □

## Quotients d'un automate de Pascal canonique

Au contraire de la sous-section précédente, les propriétés regroupées ici ont directement pour but d'être appliquées dans l'algorithme présenté dans la sous-section 3.1.4 suivante.

Dans la suite, on suppose que $\mathcal{P}_n^R$ est un automate de Pascal **canonique**, on note $\mathcal{A}$ un quotient de cet automate et $\varphi : \mathcal{P}_n^R \to \mathcal{A}$ le morphisme d'automates associé.

Puisque tout quotient d'un automate à groupe est un automate à groupe (*cf.* proposition 1.11), le lemme suivant découle de la proposition 3.9.

**Lemme 3.22** – *Tout quotient d'un automate de Pascal est un automate à groupe.*

On ajoute à $\mathcal{A}$ la lettre $g$ déjà utilisée précédemment ; son action est identique à celle de $10^{-1}$, c'est-à-dire que

$$s \xrightarrow{g} s' \quad \Longleftrightarrow \quad (s \cdot 1) \;=\; (s' \cdot 0) \;.$$

L'action de chaque lettre dans $\mathcal{A}$ est simulée par celle 0 et $g$ de la même manière que dans $\mathcal{P}_n^R$, comme l'exprime le lemme suivant.

**Lemme 3.23** – *Pour toute lettre $a \in [\![p]\!]$ et tout état $s$ de $\mathcal{A}$, $(s \cdot a) = (s \cdot (g^a 0))$.*

La proposition 3.24, ci-dessous, montre que les circuits induits par la lettre $g$ dans $\mathcal{A}$ permettent de calculer les paramètres $(n, R)$ de $\mathcal{P}_n^R$.

**Proposition 3.24** – *Soit un automate $\mathcal{A}$, quotient d'un automate de Pascal canonique $\mathcal{P}_n^R$.*

**a)** *Les circuits induits par la lettre $g$ dans $\mathcal{A}$ sont tous de longueur $n$.*

**b)** *Le mot $g^r$ est accepté par $\mathcal{A}$ si et seulement si $r \in R$.*

La démonstration de cette proposition nécessite les deux lemmes suivants qui donne des conditions suffisantes pour que deux états de $\mathcal{P}_n^R$ ne soient $\varphi$-équivalents.

**Lemme 3.25** – *Il n'existe pas $s \neq 0$ dans $\mathbb{Z}/n\mathbb{Z}$ tel que $\varphi((s,0)) = \varphi((0,0))$.*

Démonstration par l'absurde. Soit un tel $s > 0$ appartenant à $\mathbb{Z}/n\mathbb{Z}$. On suppose qu'il s'agit du plus petit.

L'équation (3.6) (définissant l'action de $g$ dans $\mathcal{P}_n^R$) appliquée à $(s', 0)$ donne :

$$\forall s' \in \mathbb{Z}/n\mathbb{Z} \qquad (s', 0) \xrightarrow[\mathcal{P}_n^R]{g} (s'+1, 0) \;,$$

ce qui implique que, puisqu'un morphisme d'automate respecte les transitions :

$$\forall h \in \mathbb{Z}/n\mathbb{Z} \qquad \varphi((s+h, 0)) \;=\; \varphi((h, 0)) \;.$$

Soit un entier $i < s$, en utilisant la formule précédente pour $h$ parcourant l'ensemble $\{i,\, (i+s),\, (i+2s), \ldots\}$, on obtient que

$$\varphi((i,0)) \;=\; \varphi((i+s,0)) \;=\; \varphi((i+2\,s,0)) \;=\; \cdots$$

ou, autrement dit, que

$$\forall i < s,\;\; \forall j \in \mathbb{N} \qquad \varphi((i,0)) \;=\; \varphi((i+j\,s,0)) \;.$$

Puisqu'un morphisme respecte le statut final/non-final des états, ceci implique que $(j \in R \iff (j+s) \in R)$ donc que $s$ est une période plus petite que $n$, et donc que $(n, R)$ n'est pas un paramètre canonique. Contradiction.

**Lemme 3.26** – *Soient deux éléments,* $(s,t)$ *et* $(s',t')$ *de* $G_n$*. Si* $s \neq s'$ *et* $\varphi((s,t)) = \varphi((s',t'))$*, alors* $t \neq t'$*.*

Démonstration par l'absurde. Supposons qu'il existe des éléments $s, s' \in \mathbb{Z}/n\mathbb{Z}$ et $t \in \mathbb{Z}/\psi\mathbb{Z}$ tels que $s \neq s'$ et $\varphi((s,t)) = \varphi((s',t))$.

Le mot $u = 0^{\psi-t}\, g^s$ étiquette les deux chemins de $\mathcal{P}_n^R$ :

$$(s,t) \xrightarrow{u} (0,0) \quad \text{et} \quad (s',t) \xrightarrow{u} (s'-s,0)\ .$$

L'hypothèse $\varphi((s,t)) = \varphi((s',t'))$ implique alors que $\varphi(s'-s,0) = \varphi(0,0)$ et l'hypothèse $s \neq s'$ contredit alors le lemme 3.25 précédent. □

Nous pouvons maintenant démontrer la proposition 3.24.

Démonstration de la proposition 3.24. **a)** On fixe $k \in \mathbb{Z}/\psi\mathbb{Z}$. Par définition, le $g$-circuit dans $\mathcal{P}_n^R$ contenant l'état $(0,k)$ est

$$(0,k) \xrightarrow[\mathcal{P}_n^R]{g} (p^k,k) \xrightarrow[\mathcal{P}_n^R]{g} (2p^k,k) \xrightarrow[\mathcal{P}_n^R]{g} \cdots \xrightarrow[\mathcal{P}_n^R]{g} ((n-1)p^k,k) \xrightarrow[\mathcal{P}_n^R]{g} (0,k)\ .$$

Le circuit respectifs dans $\mathcal{A}$ est :

$$\varphi((0,k)) \xrightarrow[\mathcal{A}]{g} \varphi((p^k,k)) \xrightarrow[\mathcal{A}]{g} \cdots \xrightarrow[\mathcal{A}]{g} \varphi((n-1)p^k,k)) \xrightarrow[\mathcal{A}]{g} \varphi((0,k))\ .$$

Ce circuit est simple (c'est–à-dire de longueur $n$) si et seulement si pour tout $s, s' \in \mathbb{Z}/n\mathbb{Z}$ distincts, $\varphi((s,k)) \neq \varphi((s',k))$, ce qui découle du lemme 3.26 précédent. Ceci conclut la preuve du point (**a**) car tout $g$-circuit $\mathcal{A}$ est nécessairement l'image par $\varphi$ d'un $g$-circuit de $\mathcal{P}_n^R$.

**b)** Le calcul du mot $g^r$ atteint un état final dans $\mathcal{P}_n^R$ si et seulement si $r \in R$, donc est accepté par tout quotient de $\mathcal{P}_n^R$ si et seulement si $r \in R$. □

Dans la suite est proposée une méthode pour caractériser le morphisme d'automate $\varphi : \mathcal{P}_n^R \to \mathcal{A}$. Il est entièrement déterminé par la classe de $\varphi$-équivalence de $(0,0)$ et en particulier par l'élément $(h,k)$ de cette classe tel que $k$ est strictement positif et minimal.

Cette classe de $\varphi$-équivalence peut être calculée à partir de $\mathcal{A}$. Elle est en effet composée des différentes intersections du 0-circuit et du $g$-circuit contenant l'état initial comme l'établit le lemme suivant.

**Lemme 3.27** – *Soit un élément* $(s,t) \in G_n$*. Le calcul du mot* $g^s\, 0^t$ *dans* $\mathcal{A}$ *atteint l'état initial si et seulement si* $\varphi((s,t)) = \varphi((0,0))$*.*

Démonstration. Les calculs du mot $g^s\, 0^t$ respectivement dans $\mathcal{P}_n^R$ (*cf.* équation (3.3)) et dans $\mathcal{A}$ sont :

$$(0,0) \xrightarrow[\mathcal{P}_n^R]{g^s} (s,0) \xrightarrow[\mathcal{P}_n^R]{0^t} (s,t)\,;$$

$$\varphi((0,0)) \xrightarrow[\mathcal{A}]{g^s} \varphi((s,0)) \xrightarrow[\mathcal{A}]{0^t} \varphi((s,t))\ .$$

Un morphisme d'automates envoie l'état initial sur l'état initial (équation (1.1a)), donc $\varphi((s,t))$ est l'état initial de $\mathcal{A}$. □

Soit $(h, k)$ un élément de $G_n$. On note $\gamma_{(h,k)}$ la permutation sur $G_n$ induite par la multiplication *à gauche* par $(h, k)$ (contrairement à $\tau_u$ qui correspond à la multiplication *à droite* par $(\pi_p(u), |u|)$).

$$\forall (s,t) \in G_n \qquad \gamma_{(h,k)}(s,t) \;=\; (h,k) \diamond (s,t) \;=\; (h + s\,p^k,\; k+t)\;. \tag{3.7}$$

Ce même élément $(h, k)$ définit également une permutation sur $\mathbb{Z}/p\mathbb{Z}$, noté $\sigma_{(h,k)}$ :

$$\forall s \in \mathbb{Z}/p\mathbb{Z} \qquad \sigma_{(h,k)}(s) \;=\; h + s\,p^k\;. \tag{3.8}$$

Dans la suite on considérera toujours les permutations $\gamma_{(h,k)}$ et $\sigma_{(h,k)}$ paramétrées par un élément particulier $(h, k)$, caractéristique du morphisme $\varphi$. On appelle (par abus de langage) *plus petit état $\varphi$-équivalent à* $(0, 0)$ l'état $(h, k)$ qui satisfait les deux conditions suivantes :

- $\varphi((h, k)) = \varphi((0, 0))$
- pour tout état $(s, t) \neq (0, 0)$ tel que $\varphi((s, t)) = \varphi((0, 0))$, alors $k < t$.

Cet élément est bien défini puisque, d'après le lemme 3.26, deux éléments distincts ayant la même seconde composante ne peuvent pas être $\varphi$-équivalents.

La permutation $\gamma_{(h,k)}$ engendre alors le noyau du morphisme d'automates $\varphi$. En fait, chaque classe d'équivalence est une orbite de $\gamma_{(h,k)}$, comme l'exprime le lemme suivant.

**Proposition 3.28** – *Chaque classe de $\varphi$-équivalence est une orbite de la permutation $\gamma_{(h,k)}$ (dans $G_n$) et $R$ est une union d'orbites de $\sigma_{(h,k)}$ (dans $\mathbb{Z}/n\mathbb{Z}$).*

Démonstration. Un morphisme d'automate commute avec la fonction de transition (*cf.* équation (1.1c)) donc avec $\diamond$. Puisque $\varphi((0, 0)) = \varphi((h, k))$, alors

$$\forall (s,t) \in G_n \qquad \varphi((0,0) \diamond (s,t)) \;=\; \varphi((h,k) \diamond (s,t))\;.$$

Autrement dit, $\varphi((s, t)) = \varphi(\gamma_{(h,k)}(s, t))$. Ceci démontre que chaque classe de $\varphi$-équivalence est stable par $\gamma_{(h,k)}$ donc que $R$ est une union d'orbites de $\sigma_{(h,k)}$.

L'assertion suivante résout le cas particulier de la classe d'équivalence de $(0, 0)$.

**Assertion 3.28.1** – *La classe de $\varphi$-équivalence contenant $(0, 0)$ est une orbite de $\gamma_{(h,k)}$.*

*Démonstration de l'assertion.* Soient $C$ la classe de $\varphi$-équivalence de $(0, 0)$ et $(s, t)$ un élément de $C$. On note $i$ l'entier tel que $\psi \leqslant (t + i\,k) < \psi + k$.

Puisque l'opération effectuée par $\gamma_{(h,k)}$ sur la seconde composante est simplement une addition par $k$ (modulo $\psi$), l'état $(s', t') = \gamma^i_{(h,k)}((s, t))$ vérifie $0 \leqslant t' < k$.

Or, nous avons vu plus tôt que les classes de $\varphi$-équivalence sont stables par $\gamma_{(h,k)}$, donc $(s', t')$ appartient à $C$. Puisque $(h, k)$ est le plus petit état $\varphi$-équivalent à $(0, 0)$, le seul état de $C$ dont la seconde composante est strictement inférieure à $k$ est $(0, 0)$, donc

$$(s', t') \;=\; \gamma^i_{(h,k)}((s, t)) \;=\; (0, 0)\;,$$

ce qui implique que $(s, t)$ est dans l'orbite de $(0, 0)$.

Traitons maintenant le cas général. Soit $(s, t)$ et $(s', t')$ deux états $\varphi$-équivalents. Il s'ensuit que les états $((s, t) \diamond (s, t)^{-1}) = (0, 0)$ et $((s', t') \diamond (s, t)^{-1})$

sont aussi $\varphi$-équivalents. D'après l'assertion précédente, il existe un entier $i$ tel que $\gamma_{(h,k)}^{i}((s',t') \diamond (s,t)^{-1}) = (0,0)$, ou en d'autres termes tel que

$$(h,k)^i \diamond (s',t') \diamond (s,t)^{-1} = (0,0) \ .$$

En multipliant (à droite) par $(s,t)$ les deux côtés de l'équation on obtient exactement $\gamma_{(h,k)}^{i}((s',t')) = (s,t)$. □

On définit maintenant l'automate

$$\mathcal{A}_{(h,k)} \ = \ \langle\, Q_{(h,k)},\, \{0,g\},\, \delta_{(h,k)},\, (0,0),\, F_{(h,k)} \,\rangle \tag{3.9a}$$

dont l'ensemble d'états est $Q_{(h,k)} = \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/k\mathbb{Z}$ (notez bien que le second membre de ce produit cartésien est $\mathbb{Z}/k\mathbb{Z}$ et non $\mathbb{Z}/\psi\mathbb{Z}$) ; dont l'ensemble des états finals est $F_{(h,k)} = \{(s,t) \mid s \in R \text{ et } t \in \mathbb{Z}/k\mathbb{Z}\}$ ; et dont la fonction de transition $\delta_{(h,k)}$ est définie par :

$$\forall (s,t) \in Q_{(h,k)} \qquad \begin{array}{rcl} (s,t)\cdot 0 & = & \begin{cases} (s,t+1) & \text{si } t < (k-1) \\ \gamma_{(h,k)}^{-1}((s,t+1)) \ = \ \left(\dfrac{s-h}{p^k},\, 0\right) & \text{si } t = (k-1) \end{cases} \\ (s,t)\cdot g & = & (s+p^t,\, t) \ . \end{array} \tag{3.9b}$$

Cet automate $\mathcal{A}_{(h,k)}$ est alors isomorphe à $\mathcal{A}$, comme énoncé par le théorème 3.31 dont la démonstration nécessite les deux prochains lemmes qui décrivent le rapport de $\mathcal{A}_{(h,k)}$ avec le morphisme $\varphi$.

**Lemme 3.29** – *Soient une lettre $x \in \{0,g\}$ et un élément $(s,t)$ de $\mathbb{Z}/n\mathbb{Z}\times\mathbb{Z}/\psi\mathbb{Z}$ donc un état de $\mathcal{A}_{(h,k)}$ et de $\mathcal{P}_n^R$. Les successeurs respectifs de $(s,t)$ dans $\mathcal{A}_{(h,k)}$ et dans $\mathcal{P}_n^R$ sont, en tant qu'états de $\mathcal{P}_n^R$, $\varphi$-équivalents.*

Démonstration. Tout état $(s,t)$ de $\mathcal{A}_{(h,k)}$ appartient à $\mathbb{Z}/n\mathbb{Z}\times\mathbb{Z}/k\mathbb{Z}$ donc est également un état de $\mathcal{P}_n^R$. La transition sortante de cet état $(s,t)$ par une lettre $x \in \{0,g\}$ est définie de la même façon dans $\mathcal{P}_n^R$ et pour $\mathcal{A}_{(h,k)}$ sauf dans le cas où $t = (k-1)$ et $g = 0$. Dans ce cas, les successeurs respectifs de $(s,k-1)$ par 0 dans $\mathcal{P}_n^R$ et $\mathcal{A}_{(h,k)}$ sont $\varphi$-équivalents car l'un est l'image de l'autre par $\gamma_{(h,k)}$. □

**Lemme 3.30** – *Chaque classe de $\varphi$-équivalence contient exactement un état de $Q_{(h,k)}$.*

Démonstration. Existence. On note $C$ une classe d'équivalence, et $(s,t)$ son plus petit élément (selon la seconde composante) qui est unique d'après le lemme 3.26. Si $t \geqslant k$ alors $\gamma_{(h,k)}^{-1}(s,t)$ est de la forme $(s',t-k)$ avec $(t-k) \geqslant 0$ ; ceci est impossible car $(s',\, t-k)$ un état de $C$ dont la seconde composante est plus petite que $(s,t)$, donc $t < k$.

Unicité. Par l'absurde. Soient $(s,t)$ et $(s',t')$ deux états distincts et $\varphi$-équivalents tels que $0 \leqslant t,\, t' < k$. D'après le lemme 3.26 $t$ et $t'$ sont différents ; on suppose sans perdre la généralité que $t < t'$ ce qui implique que $0 < t' - t < k$.

Or, l'état $(s',t') \diamond (s,t)^{-1}$ est d'une part $\varphi$-équivalent à $(0,0)$ et d'autre part de la forme $(s'',t'-t)$ pour un certain $s''$. Ceci contredit la définition de $(h,k)$ comme le plus petit état $\varphi$-équivalent à $(0,0)$. □

**Théorème 3.31** – *Soient un automate de Pascal canonique $\mathcal{P}_n^R$ et un automate $\mathcal{A}$ tel qu'il existe un morphisme d'automates $\varphi : \mathcal{P}_n^R \to \mathcal{A}$. Alors, $\mathcal{A}$ est isomorphe à $\mathcal{A}_{(h,k)}$, où $(h,k)$ est le plus petit état $\varphi$-équivalent à $(0,0)$.*

Démonstration. A chaque état $q$ de $\mathcal{A}$ est associé sa classe de $\varphi$-équivalence $\varphi^{-1}(q)$, on définit alors la fonction suivante :

$$\begin{array}{rccl} \xi : & Q_{\mathcal{A}} & \longrightarrow & Q_{(h,k)} \\ & q & \longmapsto & \text{l'unique état de } \varphi^{-1}(q) \text{ appartenant à } Q_{(h,k)} \ . \end{array}$$

Cette fonction est bien définie d'après le lemme 3.30 précédent. Elle est injective : les images inverses par $\varphi$ des états de $\mathcal{A}$ sont disjointes ; et surjective : tout état $(s,t)$ de $Q_{(h,k)}$ est l'image par $\xi$ de $\varphi((s,t))$.

Démontrons maintenant que $\xi$ est un morphisme d'automate $\mathcal{A} \to \mathcal{A}_{(h,k)}$. Dans la classe d'équivalence de l'état initial $i_{\mathcal{A}}$ de $\mathcal{A}$ se trouve $(0,0)$ qui appartient à $Q_{(h,k)}$, donc $\xi(i_{\mathcal{A}}) = (0,0)$, l'état initial de $\mathcal{A}_{(h,k)}$. Soit $f$ un état final de $\mathcal{A}$, chaque état de $\mathcal{P}_n^R$ dont l' image par $\varphi$ est égale à $f$ est donc un état final, en particulier celui appartenant à $Q_{(h,k)}$ ; le même raisonnement peut être tenu pour un état $f$ qui n'est pas final. La troisième condition découle du lemme 3.29. □

## Reconnaître un quotient d'un automate de Pascal

Soit un automate $\mathcal{A} = \langle Q, [\![p]\!], \delta, i, T \rangle$ fixé dans toute cette sous-partie. Nous allons décrire ici l'algorithme pour déterminer si $\mathcal{A}$ est le quotient d'un automate de Pascal canonique ; il se découpe en trois étapes successives :

- *la simplification*, qui modifie les transitions et l'alphabet de $\mathcal{A}$ pour donner un automate simplifié noté $\mathcal{A}'$ sur l'alphabet $\{0, g\}$ ;
- *l'analyse*, qui calcule les valeurs que doivent respectivement prendre les paramètres $n$, $R$, $(h,k)$ et $\psi$ pour que $\mathcal{A}'$ puisse être un quotient de $\mathcal{P'}_n^R$ ;
- et *la vérification* que $\mathcal{A}'$ est bel et bien le quotient de $\mathcal{P'}_n^R$ défini par $(h,k)$.

### Prérequis

Tout quotient d'un automate de Pascal doit être un automate à groupe, d'après le lemme 3.22. De plus, tout quotient d'un automate de Pascal doit accepter *par valeur*, c'est-à-dire que le successeur d'un état final (resp. non-final) par la lettre 0 doit être final (resp. non-final). Ces deux conditions peuvent être vérifiées par un simple parcours et seront supposées satisfaites dans le suite. De plus, on s'autorisera à prendre les transitions de $\mathcal{A}$ *à l'envers*[2] ; le calcul de celles-ci se fait par un simple parcours de l'automate.

### Étape 1 – Simplification

Soit un nouvel alphabet $B = \{0, g\}$. L'automate $\mathcal{A} = \langle Q, [\![p]\!], \delta, i, T \rangle$ sur l'alphabet $[\![p]\!]$ peut être transformé en l'automate $\mathcal{B} = \langle Q, A \cup B, \delta', i, T \rangle$, où $\delta'(s,a) =$

2. En fait, seules les transitions étiquetées par la lettre 1 seront prise à l'envers et une seule fois chacune.

$\delta(s,a)$ pour tout $a \in [\![p]\!]$ et $\delta'(s,g) = \delta(s, 1\,0^{-1})$ ; cette dernière équation est bien définie puisque $\mathcal{A}$ est un automate à groupe.

Pour que l'automate $\mathcal{A}$ puisse être un automate de Pascal, il est nécessaire, d'après le lemme 3.23, que l'équation suivante soit vérifiée :

$$\forall a \in [\![p]\!],\ \forall s \in Q \qquad s \cdot a \ = \ s \cdot (g^a 0) \quad \text{dans l'automate } \mathcal{B}.$$

Vérifier que cette équation est satisfaite nécessite de faire, pour chaque une lettre $a$ et chaque état $s$, c'est-à-dire pour chaque transitions de $\mathcal{A}$, un test que l'on dira *primitif*. Pour que la vérification globale soit linéaire (en le nombre de transitions de $\mathcal{A}$), il suffit donc que chaque test primitif soit en temps constant. Ceci peut être réalisé en conservant les résultats intermédiaires. En effet, calculer le point d'arrivée $s''$ de la transition étiquetée par $g^a$ partant d'un état donné $s$ se fait en temps constant si l'on connait le point d'arrivée $s'$ de celle étiquetée par $g^{(a-1)}$ (précédemment calculée) : $s''$ est le point d'arrivée de la transition étiquetée par $g$ partant de $s'$. En pratique, on enrichit successivement l'automate par toutes les transitions étiquetées par $g^2$, puis par celles étiquetées par $g^3$, et ainsi de suite jusqu'à celles étiquetées par $g^{p-1}$ ; le nombre de transitions ainsi ajoutées est égal à $(p-2) \times \mathsf{Card}(Q)$ donc inférieurs au nombre de transitions de $\mathcal{A}$.

On note $\mathcal{A}' = \langle\, Q, B, \delta', i, T \,\rangle$ l'automate qui résulte de la suppression dans $\mathcal{B}$ de toutes les transitions qui ne sont étiquetées ni par 0 ni par $g$.

### Étape 2 – Analyse

Dans toute l'étape 2, on suppose que $\mathcal{A}'$ est le quotient d'un automate de Pascal $\mathcal{P}_n^R$ pour calculer (entre autre) les paramètres $(n, R)$. Si ce n'est pas le cas, les paramètres calculés n'auront aucun sens et l'étape 3 de l'algorithme échouera.

Si $\mathcal{A}'$ est le quotient d'un automate de Pascal ${\mathcal{P}'}_n^R$, alors les paramètres $(n, R)$ peuvent être calculés à partir de $\mathcal{A}'$ comme indiqué par la proposition 3.24.

- On note $n$ la longueur de n'importe quel circuit de $g$.
- On note $R$ l'ensemble des nombres $r < n$ tels que $g^r$ est accepté par $\mathcal{A}'$.

Il reste à calculer le paramètre $(h, k)$ qui gouverne le quotient ; il s'agit de l'état (de ${\mathcal{P}'}_n^R$) qui a la plus petite seconde composante parmi ceux dont l'image par $\varphi$ est l'état initial de $\mathcal{A}'$. En appliquant le lemme 3.27, ce paramètre est fourni par le circuit mixte $i \xrightarrow{g^h 0^k} i$ tel que $k$ minimal (et $h < n$). Une méthode simple consiste à calculer les circuits de 0 et de $g$ respectifs qui contiennent l'état initial et de considérer leurs intersections.

### Étape 3 – Vérification

La dernière étape de l'algorithme consiste simplement à vérifier que $\mathcal{A}'$ est bel et bien isomorphe à $\mathcal{A}'_{(h,k)}$. Un façon simple pour cela est de parcourir $\mathcal{A}'$ en commençant par l'état initial et d'étiqueter ses états par $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/k\mathbb{Z}$ en suivant les règles de l'équation (3.9b).

- Si on doit étiqueter un même état par deux étiquettes différentes, $\mathcal{A}'$ est rejeté.
- Si après étiquetage, deux états ont la même étiquette, $\mathcal{A}'$ est rejeté.

- Si après étiquetage, un état est final mais que sa première composante n'est pas dans $R$ ou inversement si un état n'est pas final mais que sa première composante est dans $R$, $\mathcal{A}'$ est rejeté.

- Sinon, $\mathcal{A}$ est accepté.

## Exécution sur un exemple

On donne ici un exemple de l'algorithme décrit dans la sous-section précédente. L'automate considéré ci-après est originalement sur l'alphabet $[\![3]\!]$ et reconnaît un ensemble d'entiers écrits en base 3 ; pour des raisons de clarté, on ne considère ici que l'automate simplifié $\mathcal{A}'_1$ sur lequel *l'étape* 1 *a déjà été effectuée*.

## Prérequis et simplification

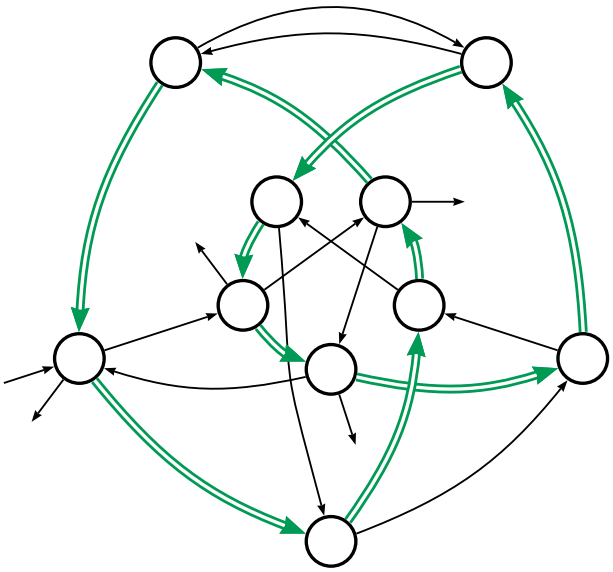


Figure 3 – l'automate simplifié $\mathcal{A}'_1$

## Analyse

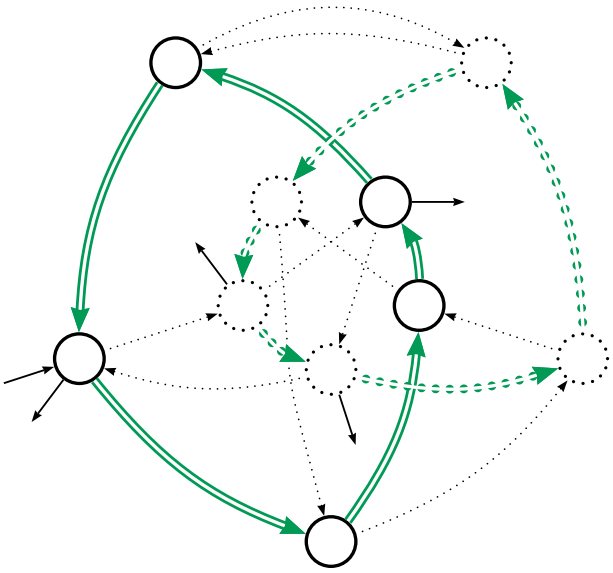


Figure 4 – Les circuits de $g$ sont tous de longueur 5 donc la période est $n = 5$. De plus, sur le circuit de $g$ partant de l'état initial, on atteint un état final par $g^0$ et $g^3$, donc l'ensemble de restes est $R = \{0, 3\}$.

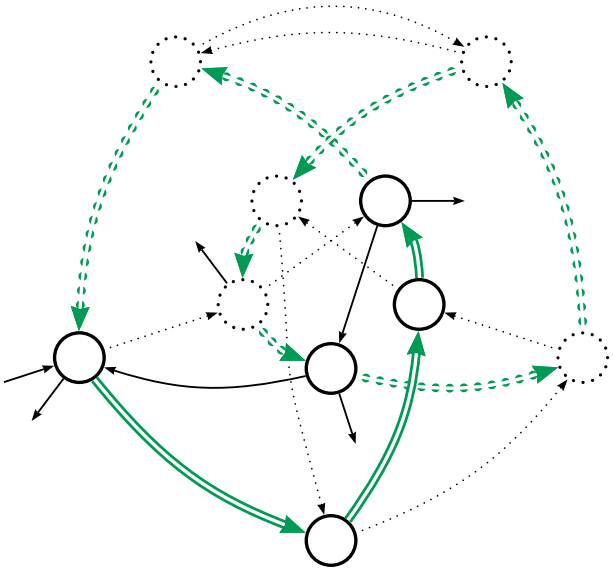


Figure 5 – Le plus petit (et, dans ce cas particulier, le seul) circuit mixte est ici mis en évidence ; il est étiqueté par $g^3\,0^2$ donc le paramètre du quotient est $(h, k) = (3, 2)$.

| | | | |
|---|---|---|---|
| Base | $p$ | = | 3 |
| Période | $n$ | = | 5 |
| Ensemble de restes | $R$ | = | {0,3} |
| Paramètre du quotient | $(h, k)$ | = | $(3, 2)$ |
| Ordre de $p$ dans $\mathbb{Z}/n\mathbb{Z}$ | $\psi$ | = | 4 |

Figure 6 – Tableau récapitulatif des paramètres

## Vérification

La figure 7 donne les règles pour étiqueter $\mathcal{A}'_1$ ; sur les figures, les transitions vérifiées sont effacées au fur et à mesure.

| | | | | |
|---|---|---|---|---|
| a) | $(s,0)\cdot 0$ | $=$ | $(s,1)$ | |
| b) | $(s,1)\cdot 0$ | $=$ | $(4\,s-2,0)$ | $=(\frac{s-h}{p^k},0)$ |
| c) | $(s,0)\cdot g$ | $=$ | $(s+1,0)$ | $=(s+p^0,0)$ |
| d) | $(s,1)\cdot g$ | $=$ | $(s+3,1)$ | $=(s+p^1,1)$ |

Figure 7 – Tableau des transitions de $\mathcal{A}_{(h,k)}$

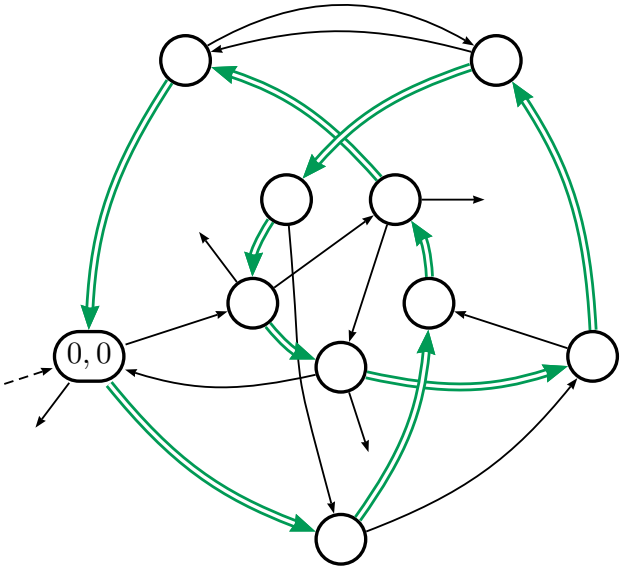



Figure 8 – L'état initial est $(0,0)$

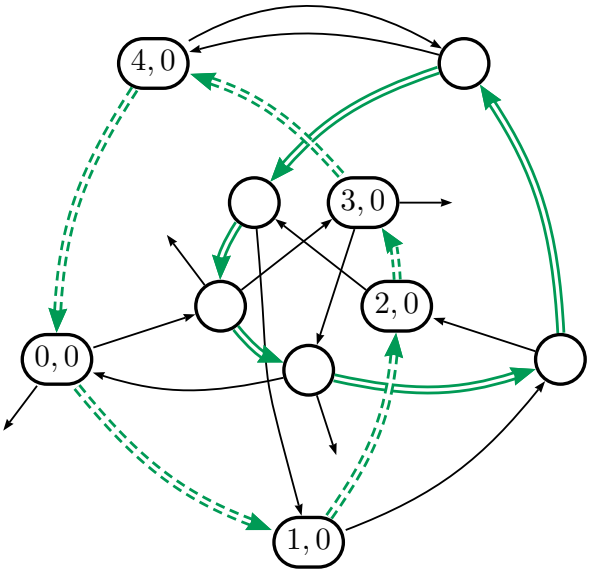



Figure 9 – Exécution de la règle 7c : $(s,0) \xrightarrow{g} (s+1,0)$

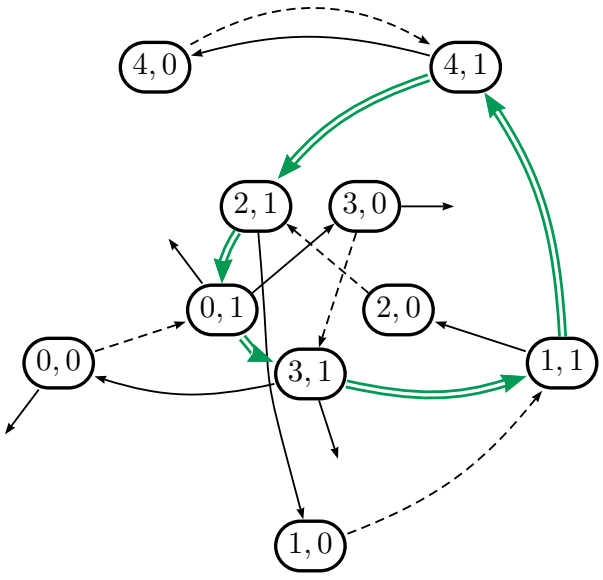



Figure 10 – Exécution de la règle 7a : $(s,0) \xrightarrow{0} (s,1)$

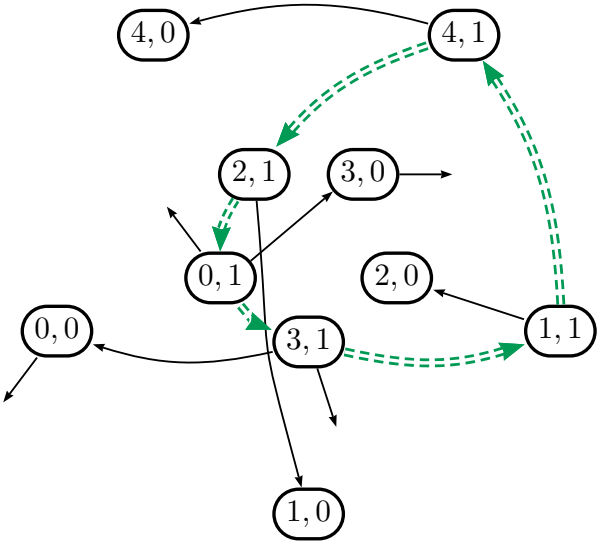



Figure 11 – Exécution de la règle 7d : $(s,1) \xrightarrow{g} (s+3,1)$

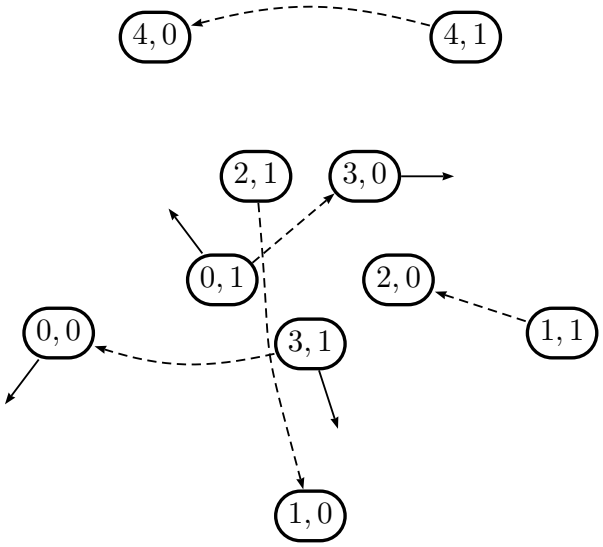



Figure 12 – Exécution de la règle 7b : $(s, 1) \xrightarrow{0} (4\,s - 2, 0)$

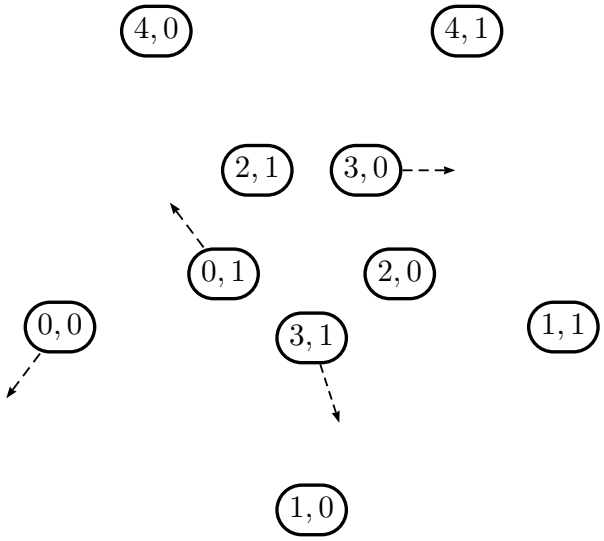



Figure 13 – Vérification qu'un état est final si et seulement si sa première composante appartient à $R = \{0, 3\}$.

# Le critère (UP) et sa décidabilité

Cette partie introduit des conditions structurelles sur les composantes fortement connexes (CFC) d'automates ; elles sont regroupées au sein du *critère (UP)*, pour ultimement périodique. Il sera prouvé plus loin qu'un automate minimal satisfait le critère (UP) si et seulement s'il accepte un ensemble ultimement périodique de nombres. Ce critère est énoncé plus loin, et nécessite deux définitions préalables.

**Définition 3.32** – *Soit $\mathcal{A} = \langle Q, A, \delta, i, F \rangle$ un automate et $\sigma$ la relation de forte connexité sur $Q$ : $s\,\sigma\,s'$ s'il existe deux mots $u, v \in A^*$ tels que $(s \cdot u) = s'$ et $(s' \cdot v) = s$.*

*On note $\gamma_{\mathcal{A}}$ l'application surjective de $Q$ vers $Q/\sigma$ et $\mathcal{C}_{\mathcal{A}}$ la* condensation *de $\mathcal{A}$ c'est-à-dire est le graphe étiqueté orienté*

$$\mathcal{C}_{\mathcal{A}} \;=\; (Q/\sigma, A, E)$$

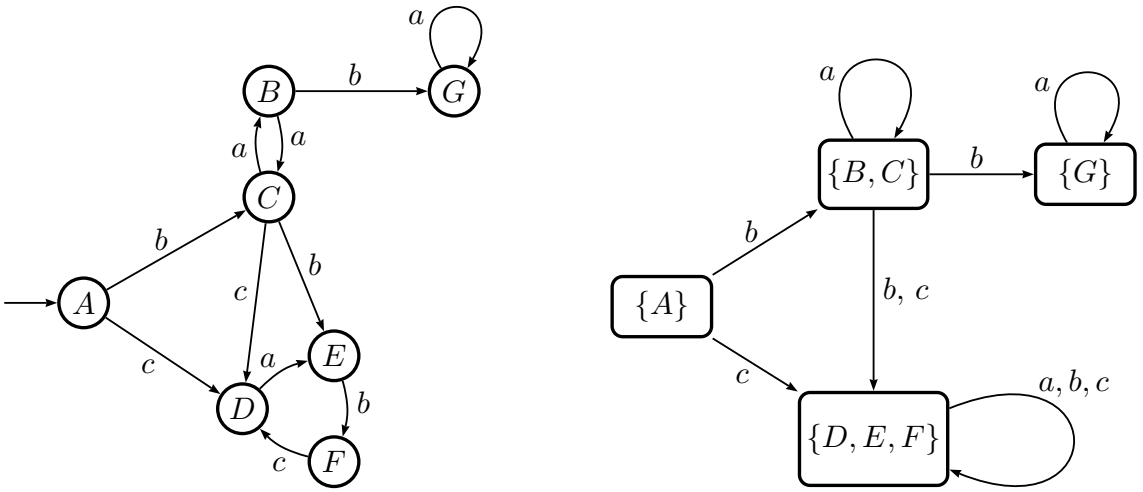
*où le couple $(x, a, y)$ est un arc de $E$ si il existe deux états $s$ et $s'$ dans $Q$ tels que $\gamma_{\mathcal{A}}(s) = x$, $\gamma_{\mathcal{A}}(s') = y$ et $s \xrightarrow[\mathcal{A}]{a} s'$.*

*De plus, on appelle* composante fortement connexe (CFC) d'un état *$s$, la classe de $\sigma$-équivalence de $s$. Si $s$ est seul dans sa classe d'équivalence et que l'unique mot $u$ tel que $s \xrightarrow{u} s$ est $u = \varepsilon$, on dit que sa CFC est triviale.*

Par exemple, la figure 14 montre la condensation d'un automate $\mathcal{A}_2$ ; $\{A\}$ est la seule CFC triviale

**Définition 3.33** – *On dit qu'une CFC $C$ d'un automate $\mathcal{A}$ se plonge dans une autre CFC $D$ si il existe une fonction $f : C \to D$ tel que pour tout $s$ de $C$ et $a$ dans $A$, alors $(s \cdot a)$ existe si et seulement si $f(s) \cdot a$ existe et, dans ce cas,*

- *si $(s \cdot a)$ appartient à $C$ alors $f(s \cdot a) = f(s) \cdot a$ ;*




(a) Un automate $\mathcal{A}_2$ (b) Sa condensation $\mathcal{C}_{\mathcal{A}_2}$

Figure 14 – Exemple de condensation

- *sinon,* $(s \cdot a)$ *appartient à* $D$ *et* $(s \cdot a) = f(s) \cdot a$.

**Définition 3.34** – *Soit* $\mathcal{A}$ *un automate déterministe,* $\mathcal{C}_A$ *sa condensation, et* $\gamma_{\mathcal{A}}$ *la fonction envoyant chaque état de* $\mathcal{A}$ *vers la CFC de* $\mathcal{C}_A$ *qui le contient. On dit que* $\mathcal{A}$ *satisfait le critère (UP) (ou que* $\mathcal{A}$ *est un UP-automate) s'il satisfait les cinq conditions suivantes.*

**UP-0** *Le successeur par la lettre* 0 *d'un état final existe et est final.*
*Le successeur par la lettre* 0 *d'un état non-final est, s'il existe, non-final.*

**UP-1** *Toute CFC de* $\mathcal{A}$ *qui contient une transition interne par une lettre différente de* 0 *est envoyée par* $\gamma_{\mathcal{A}}$ *sur une feuille de* $\mathcal{C}_{\mathcal{A}}$.
*Une telle CFC est dite* de type 1.

**UP-2** *Toute CFC non-triviale de* $\mathcal{A}$ *qui n'est pas de type 1 :*
**a)** *est un simple circuit étiqueté par 0 (aussi appelé* circuit de 0*) et*
**b)** *a pour image par* $\gamma_{\mathcal{A}}$ *un sommet de* $\mathcal{C}_{\mathcal{A}}$ *qui a au plus un successeur dans* $\mathcal{C}_{\mathcal{A}}$*, auquel cas ce successeur est une CFC de type 1.*
*Une telle CFC est dite* de type 2.

**UP-3** *Toute CFC de type 1 est le quotient d'un automate de Pascal.*

**UP-4** *Toute CFC de type 2 se plonge dans la CFC de type 1 qui lui est associée par (UP-2) si elle existe.*

**Remarque 3.35** – *La condition (UP-0) n'est pas spécifique, il s'agit en fait d'une précondition (d'où son numéro '0') qui assure qu'un UP-automate accepte par valeur (voir le début de la section 2.3).*

La figure 15 est un schéma récapitulatif des conditions (UP-1) à (UP-4). La condensation d'un UP-automate contient deux niveaux de CFCs non-triviales, représentés par des carrés et les ovales. Les carrés sont les CFCs de type 1 donc des

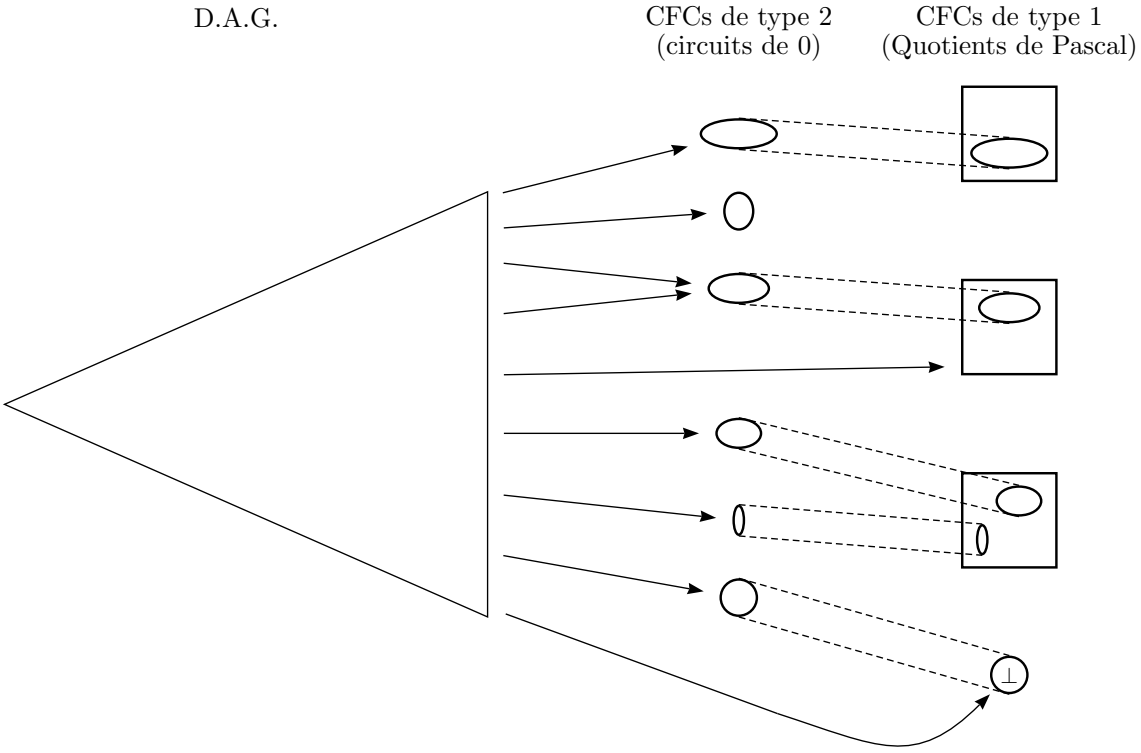



Figure 15 – Représentation schématique du critère (UP)

feuilles de $\mathcal{C}_{\mathcal{A}}$ (d'après (UP-1)) et des quotients d'automates de Pascal (d'après (UP-3)). Les ovales sont les CFCs de type 2, chacune d'entre elles se plonge dans son unique successeur s'il existe (ce qui est représenté par des traits tiretés).

**Exemple 3.36** – *La figure 16 représente un automate $\mathcal{A}_3$ qui satisfait le critère (UP). Les sous-automates encadrés sont, de haut en bas, les minimisations des automates de Pascal $\mathcal{P}_3^{\{1,2\}}$ et $\mathcal{P}_5^{\{1,2,3,4\}}$. Les trois autres CFC non-triviales, $\{B_2, C_2\}$, $\{D_2\}$ et $\{111\}$, sont des circuits de 0. La troisième n'a dans $\mathcal{C}_{\mathcal{A}_3}$ aucun successeurs, alors que chacune des deux premières en a exactement un dans lequel elle se plonge : $\{B_2, C_2\}$ se plonge dans $\{A, B, C\}$ par la fonction injective $B_2 \mapsto B$ et $C_2 \mapsto C$ ; $\{D_2\}$ se plonge dans $\{D, E, F, G, H\}$ de manière analogue.*

Montrons que le critère (UP) est décidable en temps linéaire pour les automates minimaux.

**Théorème 3.37** – *Soit $\mathcal{A}$ un automate minimal dont le nombre de transitions est $m$. Il peut être décidé en temps $O(m)$ si $\mathcal{A}$ satisfait le critère (UP) ou non.*

Démonstration. Soit $\mathcal{A}$ un automate minimal. Un simple parcours de $\mathcal{A}$ permet de vérifier s'il satisfait la condition (UP-0). Les autres conditions du critère (UP) se vérifient sur les CFCs de $\mathcal{A}$, plus particulièrement sur la condensation de $\mathcal{A}$. Or

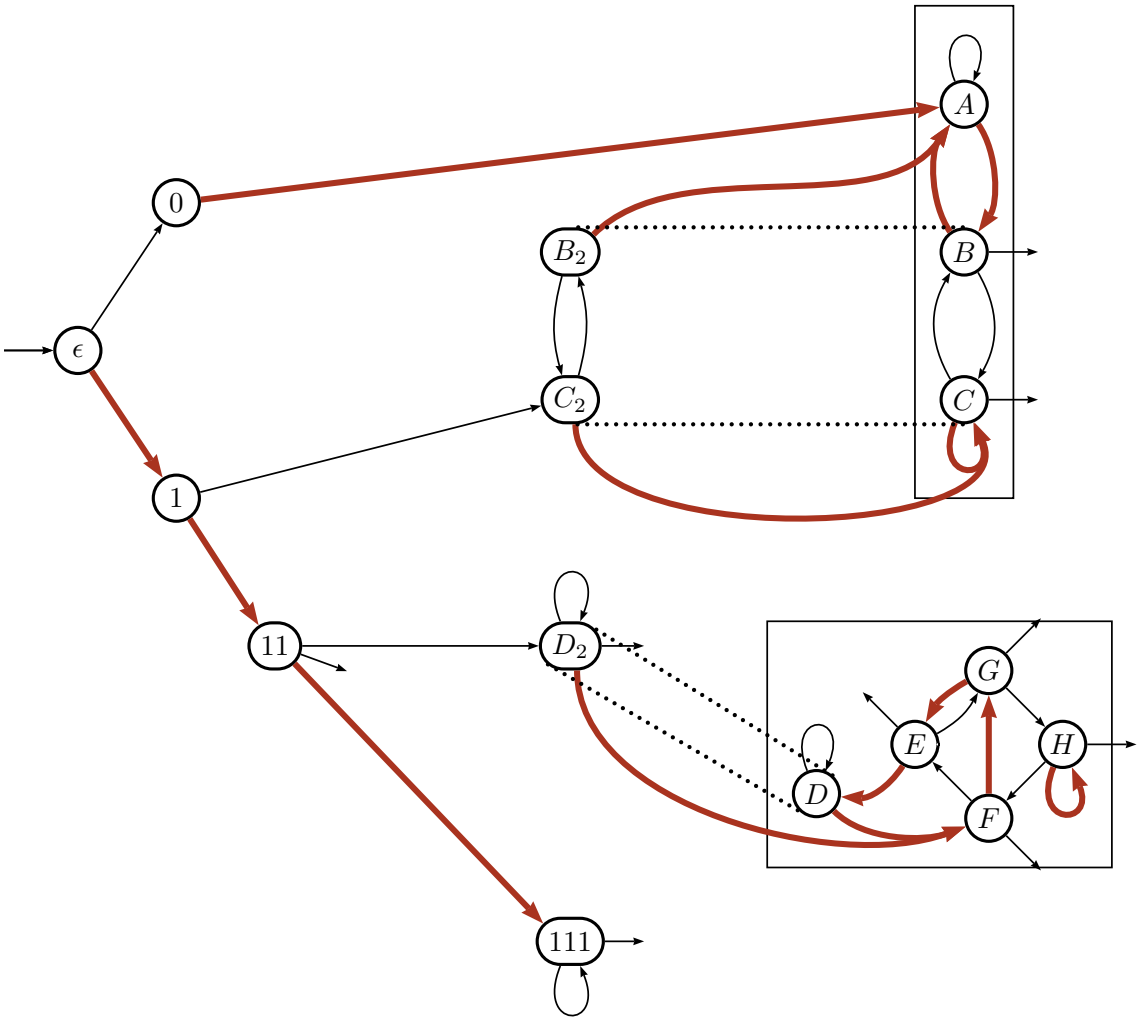



Figure 16 – L'UP-automate $\mathcal{A}_3$

celle-ci peut être calculé en temps $\mathsf{O}(m)$ par les algorithmes classiques de Tarjan (*cf.* [77]) ou de Kosaraju (*cf.* [1, 28]).

Vérifier (UP-1) et (UP-2) s'effectue au cours d'un simple parcours de chaque CFC. Vérifier (UP-3) consiste à vérifier que certaines CFC sont des quotients d'automates de Pascal ; puisque $\mathcal{A}$ est minimal cela est équivalent à vérifier que ce sont des quotients d'automates de Pascal canoniques, ce qui s'effectue en temps linéaire d'après le théorème 3.1, résultat principal de la section précédente. Enfin, vérifier (UP-4) en temps linéaire peut se faire comme expliqué ci-dessous. La fonction de plongement envoie un état $s$ d'une CFC $C$ de type 2 l'unique état $x$ tel que

$$s \xrightarrow{1} s' \quad \text{et} \quad x \xrightarrow{1} s' \quad \text{et} \quad x \notin C \ .$$

Si $s'$ existe (s'il n'existe pas, il n'y a pas de plongement), il est nécessairement dans

[77] Robert E. Tarjan, 1972, *Depth-First Search and Linear Graph Algorithms.*

[1] Alfred V. Aho, John E. Hopcroft et Jeffrey D. Ullman, 1983, *Data Structures and Algorithms.*

[28] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest et Clifford Stein, 2002, *Introduction à l'algorithmique (2ème ed.).*

l'unique CFC $D$ (de type 2) qui succède à $C$ ; or $D$ est un quotient d'un automate de Pascal, donc un automate à groupe. Il n'y a donc qu'un seul état de $D$ dont le successeur par 1 est $s'$.

Pour ne pas parcourir plusieurs fois la CFC $D$, il suffit de faire un parcours préliminaire durant lequel on note chaque prédécesseur par 1 (ce qui, normalement, a déjà été fait pendant la vérification de (UP-3)). Récupérer ultérieurement le prédécesseur se fera donc en temps constant. □

# Construction d'un UP-automate acceptant un ensemble ultimement périodique arbitraire

On rappelle que la base est notée $p$ et est fixée dans tous le chapitre. Le but de cette section est d'établir le théorème suivant.

**Théorème 3.38** – *Soient une période $n \in \mathbb{N}$, une prépériode $m \in \mathbb{N}$, un ensemble $R \subseteq \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$ et un ensemble $I \subseteq \{0, 1, \ldots, m-1\}$. Il existe un automate complet qui accepte l'ensemble d'entiers $(I \cup E^R_{n,m})$ et qui satisfait le critère (UP).*

Cet automate sera construit tout au long de cette section 3.3 en prenant des cas particuliers de plus en plus généraux. Notez que le cas particulier où la période est première avec la base est traité dans la section 3.1 : un automate de Pascal est en effet un automate complet satisfaisant le critère (UP).

## UP-automate acceptant $E^R_d$ où $d \mid p^j$

Soient deux entiers $d$ et $j$ tels que $d \mid p^j$ et un ensemble $R \subseteq \mathbb{Z}/d\mathbb{Z}$ de restes modulo $d$. On rappelle que si un automate $\mathcal{A}$ accepte par valeur, on note $N(\mathcal{A})$ l'ensemble d'entiers accepté par $\mathcal{A}$. On construit ici un UP-automate $\mathcal{A}^R_d$ qui vérifie donc $N(\mathcal{A}^R_d) = E^R_d$ (proposition 3.41).

On utilise une généralisation de la méthode usuelle pour déterminer si un nombre écrit en base 10 est un multiple de 5 : vérifier si son chiffre des unités est un 0 ou un 5. Puisque nous lisons le chiffre de poids faible en premier, il suffit de tester si le *premier* chiffre est un 0 ou un 5. Le lemme suivant donne le cas général.

**Lemme 3.39** – *Pour tout reste $r \in \mathbb{Z}/d\mathbb{Z}$, l'ensemble*

$$F_r \;=\; \left\{\, u \mid |u| \leqslant j \quad \textit{et} \quad \pi_p(u) \equiv r \;[d] \,\right\}$$

*satisfait les deux propriétés suivantes :*

**a**) $\forall v \in [\![p]\!]^{<j} \qquad \pi_p(v) \equiv r \;[d] \iff v \in F_r$

**b**) $\forall v \in [\![p]\!]^{\geqslant j} \qquad \pi_p(v) \equiv r \;[d] \iff$ *le préfixe de longueur $j$ de $v$ est dans $F_r$*

Démonstration. Le point (**a**) découle directement de la définition de $F_r$.

**b)** Soit un mot $v \in [\![p]\!]^{\geqslant j}$ dont on note $u \in [\![p]\!]^j$ son préfixe de longueur $j$. D'après l'équation (2.2c), $\pi_p(v)$ et $\pi_p(u)$ sont congrus modulo $p^j$ donc modulo $d$ (puisque par hypothèse $d \mid p^j$). Donc $\pi_p(v) \equiv r\ [p^j]$ si et seulement si $u \in F_r$. □

On note $F_R$ l'ensemble $F_R = \cup_{r\in R} F_r$ et on définit l'automate $\mathcal{A}_d^R$ de telle sorte qu'il accepte un mot $u$ s'il est dans $F_R$ ou si son préfixe (éventuel) de longueur $j$ est dans $F_R$.

$$\mathcal{A}_d^R \;=\; \langle\, [\![p]\!]^{\leqslant j},\ [\![p]\!],\ \delta, \varepsilon, F_R \,\rangle$$

Les états de $\mathcal{A}_d^R$ sont les mots sur $[\![p]\!]$ de longueurs inférieures ou égales à $j$ ; un état est final s'il appartient à $F_R$ ; la fonction de transition $\delta$ est définie par :

$$\forall u \in [\![p]\!]^{\leqslant j},\ \ \forall a \in [\![p]\!] \qquad \begin{array}{ll} u \xrightarrow{a} u\,a & \text{si } |u| < j \\ u \xrightarrow{a} u & \text{sinon.} \end{array}$$

**Exemple 3.40** – *On considère le cas où $p = 2$, $n = 4$ et $R = \{1\}$. Dans ce cas, $F_R$ est l'ensemble formé des mots 1 et 10. L'ensemble $E_4^{\{1\}}$ est accepté par l'automate $\mathcal{A}_4^{\{1\}}$ représenté par la figure 17.*

*Si de plus on ajoutait 2 à $R$, alors $F_R$ contiendrait le mot supplémentaire 01 et l'état homonyme serait final sur la figure.*

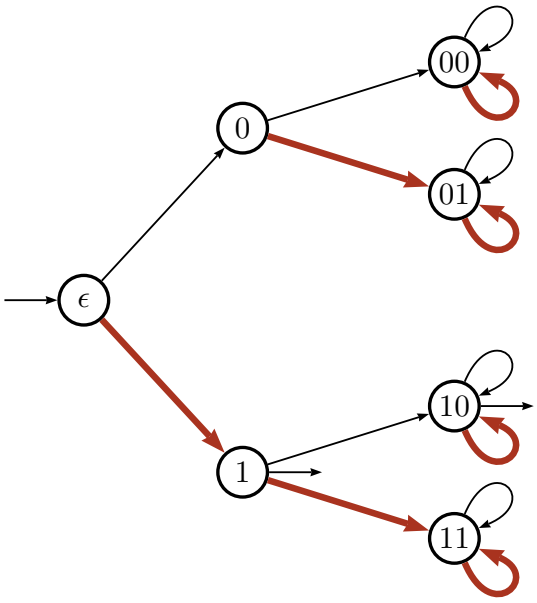



Figure 17 – L'automate $\mathcal{A}_4^{\{1\}}$ acceptant les entiers congrus à 1 modulo 4 (en base 2)

**Proposition 3.41** – *Soient deux entiers $d$ et $j$ tels que $d \mid p^j$ et un ensemble $R \subseteq \mathbb{Z}/d\mathbb{Z}$ de restes modulo $d$. Alors, l'automate $\mathcal{A}_d^R$ est complet,*

**a)** *accepte l'ensemble d'entiers $E_d^R$ et*

**b)** *satisfait le critère (UP).*

Démonstration. Le fait que $\mathcal{A}_d^R$ est complet découle directement de la définition. et le point (**a**) du lemme 3.39.

(**b**) Chaque CFC non-triviale de $\mathcal{A}_d^R$ est réduite à un état stable par toutes les lettres de $[\![p]\!]$, donc en particulier par une lettre non nulle. Ce sont donc des CFC de type 1 et elles sont bien des feuilles de la condensation $\mathcal{C}_{(\mathcal{A}_d^R)}$ de $\mathcal{A}_d^R$ ; celui-ci satisfait donc (UP-1). Ces CFC acceptent, soit tous les mots, soit aucun ; il s'agit donc des automates de Pascal triviaux (c'est-à-dire de période 1) $\mathcal{P}_1^{\emptyset}$ et $\mathcal{P}_1^{\{1\}}$ ; il s'ensuit que $\mathcal{A}_d^R$ satisfait (UP-3). Puisque $\mathcal{A}_d^R$ n'a aucune CFC de type 2, il satisfait trivialement (UP-2) et (UP-4). Enfin, il découle du point (**a**) que $\mathcal{A}_d^R$ accepte les mots par valeur, donc qu'il satisfait (UP-0). □

## UP-automate acceptant $E_n^R$

Soient une période $n \in \mathbb{N}$ et un ensemble $R \in \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$. Le but de cette sous-section est de définir un UP-automate $\mathcal{A}_n^R$ qui accepte par valeur $E_n^R$. Il existe trois (uniques) entiers $k$, $d$ et $j$ tels que

- $n = k\,d$,
- $k$ est premier avec la base $p$,
- $d$ divise $p^j$ mais pas $p^{(j-1)}$.

Il découle des deux dernières conditions que $k$ et $d$ sont premiers entre eux. On utilise dans la suite une version spécialisée du théorème *des restes chinois* donnée-ci dessous.

**Théorème 3.42** (dit des restes chinois) – *Soient deux entiers $k$ et $d$ premiers entre eux. Pour tous entiers $r_k$ et $r_d$, il existe un unique entier $r < k\,d$ tel que $r \equiv r_k\ [k]$ et $r \equiv r_d\ [d]$.*

*De plus, tout entier qui est à la fois congru à $r_k$ modulo $k$ et $r_d$ modulo $d$ est également congru à $r$ modulo $k\,d$.*

Supposons pour l'instant que $R$ est un singleton $\{r\}$ (avec donc $r \in \mathbb{Z}/n\mathbb{Z}$) et on note $r_d = (r \,\%\, d)$ et $r_k = (r \,\%\, k)$. Il découle alors du théorème 3.42 que

$$\forall i \in \mathbb{N} \qquad i \;\equiv\; r\ [n] \quad \Longleftrightarrow \quad \begin{cases} i \;\equiv\; r_k & [k] \\ i \;\equiv\; r_d & [d] \end{cases} . \tag{3.10}$$

Soit $i$ un entier. L'automate de Pascal $\mathcal{P}_k^{\{r_k\}}$ accepte (un mot de valeur) $i$ si et seulement si $i \equiv r_k\ [k]$ et l'automate $\mathcal{A}_d^{\{r_d\}}$ (défini précédemment) accepte $i$ si $i \equiv r_d\ [d]$. Le produit $\mathcal{A}_d^{\{r_d\}} \times \mathcal{P}_k^{\{r_k\}}$ de ces deux automates accepte donc $i$ si et seulement s'il satisfait les deux conditions du membre droit de l'équation (3.10) donc s'il est congru à $r$ modulo $n$.

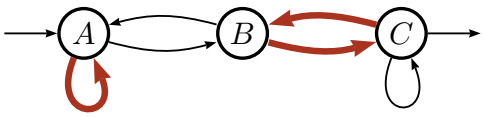



Figure 18 – La minimisation de l'automate de Pascal $\mathcal{P}_3^{\{2\}}$

**Exemple 3.43** – *Soit $n = 12$ et $R = \{r\} = \{5\}$ ; si bien que $k = 3$, $d = 4$, $r_k = (5 \% 3) = 2$ et $r_d = (5 \% 4) = 1$. La figure 18 représente l'automate de Pascal $\mathcal{P}_3^{\{2\}}$, (minimisé pour des questions de clarté) et la figure 19 représente le produit $\mathcal{A}_4^{\{1\}} \times \mathcal{P}_3^{\{2\}}$ ; voir aussi la figure 17, page 68 montrant l'automate $\mathcal{A}_4^{\{1\}}$.*

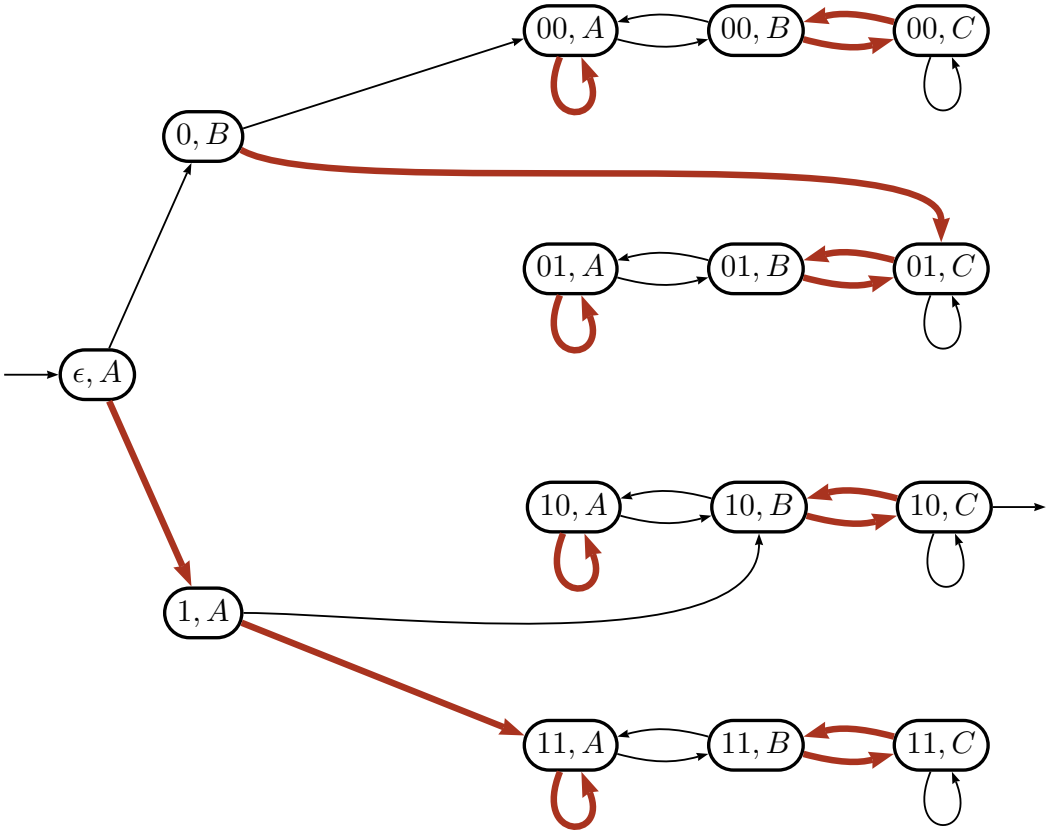



Figure 19 – Le produit $\mathcal{A}_4^{\{1\}} \times \mathcal{P}_3^{\{2\}}$, acceptant les entiers congrus à 5 modulo 12.

On ne suppose plus dorénavant que $R$ est un singleton ; on note ses éléments $R = \{ r_0, r_1, \cdots r_{q-1} \}$. On applique le théorème 3.42 comme précédemment sur chacun des $r_x$, de telle sorte qu'un entier $i$ soit congru à $r_x$ modulo $n$ si et seulement si il est congru à $r_{x,d}$ modulo $d$ et à $r_{x,k}$ modulo $k$.

Un entier $i$ est donc congru à un certain $r_x \in R$ modulo $n$ si et seulement il satisfait la formule logique suivante :

$$\bigvee_{x=0}^{q-1} (i \% d = r_{x,d}) \wedge (i \% k = r_{x,k}) \tag{3.11}$$

On va maintenant transformer cette formule en factorisant par les premiers membres quand ils sont égaux, c'est-à-dire que si $r_{z,d} = r_{x,d}$ pour certain $x, z$, alors on factorise l'expression sur le modèle suivant :

$$\begin{aligned} &((i \% d = r_{x,d}) \ \wedge \ (i \% k = r_{x,k})) \ \ \vee \ \ ((i \% d = r_{z,d}) \ \wedge \ (i \% k = r_{z,k})) \\ &= \ (i \% d = r_{x,d}) \ \ \wedge \ \ ((i \% k = r_{x,k}) \ \vee \ (i \% k = r_{z,k})) \end{aligned}$$

Pour tout $y \in [\![d]\!]$, on note $K_y$ l'ensemble contenant un certain $r_{x,k}$ si $r_{x,d} = y$ :

$$K_y = \{\, r_{x,k} \mid r_{x,d} = y \,\} \ .$$

Un entier $i$ est donc congru à un certain $r_x \in R$ modulo $n$ si et seulement si :

$$\bigvee_{y=0}^{d-1} (i \,\%\, d = y) \wedge (i \,\%\, k \in K_y) \ . \tag{3.12}$$

Le lemme suivant en découle directement.

**Lemme 3.44** – *Soit un entier $i$ dont on note $y = i \,\%\, d$ la classe de congruence modulo $d$. L'équivalence suivante est vérifiée :*

$$i \,\%\, n \in R \quad \Longleftrightarrow \quad i \,\%\, k \in K_y \ .$$

Soit un mot $u$ dont on note la valeur $i = \pi_p(u)$. La classe de congruence modulo $d$ de $i$ est donnée par les $j$ premières lettres de $u$. L'automate acceptant un entier $i$ si et seulement si il satisfait l'équation (3.12) consiste donc en un arbre de profondeur $j$ dont chaque branche (de longueur $j$) qui a pour valeur $y$ arrive dans l'automate de Pascal $\mathcal{P}_k^{K_y}$ (*au bon endroit*).

**Exemple 3.45** – *On considère le cas où $p = 3$, $n = 18$ et $R = \{0, 2, 4, 5, 9\}$ ; donc $d = 9 = 3^2$, $k = 2$ et $j = 2$. Les tableaux suivants récapitulent les différentes autres variables mentionnées précédemment.*

| $r_x$ | $r_{x,d}$ | $r_{x,k}$ |
|---|---|---|
| 0 | 0 | 0 |
| 2 | 2 | 2 |
| 4 | 4 | 0 |
| 5 | 5 | 1 |
| 9 | 0 | 1 |

| $y$ | $K_y$ |
|---|---|
| 0 | $\{0, 1\}$ |
| 1 | $\emptyset$ |
| 2 | $\{2\}$ |
| 3 | $\emptyset$ |
| 4 | $\{0\}$ |

| $y$ | $K_y$ |
|---|---|
| 5 | $\{1\}$ |
| 6 | $\emptyset$ |
| 7 | $\emptyset$ |
| 8 | $\emptyset$ |

*La figure 20 présente un schéma de l'automate acceptant les entiers en base 3 dont le reste modulo 18 appartient à $\{0, 2, 4, 5, 9\}$.*

- *La branche la plus à gauche est étiquetée par* 00 *dont la valeur en base 3 est* 0 *; elle atteint bel et bien un(e copie de l')automate de Pascal $\mathcal{P}_k^{K_0} = \mathcal{P}_2^{\{0,1\}}$. Cette branche accepte donc tous les mots qui commencent par* 00*, c'est-à-dire tous les mots dont les valeurs respectives sont divisibles par* 9 *(donc congrues à* 0 *ou* 9 *modulo* 18*).*
- *La branche la plus à droite est étiquetée par* 22*, dont la valeur en base 3 est* 8 *et atteint $\mathcal{P}_k^{K_8} = \mathcal{P}_2^{\emptyset}$, c'est-à-dire un automate qui n'accepte aucun mot. En effet, les mots qui commencent par* 22 *ont des valeurs congrues à* 2 *modulo* 9 *(donc à* 8 *ou* 17 *modulo* 18*) et ne doivent pas être acceptés.*
- *La branche verticale (au milieu) accepte les mots $u$ qui commencent par 11 (donc tels que $\pi_p(u) \equiv 4$ [9]) et qui sont acceptés par $\mathcal{P}_2^{\{0\}}$ (donc $\pi_p(u) \equiv 1$ [2]), c'est-à-dire les mots dont les valeurs respectives sont congrues à* 4 *modulo* 18*.*

*Les autres branches sont construites de manière analogue.*

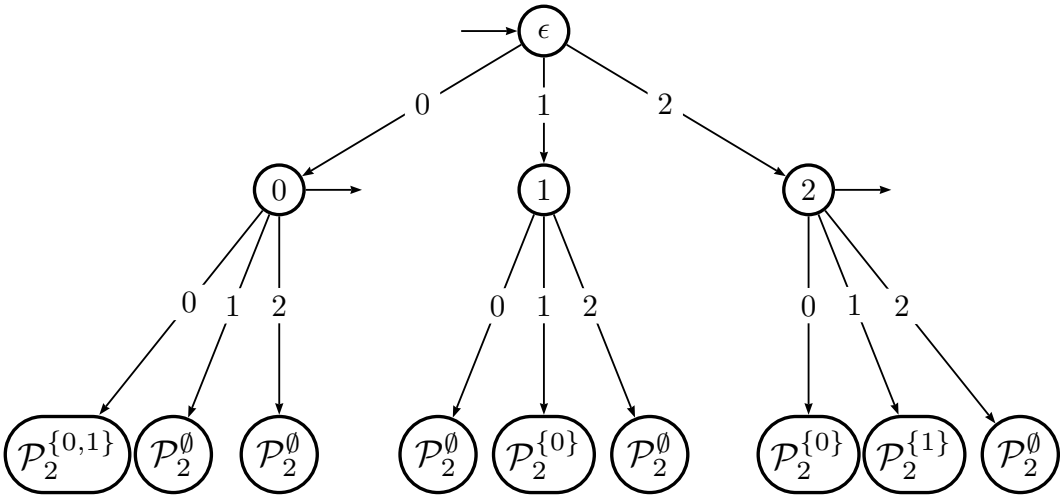



Figure 20 – Schéma de l'automate acceptant les entiers écrits en base 3 dont le reste modulo 18 appartient à $\{0, 2, 4, 5, 9\}$

Notez que quel que soit $K \subseteq \mathbb{Z}/k\mathbb{Z}$, l'automate de Pascal $\mathcal{P}_k^K$ a le même ensemble d'états, la même fonction de transition et le même état initial ; tant qu'il n'est pas question des états finals, l'ensemble de restes $K$ n'est pas pertinent et on note donc $\mathcal{P}_k^?$ cet automate dont les états finals ne sont pas définis. L'automate $\mathcal{A}_n^R$ qui accepte les entiers dont le reste modulo $n$ appartient à $R$, est le produit de $\mathcal{P}_k^?$ par un arbre complet de profondeur $j$ dans lequel les états finals ne sont pas définis comme dans un produit classique. Formellement,

$$\mathcal{A}_n^R \;=\; \langle\, [\![p]\!],\, ([\![p]\!]^{\leqslant j}) \times Q_k,\, \delta,\, (\varepsilon, i_k),\, F \,\rangle$$

- L'ensemble des états est $([\![p]\!]^{\leqslant j}) \times Q_k$, où $Q_k$ est l'ensemble d'états de l'automate de Pascal $\mathcal{P}_k^?$ ;
- L'état initial est $(\varepsilon, i_k)$, où $i_k$ l'état initial de $\mathcal{P}_k^?$ ;
- L'ensemble des états finals $F$ contient un état $(u, q)$
  - si $|u| < j$ et $(\pi_p(u) \,\%\, n) \in R$ ou
  - si $|u| = j$ et $q$ est final dans $\mathcal{P}_k^{K_y}$ où $y = \pi_p(u) \,\%\, d$.
- La fonction de transition $\delta$ est définie par :

$$\forall (u,q) \in [\![p]\!]^j \times Q_k,\ \ \forall a \in [\![p]\!]$$
$$\left.\begin{array}{llll} (u,q) & \xrightarrow[\mathcal{A}_n^R]{a} & (u\,a, q') & \text{si} \quad |u| < j \\ (u,q) & \xrightarrow[\mathcal{A}_n^R]{a} & (u, q') & \text{si} \quad |u| = j \end{array}\right\} \quad \text{et} \quad q \xrightarrow[\mathcal{P}_k^?]{a} q' \ .$$

La proposition suivante est une reformulation du théorème 3.38 dans le cas des ensembles purement périodiques.

**Proposition 3.46** – *Soient un entier $n$ et un ensemble $R \subseteq \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$. Alors, l'automate $\mathcal{A}_n^R$ est complet,*

**a)** *accepte l'ensemble d'entiers $E_n^R$ et*

**b)** *satisfait le critère (UP).*

Démonstration. Il découle directement de la définition que $\mathcal{A}_n^R$ est complet.

**a)** Soit un mot $u \in [\![p]\!]^*$. L'automate $\mathcal{A}_n^R$ est complet, donc il suffit de prouver que l'état atteint par le calcul de $u$ est final si et seulement si $(u \,\%\, n) \in R$.

Si $|u| < j$, alors l'état atteint par le calcul de $u$ est $(u, x)$ pour un certain $x$ et donc d'après la définition, il est final si et seulement si $(u \,\%\, n) \in R$.

Si $|u| \geqslant j$, alors l'état $(w, q)$, atteint par le calcul de $u$ dans $\mathcal{A}_n^R$, satisfait les deux conditions suivantes :

- $w$ est le préfixe de longueur $j$ de $u$ ;
- $q$ est l'état atteint par le calcul de $u$ dans $\mathcal{P}_k^?$.

Cet état $(w, q)$ est final si et seulement si $q$ est final dans $\mathcal{P}_k^{K_y}$ avec

$$y \;=\; \pi_p(w) \,\%\, d \;=\; \pi_p(u) \,\%\, d \;,$$

donc, si et seulement si $\pi_p(u) \,\%\, k \in K_y$ et $\pi_p(u) \,\%\, d = y$. Ceci est enfin équivalent à $\pi_p(u) \,\%\, n \in R$ d'après le lemme 3.44.

**b)** Soit $(u, q)$ un état de $\mathcal{A}_n^R$. Si $|u| < j$, il découle de la définition que $\{(u, q)\}$ est une CFC triviale. Au contraire, si $|u| = j$ alors l'ensemble $u \times Q_k$ est une CFC non-triviale dont la définition est identique à celle de $\mathcal{P}_k^{K_y}$. Toutes les CFC de $\mathcal{A}_n^R$ qui sont non-triviales sont donc des feuilles de $\mathcal{C}_{\mathcal{A}_n^R}$ et des automates de Pascal. D'autre part, $\mathcal{A}_n^R$ accepte par valeur donc satisfait (UP-0) donc le critère (UP). □

## UP-automate acceptant $E_{n,m}^R$

Soient une période $n \in \mathbb{N}$, une pré-période $m \in \mathbb{N}$ et un ensemble $R \subseteq \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$. Nous allons construire un automate $\mathcal{B}_{n,m}^R$ qui accepte l'ensemble $E_{n,m}^R$ ; il s'agit du produit $\mathcal{B}_{n,m}^R = \mathcal{A}_n^R \times \mathcal{G}_m$, où $\mathcal{G}_m$ est l'automate naïf qui accepte les entiers supérieurs à $m$.

On note $l = \lceil \mathsf{log}_p(m) \rceil$, de telle sorte que $p^l \geqslant m$. Un mot $u$ a une valeur supérieure à $m$ s'il a un chiffre non-nul à la position indexée par $k \geqslant l$ ; en effet, ce chiffre a un poids $p^k \geqslant p^l \geqslant m$. Par exemple la figure 21 montre l'automate $\mathcal{G}_{\geqslant 5}$, acceptant les entiers supérieurs à 5 en base 2 (donc $l = 3$).

Formellement, l'automate $\mathcal{G}_m$ est défini par :

$$\mathcal{G}_m \;=\; \langle\; [\![d]\!]^{\leqslant l} \uplus \top,\; [\![p]\!], \delta, \varepsilon \,,\; F \;\rangle$$

où un état $u \in ([\![d]\!]^{\leqslant l} \uplus \top)$ est final si sa valeur est supérieure ou égal à $m$ ou s'il est égal à $\top$ ; et $\delta$ est définie par :

$$\forall u \in [\![d]\!]^l,\;\; \forall a \in [\![p]\!] \qquad \begin{cases} u \xrightarrow[\mathcal{G}_m]{a} u\,a & \text{si } |u| < l \\ u \xrightarrow[\mathcal{G}_m]{a} u & \text{si } |u| = l \;\text{ et }\; a = 0 \\ u \xrightarrow[\mathcal{G}_m]{a} \top & \text{si } |u| = l \;\text{ et }\; a \neq 0 \end{cases} \tag{3.13a}$$

$$\forall a \in [\![p]\!] \qquad \top \xrightarrow[\mathcal{G}_m]{a} \top \tag{3.13b}$$

Les figures 22 à 24 montrent le produit $\mathcal{G}_1 \times \mathcal{A}_{24}^{\{0\}}$ qui accepte les entiers congrus à 0 modulo 24 à l'exception de 0, c'est-à-dire l'ensemble $E_{24,1}^{\{0\}}$

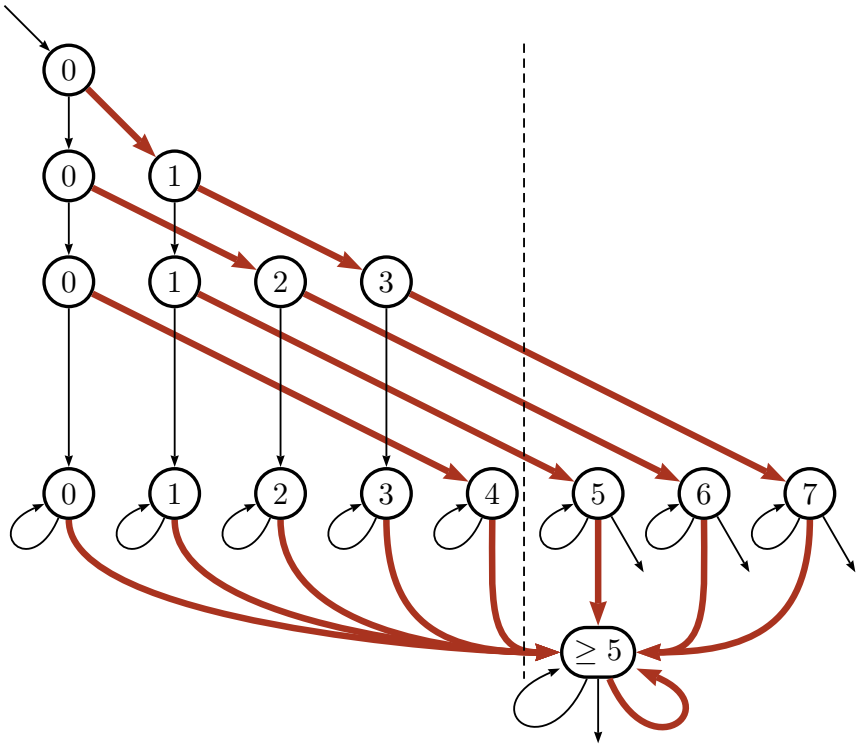



Figure 21 – L'automate $\mathcal{G}_{\geqslant 5}$ acceptant les entiers supérieurs à 5 (en base 2)

Il découle directement de la définition que l'automate $\mathcal{G}_m$ accepte l'ensemble d'entier $\{i \in \mathbb{N} \mid i \geqslant m\}$. On note $\mathcal{B}^R_{n,m}$ le produit d'automates $G_m \times \mathcal{A}^R_n$ et la proposition suivante montre qu'il réalise le théorème 3.38 dans le cas considéré ici.

**Proposition 3.47** – *Soient une période $n \in \mathbb{N}$, une pré-période $m \in \mathbb{N}$ et un ensemble $R \subseteq \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$. Alors, l'automate $\mathcal{B}^R_{n,m}$ est complet,*

**a)** *accepte l'ensemble d'entiers $E^R_{n,m}$ et*

**b)** *satisfait le critère (UP).*

Démonstration. L'automate $\mathcal{B}^R_{n,m}$ est un produit de deux automates complets, il est donc complet. Le point (**a**) découle directement de la correction de $\mathcal{G}_m$ (ci-dessus) et de $\mathcal{A}^R_n$ (proposition 3.46a).

**b)** Il découle du point (**a**) que $\mathcal{B}^R_{n,m}$ satisfait (UP-0), puisqu'il accepte par valeur. On note simplement $\mathcal{C}_\mathcal{B}$ la condensation de $\mathcal{B}^R_{n,m}$ (au lieu de $\mathcal{C}_{\mathcal{B}^R_{n,m}}$).

Une simple vérification montre les CFC non-triviale de $\mathcal{B}^R_{n,m}$ sont d'une des deux formes suivantes :

(i) $\top \times C$, où $C$ est une CFC de $\mathcal{A}^R_n$.

(ii) $u \times X$, où $u$ est un mot de longueur $l$ et $X$ est contenu dans une CFC de $\mathcal{A}^R_n$.

Nous verrons dans la suite que les CFC de type (i) et (ii) correspondent respectivement aux CFC de type 1 et 2 décrites dans le critère (UP). La preuve se divise en trois assertions, dont la première découle directement des définitions.

**Assertion 3.47.1** (UP-1&3) – *Toute CFC de type (i) est stable par toutes les lettres, est une feuille de $\mathcal{C}_\mathcal{B}$ et est un automate de Pascal.*

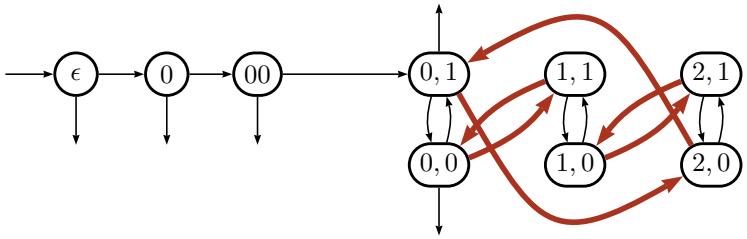



Figure 22 – (Partie co-accessible de) $\mathcal{A}_{24}^{\{0\}}$

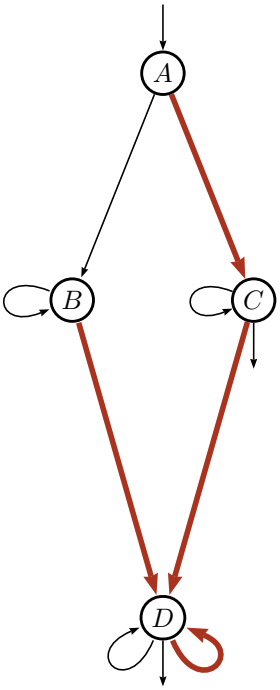



Figure 23 – $\mathcal{G}_1$

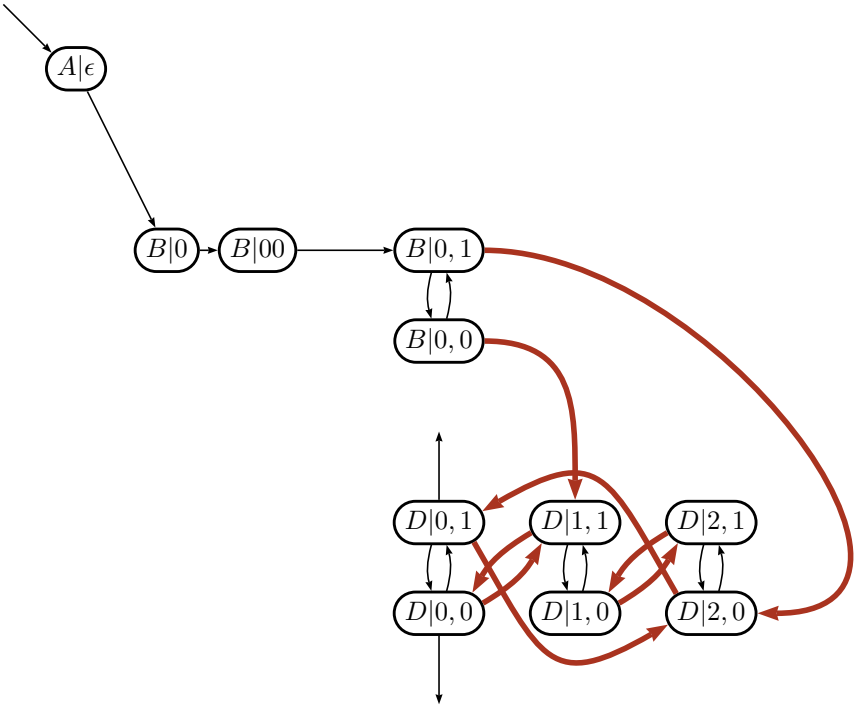



Figure 24 – (Partie co-accessible de) $\mathcal{G}_1 \times \mathcal{A}_{24}^{\{0\}}$

**Assertion 3.47.2** (UP-2) – *Toute CFC de type (ii) est un circuit simple de 0 et elle possède dans $\mathcal{C}_{\mathcal{B}}$ un unique successeur qui est de type (i).*

*Démonstration de l'assertion.* Soit une CFC de $\mathcal{B}_{n,m}^R$ : $u \times X$, où $u$ est un mot de longueur $l$ et $X$ est contenu dans une CFC $C$ de $\mathcal{A}_n^R$. Toutes les CFC de $\mathcal{A}_n^R$ sont des automates de Pascal (proposition 3.46b) donc $C$ est stable par toutes les lettres.

Soient un état $(u, q)$ de la CFC $u \times X$, une lettre $a \in [\![p]\!]$ ; on note $q'$ l'état tel que $q \xrightarrow{a} q'$ est une transition de $\mathcal{A}_n^R$ ; $q'$ est donc dans $C$. D'après l'équation (3.13a),

$$\begin{aligned} a = 0 \quad &\Longrightarrow \quad (u, q) \xrightarrow{a} (u, q') \text{ dans } \mathcal{B}_{n,m}^R \text{ et} \\ a \neq 0 \quad &\Longrightarrow \quad (u, q) \xrightarrow{a} (\top, q') \text{ dans } \mathcal{B}_{n,m}^R ; \end{aligned}$$

notez que dans le deuxième cas $(\top, q')$ est un état de $\top \times C$. Puisque $(u, q)$ n'a qu'un seul successeur (celui par 0) qui est dans $u \times C$ (alors que les autres sont

dans $\top \times C$) alors il est nécessaire que ce successeur soit dans $u \times X$ sinon $\{(u, q)\}$ serait une CFC triviale ; ceci implique que $u \times X$ est réduite à un circuit de 0.

Puisque d'après l'équation (3.13a), $\top$ est un état-puits (de $\mathcal{G}_m$). Il s'ensuit que $\top \times C$ est une CFC de $\mathcal{B}^R_{n,m}$, l'unique CFC successeur de $u \times X$.

**Assertion 3.47.3** (UP-4) – *Chaque CFC de type (ii) se plonge dans la CFC de type (i) qui lui est associée.*

*Démonstration de l'assertion.* Soit une CFC de $\mathcal{B}^R_{n,m}$ : $u \times X$, où $u$ est un mot de longueur $l$ et $X$ est contenu dans une CFC $C$ de $\mathcal{A}^R_n$. On définit la fonction $f$ comme $(u, q) \mapsto (\top, q)$ pour tout $q \in X$. La transition $(u, q) \xrightarrow{0} (u, q')$ existe si et seulement si $q \xrightarrow{0} q'$ existe dans $\mathcal{A}^R_n$, si et seulement si $(\top, q) \xrightarrow{0} (\top, q')$ existe. D'autre part, pour toute lettre $a$ différente de 0, il existe $q'$ telle que les transitions $q \xrightarrow{a} q'$ dans $\mathcal{A}^R_n$, auquel cas les deux transitions suivantes existent :

$$(u, q) \xrightarrow{0} (\top, q') \quad \text{et} \quad (\top, q) \xrightarrow{0} (\top, q') \ . \qquad \square$$

**Remarque 3.48** – *Bien que la figure 24 puisse induire en erreur, l'automate $\mathcal{B}^R_{n,m}$ est bel et bien complet car il est construit par produits successifs d'automates complets.*

## UP-automate acceptant $I \cup E^R_{n,m}$

Soient une période $n \in \mathbb{N}$, une pré-période $m \in \mathbb{N}$, un ensemble $R \subseteq \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$ et $I$ un sous-ensemble de $\{0, 1, \ldots, m-1\}$. On reprend la notation $l = \lceil \log_b(m) \rceil$.

Nous allons maintenant modifier l'automate $\mathcal{B}^R_{n,m}$ pour qu'il accepte les entiers de $I$.

Soit un entier $i \in I$ dont on note la représentation $w = \langle i \rangle_p$. Puisque $i$ est strictement inférieur à $m$, il n'est pas dans $E^R_{n,m}$, et donc pas non plus dans $N(\mathcal{B}^R_{n,m})$.

On note $k \geqslant 0$ le plus petit entier tel que le calcul $w\,0^k$ atteigne dans $\mathcal{B}^R_{n,m}$ une CFC non-triviale que l'on note $C$. La longueur du mot $w\,0^k$ est supérieure ou égale à $l$ (mais pas nécessairement égale à $l$). Par exemple, dans le cas de l'automate $\mathcal{A}^0_{24,1}$ représenté par la figure 24, si l'on fixe $i = 0$, le mot $w = \varepsilon = \langle 0 \rangle_p$ n'atteint pas de CFC, et il faut lui concaténer trois 0 pour atteindre une CFC. Ce mot, $w\,0^3$ est de longueur 3 alors que $l = 1$.

**Lemme 3.49** – *La CFC $C$ atteinte par le calcul de $w\,0^k$ dans $\mathcal{B}^R_{n,m}$ est de type 2 (un circuit de 0).*

Démonstration. Puisque $i < m$, $|\langle i \rangle_p| \leqslant |\langle m \rangle_p|$ ce qui implique que $|w| \leqslant l$. Le calcul de $w$ dans $\mathcal{G}_m$ atteint donc l'état $w$ donc pas l'état $\top$ ; Or, $\top$ n'est accessible qu'en lisant une lettre non-nulle (équation (3.13a)), il n'est donc en particulier pas atteint par le calcul du mot $w\,0^k$. Puisque toutes les CFC de type 1 de $\mathcal{B}^R_{n,m}$ sont formés d'états dont la seconde composante est $\top$, cela conclut la démonstration. $\square$

**Corollaire 3.50** – *Soit un entier $j < k$. L'état atteint par le calcul du mot $w\,0^j$ dans $\mathcal{B}^R_{n,m}$ n'est atteint par le calcul d'aucun autre mot.*

**Lemme 3.51** – *Les mots dont les calculs respectifs atteignent $C$ forment l'ensemble $\{w\,0^j \mid j \geqslant k\}$.*

Démonstration. Il découle du corollaire précédent que tous les mots de cet ensemble $\{w\,0^j \mid j \geqslant k\}$ atteignent $C$.

La CFC de type 2 atteinte par $w\,0^k$ est nécessairement de la forme $u \times X$ où $u = w\,0^{k'}$ avec $k' = l - |w|$, qui vérifient donc $|u| = l$ et $k' \leqslant k$. Or, $\mathcal{G}_m$ est un arbre jusqu'aux états étiquetés par des mots de longueur $l$ donc les mots dont les calculs respectifs atteignent l'état $u$ dans $\mathcal{G}_m$ forment le langage $u0^*$ (lire une lettre non-nulle aboutirait dans l'état $\top$). Donc tous les mot qui n'appartiennent pas $u\,0^*$ n'atteignent donc pas $C$ dans $\mathcal{B}^R_{n,m}$. Puisque de plus $k$ est le plus petit entier tel que le calcul de $w\,0^k = u\,0^{k-k'}$ atteint $C$, tous les mots qui atteignent $C$ appartiennent donc à $\{w\,0^j \mid j \geqslant k\}$. □

Pour construire un automate qui accepte $E^R_{n,m} \cup I$, il suffit donc de rendre final dans $\mathcal{B}^R_{n,m}$ tous les états atteints par les calculs des mots du langage $\langle I \rangle_p 0^*$. Ce nouvel automate satisfait évidemment (UP-0) et puisque toutes les autres conditions sont indépendantes du statut final/non-final des états, il satisfait le critère (UP), ce qui conclut la démonstration du théorème 3.38.

# Correction et complétude du critère (UP)

Cette section conclut le chapitre 3. Elle consiste d'abord à établir le théorème suivant qui énonce la complétude et la correction (resp. démontrées dans les section 3.4.1 et section 3.4.2) du critère UP sur les automates minimal ; puis à utiliser celles-ci pour démontrer le théorème I (sous-section 3.4.3).

**Théorème 3.52** – *Soit $\mathcal{A}$ un automate minimal et, complet ou émondé. L'automate $\mathcal{A}$ accepte par valeur un ensemble ultimement périodique d'entiers si et seulement si il satisfait le critère (UP).*

## Quotients d'un UP-automate

Le but de cette sous-partie est de prouver le sens direct du théorème 3.52. Cela consiste à démontrer que tout quotient d'un UP-automate est un UP-automate (proposition 3.54, plus loin). Pour cela, nous montrons d'abord que chaque CFC d'un quotient est l'image d'une CFC de son revêtement, comme l'exprime le lemme suivant.

**Lemme 3.53** – *Soient deux automates (finis) $\mathcal{A}$ et $\mathcal{M}$ tels qu'il existe un morphisme d'automates $\varphi : \mathcal{A} \to \mathcal{M}$. Pour toute CFC non-triviale de $\mathcal{M}$ noté $Y$, alors il existe un ensemble d'états $X$ de $\mathcal{A}$ tel que*

- **a)** *$X$ est une CFC non-triviale,*
- **b)** *$\varphi(X) = Y$ et*
- **c)** *pour tout $x \in X$ et $a \in [\![p]\!]$, $(x \cdot a)$ appartient à $X$ si et seulement si $(\varphi(x) \cdot a)$ appartient à $Y$*

Démonstration. Soit $Y$ une CFC non-triviale de $\mathcal{M}$ et $S = \varphi^{-1}(Y)$. On munit $S$ de la relation accessibilité ; cette relation possède (au moins) une classe d'équivalence minimale dont les éléments forment un ensemble noté $X$. Il s'ensuit donc qu'aucun état de $X$ ne peut atteindre un état de $S \setminus X$.

**a)** Soit $z$ un état de $\mathcal{A}$ et $x$ un état de $X$ tel qu'il existe deux mots $u, v$ satisfaisant

$$x \xrightarrow[\mathcal{A}]{u} z \xrightarrow[\mathcal{A}]{v} x \ .$$

Puisque $\varphi$ est morphisme, il s'ensuit que

$$\varphi(x) \xrightarrow[\mathcal{M}]{u} \varphi(z) \xrightarrow[\mathcal{M}]{u} \varphi(x)$$

et puisque $Y$ est un CFC contenant $\varphi(x)$, elle contient également $\varphi(z)$. Ceci implique que $z$ est dans $\varphi^{-1}(Y) = S$. Or $z$ est atteignable par $x$, un état de $X$ donc $z$ est dans $X$. Il s'ensuit que $X$ est une CFC.

Soit $x$ un état de $X$. Son image $\varphi(x)$ appartient à $Y$ qui est une CFC non-triviale donc il existe $u \neq \varepsilon$ tel que $\varphi(x) \xrightarrow[\mathcal{M}]{u} \varphi(x)$ existe. Si bien que $x \xrightarrow[\mathcal{A}]{u} x$ donc $X$ n'est pas une CFC triviale.

**b)** Soit un état $y' \in Y$ de $\mathcal{M}$. On note $x \in X$ un état de $\mathcal{A}$, ce qui implique que $\varphi(x)$ appartient à $Y$. Puisque $Y$ est une CFC non-triviale, les éléments $\varphi(x)$ et $y'$ (de $Y$) peuvent s'atteindre l'un l'autre ; on note $u$ le mot tel que $\varphi(x) \xrightarrow[\mathcal{M}]{u} y'$. On appelle $x'$ l'élément tel que $x \xrightarrow[\mathcal{A}]{u} x'$ qui existe puisque $\varphi$ est un morphisme ; il vérifie (pour la même raison) $\varphi(x') = y'$. Il s'ensuit que $x'$ appartient à $S$ donc à $X$ car atteignable depuis un élément de $X$. Donc $Y \subseteq \varphi(X)$ et puisque $X \subseteq S = \varphi^{-1}(Y)$, il en découle que $Y = \varphi(X)$.

**c)** Soit un état $x \in X$ et une lettre $a \in [\![p]\!]$. Si $(x \cdot a)$ est dans $X$, il découle du point (**b**) que $\varphi(x \cdot a) = (\varphi(x) \cdot a)$ est dans $Y$. Inversement, si $(\varphi(x) \cdot a)$ est dans $Y$, alors $(x \cdot a)$ existe et est dans $S(= \varphi^{-1}(Y))$ mais puisque aucun état de $X$ ne peut atteindre un état de $S \setminus X$, $(x \cdot a)$ est nécessairement dans $X$. □

Le critère (UP) est constitué de propriétés de CFC qui sont toutes stables par quotient, ce qui permet d'établir la proposition suivante.

**Proposition 3.54** – *Soit un automate $\mathcal{A}$. Si $\mathcal{A}$ satisfait le critère (UP) alors tous les quotients de $\mathcal{A}$ le satisfont aussi.*

Démonstration. Soit un automate $\mathcal{M}$ tel qu'il existe un morphisme d'automates $\varphi : \mathcal{A} \to \mathcal{M}$. Il s'ensuit que $\mathcal{A}$ et $\mathcal{M}$ acceptent le même langage, donc puisque $\mathcal{A}$ accepte par valeur, c'est aussi le cas pour $\mathcal{M}$, donc $\mathcal{M}$ satisfait la condition (UP-0).

Soit $Y$ une CFC non-triviale de $\mathcal{M}$. On note $X$ la CFC non-triviale de $\mathcal{A}$ qui vérifie les conditions du lemme 3.53 précédent.

• (UP-1) et (UP-3) – Si $Y$ comprend une transition interne étiquetée par une lettre différente de 0, il en est de même pour $X$ (condition 3.53c) , donc $X$ une CFC de type 1. D'après (UP-3), $X$ est donc le quotient d'un automate de Pascal, donc $Y = \varphi(X)$ est le quotient du même automate de Pascal ; puisque les automates

de Pascal sont complets, leurs quotients le sont aussi donc $Y$ est une feuille de $\mathcal{C}_{\mathcal{M}}$. L'automate $\mathcal{M}$ satisfait donc (UP-1) et (UP-3).

• (UP-2) – Si toutes les transitions internes de $Y$ sont étiquetées par 0, il s'agit d'un simple circuit de 0 auquel cas $X$ est également un simple circuit de 0 (condition 3.53c). Prouvons par l'absurde que $Y$ a au plus un unique successeur dans $\mathcal{C}_{\mathcal{M}}$. Si $X$ n'a pas de successeur dans $\mathcal{C}_{\mathcal{A}}$, alors $Y$ ne peut pas en avoir dans $\mathcal{C}_{\mathcal{M}}$.

On suppose dorénavant que $X$ a un unique successeur $X'$ dans $\mathcal{C}_{\mathcal{A}}$ ; on note $Y' = \varphi(X')$ qui est évidemment fortement connexe. De plus, pour tout $y \in Y$, $u \in [\![p]\!]^*$ et $y \notin Y$ tel que $y \xrightarrow{u} y'$, il existe $x \in X$ et $x' \notin X$ tels que $x \xrightarrow[\mathcal{A}]{u} x'$ et $\varphi(x) = y$ donc $\varphi(x') = y'$. Puisque $\mathcal{A}$ satisfait (UP-2), la seule CFC atteignable depuis $X$ dans $\mathcal{C}_{\mathcal{A}}$ est $X'$ donc $x' \in X'$ et donc $y' \in Y'$. Il s'ensuit que $Y'$ est une CFC ; de plus, puisque $X'$ est une CFC de type 1 alors elle admet une transition interne étiquetée par une lettre non-nulle et puisque $Y' = \varphi(X')$, c'est aussi le cas de $Y'$. L'unique successeur de $Y$ dans $\mathcal{C}_{\mathcal{M}}$ est donc de type 1 ; l'automate $\mathcal{M}$ satisfait donc (UP-2).

• (UP-4) – On conserve les notations utilisées pour le point précédent. La CFC $X$ est un circuit de 0 que l'on note

$$x_0 \xrightarrow{0} x_1 \xrightarrow{0} x_2 \longrightarrow \cdots \xrightarrow{0} x_{k-1} \xrightarrow{0} x_0$$

Donc $Y$ est le circuit de 0 noté

$$y_0 \xrightarrow{0} y_1 \xrightarrow{0} y_2 \longrightarrow \cdots \xrightarrow{0} y_{i-1} \xrightarrow{0} y_0$$

qui vérifie nécessairement que $i \mid k$ et

$$\forall j < k \qquad \varphi(x_j) \;=\; y_{j\,\%\,i} \;=\; \varphi(x_{j\,\%\,i}) \;.$$

D'après $(UP-4)$ il existe une fonction de plongement $f : X \to X'$, c'est-à-dire vérifiant (définition 3.33)

$$\begin{array}{ll} \forall j < k,\ \forall a \in [\![p]\!],\ a \neq 0 & f(x_j) \cdot a \;=\; x_j \cdot a \\ \forall j < k & f(x_j) \cdot 0 \;=\; f(x_j \cdot 0) \;. \end{array} \qquad (*)$$

On va démontrer que la fonction $g : Y \to Y'$ qui envoie chaque $y_j \in Y$ sur $\varphi(f(x_j))$ est une fonction de plongement.

Soit une lettre $a \in [\![p]\!]$ différente de 0 et un entier $j < k$. Puisqu'un morphisme d'automates permute avec la fonction de transition, la définition $(*)$ de $f$ implique donc que

$$\varphi(f(x_j)) \cdot a \;=\; \varphi(f(x_j) \cdot a) \;=\; \varphi(x_j \cdot a) \;=\; (\varphi(x_j) \cdot a) \;=\; (\varphi(x_{j\,\%\,i}) \cdot a) \;.$$

En utilisant deux fois l'équation précédente (une fois pour $j$, une fois pour $j\,\%\,i$), on obtient que

$$\forall j < k \qquad \varphi(f(x_j)) \cdot a \;=\; \varphi(f(x_{j\,\%\,i})) \cdot a \;.$$

Puisque $Y'$ est à groupe (étant le quotient d'un automate de Pascal), il s'ensuit que

$$\forall j < k \qquad \varphi(f(x_j)) \;=\; \varphi(f(x_{j\,\%\,i})) \;. \qquad (**)$$

Il découle de $(*)$ et de la définition de $g$ que :

$$\forall j < i,\ \forall a \in [\![p]\!]^*,\ a \neq 0 \qquad g(y_j) \cdot a \;=\; \varphi(f(x_j)) \cdot a \;=\; \varphi(x_j) \cdot a \;=\; y_j \cdot a$$

et des différentes équations précédentes que

$$\begin{array}{lllll} \forall j < i \quad g(y_j)\cdot 0 & = & \varphi(f(x_j))\cdot 0 & = & \varphi(f(x_j\cdot 0)) \quad \text{d'après } (*) \\ & & & = & \varphi(f(x_{(j+1)\,\%\,k})) \quad X \text{ est un circ. de } 0 \\ & & & = & \varphi(f(x_{(j+1)\,\%\,i})) \quad \text{d'après } (**) \\ & & & = & g(y_{(j+1)\,\%\,i}) \quad \text{d'après la def. de } g \\ & & & = & g(y_j\cdot 0) \quad Y \text{ est un circ. de } 0 \end{array}$$

La fonction $g$ est donc effectivement une fonction de plongement $Y \to Y'$ donc $\mathcal{M}$ satisfait (UP-4). □

Nous pouvons maintenant démontrer la complétude du critère (UP) pour la classe des automates complets et minimaux.

**Proposition 3.55** – *Soit un automate $\mathcal{M}$ minimal, complet et qui accepte par valeur un ensemble ultimement périodique d'entiers. Alors $\mathcal{M}$ satisfait le critère (UP).*

Démonstration. D'après le théorème 3.38, il existe un automate complet $\mathcal{A}$ qui accepte par valeur le même ensemble ultimement périodique. Il s'ensuit que $\mathcal{M}$ est un quotient de $\mathcal{A}$ donc d'après la proposition 3.54 précédente qu'il satisfait le critère (UP). □

La complétude du critère UP pour la classe des automates émondés et minimaux est une conséquence de la proposition précédente et du lemme suivant.

**Lemme 3.56** – *Soit un automate $\mathcal{A}$ satisfaisant le critère (UP). Alors, la partie émondée de $\mathcal{A}$ satisfait le critère (UP).*

Démonstration. On note $\mathcal{B}$ la partie émondée de $\mathcal{A}$.

Le processus d'émondage retire nécessairement des CFC toute entière. Le type (1 ou 2) d'une CFC ne dépend que d'elle même (stable par une lettre non-nulle ou non), donc chaque CFC de $\mathcal{B}$ a le même type que son pendant dans $\mathcal{A}$. Une simple lecture des conditions (UP-1) à (UP-4) montre qu'elles sont stable par émondage.

Il reste à démontrer qu'il satisfait (UP-0). Ceci est simplement une conséquence du fait qu'elle est équivalente au fait d'accepter par valeur : $\mathcal{A}$ et $\mathcal{B}$ sont équivalents donc puisque le premier accepte par valeur, le second aussi. □

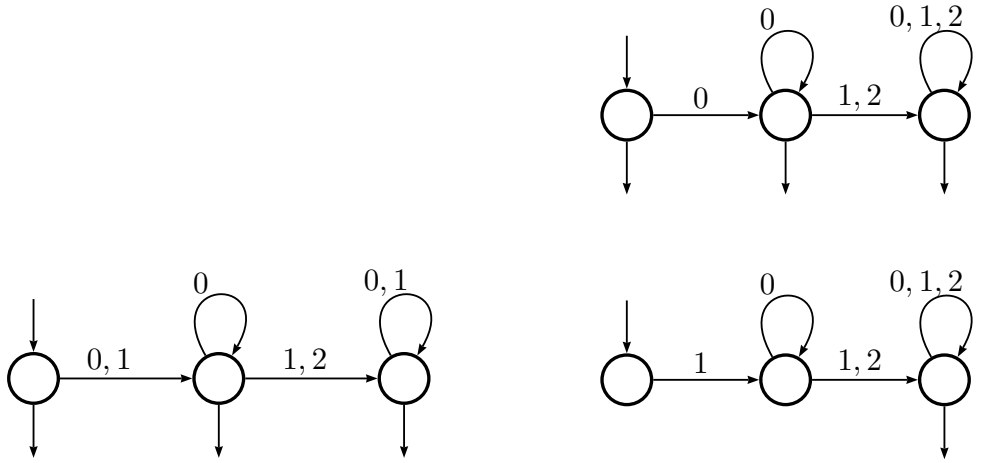
Ceci conclut la démonstration du sens direct du théorème 3.52.

## Correction du critère (UP)

Dans cette sous-section, nous allons démontrer la proposition suivante ; le sens réciproque du théorème 3.52 en est un corollaire.

**Proposition 3.57** – *Tout automate satisfaisant le critère (UP) accepte un ensemble ultimement périodique d'entiers.*

Soit un automate $\mathcal{A}$ satisfaisant le critère (UP). Pour démontrer que $\mathcal{A}$ accepte un ensemble d'entiers ultimement périodique, nous allons séparer $\mathcal{A}$ en ses *branches* (des sous-automates de $\mathcal{A}$ définis ci-après) et montrer que chacune accepte un ensemble ultimement périodique. L'automate $\mathcal{A}$ accepte alors l'union de ces ensembles ultimement périodiques c'est-à-dire un ensemble ultimement périodique.




(a) Un UP-automate $\mathcal{A}_5$ (b) Les branches de $\mathcal{A}_5$

Figure 25 – Une branche de l'automate ne correspond pas à une branche de la condensation

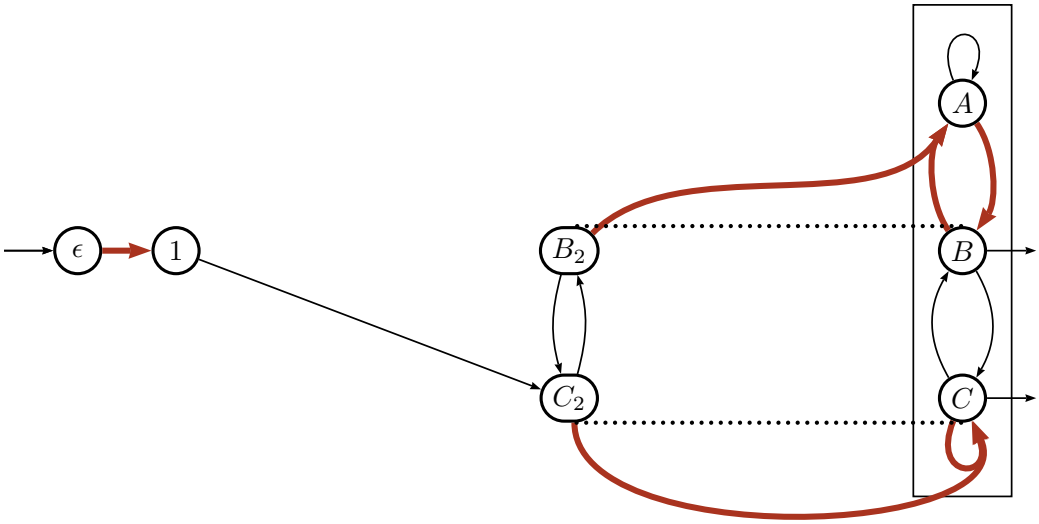
Une branche d'un UP-automate $\mathcal{A}$ est le sous-automate résultant du choix d'une branche du DAG initial de CFC triviales de $\mathcal{A}$ ; il ne s'agit pas des branches de la condensation, comme le montre la figure 25. On choisit un mot $u$ qui atteint un état $q$ d'une CFC non-triviale et dont tous les préfixes strictes atteignent des CFC triviales ; la branche induite par $u$ contient tous les états du chemin $i_{\mathcal{A}} \xrightarrow{u} q$ plus tous les états accessibles depuis $q$. L'automate ainsi défini ne reconnaît plus nécessairement par valeur, il faut donc retirer le statut final de certains états : on factorise $u$ comme $v0^i$ de telle sorte que $v$ ne finisse pas par un 0 et on retire de l'ensemble des états finals tous les états atteints par les préfixes strictes de $v$ (car ils n'ont plus de successeur par 0).

Par exemple, la figure 26 représente les différentes branches de l'automate $\mathcal{A}_3$ figurant à la page 66. La définition formelle d'une branche est donnée ci-dessous.

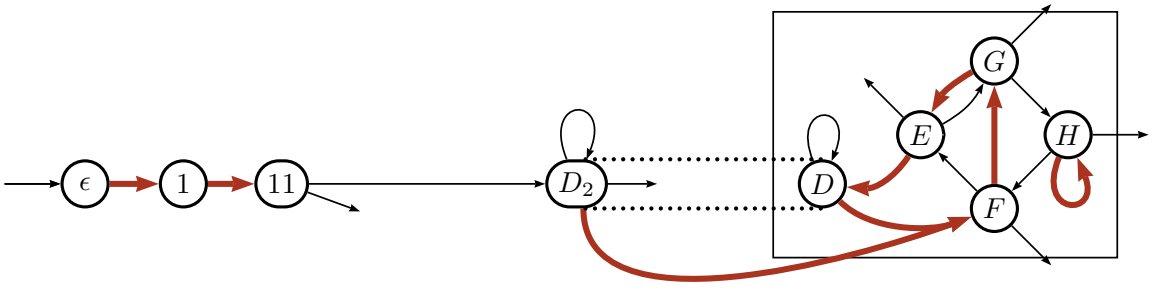
**Définition 3.58** – *Un sous-automate $\mathcal{B}$ de $\mathcal{A}$ est appelé* branche *de $\mathcal{A}$ si chaque état $s$ de $\mathcal{B}$ satisfait les conditions suivantes.*

- **a)** *L'état $s$ est accessible depuis l'état initial.*
- **b)** *Si $s$ est dans une CFC triviale, alors $s$ admet exactement une transition sortante (dans $\mathcal{B}$).*
- **c)** *Si $s$ est dans une CFC non-triviale (c'est-à-dire de type 1 ou 2), alors les successeurs et les transitions sortantes de $s$ qui existent dans $\mathcal{A}$ existent aussi dans $\mathcal{B}$.*
- **d)** *Si $s$ n'a pas de successeur par 0 dans $\mathcal{B}$ (mais en a une dans $\mathcal{A}$), alors il est non-final ; sinon il a le même statut final/non-final que dans $\mathcal{A}$.*

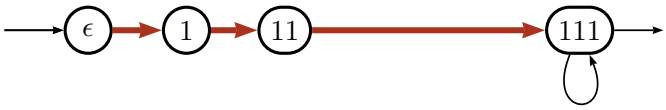

(a) Branche 1 : accepte les entiers congrus à $\{6, 10\}$ modulo 12 (et supérieurs à 4)




(b) Branche 2 : accepte les entiers congrus à $\{1, 5\}$ modulo 12 et supérieurs à 4




(c) Branche 3 : accepte l'entier 3 et les entiers congrus à $\{11, 19, 27, 35\}$ modulo 40 (et supérieurs à 8)




(d) Branche 4 : accepte uniquement l'entier 7

Figure 26 – Les branches de l'automate $\mathcal{A}_3$

La prochaine proposition énonce que tout UP-automate qui est réduit à une seule branche accepte un ensemble ultimement périodique d'entiers.

**Lemme 3.59** – *Soit un UP-automate $\mathcal{B}$. Si $\mathcal{B}$ est réduit à une branche, alors $N(\mathcal{B})$ est un ensemble ultimement périodique.*

Démonstration. Il existe un mot $u \in [\![p]\!]^*$ dont le calcul atteint une CFC non-triviale mais tel ceux de tous ses préfixes strictes atteignent des CFCs triviales. On écrit $r_d = \pi_p(u)$, $j = |u|$ et on définit

$$I \;=\; \{\pi_p(v) \mid v \text{ est un préfixe de } u \text{ accepté par } \mathcal{B}\}\ .$$

($I$ est soit vide, soit le singleton $\{\pi_p(u)\}$ car $\mathcal{B}$ satisfait (UP-0).)

Puisque $\mathcal{B}$ est réduit à une seule branche et qu'il satisfait le critère (UP), alors il possède au plus deux CFC non-triviales. On ne va traiter ici que le cas le plus difficile, à savoir le cas où $\mathcal{B}$ a deux CFC, l'une de type 1 et l'autre de type 2.

On note $C$ (resp. $D$) la CFC de type 2 (resp. 1) de $\mathcal{B}$ qui est donc un circuit de 0 (resp. le quotient d'un automate de Pascal). On note de plus $s_c$ l'état (de $C$) atteint par $u$. D'après (UP-4), $C$ se plonge dans $D$, il existe donc une fonction $f : C \to D$ telle que tout état $q$ de $C$,

- $f(q \cdot 0) = f(q) \cdot 0$ et
- pour toute lettre $a \neq 0$, $(q \cdot a) = f(q) \cdot a$ .

On note $i_D$ l'unique l'état de $D$ tel que $i_D \xrightarrow{u} f(s_c)$ ; cet état existe et est unique car $D$ est le quotient d'un automate de Pascal (d'après (UP-3)) et est donc à groupe (*cf.* lemme 3.22).

**Assertion 3.59.1** – *Soit un mot $w \in [\![p]\!]^*$ dont le calcul atteint $D$. Ce mot $w$ est accepté par $\mathcal{B}$ si et seulement si $(i_D \cdot w)$ est final.*

*Démonstration de l'assertion.* Puisque $w$ atteint $D$, il est de la forme $w = u\,0^i\,a\,v$ pour un certain mot $v$, une certaine lettre $a \neq 0$ et un certain exposant $i$. Les chemins étiquetés par $w = u\,0^i\,a$ partant respectivement de $i_{\mathcal{B}}$ et de $i_D$ sont donc

$$i_{\mathcal{B}} \xrightarrow{u} s_c \xrightarrow{0^i} [s_c \cdot 0^i] \xrightarrow{a} [s_c \cdot 0^i \cdot a] \quad \text{et}$$
$$i_D \xrightarrow{u} f(s_c) \xrightarrow{0^i} [f(s_c) \cdot 0^i] \xrightarrow{a} [f(s_c) \cdot 0^i \cdot a] \ .$$

Or, la définition de $f$ comme fonction de plongement implique que

$$[f(s_c) \cdot 0^i \cdot a] \;=\; [f(s_c \cdot 0^i) \cdot a] \;=\; [(s_c \cdot 0^i) \cdot a]\ ,$$

donc que l'état atteint en lisant $w$ depuis $i_{\mathcal{B}}$ ou depuis $i_D$ est le même, ce qui conclut la démonstration.

On note $(k, R_k)$ les paramètres de l'automate de Pascal dont $D$ est le quotient quand on fixe $i_D$ comme état initial (le paramètre $R_k$ dépend de l'état initial choisi).

Soit un mot $w \in [\![p]\!]^*$.

- Si le calcul de $w$ n'atteint pas $D$, alors $w$ est accepté par $\mathcal{B}$ si et seulement si $\pi_p(w)$ appartient à $I$ ($w$ est alors soit un préfixe de $u$, soit de la forme $u\,0^i$ pour un certain exposant $i$).

- Si au contraire le calcul de $w$ atteint $D$, alors d'une part $w$ commence par $u$. De plus, le mot $w$ est accepté par $\mathcal{B}$ si et seulement si $(i_D \cdot w)$ est final d'après l'assertion précédente. Puisque $|u| = j$, alors $w$ est accepté par $\mathcal{B}$ si et seulement si

$$\pi_p(w) \geqslant p^j \quad , \quad \pi_p(w) \,\%\, p^j = \pi_p(u) \quad \text{et} \quad \pi_p(w) \,\%\, k \in R_k \ .$$

  D'après le théorème 3.42 des restes chinois, il existe $R \subseteq \mathbb{Z}/(k \times p^j)\mathbb{Z}$ tel que la conjonction précédente est équivalente à

$$\pi_p(w) \geqslant p^j \quad \text{et} \quad \pi_p(w) \,\%\, (k \times p^j) \in R \ .$$

En résumé, en posant $m = p^j$, et $n = (k \times p^j)$, un mot $w$ est accepté par $\mathcal{B}$ si et seulement si $\pi_p(w)$ est dans $I \cup E^R_{n,m}$. □

Montrons maintenant que le critère (UP) est distribué aux branches.

**Lemme 3.60** – *Toute branche d'un UP-automate est un UP-automate.*

Démonstration. Soit $\mathcal{A}$ un UP-automate et $\mathcal{B}$ une branche de $\mathcal{A}$. Dès qu'un état $s$ de $\mathcal{B}$ est dans une CFC non-triviale dans $\mathcal{A}$, tous les chemins qui partent de $s$ (et tous les états ainsi atteignables) dans $\mathcal{A}$ sont aussi dans $\mathcal{B}$ (condition 3.58c). Il s'ensuit qu'une CFC de $\mathcal{B}$ est une CFC de $\mathcal{A}$ et que l'unique CFC qui lui succède dans $\mathcal{C}_\mathcal{A}$ est toute-entière dans $\mathcal{B}$. L'automate $\mathcal{B}$ hérite donc les propriétés (UP-1) à (UP-4) de $\mathcal{A}$.

Puisque $\mathcal{A}$ satisfait la condition (UP-0), si $\mathcal{B}$ ne la satisfait pas alors il existe deux états finals $s, s'$ de $\mathcal{A}$ tels que $s \xrightarrow[\mathcal{A}]{0} s'$ et $s$ existe dans $\mathcal{B}$ mais n'admet pas de successeur par par 0 dans $\mathcal{B}$ ; dans ce cas la condition (**d**) de la définition 3.58 assure que l'état $s$ n'est pas final dans $\mathcal{B}$. □

Le lemme suivant exprime que le langage d'un UP-automate est l'union des langages de ses branches.

**Lemme 3.61** – *Un mot est accepté par un UP-automate $\mathcal{A}$ si et seulement si il est accepté par l'une de ses branches.*

Démonstration. Soit un mot $w \in [\![p]\!]^*$ accepté par $\mathcal{A}$.

Si le calcul de $w$ atteint une CFC non-triviale de $\mathcal{A}$ on note $u$ son plus petit préfixe qui atteint une CFC non-triviale et la branche définie par $u$ doit accepter $w$ car aucun état final d'une CFC non-triviale dans $\mathcal{A}$ ne devient non-final dans une branche de $\mathcal{A}$.

Si au contraire, le calcul de $w$ atteint une CFC triviale, on pose $w' = w0^i$ où $i$ est suffisamment grand (par exemple le nombre d'états de $\mathcal{A}$) ; puisque $\mathcal{A}$ satisfait (UP-0), il accepte $w'$ et le calcul de celui-ci atteint nécessairement une CFC non-triviale, donc il existe une branche de $\mathcal{A}$ qui accepte $w'$ d'après le cas précédent. Or d'après le lemme 3.60, cette même branche satisfait le critère (UP) donc en particulier la condition (UP-0) et accepte donc $w$.

Le sens réciproque découle du fait que chaque branche de $\mathcal{A}$ est un sous-automate de $\mathcal{A}$ et accepte donc un langage inclus dans $L(\mathcal{A})$. □

La démonstration de la proposition 3.57 est alors une simple combinaison des deux lemmes précédents.

Démonstration du proposition 3.57. L'ensemble d'entiers accepté par un UP-automate est donc l'union des ensembles d'entiers acceptés par ses branches (lemme 3.61) ; or chacune de ces branches satisfait le critère (UP) (lemme 3.60) donc accepte un ensemble ultimement périodique d'entiers (lemme 3.59). En définitive, un UP-automate accepte une union d'ensembles d'entiers tous ultimement périodiques, c'est-à-dire un ensemble d'entiers ultimement périodique, ce qui conclut la démonstration du proposition 3.57. □

Le sens réciproque du théorème 3.52 étant un cas particulier de la proposition 3.57, ceci conclut la démonstration du théorème 3.52 tout entier.

**Remarque 3.62** – *Un automate $\mathcal{A}$ peut avoir un nombre de branches exponentiel en son nombre d'états, comme l'illustre la figure 27. Ceci n'est néanmoins pas contradictoire avec le théorème principal de cette section, le calcul des branches n'est pas nécessaire pour vérifier qu'un automate satisfait le critère (UP). Les branches donnent cependant une manière explicite de calculer l'expression de l'ensemble ultimement périodique accepté par l'automate. Cette manière n'est d'ailleurs pas très efficace ; il est donné dans* [46] *une méthode pour calculer la formule logique (de l'arithmétique de Presburger) correspondante en temps polynomial.*

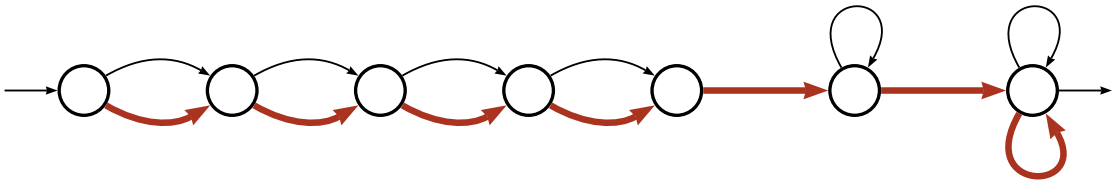


Figure 27 – UP-automate minimal ayant 7 états et $2^4$ branches

## Ultime périodicité est décidable en temps quasi-linéaire

**Théorème I** – *Soient une base entière $p$ et un automate déterministe $\mathcal{A}$ sur l'alphabet $[\![p]\!]$. On note $n$ le nombre d'états de $\mathcal{A}$ et $m$ son nombre de transitions. Le problème* ULTIME PÉRIODICITÉ *peut être décidé en temps $O(n \log(n) + m)$.*

Démonstration. On commence par émonder $\mathcal{A}$ puis on minimise le résultat, ce qui donne un automate minimal émondé, noté $\mathcal{M}$, équivalent à $\mathcal{A}$. Calculer la partie émondé d'un automate se fait en temps $O(m)$ et sa minimisation en temps $O(n \log(n))$ avec l'algorithme classique de Hopcroft (*cf.* [41, 28]).

[46] Jérôme Leroux, 2005, *A polynomial time Presburger criterion and synthesis for number decision diagrams.*

[41] John E. Hopcroft, 1971, *An nlogn algorithm for minimizing states in a finite automaton.*

[28] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest et Clifford Stein, 2002, *Introduction à l'algorithmique (2ème ed.).*

Puisque $\mathcal{M}$ est minimal et émondé, il découle du théorème 3.52 que $\mathcal{M}$ accepte un ensemble ultimement périodique si et seulement si il satisfait le critère (UP). Vérifier si $\mathcal{M}$ satisfait le critère (UP) se fait en temps $\mathsf{O}(m')$ (théorème 3.37, page 65) où $m'$ est le nombre de transitions de $\mathcal{M}$, donc vérifiant $m' \leqslant m$. Puisque $\mathcal{M}$ et $\mathcal{A}$ sont équivalents, ceci conclut la démonstration. □

# Conclusion de la première partie

Les résultats présentés dans cette partie apportent presque une réponse définitive à la question posé par l'article d'Honkala [40]. Deux améliorations peuvent encore être apportées : obtenir une complexité linéaire ou étendre le résultat aux automates non-déterministes.

Notre algorithme n'est pas linéaire uniquement car il nécessite une minimisation préalable. Bien que l'on ne puisse probablement pas retirer totalement la condition de minimalité, une minimisation partielle pourrait suffire. En effet, il est seulement nécessaire que l'automate en entrée soit le quotient de l'UP-automate canonique (décrit dans la section 3.3) reconnaissant un ensemble ultimement périodique arbitraire. Or, il existe déjà des algorithmes linéaires pour minimiser des automates dont les CFCs sont simples (voir par exemple [6]). Le principal obstacle semble donc de définir une minimisation partielle d'un automate équivalent à un automate de Pascal qui puisse être réalisée en temps linéaire mais qui produise néanmoins un automate suffisamment "petit" pour pouvoir utiliser l'algorithme de la section 3.1.

En revanche, définir un nouveau critère (UP′) pour les automates non-déterministes semble bien plus ardu. En effet, le critère actuel s'appuie fortement sur les formes et les relations entre les CFCs et nos expérimentations suggèrent que la procédure de déterminisation produit un effet erratique sur celles-ci.

D'autre part, certaines méthodes développées dans ce chapitre sont très probablement généralisables à d'autres systèmes de numération. C'est le cas, notamment, pour l'introduction de la lettre $10^{-1}$ qui permet dans la section 3.1 de "deviner" immédiatement la période (si elle existe).

[40] Juha Honkala, 1986, *A Decision Method for The Recognizability of Sets Defined by Number Systems.*

[6] Jorge Almeida et Marc Zeitoun, 2008, *Description and analysis of a bottom-up DFA minimization algorithm.*

# Deuxième partie

# Sur la base rationnelle

Chapitre 4

# Systèmes de numération à base rationnelle

Dans ce chapitre sont définis les *systèmes de numération à base rationnelle*, ou plus simplement, *les bases rationnelles*. Il s'agit d'une généralisation de la base entière précédemment décrite au chapitre 2.

Contrairement à la convention prise pour la base entière, nous allons dorénavant représenter les nombres de façon usuelle, c'est-à-dire avec le chiffre de poids fort en premier. En particulier, la première lettre d'un mot est celle qui est la plus à gauche et un mot $u$ commence par une lettre $a$ s'il s'écrit $u = au'$ pour un certain $u'$ ; l'indice 0 sera quand même utilisé pour la lettre la plus à droite, comme par exemple $u = a_k \cdots a_1 a_0$.

Certains transducteurs (comme l'additionneur, le normalisateur ou, dans le chapitre 5, l'incrémenteur) sont plus simples s'ils lisent le chiffre de poids faible en premier. Dans ce cas, ils liront les mots de *droite à gauche*, seront appelés *des transducteurs droits* et leurs transitions seront notées dans le texte au moyen d'une flèche qui pointe vers la gauche, comme par exemple $s \xleftarrow{a \mid b} s'$. Par exemple, les figures 1a et 1b représentent deux transducteurs ayant la même définition (ensemble d'états, de transitions, etc.) mais le second est déclaré *droit* alors que le premier ne l'est pas. Les calculs respectifs du mots 012 dans ces transducteurs sont alors

$$A \xrightarrow[\mathcal{T}_g]{0 \mid 0} A \xrightarrow[\mathcal{T}_g]{1 \mid 1} B \xrightarrow[\mathcal{T}_g]{2 \mid 0} B \quad \text{et} \quad B \xleftarrow[\mathcal{T}_d]{0 \mid 2} B \xleftarrow[\mathcal{T}_d]{1 \mid 1} A \xleftarrow[\mathcal{T}_d]{2 \mid 2} A \ ,$$

donc ses images sont $\mathcal{T}_g(012) = 010$ et $\mathcal{T}_d(012) = 212$, respectivement.

La base rationnelle a été introduite dans [2] (voir aussi [36]) et a apporté un nouvel éclairage sur les problèmes de Mahler (*cf.* [50]) et de Josephus (*cf.* [65]). Sauf mentions contraires, les résultats présentés dans ce chapitre sont tirés de [2], souvent sous une formulation différente.

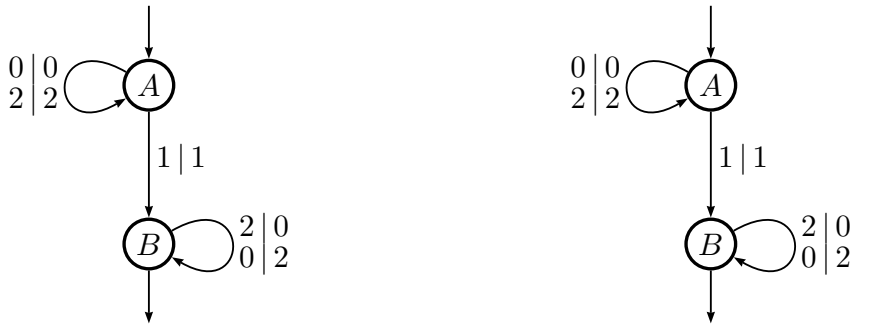
---

[2] Shigeki Akiyama, Christiane Frougny et Jacques Sakarovitch, 2008, *Powers of rationals modulo 1 and rational base number systems.*

[36] Christiane Frougny et Jacques Sakarovitch, 2010, *Number representation and finite automata.*

[50] Kurt Mahler, 1968, *An unsolved problem on the powers of 3/2.*

[65] Andrew M. Odlyzko et Herbert S. Wilf, 1991, *Functional iteration and the Josephus problem.*




(a) Un transducteur (gauche) $\mathcal{T}_g$ (b) Un transducteur droit $\mathcal{T}_d$

Figure 1 – Exemple de transducteurs gauche et droit

# Représentation et évaluation en base $\frac{p}{q}$

Dans toute la suite, $p$ et $q$ désignent deux entiers premiers entre eux tels que $p > q > 1$ ; si bien que la fraction $\frac{p}{q}$, appelée *la base*, est irréductible et supérieure à 1.

La $\frac{p}{q}$*-représentation d'un entier* $N$ est le mot résultant de *l'algorithme d'Euclide modifié* décrit ci-dessous. On pose $N_0 = N$ et

$$\forall i \in \mathbb{N} \qquad q\,N_i \;=\; p\,N_{(i+1)} + a_i \;, \tag{4.1}$$

où $a_i$ est le reste de la division euclidienne de $q\,N_i$ par $p$, donc appartient à l'alphabet *canonique* $[\![p]\!]$. Puisque $p > q$, la suite $(N_i)_i$ est d'abord strictement décroissante jusqu'à devenir stationnaire à $N_{(k+1)} = 0$. La $\frac{p}{q}$*-représentation* de $N$ est alors le mot $a_k \cdots a_1\, a_0$, noté $\langle N \rangle_{\frac{p}{q}}$ ; en particulier la $\frac{p}{q}$-représentation de 0 est le mot vide $\varepsilon$, un mot de longueur $k = 0$, puisque dans ce cas l'algorithme s'arrête à $k = -1$.

Le même algorithme peut être décrit de manière explicitement récursive et plus compacte par :

$$\langle 0 \rangle_{\frac{p}{q}} \;= \varepsilon \tag{4.2a}$$

$$\forall m > 0 \qquad \langle m \rangle_{\frac{p}{q}} \;= \langle n \rangle_{\frac{p}{q}}\, a \quad \text{où} \quad n \in \mathbb{N},\ a \in [\![p]\!] \ \text{ et } \ q\,m \;=\; p\,n + a \tag{4.2b}$$

**Exemple 4.1** – *Soient $p = 3$ et $q = 2$. La représentation d'un entier en base $\frac{3}{2}$ est donc écrite sur l'alphabet $[\![3]\!] = \{0, 1, 2\}$. La figure 2 montre l'application de l'algorithme d'Euclide modifié au calcul de $\langle 13 \rangle_{\frac{3}{2}}$ : passer d'une ligne entourée à la suivante correspondant à une application de (4.2b) et consiste en une multiplication par $2(= q)$ suivie d'une division euclidienne par $3(= p)$ ; la ligne entourée en trait plain correspond à la fin de l'algorithme après une application de (4.2a).*

*La figure 3 montre la table des $\frac{3}{2}$-représentations des petits entiers.*

Une simple récurrence montre que si $\langle n \rangle_{\frac{p}{q}} = a_k \cdots a_1\, a_0$, alors

$$n \;=\; \sum_{i=0}^{k} \frac{a_i}{q} \left(\frac{p}{q}\right)^i \;. \tag{4.3}$$

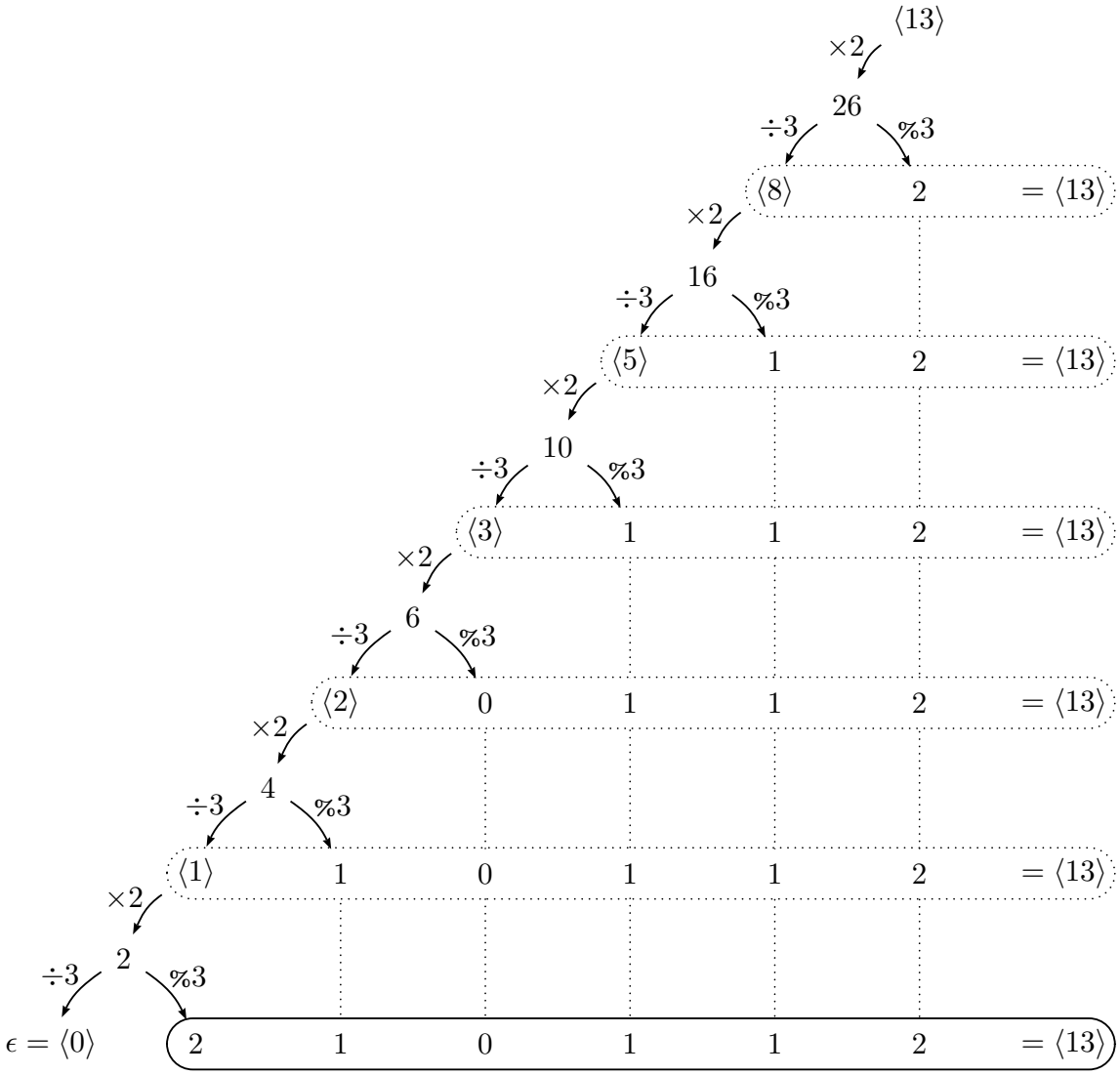



Figure 2 – Calcul de $\langle 13 \rangle_{\frac{p}{q}}$ en base $\frac{3}{2}$ avec l'algorithme d'Euclide modifié

| $N$ | $\langle N \rangle_{\frac{p}{q}}$ |
|---|---|
| 0 | $\varepsilon$ |
| 1 | 2 |
| 2 | 21 |
| 3 | 210 |
| 4 | 212 |
| 5 | 2101 |
| 6 | 2120 |
| 7 | 2122 |
| 8 | 21011 |
| 9 | 21200 |

| $N$ | $\langle N \rangle_{\frac{p}{q}}$ |
|---|---|
| 10 | 21202 |
| 11 | 21221 |
| 12 | 210110 |
| 13 | 210112 |
| 14 | 212001 |
| 15 | 212020 |
| 16 | 212022 |
| 17 | 212211 |
| 18 | 2101100 |
| 19 | 2101102 |

| $N$ | $\langle N \rangle_{\frac{p}{q}}$ |
|---|---|
| 20 | 2101121 |
| 21 | 2120010 |
| 22 | 2120012 |
| 23 | 2120201 |
| 24 | 2120220 |
| 25 | 2120222 |
| 26 | 2122111 |
| 27 | 21011000 |
| 28 | 21011002 |
| 29 | 21011021 |

Figure 3 – Les $\frac{3}{2}$-représentations des trentes premiers nombres entiers

La *fonction* $\pi_{\frac{p}{q}}$ *d'évaluation en base rationnelle* est dérivée de cette formule ; la *valeur* d'un mot $u = a_k \cdots a_1\, a_0 \in [\![p]\!]^*$ est

$$\pi_{\frac{p}{q}}(u) \;=\; \pi_{\frac{p}{q}}(a_k \cdots a_1\, a_0) \;=\; \sum_{i=0}^{k} \frac{a_i}{q} \left(\frac{p}{q}\right)^i \;. \tag{4.4}$$

Contrairement à la base entière, tous les mots de $[\![p]\!]^*$ n'ont pas une valeur entière, comme par exemple le mot $10^k$ qui s'évalue à $\frac{p^k}{q^{(k+1)}}$ en base $\frac{p}{q}$. On peut toutefois calculer la valeur d'un mot récursivement, en utilisant l'une des formules suivantes :

$$\forall a \in [\![p]\!]\ ,\ \forall u \in [\![p]\!]^* \qquad \pi_{\frac{p}{q}}(a\, u) \;=\; \frac{a}{q}\left(\frac{p}{q}\right)^{|u|} +\; \pi_{\frac{p}{q}}(u) \tag{4.5a}$$

$$\forall v \in [\![p]\!]^*,\ \forall a \in [\![p]\!] \qquad \pi_{\frac{p}{q}}(v\, a) \;=\; \pi_{\frac{p}{q}}(v)\left(\frac{p}{q}\right) +\; \frac{a}{q} \tag{4.5b}$$

$$\forall u, v \in [\![p]\!]^* \qquad \pi_{\frac{p}{q}}(v\, u) \;=\; \pi_{\frac{p}{q}}(v)\left(\frac{p}{q}\right)^{|u|} +\; \pi_{\frac{p}{q}}(u) \tag{4.5c}$$

Tout mot de valeur $n$ est égal à $\langle n \rangle_{\frac{p}{q}}$ à des zéros de tête près, comme l'exprime le théorème suivant.

**Théorème 4.2** – *Soient deux mots $u, v \in [\![p]\!]^*$. Si $\pi_{\frac{p}{q}}(u) = \pi_{\frac{p}{q}}(v)$, alors $u = 0^i\, v$ ou $v = 0^i\, u$ pour un certain entier $i$.*

Démonstration. On suppose d'abord que $u$ et $v$ ont la même longueur ; on note cette longueur $k = |u| = |v|$, $u = a_{(k-1)} \cdots a_1\, a_0$ et $v = b_{(k-1)} \cdots b_1\, b_0$.

D'après les hypothèses, $(\pi_{\frac{p}{q}}(v) - \pi_{\frac{p}{q}}(u)) = 0$, donc

$$\forall j \in \{0, 1, \ldots, k-1\}$$

$$0 \;\equiv\; \left(q^{k+1}(\pi_{\frac{p}{q}}(v) - \pi_{\frac{p}{q}}(u))\right) \quad [p^{(j+1)}]$$

$$0 \;\equiv\; \left(\sum_{i=0}^{j} p^i\, q^{(k-i)}(b_i - a_i)\right) \quad [p^{(j+1)}] \tag{$*$}$$

Démontrons par récurrence sur $j \leqslant k$ que pour tout entier $i < j$, $a_i = b_i$ ; le cas $j = 0$ ne demande donc aucune démonstration.

Soit un entier $(j+1)$ tel que $0 < (j+1) \leqslant k$. Par hypothèse de récurrence, pour tout $i < j$, $a_i = b_i$ ; il reste donc à démontrer que $a_j = b_j$. L'équation $(*)$ devient donc (puisque tous les autres termes de la somme sont nuls)

$$\left[p^j q^{(k-j)}(b_j - a_j)\right] \;\equiv\; 0 \quad [p^{(j+1)}]\ .$$

Il s'ensuit que $q^{(k-j)}(b_j - a_j) \equiv 0\ [p]$, et puisque $q^{(k-j)}$ est premier avec $p$, que $b_j \equiv a_j\ [p]$. Enfin, puisque $0 \leqslant a_j, b_j < p$, $a_j = b_j$ ce qui conclut la récurrence.

Les mot $u$ et $v$ ayant un rôle symétrique dans le théorème, s'ils ont une longueur différente, on peut supposer que $|u| > |v|$. On note $i = |u| - |v|$, si bien que les mots $u$ et $0^i\, v$ ont la même longueur. Par hypothèse $\pi_{\frac{p}{q}}(u) = \pi_{\frac{p}{q}}(v)$, et il découle de

l'équation (4.5c) que $\pi_{\frac{p}{q}}(0^k\, v) = \pi_{\frac{p}{q}}(v)$ (puisque $\pi_{\frac{p}{q}}(0^k) = 0$) donc les mots $u$ et $0^i\, v$ ont la même valeur. Appliquer le cas précédent implique donc que $u = 0^i\, v$. □

On étend, en utilisant le théorème 4.2, la notion de $\frac{p}{q}$-représentation aux nombres non-entiers (sur lesquels l'algorithme d'Euclide modifié n'a pas de sens). Un mot $u$ est la $\frac{p}{q}$-représentation d'un nombre $x$ si $u$ ne commence pas par un 0 et si $\pi_{\frac{p}{q}}(u) = x$. Inversement, un nombre $x$ est dit *représentable* en base $\frac{p}{q}$ s'il possède une $\frac{p}{q}$-représentation et similairement un ensemble $X$ est dit *représentable* s'il existe un langage $L$ tel que $\pi_{\frac{p}{q}}(L) = X$.

**Remarque 4.3** – *Si l'on fixe $q = 1$ (bien que cela soit interdit par hypothèse), la base rationnelle $\frac{p}{1}$ et la base entière $p$ ont la même définition. En effet, les fonction d'évaluation $\pi_{\frac{p}{1}}$ et $\pi_p$ ont alors exactement la même expression tout comme l'algorithme d'Euclide (page 34) et l'algorithme d'Euclide modifié coïncident.*

*En revanche, il n'existe pas en base rationnelle d'équivalent à l'algorithme glouton (page 34 ou page 236). C'est pourquoi la base rationnelle n'est pas une $\beta$-numération où $\beta$ serait égal à $\frac{p}{q}$ (cf.* [36]*), ni un système positionnel où la base serait $\left(U_i = \frac{1}{q}\left(\frac{p}{q}\right)^i\right)_{i\in\mathbb{N}}$ (ibidem).*

# Le langage $L_{\frac{p}{q}}$

Le langage $L_{\frac{p}{q}}$, défini ci-dessous, est le point focal de nos travaux sur la base rationnelle. Nous verrons tout au long de ce mémoire qu'il possède des propriétés qui révèlent une certaine régularité (comme par exemple la manière de le construire présenté dans le chapitre 8). En revanche, $L_{\frac{p}{q}}$ est paradoxalement à une position élevé dans la théorie des langages et on observe en son sein des suites qui semblent relever de répartitions aléatoires (sans toutefois qu'aucune propriété de ce type ne soit démontrée à notre connaissance).

**Définition 4.4** – *On note $L_{\frac{p}{q}} \subseteq [\![p]\!]^*$ le langage des $\frac{p}{q}$-représentations des entiers :*

$$L_{\frac{p}{q}} \;=\; \{\, \langle n\rangle_{\frac{p}{q}} \mid n \in \mathbb{N} \,\}\ .$$

La prochaine propriété est essentiellement une conséquence de l'algorithme d'Euclide modifié (équation (4.2)).

**Propriété 4.5** – *Le langage $L_{\frac{p}{q}}$ est clos par préfixe et prolongeable (à droite).*

Démonstration. Soient un mot $u \in L_{\frac{p}{q}}$ tel que $u \neq \varepsilon$ ; on note $n > 0$ l'entier tel que $\langle n\rangle_{\frac{p}{q}} = u$.

D'après l'équation (4.2b), $\langle n\rangle_{\frac{p}{q}} = \langle n'\rangle_{\frac{p}{q}} a$ pour un certain entier $n'$ et une certaine lettre $a \in [\![p]\!]$ ; donc $L_{\frac{p}{q}}$ est clos par préfixe.

[36] Christiane Frougny et Jacques Sakarovitch, 2010, *Number representation and finite automata.*

Puisque $[\![p]\!]$ est un intervalle de longueur $p > q$, il existe (au moins) une lettre $a \in [\![p]\!]$ telle que $(pn + a) \equiv 0\ [q]$. On note $m = \frac{pn+a}{q}$, qui est donc un entier supérieur à $n$, donc non-nul. Il découle de l'équation (4.2b) que $\langle m \rangle_{\frac{p}{q}} = \langle n \rangle_{\frac{p}{q}} a$ ; donc $L_{\frac{p}{q}}$ est prolongeable. □

**Exemple 4.6** – *La figure 4 représente le langage $L_{\frac{3}{2}}$ comme un arbre étiqueté. Si un mot $u$ étiquette le chemin $0 \xrightarrow{u} n$ alors $u = \langle n \rangle_{\frac{p}{q}}$. Par exemple on peut vérifier que $\langle 13 \rangle_{\frac{p}{q}} = 2\,1\,0\,1\,1\,2$ étiquette le chemin $0 \longrightarrow 13$.*

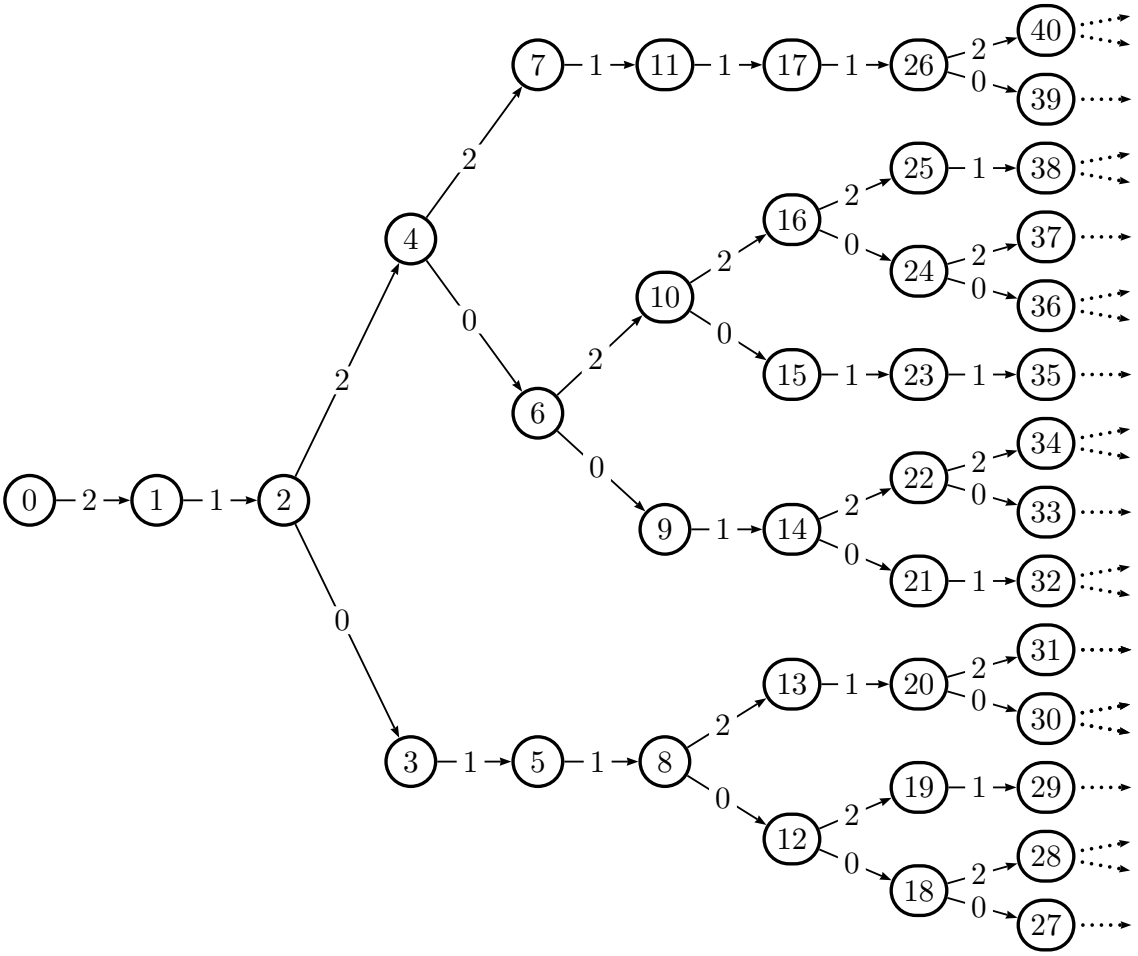



Figure 4 – Le langage $L_{\frac{3}{2}}$, représenté comme un arbre étiqueté

La base $\frac{p}{q}$ est un système de numérations abstrait (ANS, *cf.* section 1.7, page 27) ; en effet, les représentations des entiers respectent l'ordre radiciel (noté $<_{\text{rad}}$, *cf.* définition 1.16 dans la même section), comme l'énonce la proposition suivante.

**Proposition 4.7** – *Soient deux entiers $n$ et $m$. Ils vérifient $n < m$ si et seulement si $\langle n \rangle_{\frac{p}{q}} <_{rad} \langle m \rangle_{\frac{p}{q}}$.*

Démonstration. On note $\langle n\rangle_{\frac{p}{q}} = a_k \cdots a_1\, a_0$ et $\langle m\rangle_{\frac{p}{q}} = b_i \cdots b_1\, b_0$.

Sens réciproque. Puisque $\langle n\rangle_{\frac{p}{q}} <_{\text{rad}} \langle m\rangle_{\frac{p}{q}}$, le premier est plus court ou aussi long que le second : $k \leqslant i$.

Démontrons par récurrence sur $k$ que $n < m$. Si $k = 0$, alors $\langle n\rangle_{\frac{p}{q}} = \varepsilon$ donc $n = 0$ ; puisque $\langle n\rangle_{\frac{p}{q}} <_{\text{rad}} \langle m\rangle_{\frac{p}{q}}$, $\langle m\rangle_{\frac{p}{q}}$ n'est pas le mot vide donc

$$m \;=\; \pi_{\frac{p}{q}}(\langle m\rangle_{\frac{p}{q}}) \;>\; 0 \;=\; n\ .$$

Supposons maintenant que $k \geqslant 1$ (donc $i \geqslant 1$). Puisque $L_{\frac{p}{q}}$ est clos par préfixe les nombres $n' = \pi_{\frac{p}{q}}(a_k \cdots a_1)$ et $m' = \pi_{\frac{p}{q}}(b_k \cdots b_1)$ sont des entiers. D'autre part, il découle de l'équation (4.5b) que

$$(n-m) \;=\; \frac{p}{q}(n'-m') + \frac{1}{q}(a_0 - b_0)\ . \qquad (*)$$

Puisque $\langle n\rangle_{\frac{p}{q}} <_{\text{rad}} \langle m\rangle_{\frac{p}{q}}$, on est dans l'un des deux cas suivants.

- Si $\langle n'\rangle_{\frac{p}{q}} <_{\text{rad}} \langle m'\rangle_{\frac{p}{q}}$, il découle alors de l'hypothèse de récurrence que $(n'-m') > 0$ et donc $(n'-m') \geqslant 1$. Quelles que soient les lettres $a_0$ et $b_0$, l'entier $(a_0 - b_0)$ vérifie $-(p-1) \leqslant (a_0 - b_0) \leqslant (p-1)$ ce qui implique que
$$\mathsf{abs}\left(\frac{1}{q}(a_0 - b_0)\right) \;<\; \frac{p}{q} \;\leqslant\; \frac{p}{q}(n'-m')\ .$$
Il découle alors de $(*)$ que $(n-m) < 0$.
- Si $\langle n'\rangle_{\frac{p}{q}} = \langle m'\rangle_{\frac{p}{q}}$ et $a_0 < b_0$, alors $n' = m'$ ; il découle donc de $(*)$ que
$$(n-m) \;=\; \frac{1}{q}\,(a_0 - b_0) \;<\; 0\ .$$

Le sens direct découle du sens réciproque car l'ordre radiciel est total. □

**Corollaire 4.8** – *Le système à base $\frac{p}{q}$ est le système de numération abstrait calable $0^*L_{\frac{p}{q}}$.*

**Corollaire 4.9** – *Soient deux mots $u, v \in 0^*L_{\frac{p}{q}}$ de la même longueur. Alors $u <_{rad} v$ si et seulement si $\pi_{\frac{p}{q}}(u) < \pi_{\frac{p}{q}}(v)$.*

Le langage $L_{\frac{p}{q}}$ est complexe eu égard à la hiérarchie usuelle des langages (dite de Chomsky, voir par exemple [19]), comme l'exprime le théorème suivant.

**Théorème 4.10** – *Le langage $L_{\frac{p}{q}}$ n'est pas algébrique.*

En réalité, $L_{\frac{p}{q}}$ est encore bien plus compliqué que cela, il satisfait la *propriété de l'itération préfixe finie (FLIP, pour* Finite Left Iteration Property*)* qui sera définie dans la section 5.1 du chapitre 5. La démonstration du théorème 4.10 demande le reste de cette sous-partie ; elle est de plus volontairement segmentée en des résultats qui seront utiles par la suite.

[19] Olivier Carton, 2008, *Langages formels, calculabilité et complexité*.

**Définition 4.11** – *On note $\mathcal{T}_{\frac{p}{q}}$ l'automate infini*

$$\mathcal{T}_{\frac{p}{q}} \;=\; \langle\, \mathbb{N},\, [\![p]\!], \delta, 0, \mathbb{N} \,\rangle \;,$$

*où la fonction de transition $\delta$ est définie par :*

$$\forall n, m \in \mathbb{N},\ \forall a \in [\![p]\!] \qquad n \xrightarrow[\mathcal{T}_{\frac{p}{q}}]{a} m \qquad \textit{si et seulement si} \qquad q\,m \;=\; p\,n + a \;. \tag{4.6}$$

Par exemple, si on ajoute à la figure 4 une boucle étiquetée par 0 sur le sommet 0 et qu'on le rend initial, alors on obtient l'automate $\mathcal{T}_{\frac{3}{2}}$. La différence entre $\mathcal{T}_{\frac{p}{q}}$ et $L_{\frac{p}{q}}$ est essentiellement une question de formalisme ; au chapitre 7, ces deux objets seront même identifiés. Pour l'heure, $\mathcal{T}_{\frac{p}{q}}$ permet d'une part l'utilisation de la terminologie des automates et d'autre part de retirer la singularité de $L_{\frac{p}{q}}$ en (l'état) 0.

**Propriété 4.12** – *Soit un mot $u \in [\![p]\!]^*$ étiquetant un chemin $n \xrightarrow{u} m$ de $\mathcal{T}_{\frac{p}{q}}$.*

**a)** *Si $n = 0$, alors $u = 0^k \langle m \rangle_{\frac{p}{q}}$, pour un certain $k \in \mathbb{N}$.*

**b)** *Si $n > 0$ ou si $u$ ne commence par par un 0, alors $\langle n \rangle_{\frac{p}{q}}\, u = \langle m \rangle_{\frac{p}{q}}$ .*

**c)** $L(\mathcal{T}_{\frac{p}{q}}) = 0^* L_{\frac{p}{q}}$

Démonstration. Le point (**a**) est la conséquence du point (**b**) qui découle de la simple comparaison de l'équation (4.6) avec l'équation (4.2) qui donne une définition récursive de l'algorithme d'Euclide modifié. Le point (**c**) découle du point (**a**). □

Puisque $p$ et $q$ sont premiers entre eux, les propriétés suivantes sont des conséquences directes de l'équation (4.6).

**Propriété 4.13** – *Soient deux transitions $n \xrightarrow{a} m$, $n' \xrightarrow{a} m'$ de $\mathcal{T}_{\frac{p}{q}}$ et un entier $k$.*

**a)** *Les entiers $n$ et $n'$ sont congrus modulo $q^{(k+1)}$ si et seulement si $m$ et $m'$ sont congrus modulo $q^k$.*

**b)** *Les entiers $n$ et $n'$ sont congrus modulo $p^k$ si et seulement si $m$ et $m'$ sont congrus modulo $p^{(k+1)}$.*

Le résultat suivant peut être exprimé intuitivement par : *la classe de congruence d'un entier modulo $q^k$ détermine (et est déterminée par) son futur de longueur $k$ dans $\mathcal{T}_{\frac{p}{q}}$.*

**Lemme 4.14** – *Soient deux entiers $n$, $n'$ et un mot $u \in [\![p]\!]^*$ dont on note la longueur $k$.*

**a)** *Si $n$ et $n'$ sont congrus modulo $q^k$, alors $(n \cdot u)$ existe si et seulement si $(n' \cdot u)$ existe.*

**b)** *Si $(n \cdot u)$ et $(n' \cdot u)$ existent, alors $n$ et $n'$ sont congrus modulo $q^k$.*

Démonstration. **a)** Par récurrence sur $k$ ; le résultat est trivialement vérifié pour $k = 0$. Soit un entier $(k+1)$ tel que $n$ et $n'$ sont congrus modulo $q^{(k+1)}$. Soit un mot $u = a\,v \in [\![p]\!]^+$ de longueur $(k+1)$.

Si $(p\,n + a)$ n'est pas divisible par $q$, alors $(p\,n' + a)$ ne l'est pas non plus (car $n \equiv n'\ [q]$) et donc ni $(n \cdot a)$ ni $(n' \cdot a)$ n'existe.

Autrement, $(pn+a)$ est divisible par $q$ ce qui implique que $(pn'+a)$ l'est aussi. On note $m = (n \cdot a)$ et $m' = (n' \cdot a)$ ; d'après la propriété 4.13a, il s'ensuit que $m \equiv m'[q^k]$. Donc, en appliquant l'hypothèse de récurrence, $(m \cdot v)$ existe si et seulement si $(m' \cdot v)$ existe, ce qui implique que $(n \cdot a\,v)$ existe et seulement si $(n' \cdot a\,v)$ existe.

**b)** Par récurrence sur $k$ ; le résultat est trivialement vérifié pour $k = 0$. Soient un entier $(k+1)$ et un mot $u = a\,v \in [\![p]\!]^+$ de longueur $(k+1)$ tels que $(n' \cdot a\,v)$ et $(n \cdot a\,v)$ existent On note $m = (n \cdot a)$ et $m' = (n' \cdot a)$, ce qui implique que $(m \cdot v)$ et $(m \cdot v)$ existent. Il découle donc de l'hypothèse de récurrence que $m \equiv m'\ [q^k]$. Si bien que, d'après la propriété 4.13a, $n \equiv n\ [q^k + 1]$. □

On montre symétriquement que la classe de congruence d'un entier modulo $p^k$ détermine (et est déterminée par) son *passé* de longueur $k$ dans $\mathcal{T}_{\frac{p}{q}}$.

**Lemme 4.15** – *Soient deux entiers $m$, $m'$ et un mot $u \in [\![p]\!]^*$ dont on note la longueur $k$.*

**a)** *Si $m$ et $m'$ sont congrus modulo $p^k$, alors $n \xrightarrow{u} m$ pour un certain entier $n$ si et seulement si $n' \xrightarrow{u} m'$, pour un certain $n'$.*

**b)** *Si $n \xrightarrow{u} m$ et $n' \xrightarrow{u} m'$, pour certains $n$ et $n'$, alors $m$ et $m'$ sont congrus modulo $p^k$.*

Les deux lemmes précédents impliquent les deux corollaires suivants. Le premier découle du lemme 4.14 et signifie que $\mathcal{T}_{\frac{p}{q}}$ est un automate "minimal", dans le sens où deux états distincts ne sont pas Nérode-équivalents ; autrement dit, la seule équivalence régulière à droite qui sature $L_{\frac{p}{q}}$ est l'identité. Le second est déduit par un simple argument de comptage à partir du lemme 4.15.

**Corollaire 4.16** – *Pour tous états distincts $n$ et $m$ de $\mathcal{T}_{\frac{p}{q}}$, il existe un mot $u \in [\![p]\!]^*$ tel que $(n \cdot u)$ existe et $(m \cdot u)$ n'existe pas.*

**Corollaire 4.17** – *Tout mot $u \in [\![p]\!]^*$ est le suffixe d'un mot de $L_{\frac{p}{q}}$.*

D'autre part, le lemme 4.14 permet de montrer que $L_{\frac{p}{q}}$ ne contient pas de langage régulier infini.

**Proposition 4.18** – *Soient trois mots $u, v, w \in [\![p]\!]^*$ tels que $v \neq \varepsilon$. Le langage $u\,v^*\,w$ n'est pas inclus dans $L_{\frac{p}{q}}$.*

Démonstration par l'absurde. Supposons qu'il existe un tel triplet $u, v, w$, tel que $u\,v^*\,w \subseteq L_{\frac{p}{q}}$ ; donc pour tout entier $i \in \mathbb{N}$, le calcul de $u\,v^i\,w$ existe dans $\mathcal{T}_{\frac{p}{q}}$. On note $n = \pi_{\frac{p}{q}}(u)$ et $n' = \pi_{\frac{p}{q}}(u\,v)$.

Puisque $(n \cdot v^i)$ et $(n' \cdot v^i)$ existent pour tout $i$, il découle du lemme 4.14b,

$$\forall i \in \mathbb{N} \qquad n \;\equiv\; n' \quad [q^{(i\,|v|)}]\ . \tag{$*$}$$

On fixe $i$ de telle sorte que $q^{(i\,|v|)}$ soit supérieur à $n$ et $n'$ ; il découle donc de l'équation précédente que $n = n'$.

Le seul circuit de $\mathcal{T}_{\frac{p}{q}}$ est $0 \xrightarrow{0} 0$. L'existence du chemin

$$0 \xrightarrow{u} n \xrightarrow{v} n$$

implique donc que $n = 0$ et que $u\,v = 0^k$ pour un certain entier $k$ qui vérifie $k > 0$ puisque $|v| > 0$. En définitive, le mot $u\,v\,w$ commence par $k$ (donc au moins un) 0, donc n'appartient pas à $L_{\frac{p}{q}}$. Contradiction. □

Puisque $L_{\frac{p}{q}}$ est clos par préfixe et infini, le théorème 4.10 est une conséquence de cette proposition.

Démonstration du théorème 4.10. Par l'absurde. Supposons que $L_{\frac{p}{q}}$ est un langage algébrique. Puisque $L_{\frac{p}{q}}$ est infini, d'après le lemme d'itération pour les langages algébriques (*cf.* lemme 1.15), il existe cinq mots $u, x, v, y, w \in [\![p]\!]^*$ avec $xy \neq \varepsilon$ tels que pour tout $n \in \mathbb{N}$ $(u\,x^n\,v\,y^n\,w) \in L_{\frac{p}{q}}$.

Puisque $L_{\frac{p}{q}}$ est clos par préfixe (propriété 4.5), il contient le langage $(u\,x^*)$ ce qui implique, d'après la proposition 4.18, que $x = \varepsilon$. Il s'ensuit d'une part (puisque $x\,y \neq \varepsilon$) que $y \neq \varepsilon$ et d'autre part que $L_{\frac{p}{q}}$ contient le langage $(u\,v\,y^*)$, ce qui contredit la proposition 4.18. □

# Additionneur, normalisateur

Comme c'est le cas en base entière (voir section 2.2), l'addition est réalisée en base rationnelle par un transducteur (droit,) fini, lettre-à-lettre et séquentiel, appelé *additionneur*. Il s'agit une fois encore, d'un cas particulier du *normalisateur* qui prend en entrée un mot sur un alphabet fini arbitraire et renvoie en sortie un mot de même valeur sur l'alphabet canonique $[\![p]\!]$.

Dans la suite, on fixe un alphabet $D \subset \mathbb{Z}$ fini. La fonction $\pi_{\frac{p}{q}}$ est naturellement étendue sur $D^*$ et plus généralement, elle pourra dans la suite s'appliquer à n'importe quel mot sur un alphabet de chiffres. On appelle *fonction de normalisation* $\chi_D$ la fonction partielle définie par

$$\begin{aligned} \mathsf{Dom}(\chi_D) &= \{\, u \mid \exists v \in [\![p]\!]^* \quad \pi_{\frac{p}{q}}(u) = \pi_{\frac{p}{q}}(v) \,\} \\ \chi_D : \; D^* &\to [\![p]\!]^* \\ u &\mapsto v, \text{ le mot le plus court tel que } |v| \geqslant |u| \text{ et } \pi_{\frac{p}{q}}(u) = \pi_{\frac{p}{q}}(v). \end{aligned} \tag{4.7}$$

**Théorème 4.19** – *La fonction $\chi_D$ est réalisée par un transducteur (droit,) fini, lettre-à-lettre et séquentiel $\mathcal{C}_D$.*

Le transducteur dont il est question dans le théorème est la partie accessible du transducteur $\mathcal{C}_D$, défini ci-dessous.

**Définition 4.20** – *Pour tout alphabet $D \subset \mathbb{Z}$ fini, le normalisateur*

$$\mathcal{C}_D = \langle\, \mathbb{Z},\, D,\, [\![p]\!],\, 0,\, \delta,\, \eta,\, \omega \,\rangle \tag{4.8a}$$

*est le transducteur droit et infini dont l'alphabet d'entrée est $D$, l'alphabet de sortie est $[\![p]\!]$, les transitions sont définies par*

$$\forall s, s' \in \mathbb{Z},\ \forall d \in D,\ \forall a \in [\![p]\!] \qquad s' \xleftarrow{d \mid a} s \iff q\,s + d = p\,s' + a \tag{4.8b}$$

*et la fonction finale par*

$$\forall s \in Z \quad \begin{cases} \omega(s) \;=\; \langle s \rangle_{\frac{p}{q}} & \text{si} \;\; s \geqslant 0 \\ \omega(s) \text{ n'est pas défini} & \text{si} \;\; s < 0 \end{cases} \tag{4.8c}$$

La figure 5 représente le normalisateur de $\mathcal{C}_{D_1}$, où $D_1 = \{-1, 0, 1, 3\}$ ; les états inaccessibles (et leurs transitions sortantes) sont tiretés et grisés.

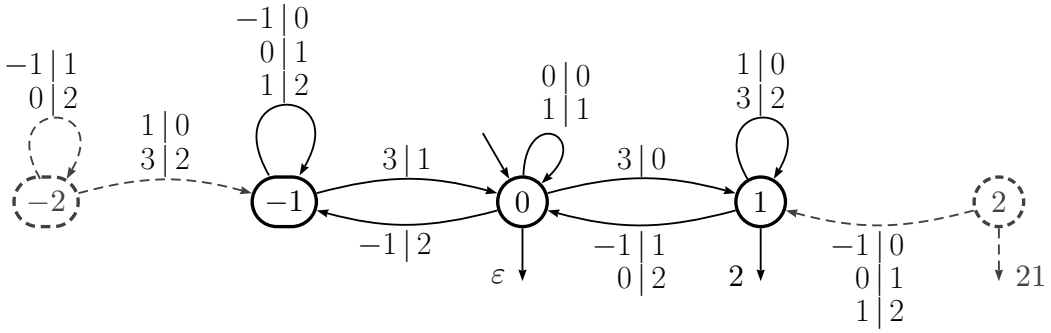



Figure 5 – Le normalisateur $\mathcal{C}_{D_1}$ en base $\frac{3}{2}$, où $D_1 = \{-1, 0, 1, 3\}$

La démonstration du théorème 4.19 consiste à montrer que $\mathcal{C}_D$ est un transducteur lettre-à-lettre et séquentiel (proposition 4.21) qui réalise la fonction $\chi_D$ (proposition 4.23) et que sa partie accessible est finie (proposition 4.24).

**Proposition 4.21** – *Le normalisateur est un transducteur lettre-à-lettre, séquentiel et complet en entrée.*

Démonstration. Il découle directement de la définition 4.20 que $\mathcal{C}_D$ est lettre-à-lettre.

Soit un état $s \in \mathbb{Z}$ et une lettre $d \in D$. La division euclidienne de $(q\,s + d)$ par $p$ donne l'unique couple $(s', a) \in \mathbb{Z} \times [\![p]\!]^*$ tel que $q\,s + d = p\,s + a$. Il existe donc une unique lettre $a \in [\![p]\!]^*$ et un unique état $s' \in \mathbb{Z}$ tels que $s' \xleftarrow{d \mid a} s$ (équation (4.8b)). En d'autres termes, le normalisateur est séquentiel et complet en entrée. □

Démontrer que le normalisateur $\mathcal{C}_D$ réalise la fonction $\chi_D$ nécessite le lemme 4.22 suivant qui décrit son fonctionnement interne.

**Lemme 4.22** – *Soient deux états $s, t$ de $\mathcal{C}_D$ et deux mots $u \in D^*$ et $v \in [\![p]\!]$. Le chemin $t \xleftarrow{u \mid v} s$ existe dans $\mathcal{C}_D$, si et seulement si $\pi_{\frac{p}{q}}(u) + s = \pi_{\frac{p}{q}}(v) + t\,(\frac{p}{q})^{|u|}$ et $|u| = |v|$.*

Démonstration. Sens direct. Puisque $u \mid v$ étiquette un chemin de $\mathcal{C}_D$, un transducteur lettre à lettre, $|u| = |v|$. L'autre équation se démontre par récurrence sur la longueur de $u$ et de $v$. Le résultat est évidemment vrai si $u = \varepsilon$ (et donc $v$ aussi).

Soit un chemin de $\mathcal{C}_D$

$$t \xleftarrow{u \mid v} s' \xleftarrow{d \mid a} s \quad ; \tag{$*$}$$

il s'ensuit, d'après l'hypothèse de récurrence, que

$$\pi_{\frac{p}{q}}(u) + s' \;=\; \pi_{\frac{p}{q}}(v) + t\left(\frac{p}{q}\right)^{|u|} \;. \qquad (**)$$

D'autre part, d'après l'équation (4.8b), la transition droite du chemin $(*)$ est équivalente (après division des deux côté par $q$) à

$$s + \frac{d}{q} \;=\; \frac{p}{q}s' + \frac{a}{q} \;. \qquad (\overset{*}{**})$$

La suite de calculs suivante conclut la démonstration du sens direct.

$$\begin{array}{rcll}
\pi_{\frac{p}{q}}(ud) + s & = & \frac{p}{q}\pi_{\frac{p}{q}}(u) + \frac{d}{q} + s & \text{d'après eq. (4.5b)} \\
& = & \frac{p}{q}\pi_{\frac{p}{q}}(u) + \quad s + \frac{d}{q} & \text{d'après eq. } (\overset{*}{**}) \\
& = & \frac{p}{q}\,\pi_{\frac{p}{q}}(u) + \;\frac{p}{q}\,s' + \frac{a}{q} & \\
& = & \frac{p}{q}(\quad \pi_{\frac{p}{q}}(u) + s' \quad) + \frac{a}{q} & \text{d'après eq. } (**) \\
& = & \frac{p}{q}(\;\pi_{\frac{p}{q}}(v) + t(\frac{p}{q})^{|u|}\;) + \frac{a}{q} & \\
& = & (\frac{p}{q}\pi_{\frac{p}{q}}(v) + \frac{a}{q}) + t(\frac{p}{q})^{(|u|+1)} & \text{d'après eq. (4.5b)} \\
& = & \quad \pi_{\frac{p}{q}}(va) \quad + t(\frac{p}{q})^{|ud|} &
\end{array}$$

Sens réciproque. Soient deux états $s, t$ de $\mathcal{C}_D$ et deux mots $u \in D^*$ et $v \in [\![p]\!]$ de même longueur vérifiant

$$\pi_{\frac{p}{q}}(u) + s \;=\; \pi_{\frac{p}{q}}(v) + t\left(\frac{p}{q}\right)^{|u|} \;.$$

Puisque le normalisateur est séquentiel et complet en entrée (proposition 4.21), il existe un unique $v' \in [\![p]\!]^*$ et un unique état $t' \in \mathbb{Z}$ tels que $|v'| = |u| (= |v|)$ et

$$t' \xleftarrow{u \mid v'} s \;.$$

Le sens direct du lemme implique alors que

$$\pi_{\frac{p}{q}}(u) + s \;=\; \pi_{\frac{p}{q}}(v') + t'\left(\frac{p}{q}\right)^{|u|} \;.$$

Il s'ensuit, en multipliant les deux membres par $q^{|u|}$, que

$$q^{|u|}\pi_{\frac{p}{q}}(v') + t'\,p^{|u|} \;=\; q^{|u|}(\pi_{\frac{p}{q}}(u) + s) \;=\; q^{|u|}\pi_{\frac{p}{q}}(v) + t\,p^{|u|}$$

si bien que

$$\forall i \leqslant |u| \qquad q^{|u|}\pi_{\frac{p}{q}}(v') \;\equiv\; q^{|u|}\pi_{\frac{p}{q}}(v) \quad [p^i]$$

Un raisonnement par récurrence (identique à celui utilisé dans la démonstration du théorème 4.2 page 94) montre que $v = v'$, ce qui implique que $t = t'$. □

**Proposition 4.23** – *Le transducteur $\mathcal{C}_D$ réalise la fonction $\chi_D$.*

Démonstration. Soit un mot $u \in D^*$ dont on note $w = \mathcal{C}_d(u) \in [\![p]\!]^*$ l'image par $\mathcal{C}_D$. On note le calcul de $u$ dans $\mathcal{C}_D$ par

$$t \xleftarrow{u \mid v} 0 \;,$$

ce qui implique que

$$w \;=\; \langle t \rangle_{\frac{p}{q}}\, v \quad \text{et} \quad |v| = |u| \;. \tag{$*$}$$

Il découle du lemme 4.22 précédent appliqué à $s = 0$ que

$$\pi_{\frac{p}{q}}(u) \;=\; \pi_{\frac{p}{q}}(v) + t \left(\frac{p}{q}\right)^{|u|}$$

dont le membre de droite est égal à $\pi_{\frac{p}{q}}(\langle t \rangle_{\frac{p}{q}}\, v)$ (équation (4.5c)). Il découle de $(*)$ que

$$\pi_{\frac{p}{q}}(w) \;=\; \pi_{\frac{p}{q}}(\langle t \rangle_{\frac{p}{q}}\, v) \;=\; \pi_{\frac{p}{q}}(u) \quad \text{et} \quad |w| \geqslant |v| = |u| \;,$$

donc que le mot $w$ appartient à l'ensemble

$$\left\{ w' \in [\![p]\!]^* \;\middle|\; \pi_{\frac{p}{q}}(w') \;=\; \pi_{\frac{p}{q}}(u) \quad \text{et} \quad |w'| \geqslant |u| \right\} \;.$$

Si $t \neq 0$, alors $w$ ne commence pas par un 0 (car $\langle t \rangle_{\frac{p}{q}} \neq \varepsilon$) et puisque deux mots de même valeur ne diffèrent que par des 0 de tête (théorème 4.2) tous les mots qui ont pour valeur $\pi_{\frac{p}{q}}(w)(= \pi_{\frac{p}{q}}(u))$ sur l'alphabet $[\![p]\!]^*$ sont de la forme $0^i\, w$, donc plus long que $w$. Au contraire si $t = 0$, alors $\langle t \rangle_{\frac{p}{q}} = \varepsilon$, donc $w = v$, ce qui implique que $|w| = |v| = |u|$ (équation droite de $(*)$). Dans les deux cas, $w$ est le mot le plus court de l'ensemble

$$\left\{ w' \in [\![p]\!]^* \;\middle|\; \pi_{\frac{p}{q}}(w') \;=\; \pi_{\frac{p}{q}}(u) \quad \text{et} \quad |w'| \geqslant |u| \right\} \;. \qquad \square$$

La proposition suivante est la dernière étape de la démonstration du théorème 4.19.

**Proposition 4.24** – *La partie accessible de $\mathcal{C}_D$ est finie.*

Démonstration. On note $e$ le plus grand chiffre de $D$. Soit $u$ un mot d'une certaine longueur $k$ dont le chemin dans $\mathcal{C}_D$ est $s \xleftarrow{u \mid v} 0$. D'après le lemme 4.22, $\pi_{\frac{p}{q}}(u) = \pi_{\frac{p}{q}}(v) + s \left(\frac{p}{q}\right)^k$, ce qui implique, puisque $v \in [\![p]\!]^*$, que

$$s \left(\tfrac{p}{q}\right)^k \;\leqslant\; \pi_{\frac{p}{q}}(u) \;. \tag{$*$}$$

D'autre part, chacun des chiffres de $u$ est inférieur à $e$ donc

$$\pi_{\frac{p}{q}}(u) \;\leqslant\; \pi_{\frac{p}{q}}(e^k) \;. \tag{$**$}$$

Il découle des équations $(*)$ et $(**)$ que

$$s \;\leqslant\; \left(\frac{p}{q}\right)^{-k} \pi_{\frac{p}{q}}(e^k) \;=\; \left(\frac{p}{q}\right)^{-k} \frac{e}{q} \left( \sum_{i=0}^{k-1} \left(\frac{p}{q}\right)^i \right) \;=\; \frac{e}{q} \left( \sum_{i=0}^{k-1} \left(\frac{p}{q}\right)^{-(i+1)} \right) \;.$$

En résumé, un mot $u$ de longueur $k$ ne peut atteindre depuis l'état initial un état plus grand que $\frac{e}{q} \sum_{i=0}^{k-1} \left(\frac{p}{q}\right)^{-(i+1)}$ ; or, cette série converge, donc les états accessibles sont majorés. Un raisonnement similaire utilisant la plus petite lettre de $D$ montre que les états accessibles sont également minorés. $\square$

**L'additionneur** est le cas particulier du normalisateur où $D = [\![2p-1]\!]$ en utilisant une addition chiffre-à-chiffre préalable (voir section 2.2). Par exemple, celui de la base $\frac{3}{2}$ est représenté par la figure 6.

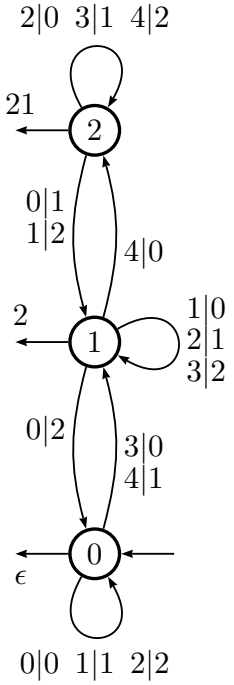



Figure 6 – Le normalisateur $\mathcal{C} = \mathcal{C}_{[\![2\,p-1]\!]}$ en base $\frac{3}{2}$, duquel est tiré l'additionneur

On utilisera bien plus souvent l'additionneur que le normalisateur dans les preuves mais on représentera uniquement le normalisateur (à l'instar de la figure 6) car l'additionneur possède beaucoup trop de transitions ; une définition directe de l'additionneur (en tant que transducteur de $([\![p]\!] \times [\![p]\!])^*$ vers $[\![p]\!]^*$) est donc donnée ci-dessous. (L'ensemble de ses états est $\mathbb{N}$ plutôt que $\mathbb{Z}$ car $[\![2p-1]\!]$ ne contient pas de chiffre négatif, donc aucun état négatif ne peut être accessible.)

**Définition 4.25** – *L'additionneur $\mathcal{C}$ est le transducteur*

$$\mathcal{C} \;=\; \langle\, \mathbb{N},\, [\![p]\!] \times [\![p]\!],\, [\![p]\!]^*,\, 0,\, \delta,\, \eta,\, \omega \,\rangle$$

*dont les transitions sont définies par*

$$\forall s \in \mathbb{N},\;\; \forall d, d' \in [\![p]\!] \qquad s' \xleftarrow[\mathcal{C}]{(d,\,d')\;\mid\; a} s \iff q\,s + (d+d') = p\,s' + a \tag{4.9a}$$

*et la fonction finale est définie par*

$$\forall s \in \mathbb{N} \qquad \omega(s) \;=\; \langle s \rangle_{\frac{p}{q}}\ . \tag{4.9b}$$

L'additionneur étant une spécialisation du convertisseur, le théorème suivant est donc une reformulation et une conséquence du théorème 4.19.

**Théorème 4.26** – *L'additionneur $\mathcal{C}$ est un transducteur droit, fini, lettre-à-lettre et séquentiel qui réalise la fonction $\chi_+$ définie par :*

$$\begin{aligned} \chi_+ :\; ([\![p]\!] \times [\![p]\!])^* &\longrightarrow [\![p]\!]^* \\ u,\; v &\longmapsto w,\ \textit{le mot le plus court tel que } \begin{cases} |w| \;\geqslant\; |u|\ (=\ |v|) \\ \pi_{\frac{p}{q}}(w) \;=\; \pi_{\frac{p}{q}}(u) + \pi_{\frac{p}{q}}(v) \end{cases} \end{aligned}$$

De plus, son fonctionnement interne est décrit par le lemme suivant, qui est une reformulation du lemme 4.22.

**Lemme 4.27** – *Soient deux états $s, t$ de $\mathcal{C}_D$ et trois mots $u, v, w \in [\![p]\!]^*$ de la même longueur. Le chemin $t \xleftarrow{(u,v) \mid w} s$ existe dans $\mathcal{C}$, si et seulement si*

$$\pi_{\frac{p}{q}}(u) + \pi_{\frac{p}{q}}(v) + s \;=\; \pi_{\frac{p}{q}}(w) + t\left(\frac{p}{q}\right)^{|u|} .$$

Enfin, la proposition suivante montre que l'on peut relever l'action de l'additionneur aux représentations.

**Proposition 4.28** – *Soient deux nombres $x, y \in \mathbb{Q}$ représentables en base $\frac{p}{q}$ ; on note $i$ la différence de leurs longueurs : $i = |\langle y\rangle_{\frac{p}{q}}| - |\langle x\rangle_{\frac{p}{q}}|$. Si $i > 0$, alors, le nombre $(x + y)$ est représentable en base $\frac{p}{q}$ et*

$$\langle x + y\rangle_{\frac{p}{q}} \;=\; \mathcal{C}\left(0^i\langle x\rangle_{\frac{p}{q}},\ \langle y\rangle_{\frac{p}{q}}\right) .$$

Démonstration. On note respectivement $u = 0^i\langle x\rangle_{\frac{p}{q}}$ et $v = \langle y\rangle_{\frac{p}{q}}$ (de telle sorte que $|u| = |v|$). L'additionneur est complet en entrée et tous ses états sont finals, si bien que $w = \mathcal{C}(u, v)$ existe ; il découle du théorème 4.26 que $\pi_{\frac{p}{q}}(w) = x + y$ ce qui implique que $(x + y)$ est représentable.

L'assertion suivante conclut la démonstration.

**Assertion 4.28.1** – *Le mot $w$ ne commence pas par un 0.*

*Démonstration de l'assertion.* On note le calcul de $(u, v)$ dans $\mathcal{C}$ par

$$t \xleftarrow{(u,v) \mid w'} 0 ,$$

de telle sorte que $w = \langle t\rangle_{\frac{p}{q}}\, w'$.

Si $t \neq 0$, alors $w$ commence par $\langle t\rangle_{\frac{p}{q}} \neq \varepsilon$ donc $w$ ne commence par par un 0 ; on suppose dans la suite que $t = 0$. Si $u = \varepsilon$, alors $v = \varepsilon$ (car $|u| = |v|$) et donc $w = \mathcal{C}_{\frac{p}{q}}(\varepsilon, \varepsilon) = \varepsilon$, un mot qui ne commence pas par un 0 ; on suppose dans la suite que $|u| = |v| > 0$.

On note $d$ et $d$' les premières lettres respectives de $u$ et $v$. Il existe $s \in \mathbb{N}$ et $a \in [\![p]\!]$ tel que

$$(t =)0 \xleftarrow{(d,d') \mid a} s$$

est la dernière transition prise par le calcul de $(u, v)$ dans $\mathcal{C}$. Il découle de la définition 4.25 que

$$q\, s + d + d' \;=\; p \times 0 + a .$$

Or,

- $d' > 0$ car c'est la lettre qui commence le mot $v = \langle y\rangle_{\frac{p}{q}}$ ;
- $s \geqslant 0$ car c'est un état de $\mathcal{C}$ ;
- $d \geqslant 0$ car c'est une lettre de $[\![p]\!]^*$.

Aux vues de ces trois inéquations, il découle de l'équation précédente que $a > 0$. Or $a$ est la première lettre de $w'$ donc également la première lettre de $w$, puisque $\langle t\rangle_{\frac{p}{q}} = \langle 0\rangle_{\frac{p}{q}} = \varepsilon$. □

## Mots minimaux, évaluation après la virgule

Dans cette section, on manipule des mots infinis. Leurs lettres sont indexés de gauche à droite, au contraire des mots finis.

À chaque mot infini sur l'alphabet $[\![p]\!]$ est donnée une valeur réelle par la fonction d'évaluation *après la virgule*, définie par :

$$\begin{array}{rccc} \rho_{\frac{p}{q}} : & [\![p]\!]^\omega & \longrightarrow & \mathbb{R} \\ & a_0\, a_1 \cdots a_k \cdots & \longmapsto & \displaystyle\sum_{i\geqslant 0} \frac{a_i}{q} \left(\frac{p}{q}\right)^{-(i+1)} \end{array} \tag{4.10}$$

Si $\mathcal{A}$ est un automate (fini ou infini), on note $\mathfrak{L}(\mathcal{A})$ l'ensemble des mots **infinis** acceptés par $\mathcal{A}$, c'est-à-dire l'ensemble des mots infinis dont chaque préfixe admet un calcul (pas nécessairement acceptant) dans $\mathcal{A}$. On rappelle que l'ordre préfixe sur les mots est noté $\sqsubseteq$ (ou $\sqsubset$ s'il est strict).

On note $W_{\frac{p}{q}}$ la clôture topologique de $(0^* L_{\frac{p}{q}})$, c'est-à-dire :

$$W_{\frac{p}{q}} \;=\; \mathfrak{L}(\mathcal{T}_{\frac{p}{q}}) \;=\; \{\, w \in [\![p]\!]^\omega \mid \forall u \sqsubset w,\ u \in 0^* L_{\frac{p}{q}} \,\}\ . \tag{4.11}$$

**Exemple 4.29** – *La figure 7 est une représentation de $W_{\frac{3}{2}}$ comme un arbre fractal, c'est-à-dire l'arbre résultant du dépliage de la boucle sur l'état 0 de $\mathcal{T}_{\frac{p}{q}}$. Sur la figure, si un chemin est $0 \xrightarrow{u} n$ alors $n = \pi_{\frac{p}{q}}(u)$ et l'ordonnée du nœud atteint par ce chemin est $\rho_{\frac{p}{q}}(u\,0^\omega)$. Par exemple, les mots 21 et 210 atteignent respectivement des sommets étiquetés par $2 = \pi_{\frac{p}{q}}(21)$ et $3 = \pi_{\frac{p}{q}}(210)$ qui ont la même ordonnée $0.888\cdots = \rho_{\frac{p}{q}}(210\,0^\omega) = \rho_{\frac{p}{q}}(21\,0^\omega)$.*

Les deux prochains lemmes donnent des propriétés élémentaires de $\rho_{\frac{p}{q}}$.

**Lemme 4.30** – *La fonction $\rho_{\frac{p}{q}}$ est continue de $[\![p]\!]^\omega$ vers $\mathbb{R}$.*

Démonstration. Soient un mot $w \in [\![p]\!]^\omega$ et un voisinage $V \subseteq [\![p]\!]^\omega$ de $w$ ; on note $x = \rho_{\frac{p}{q}}(w)$. Pour tout entier $i$ on note $u_i$ le préfixe de $w$ de longueur $i$. Tout mot $w' \in [\![p]\!]^\omega$ qui admet $u_i$ comme préfixe s'évalue après la virgule dans l'intervalle $[\rho_{\frac{p}{q}}(u_i 0^\omega),\ \rho_{\frac{p}{q}}(u_i (p-1)^\omega)]$. Cet intervalle est de taille $(\sum_{k=i}^{\infty} \frac{p-1}{q} \left(\frac{p}{q}\right)^{-(k+1)})$ donc tend vers 0 quand $i$ tend vers $+\infty$. □

**Lemme 4.31** – *La fonction $\rho_{\frac{p}{q}}$ préserve l'ordre de $\left(W_{\frac{p}{q}}, \leqslant_{rad}\right)$ vers $(\mathbb{R}, \leqslant)$.*

Démonstration. Soient deux mots $u, v \in W_{\frac{p}{q}}$ tels que $u <_{\text{rad}} v$. Pour tout entier $i$, on note $u_i$ et $v_i$ les préfixes de longueurs $i$ respectifs de $u$ et $v$. Il s'ensuit que $u_i \leqslant_{\text{rad}} v_i$ donc que $\pi_{\frac{p}{q}}(u_i) \leqslant \pi_{\frac{p}{q}}(v_i)$ (corollaire 4.9). Or, puisque

$$\rho_{\frac{p}{q}}(u_i\, 0^\omega) \;=\; \left(\frac{p}{q}\right)^{-|u_i|} \pi_{\frac{p}{q}}(u_i)$$

il s'ensuit que $\rho_{\frac{p}{q}}(u_i\, 0^\omega) \;\leqslant\; \rho_{\frac{p}{q}}(v_i\, 0^\omega)$. Le lemme découle alors de la continuité de $\rho_{\frac{p}{q}}$. □

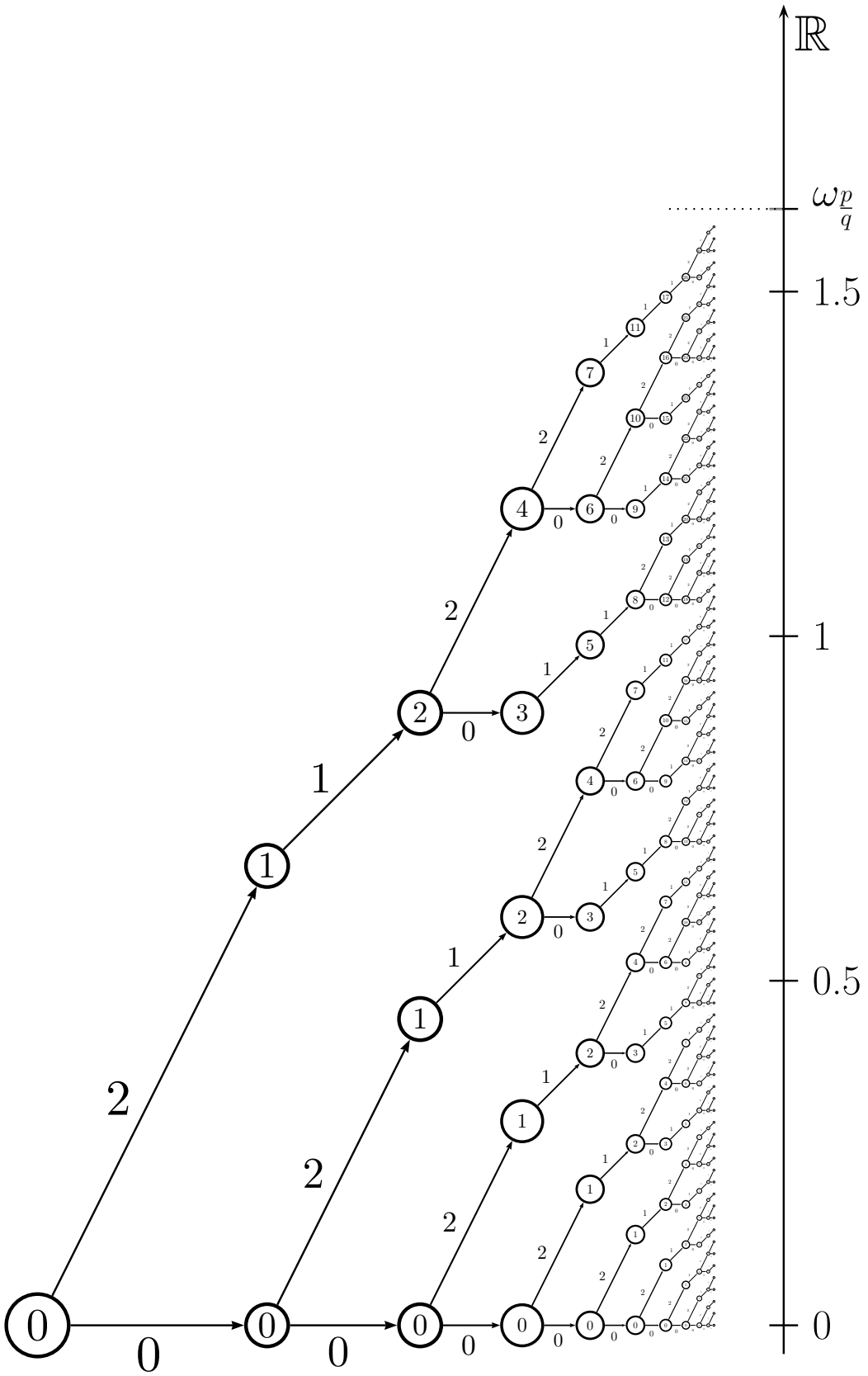



Figure 7 – L'ensemble $W_{\frac{3}{2}}$ et l'évaluation après la virgule

On dit qu'un mot infini $w$ *étiquette une branche* de $\mathcal{T}_{\frac{p}{q}}$ si il existe un entier $n$ tel que, pour tout préfixe (fini) $u$ de $w$, $(n \cdot u)$ existe. En particulier, les mots de $W_{\frac{p}{q}}$ sont les mots qui étiquettent les branches de $\mathcal{T}_{\frac{p}{q}}$ partant de la racine 0.

On appelle *alphabet minimal* l'ensemble $[\![q]\!]$ formé des $q$ plus petits entiers (c'est-à-dire formé des $q$ plus petites lettres de $[\![p]\!]$). Il découle de la définition même de $\mathcal{T}_{\frac{p}{q}}$ (équation (4.6)) que pour tout état $n > 0$, il existe une unique lettre $a \in [\![q]\!]$ telle que $(n \cdot a)$ existe ; on appelle cette lettre *l'étiquette minimale sortante* de $n$.

**Définition 4.32** – *Soit un entier $n$. On appelle* mot minimal de $n$, *noté $w_n^-$, l'unique mot infini de $[\![q]\!]^\omega$ qui étiquette une branche de $\mathcal{T}_{\frac{p}{q}}$ partant de $n$.*

*On note $\Omega_{\frac{p}{q}}$ l'ensemble de tous les mots minimaux.*

Puisque le mot minimal de $n$ est unique, tout mot $u$ fini sur l'alphabet $[\![q]\!]$ tel que $(n \cdot u)$ existe est un préfixe de $w_n^-$ ou, autrement dit,

$$\forall n \in \mathbb{N},\ \forall q \in [\![p]\!]^* \qquad (n \cdot u) \text{ existe dans } \mathcal{T}_{\frac{p}{q}} \quad \Longrightarrow \quad u \sqsubset w_n^- \,. \tag{4.12}$$

Quelques propriétés élémentaires de l'ensemble $\Omega_{\frac{p}{q}}$ sont rassemblées ci-dessous.

**Propriété 4.33** –

**a)** *L'ensemble $\Omega_{\frac{p}{q}}$ est dénombrable.*

**b)** *La fonction qui associe son mot minimal à chaque entier $n$ est injective*

**c)** *La clôture topologie de $\Omega_{\frac{p}{q}}$ est $[\![q]\!]^\omega$ tout entier.*

Démonstration. Chaque mot minimal est défini par un entier $n$, ce qui implique immédiatement le point (**a**).

**b)** Supposons au contraire qu'il existe deux entiers $n$ et $m$ tels que $w_n^- = w_m^-$. Il alors découle du lemme 4.14b que $n$ et $m$ sont congrus modulo $q^k$, pour tout entier $k$, ce qui est impossible.

**c)** Soit $w$ un mot infini de $[\![q]\!]^\omega$ et $u$ le préfixe de $w$ de longueur $i$. D'après le corollaire 4.17, il existe un entier $m$ tel que $u$ est le suffixe d'un mot de $L_{\frac{p}{q}}$. On note $n$ l'entier tel que $n \xrightarrow{u} m$ ; il s'ensuit que $u$ est un préfixe de $w_n^-$. □

On définit symétriquement l'*alphabet maximal* $\{\,(p-q),\,(p-q+1),\,\ldots,(p-1)\,\}$, l'*étiquette maximale sortante* et le *mot maximal de $n$*, noté $w_n^+$. Étudier les mots minimaux ou les mots maximaux revient au même, comme le suggère la relation exprimé par le lemme suivant.

**Lemme 4.34** – *Soit un entier $n$. Pour tout entier $i$, si l'on note $a$ la $i$-ème lettre de $w_n^+$ et $b$ la $i$-ème lettre de $w_{(n+1)}^-$, alors $b = a - (p - q)$.*

Démonstration. Le résultat est obtenu par une induction basée sur l'assertion suivante.

**Assertion 4.34.1** – *Soit $n \xrightarrow{a} m$ une transition de $\mathcal{T}_{\frac{p}{q}}$. Si $a$ est une lettre maximale, alors $b = (a - p + q)$ est une lettre minimale et $(n+1) \xrightarrow{b} (m+1)$ est une transition de $\mathcal{T}_{\frac{p}{q}}$.*

*Démonstration de l'assertion.* D'après l'équation (4.6), l'existence de la transition $n \xrightarrow{a} m$ implique que

$$q\,m \;=\; p\,n + a\;.$$

Une simple réécriture montre que

$$q\,(m+1) \;=\; p\,(n+1) + (a - p + q) \;=\; p\,(n+1) + b\;,$$

si bien que la transition $(n+1) \xrightarrow{b} (m+1)$ existe à condition que $b \in [\![p]\!]$. Puisque $a$ est une lettre minimale, elle appartient à $\{(p-q), (p-q+1), \ldots, (p-1)\}$ donc $b = (a - p + q)$ appartient à $[\![q]\!] \subseteq [\![p]\!]$. □

Ce lemme permet de démontrer la proposition suivante. Elle exprime que des couples spécifiques de branches de $\mathcal{T}_{\frac{p}{q}}$ sont étiquetées par des mots qui ont la même valeur.

**Proposition 4.35** – *Soient deux entiers $n, m$ et une lettre $a \in [\![p]\!]$ tels que $n \xrightarrow{a} m$ et $n \xrightarrow{a+q} (m+1)$ dans $\mathcal{T}_{\frac{p}{q}}$. Alors $\rho_{\frac{p}{q}}((a+q)\,w^-_{m+1}) = \rho_{\frac{p}{q}}(a\,w^+_m)$*

Démonstration. On note $w^+_m = a_0\,a_1\,a_2 \cdots a_k \cdots$ le mot maximal de $m$. D'après le lemme 4.34, le mot minimal de $(m+1)$ est égal à

$$w^-_{(m+1)} \;=\; (a_0 - (p-q))\,(a_1 - (p-q))\,(a_2 - (p-q)) \cdots (a_k - (p-q)) \cdots$$

Les évaluations après la virgule de ces deux mots sont donc les suivantes :

$$\rho_{\frac{p}{q}}(w^+_m) \;=\; \frac{1}{q} \sum_{i=0}^{\infty} a_i \left(\frac{p}{q}\right)^{-(i+1)}$$

$$\rho_{\frac{p}{q}}(w^-_{(m+1)}) \;=\; \frac{1}{q} \sum_{i=0}^{\infty} (a_i - (p-q)) \left(\frac{p}{q}\right)^{-(i+1)} \;=\; \rho_{\frac{p}{q}}(w^+_m) - \left(\frac{p-q}{q} \sum_{i=0}^{\infty} \left(\frac{p}{q}\right)^{-(i+1)}\right)\;.$$

Or, $\sum_{i=1}^{\infty} \left(\frac{p}{q}\right)^{i}$ est la somme d'une suite géométrique, égale à $\frac{q}{p-q}$, ce qui implique que

$$\rho_{\frac{p}{q}}(w^-_{(m+1)}) \;=\; \rho_{\frac{p}{q}}(w^+_m) - 1\;.$$

Il s'ensuit la suite de calculs suivante qui conclut la démonstration.

$$\begin{array}{rcl}
\rho_{\frac{p}{q}}((a+q)\,w^-_{m+1}) & = & (\frac{a+q}{q})(\frac{p}{q})^{-1} \;+\; \rho_{\frac{p}{q}}(w^-_{m+1}) \times \left(\frac{p}{q}\right)^{-1} \\
& = & \frac{a}{q}(\frac{p}{q})^{-1} \;+\; (\frac{p}{q})^{-1} \;+\; \rho_{\frac{p}{q}}(w^-_{m+1}) \times \left(\frac{p}{q}\right)^{-1} \\
& = & \frac{a}{q}(\frac{p}{q})^{-1} \;+\; (\rho_{\frac{p}{q}}(w^-_{m+1}) + 1) \times \left(\frac{p}{q}\right)^{-1} \\
& = & \frac{a}{q}(\frac{p}{q})^{-1} \;+\; \rho_{\frac{p}{q}}(w^+_m) \times \left(\frac{p}{q}\right)^{-1} \\
& = & \rho_{\frac{p}{q}}(a\,w^+_m)
\end{array}$$

□

On note $\omega_{\frac{p}{q}}$ le plus grand nombre réel de $\rho_{\frac{p}{q}}(W_{\frac{p}{q}})$, c'est-à-dire $\omega_{\frac{p}{q}} = \rho_{\frac{p}{q}}(w^+_0)$, puisque $\rho_{\frac{p}{q}}$ préserve l'ordre (lemme 4.31).

**Théorème 4.36** – *L'image de $W_{\frac{p}{q}}$ par la fonction $\rho_{\frac{p}{q}}$ est l'intervalle $[0, \omega_{\frac{p}{q}}]$.*

Démonstration. Tout d'abord, $W_{\frac{p}{q}}$ contient $w^-_0$ et $w^+_0$ dont les évaluations après la virgule sont respectivement 0 et $\omega_{\frac{p}{q}}$.

Soit un mot de $w \in W_{\frac{p}{q}}$ ; il vérifie $w_0^- \leqslant_{\text{rad}} w \leqslant_{\text{rad}} w_0^+$ donc, puisque $\rho_{\frac{p}{q}}$ préserve l'ordre (lemme 4.31) $0 \leqslant \pi_{\frac{p}{q}}(w) \leqslant \omega_{\frac{p}{q}}$. Il s'ensuit que $\rho_{\frac{p}{q}}(W_{\frac{p}{q}}) \subseteq [0, \omega_{\frac{p}{q}}]$.

**Assertion 4.36.1** – *L'ensemble $W_{\frac{p}{q}}$ est topologiquement fermé.*

*Démonstration de l'assertion.* Soit $(w_i)_{i\in\mathbb{N}}$ une suite de mots à valeurs dans $W_{\frac{p}{q}}$ qui converge vers $w$. Par l'absurde. Supposons que $w$ ne soit pas dans $W_{\frac{p}{q}}$, il existe donc un entier $k$ tel que le préfixe $u$ de longueur $k$ de $w$ n'est pas dans $0^*L_{\frac{p}{q}}$. Puisque $(w_i)_i$ converge vers $w$, il existe $i$ tel que $w_i$ et $w$ ont le même préfixe de longueur $k$. Il s'ensuit que $u$ est le préfixe de $w_i$ qui appartient à $W_{\frac{p}{q}}$ donc $u \in 0^*L_{\frac{p}{q}}$. Contradiction.

Puisque $W_{\frac{p}{q}}$ est fermé, la topologie sur les mots infinis implique qu'il est compact (lemme 1.2 page 20) ; et puisque $\rho_{\frac{p}{q}}$ est continue, $\rho_{\frac{p}{q}}(W_{\frac{p}{q}})$ est un ensemble fermé de $\mathbb{R}$.

Par l'absurde. Supposons qu'il existe un réel $x$ dans $]0, \omega_{\frac{p}{q}}[$ qui n'appartient pas à $\pi_{\frac{p}{q}}(W_{\frac{p}{q}})$. On note

$$x^+ \;=\; \inf\{y > x \mid y \in \rho_{\frac{p}{q}}(W_{\frac{p}{q}})\} \quad \text{et} \quad x^- \;=\; \sup\{y < x \mid y \in \rho_{\frac{p}{q}}(W_{\frac{p}{q}})\}$$

qui vérifient donc

$$x^- \;<\; x \;<\; x^+ \,. \tag{$*$}$$

(Les inégalités sont strictes parce que $x \notin \rho_{\frac{p}{q}}(W_{\frac{p}{q}})$.)

Puisque $\pi_{\frac{p}{q}}(W_{\frac{p}{q}})$ est fermé, il existe deux mots infinis de $W_{\frac{p}{q}}$ qui s'évaluent respectivement à $x^+$ et $x^-$ ; on note $w^+$ (resp. $w^-$) le plus petit mot (resp. le plus grand) dans l'ordre radiciel tel que que $\rho_{\frac{p}{q}}(w^+) = x^+$ (resp. $\rho_{\frac{p}{q}}(w^-) = x^-$), donc

$$w^- \;<_{\text{rad}}\; w^+ \,.$$

De plus, il s'agit de deux mots consécutifs de $W_{\frac{p}{q}}$ dans l'ordre radiciel : s'il existait un mot $w' \in W_{\frac{p}{q}}$ tel que $w^- <_{\text{rad}} w' <_{\text{rad}} w^+$ alors $\pi_{\frac{p}{q}}(w')$ appartiendrait à $[x^-, x^+]$ (car $\rho_{\frac{p}{q}}$ respecte l'ordre) donc à $\{x^-, x^+\}$ (d'après les définitions de $x^-$ et $x^+$), ce qui contredirait la définition de $w^-$ ou $w^+$.

On note $u$ le plus grand préfixe commun de $w^+$ et $w^-$ ainsi que $a\,v^+$ et $b\,v^-$ les suffixes correspondant (donc vérifiant $a \neq b$), de telle sorte que

$$w^+ \;=\; u\,b\,v^+ \quad \text{et} \quad w^- \;=\; u\,a\,v^- \,.$$

On note de plus $n$ et $m$ les entiers tels que $0 \xrightarrow{u} n \xrightarrow{a} m$.

Puisque $b$ est une étiquette sortante de $n$ plus grande que $a$, il existe un entier $k > 0$ tel que $b = a + k\,q$. En particulier, $a + k\,q < p$ donc $(a + q) < p$ ce qui implique que la transition $n \xrightarrow{a+q} (m+1)$ existe. Puisque de plus $L_{\frac{p}{q}}$ est prolongeable, il existe un mot infini $v$ tel que $u\,(a+q)\,v$ appartient à $W_{\frac{p}{q}}$. Or,

$$w^- = u\,a\,v^- <_{\text{rad}} u\,(a+q)\,v$$

ce qui implique que $b = a + q$ ; en effet, dans le cas contraire $u(a+q)v$ serait strictement inférieur à $u\,b\,v^+ = w^+$ dans l'ordre radiciel, ce qui contredirait le fait que $w^-$ et $w^+$ sont deux mots consécutifs de $W_{\frac{p}{q}}$.

Un raisonnement analogue montre que $v^-$ est le mot maximal partant de $m$ et que $v^+$ est le mot minimal partant de $(m+1)$. Notez bien ici que les $+$ et les $-$ s'inversent dans les notations : $w^-_{(m+1)} = v^+$ et $w^+_m = v^-$.

Il découle donc de la proposition 4.35 que $\rho_{\frac{p}{q}}(av^-) = \rho_{\frac{p}{q}}(bv^+)$ donc que $\pi_{\frac{p}{q}}(ubv^+) = \pi_{\frac{p}{q}}(uav^-)$. Autrement, dit $x^- = x^+$, ce qui contredit $(*)$. □

# La variante FK

Une variante (que nous appellerons FK en l'honneur de ses auteurs) des systèmes de numération à base rationnelle est présentée dans [35]. Dans la suite, on utilisera presque exclusivement la version décrite plus tôt (appelée AFS pour les même raisons) ; la variante FK sera néanmoins parfois utile pour mettre en perspective certains résultats. Nous n'en donnons ici qu'une présentation succincte, voir l'article original pour plus de détails.

Soient deux entiers $p$ et $q$ premiers entre eux et qui vérifient $p > q > 1$. Dans la variante FK de la base $\frac{p}{q}$, la représentation d'un entier $N$ est le mot $\langle N \rangle_{\frac{p}{q}}^{\mathrm{FK}} = a_k\, a_{k-1} \cdots a_0$ dont les lettres sont calculées par l'algorithme suivant : $N_0 = N$ et

$$\forall i \qquad qN_i = pN_{(i+1)} + q\,a_i \qquad (\text{ au lieu de } q\,N_i = pN_{(i+1)} + a_i\ ) \ . \tag{4.13}$$

La figure 8 montre le langage $\Theta_{\frac{p}{q}}$ des représentations des entiers dans la variante FK de la base $\frac{3}{2}$. La première différence avec son homologue $L_{\frac{p}{q}}$ de la variante AFS (voir par exemple $L_{\frac{3}{2}}$ à la figure 4 page 96) est l'apparition de nombreuses "feuilles". En effet $\Theta_{\frac{p}{q}}$, bien que clos par préfixe, n'est pas prolongeable : $(q-1)$ sommets sur $q$ sont des feuilles.

La fonction d'évaluation $\theta_{\frac{p}{q}}$ induite par l'algorithme décrit par l'équation (4.13) est alors la suivante

$$\theta_{\frac{p}{q}}(a_k \cdots a_1\, a_0) \ = \ \sum_{i=0}^{k} a_i \left(\frac{p}{q}\right)^i \qquad (= q\,\pi_{\frac{p}{q}}(a_k \cdots a_1\, a_0)) \ . \tag{4.14}$$

Il s'ensuit que la variante FK est *un système de numération canonique* (*cf.* [36]) alors que la variante AFS ne l'est pas. Il découle également de l'équation (4.14) que si un mot $u$ est la représentation d'un nombre $x$ dans la variante AFS, alors c'est la représentation de $\frac{x}{q}$ dans la variante FK.

Un deuxième avantage de la variante FK est la simplicité de calculer (pour un être humain) la représentation d'un entier. On écrit un nombre $n$ dans la colonne des unités puis on applique une règle de "retenue" (qui généralise celle utilisée pour les bases entières) aussi longtemps que possible : si un chiffre est plus grand que $p$, on lui soustrait $p$ et on ajoute $q$ dans la colonne qui est à sa gauche. Par exemple, il faut appliquer cette règle sept fois pour calculer la représentation de 13, comme l'illustre la figure 9. Tous les résultats intermédiaires sont des mots sur un alphabet non-canonique mais dont la valeur est égale à 13 :

$$13 \ = \ 2\times\frac{3}{2}+10 \ = \ \cdots \ = \ 2\times\frac{9}{4}+1\times\frac{3}{2}+7 \ = \ \cdots \ = \ 2\times\frac{27}{8}+1\times\frac{9}{4}+2\times\frac{3}{2}+1 \ .$$

[35] Christiane Frougny et Karel Klouda, 2012, *Rational base number systems for p-adic numbers.*

[36] Christiane Frougny et Jacques Sakarovitch, 2010, *Number representation and finite automata.*

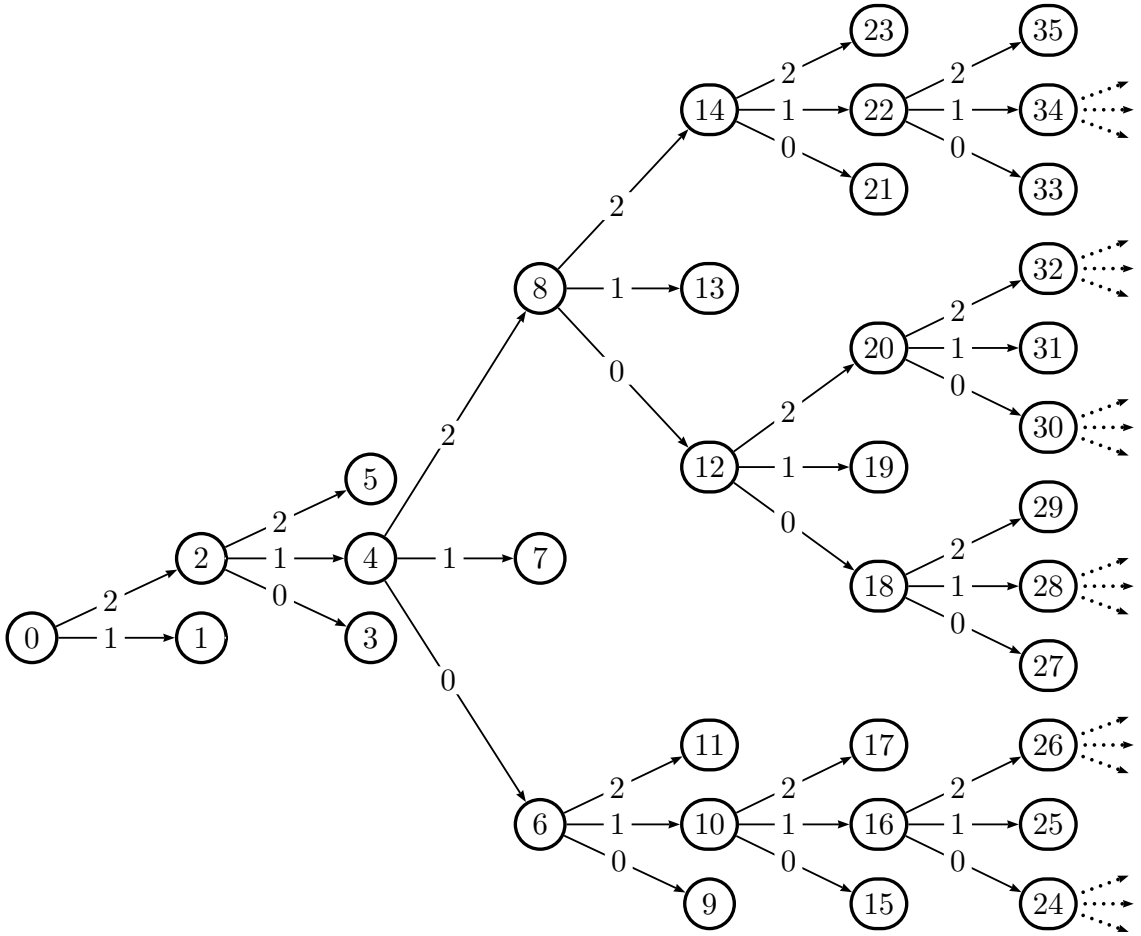



Figure 8 – Le langage $\Theta_{\frac{3}{2}}$

Cet algorithme montre bien qu'il est facile de calculer la représentation de $(n+1)$ quand on a celle de $n$ : on augmente le chiffre des unités de 1 et on propage la retenue éventuelle. La même remarque peut être faite pour la somme, qui est d'ailleurs réalisée par l'exact même transducteur $\mathcal{C}$ que pour la variante AFS.

Le même algorithme fonctionne pour la variante AFS, mais il faut inscrire le nombre $(q\,n)$ dans la colonne des unités au lieu de $n$. On calcule en fait la représentation de $(q\,n)$ dans la variante FK, qui est égale à celle de $n$ dans la variante AFS.

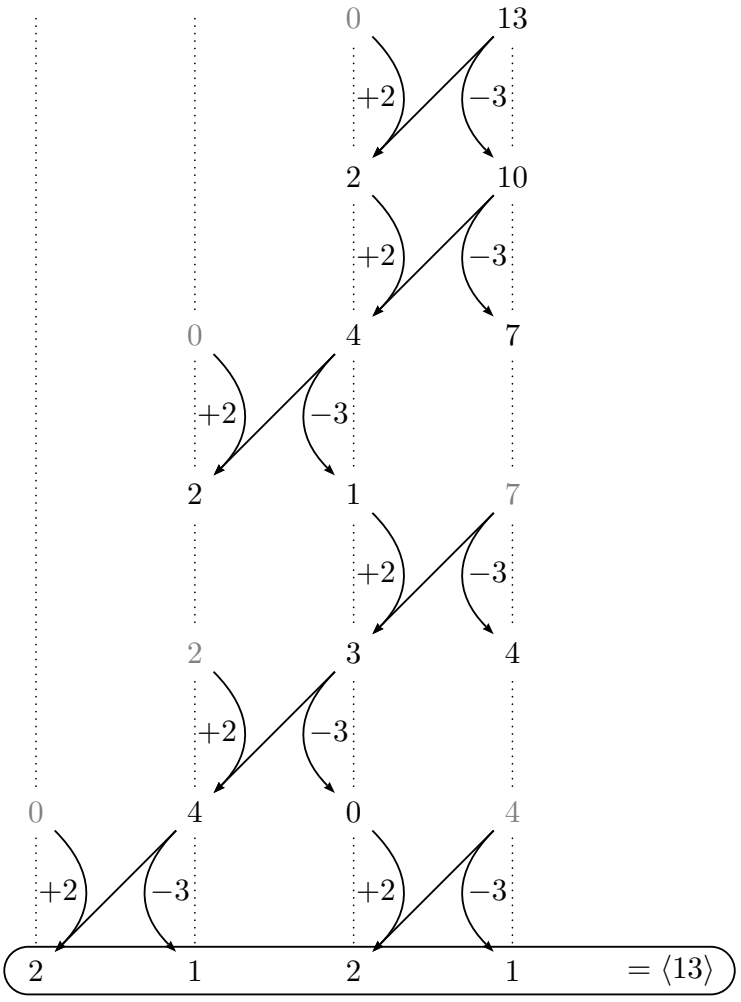



Figure 9 – Calcul de la représentation de 13 dans la variante FK de la base $\frac{3}{2}$

## Chapitre 5

# Monoïdes additifs en base rationnelle

Une des propriétés essentielles de la base rationnelle est l'existence d'un transducteur simple qui réalise l'addition (*cf.* section 4.3, page 100). Il est donc naturel d'étudier les ensembles de nombres qui sont à la fois représentables en base $\frac{p}{q}$ et stables par addition ; c'est-à-dire d'étudier les monoïdes additifs représentables en base $\frac{p}{q}$.

Par exemple, le monoïde $\mathbb{N}$ est représenté par le langage $L_{\frac{p}{q}}$. Comme nous l'avons vu précédemment, ce langage n'est ni régulier ni algébrique. Nous verrons dans la section 5.1 qu'il satisfait la *propriété FLIP*, une propriété de langage très robuste (stable entre autres par union, intersection et inclusion) qui exprime que le langage contredit toute forme de lemme d'itération.

D'autre part, on présente dans la section 5.2 l'ensemble $V_{\frac{p}{q}} = \pi_{\frac{p}{q}}([\![p]\!]^*)$ des nombres représentables en base $\frac{p}{q}$. Cet ensemble, bien que représenté par le langage régulier $[\![p]\!]^* \setminus (0\,[\![p]\!]^*)$, est un monoïde additif qui n'admet pas de famille génératrice finie et s'avère difficile à décrire sans utiliser la fonction d'évaluation $\pi_{\frac{p}{q}}$. On s'intéresse aux (sous-)monoïdes (additifs de $V_{\frac{p}{q}}$) et à leurs langages de représentations en base rationnelle.

On a vu qu'un certain monoïde simple ($\mathbb{N}$) est représenté par un langage complexe ($L_{\frac{p}{q}}$) alors qu'un monoïde complexe ($V_{\frac{p}{q}}$) est représenté par un langage simple. La section 5.3 est consacrée à montrer que ce paradoxe s'étend à toute une classe de monoïdes simples.

**Théorème II** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q > 1$. Tous les sous-monoïdes additifs et finiment engendrés de $V_{\frac{p}{q}}$ sont représentés en base $\frac{p}{q}$ par des langages FLIP.*

La démonstration du théorème II se fait en deux étapes. D'abord, on définit l'incrémenteur, un transducteur (droit, lettre-à-lettre et séquentiel) qui réalise l'addition par une constante (ou translation) et conserve la propriété FLIP. On montre ensuite que chaque monoïde finiment engendré est inclus dans une union de translations de $\mathbb{N}$ ; la classe des langages FLIP étant stable par ces opérations, cela conclut la démonstration.

---

Une partie des résultats présentés dans ce chapitre (sections 5.1 à 5.3) ont été publiés dans les actes de CAI 2013, voir `[53]`. Les autres résultats sont inédits.

Dans la section 5.4, on s'intéresse aux ensembles périodiques d'entiers. Ceux-ci sont tous représentés par des sous-langages de $L_{\frac{p}{q}}$ et sont donc tous FLIP.

Néanmoins, quand la période $n$ et le dénominateur $q$ de la base sont premiers entre eux, $q$ est un élément inversible de $\mathbb{Z}/n\mathbb{Z}$. Puisque tous les éléments de $V_{\frac{p}{q}}$ sont des nombres rationnels dont le dénominateur divise une puissance de $q$, on peut donc attribuer à chacun une classe de congruence modulo $n$ (c'est-à-dire un élément de $\mathbb{Z}/n\mathbb{Z}$). Pour tout ensemble $R \in \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$, on appelle *ensemble pseudo-périodique* de paramètre $(n, R)$ l'ensemble des nombres de $V_{\frac{p}{q}}$ dont la classe de congruence modulo $n$ appartient à $R$.

**Théorème (5.34)** – *Soient deux entiers $p$ et $q$ premiers entre eux tels que $p > q > 1$. Soient une période $n \in \mathbb{N}$ et un ensemble $R \in \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$. Si $n$ et $q$ sont premiers entre eux alors il existe un automate qui accepte par valeur l'ensemble pseudo-périodique de paramètre $(n, R)$.*

L'automate en question se construit de façon analogue à celui qui accepte l'ensemble périodique d'entiers correspondant en base entière.

On considère ensuite le cas où la période est égale à $q$. Dans ce cas, selon notre méthode, les nombres non-entiers de $V_{\frac{p}{q}}$ n'ont pas de classe de congruence modulo $q$. On démontre que les entiers d'une certaine classe de congruence modulo $q$ sont inséparables des autres.

**Théorème (5.36)** – *Soient deux entiers $p$ et $q$ premiers entre eux tels que $p > q > 1$. Soient $S$ un ensemble d'entiers périodique de plus petite période $q$ et $S'$ son complémentaire dans $\mathbb{N}$, respectivement représentés par $L = \langle S \rangle_{\frac{p}{q}}$ et $L' = \langle S' \rangle_{\frac{p}{q}}$. Alors, il n'existe pas de langage régulier $K \subseteq [\![p]\!]^*$ tel que $L \subseteq K$ et $L' \cap K = \emptyset$.*

Le théorème 5.36 provient intuitivement du fait que la classe de congruence modulo $q$ d'un entier $n$ détermine son futur, c'est-à-dire détermine les lettres $a$ que l'on peut concaténer à sa représentation tout en conservant une valeur entière ($\langle n \rangle_{\frac{p}{q}} a \in L_{\frac{p}{q}}$). Si bien que l'existence d'un oracle (l'automate $\mathcal{A}$) donnant la classe de congruence des entiers permet de construire un automate qui accepte $L_{\frac{p}{q}}$ et mène donc à une contradiction.

Nous conjecturons que le théorème 5.36 peut être généralisé à toute période qui n'est pas première avec $q$. Néanmoins, le raisonnement utilisé pour le démontrer est très spécifique et ne semble pas être facilement généralisable.

Dans la dernière section 5.5, on se demande combien et quels mots il faudrait ajouter à $L_{\frac{p}{q}}$ pour obtenir un langage régulier, tout en conservant la stabilité par addition. On appelle *approximation (par excès)* de $L_{\frac{p}{q}}$ un langage qui contient $L_{\frac{p}{q}}$ et qui est la représentation d'un (sous-)monoïde (additif de $V_{\frac{p}{q}}$).

Dans le cas où $q$ se factorise comme $q = k\,d$ avec $k$ et $d$ premiers entre eux, on définit un monoïde $M_d$ ; il est formé des nombres $x$ de $V_{\frac{p}{q}}$ dont la fraction irréductible $x = \frac{n}{m}$ possède un dénominateur $m$ premier avec $d$. Ce monoïde contient $\mathbb{N}$, donc sa représentation est une approximation de $L_{\frac{p}{q}}$ qui s'avère être un langage FLIP (quand $d \neq 1$) proche de $L_{\frac{p}{d}}$, la représentation des entiers en base $\frac{p}{\boldsymbol{d}}$ (et non $\frac{p}{q}$).

On présente ensuite une hiérarchie infinie d'approximations régulières (dans le sens où ce sont des langages réguliers) de $L_{\frac{p}{q}}$. Ces approximations sont triviales : elles ne conservent qu'une partie finie de la structure de $L_{\frac{p}{q}}$. Néanmoins, au vu de tous les résultats développés dans ce chapitre, nous conjecturons que chaque approximation régulière de $L_{\frac{p}{q}}$ est moins fine qu'un des membres de cette hiérarchie.

# Langage FLIP

Cette partie décrit une propriété de langages qui exprime intuitivement l'idée d'une irrégularité homogène.

**Définition 5.1** [1] – *Un langage $L$ sur un alphabet $A$ satisfait la* propriété d'itération préfixe finie*, ou est dit* FLIP *(pour Finite Left Iteration Property), si*

$$\forall u \in A^*,\ \forall v \in A^+,\ \exists i \in \mathbb{N} \qquad u\,v^i \notin \mathsf{Pre}(L)\ .$$

La proposition suivante donne une caractérisation de la propriété FLIP ; ce n'en est en fait qu'une simple réécriture.

**Proposition 5.2** – *Un langage $L$ est FLIP si et seulement si sa clôture topologique ne contient que des mots apériodiques.*

Il s'ensuit que tout langage formés de préfixes d'un mot apériodique est nécessairement FLIP.

**Exemple 5.3** – *Le langage $L_1 = \{\sigma^i(a) \mid i \in \mathbb{N}\}$ est constitué des itérations du morphisme de Fibonacci $\sigma : \{a, b\} \to \{a, b\}$ défini par $\sigma(a) = ab$ et $\sigma(b) = a$ :*

$$L_1 \;=\; \{\ \varepsilon,\ a,\ ab,\ aba,\ abaab,\ abaabab,\ abaababa,\ \dots\ \}\ .$$

*Ce langage satisfait même une propriété plus forte que FLIP, car il n'existe pas de mots $u \in A^*$ et $v \in A^+$ tel que $u\,v^4$ est dans $\mathsf{Pre}(L_1)$ (voir par exemple* [48]*).*

**Exemple 5.4** – *Le langage $L_2$, inspiré de la* constante de Champernowne [21] *et dont l'idée remonte à des questions étudiées par Barbier dès 1887* [10, 11]*, est sur l'alphabet $\{0, 1\}$ et est constitué des mots $u_i$, définis par : $u_0 = \varepsilon$, pour tout $i \in \mathbb{N}$, $u_{(i+1)} = u_i\langle i+1\rangle_2$ (c'est-à-dire la représentation de $i+1$ en base 2) et*

$$L_2 \;=\; \{\ \varepsilon,\ 1,\ 1\,1\,0,\ 1\,1\,0\,1\,1,\ \overbrace{\underbrace{1\,1\,0\,1\,1}_{u_3}\,\underbrace{1\,0\,0}_{\langle 4\rangle_2}}^{u_4},\ 1\,1\,0\,1\,1\,1\,0\,0\,1\,0\,1,\ 1\,1\,0\,1\,1\,1\,0\,0\,1\,0\,1\,1\,1\,0,\ \dots\ \}\ .$$

1. Cette notion a été introduite dans [53] sous l'appellation *Bounded Left Iteration Property* (BLIP). En réalité, le nombre maximal $(i-1)$ d'itérations du facteur $v$ dépend du préfixe $u$ considéré, il s'agit donc d'une itération *finie* et pas nécessairement *bornée*, ce qui donne en anglais *Finite Left Iteration Property* (FLIP).

[53] Victor Marsault et Jacques Sakarovitch, 2013, *On Sets of Numbers Rationally Represented in a Rational Base Number System.*
[48] M. Lothaire (collective), 2002, *Algebraic Combinatorics on Words.*
[21] David G. Champernowne, 1933, *The Construction of Decimals Normal in the Scale of Ten.*
[10] Èmile Barbier, 1887, *On suppose écrite la suite naturelle des nombres ; quel est le $(10^{1000})^{\text{ième}}$ chiffre écrit ?.*

La propriété FLIP est inspirée d'une propriété connue de $L_{\frac{p}{q}}$ (*cf.* [2] ou proposition 4.18, page 99) qui est donc sans surprise un langage FLIP.

**Proposition 5.5** [2] – *Pour tous entiers $p$ et $q$ premiers entre eux tels que $p > q > 1$, le langage $L_{\frac{p}{q}}$ est FLIP.*

Hormis les différents langages $L_{\frac{p}{q}}$ (et $\Theta_{\frac{p}{q}}$, leurs homologues dans la variante FK, *cf.* section 4.5, page 111), nous n'avons trouvé dans la littérature aucun exemple de langage qui est FLIP et dont le nombre de mots par longueur croît exponentiellement. L'exemple suivant montre que l'on peut néanmoins en construire sans trop de difficulté.

**Exemple 5.6** – *On considère l'alphabet $\{0, 1, \#\}$ et on définit récursivement les langages $K_i$ comme suit : $K_0 = \{\varepsilon\}$ et*

$$\forall i \in \mathbb{N} \qquad K_{(i+1)} = K_i \# \{0,1\}^i \ .$$

*Le langage $L_3 = \cup_i K_i$ est la réunion des $K_i$. La clôture topologique de $L_3$ est donc l'ensemble des mots infinis de la forme $u_0\#u_1\#u_2\#\cdots$ où, pour tout $i \in \mathbb{N}$, le mot $u_i \in \{0,1\}^*$ est de longueur $i$. Ces mots sont tous apériodiques, donc $L_3$ est FLIP.*

Un langage FLIP contredit toute forme raisonnable de lemme d'itération. En particulier, un langage FLIP (infini) ne peut-être rationnel, ni même algébrique.

**Proposition 5.7** – *Un langage ne peut être à la fois FLIP, algébrique et infini.*

Démonstration. Soit $L \subseteq A^*$ un langage algébrique et infini. D'après le lemme d'itération pour les langages algébriques (lemme 1.15, page 27), il existe cinq mots $u, v, w, x, y \in [\![p]\!]^*$ tels que $u\,x^i\,v\,y^i\,w \in L$ pour tout $n$ et que $|x\,y| \neq 0$. Si $x \neq \varepsilon$ alors $u\,x^* \subseteq \mathsf{Pre}\,(L)$ ; sinon, $x = \varepsilon$, ce qui implique que $y \neq \varepsilon$ et que $u\,v\,y^* \subseteq \mathsf{Pre}\,(L)$. Dans les deux cas, $L$ n'est pas un langage FLIP. □

La propriété FLIP est bien plus forte que ne le suggère la proposition précédente. Un langage qui n'est pas algébrique est un langage qui se comporte de façon non-algébrique *quelque part* alors qu'un langage FLIP (et infini) se comporte de façon non-algébrique *partout*.

La propriété FLIP se rapproche de la condition IRS (pour *infinite rational set*) introduite par Greibach dans [38] (voir aussi [8]) comme outil de classification des langages algébriques. Un langage est dit *IRS* s'il ne contient aucun langage régulier infini.

Un langage FLIP est nécessairement IRS, mais le contraire n'est pas vrai, comme illustré par l'exemple suivant.

---

[2] Shigeki Akiyama, Christiane Frougny et Jacques Sakarovitch, 2008, *Powers of rationals modulo 1 and rational base number systems.*

[38] Sheila A. Greibach, 1975, *One Counter Languages and the IRS Condition.*

[8] Jean-Michel Autebert, Joffroy Beauquier, Luc Boasson et Michel Latteux, 1982, *Indécidabilité de la Condition IRS.*

**Exemple 5.8** – *Le langage $\{a^i \mid i \text{ est un nombre premier}\}$ est IRS mais n'est pas FLIP, car sa clôture topologique est $a^\omega$. De la même manière, le langage $\{a^n b^n \mid n \in \mathbb{N}\}$ est IRS mais n'est pas FLIP alors que $\{a^n b^{\leqslant n} \mid n \in \mathbb{N}\}$ n'est ni IRS ni FLIP.*

La proposition suivante donne une caractérisation des langages FLIP en utilisant la condition IRS.

**Proposition 5.9** – *Un langage $L$ est FLIP si et seulement si $\mathsf{Pre}(L)$ est IRS.*

Démonstration. Soit un langage $L \subseteq A^*$. Le langage $L$ est fini si et seulement si $\mathsf{Pre}(L)$ est fini ; or les langages finis sont tous IRS et FLIP, ce qui conclut la preuve dans le cas où $L$ est fini. On supposera dans la suite que $L$ est infini.

Si $L$ n'est pas FLIP, il existe deux mots $u \in A^*$ et $v \in A^+$ tel que pour tout entier $i$, $uv^i \in \mathsf{Pre}(L)$ donc que $uv^*$ est un langage régulier inclus dans $\mathsf{Pre}(L)$ ; ce dernier n'est donc pas IRS.

Inversement, si le langage $\mathsf{Pre}(L)$ n'est pas IRS, alors il contient un langage infini régulier $K$, qui contient $uv^*w$ pour certains mots $u, w \in A^*$ et $v \in A^+$. Puisque $\mathsf{Pre}(L)$ est clos par préfixe, il contient le langage $uv^*$ donc il n'existe pas $i$ tel que $uv^i \notin \mathsf{Pre}(L)$. □

Par ailleurs, la classe des langages FLIP est stable par un certain nombre d'opérations qui sont répertoriées dans le lemme suivant.

**Lemme 5.10** –

**a)** *Tout langage fini est FLIP.*
**b)** *La clôture préfixe d'un langage FLIP est FLIP.*
**c)** *Tout langage inclus dans un langage FLIP est FLIP.*
**d)** *Une intersection quelconque de langages FLIP est FLIP.*
**e)** *Toute union finie de langages FLIP est FLIP.*
**f)** *L'image inverse d'un langage FLIP par un morphisme non-effaçant de mots est FLIP.*
**g)** *La concaténation de deux langages FLIP est FLIP.*

Démonstration. Les points (**a**) et (**b**) découlent immédiatement de la définition et le point (**d**) découle du point (**c**).

**c)** Soient deux langages $L$ et $K$ tels que $K \subseteq L$. Il s'ensuit que $\mathsf{Pre}(K) \subseteq \mathsf{Pre}(L)$ donc tout langage de la forme $uv^*$ contenu dans $\mathsf{Pre}(K)$ est aussi inclus dans $\mathsf{Pre}(L)$. Donc si $K$ n'est pas un langage FLIP, $L$ ne l'est pas non plus.

**e)** Il suffit de le montrer pour l'union de deux langages. Soient deux langages FLIP $L, K \subseteq A^*$ et deux mots $u \in A^*$ et $v \in A^+$. Il existe $i$ et $j$ tels que $uv^i \notin \mathsf{Pre}(L)$ et $uv^j \notin \mathsf{Pre}(K)$. Puisque $\mathsf{Pre}(L \cup K) = \mathsf{Pre}(L) \cup \mathsf{Pre}(K)$, alors le mot $uv^k$ n'appartient pas à $\mathsf{Pre}(L \cup K)$ où $k > \max(i, j)$.

**f)** Soient un langage FLIP $L \subseteq A^*$ et un morphisme de mots $\sigma : B^* \to A^*$ non-effaçant. On note $K = \sigma^{-1}(L) \subseteq B^*$. Supposons que $K$ ne soit pas FLIP, alors il existe deux mots $u \in A^*$ et $v \in A^+$ tel $uv^* \in \mathsf{Pre}(K)$ donc $\sigma(u)(\sigma(v))^*$ est dans $\mathsf{Pre}(L)$.

**g)** Soient deux langages FLIP $L, K \subseteq A^*$. Supposons que $LK$ ne soit pas FLIP, il existe donc deux mots $u \in A^*$ et $v \in A^+$ tel que $uv^* \in \mathsf{Pre}(LK)$. On note $i$ l'exposant tel que $uv^i \notin \mathsf{Pre}(L)$ ; pour tout $k < |uv^i|$ on note $w_k$ le préfixe de longueur $k$ de $uv^i$. Pour tout entier $j \geqslant i$, $uv^j$ appartient à $\mathsf{Pre}(LK)$ donc il existe un entier $k$ et un mot $x \in A^*$ tel que $uv^j = w_k x$ et $x \in \mathsf{Pre}(K)$. Il s'ensuit que pour un certain $k$ l'intersection de $(w_k)^{-1}uv^*$ avec $\mathsf{Pre}(K)$ est infinie donc $K$ n'est pas FLIP. Contradiction. □

# Le monoïde additif $V_{\frac{p}{q}}$

Dans toute la suite, $p$ et $q$ désignent deux entiers premiers entre eux tels que $p > q > 1$. Ils définissant le système de numération à base $\frac{p}{q}$ (*cf.* chapitre 4).

**Définition 5.11** – *On note $V_{\frac{p}{q}}$ l'ensemble des nombres représentables en base $\frac{p}{q}$ :*

$$V_{\frac{p}{q}} \;=\; \{\, \pi_{\frac{p}{q}}(u) \mid u \in [\![p]\!]^* \,\} \;=\; \pi_{\frac{p}{q}}([\![p]\!]^*) \;.$$

Soient deux nombres $x, y \in V_{\frac{p}{q}}$. Il découle de la proposition 4.28 que $(x + y)$ est représentable (donc appartient à $V_{\frac{p}{q}}$) et que $\langle x + y \rangle_{\frac{p}{q}}$ peut être calculé à partir de $\langle x \rangle_{\frac{p}{q}}$ et $\langle y \rangle_{\frac{p}{q}}$ en utilisant l'additionneur (définition 4.25). Il s'ensuit que $V_{\frac{p}{q}}$ est stable par addition, comme l'exprime la propriété suivante.

**Propriété 5.12** – *L'ensemble $V_{\frac{p}{q}}$ est un monoïde (additif).*

L'ensemble des entiers $\mathbb{N}$ est représenté en base $\frac{p}{q}$ par le langage $L_{\frac{p}{q}} \subseteq [\![p]\!]^*$ donc est inclus dans $V_{\frac{p}{q}}$. D'autre part, il découle de l'expression de la fonction d'évaluation que $V_{\frac{p}{q}} \subseteq \mathbb{Q}$ et même que tous les nombres représentables ont une forme particulière :

**Propriété 5.13** – *Tout nombre $x$ de $V_{\frac{p}{q}}$ est de la forme $\frac{n}{q^k}$ pour certains entiers $n$ et $k$.*

Puisque pour tout entier $k$, $\pi_{\frac{p}{q}}(10^k) = \frac{p^k}{q^{k+1}}$, la propriété en découle immédiatement.

**Propriété 5.14** – *$V_{\frac{p}{q}}$ est engendré par la famille $\left\{\frac{p^k}{q^{k+1}}\right\}_{k\in\mathbb{N}}$.*

De plus, $V_{\frac{p}{q}}$ contient tous les éléments de la forme $\frac{n}{q^k}$ pour $n$ assez grand, comme l'exprime le lemme suivant.

**Lemme 5.15** – *Pour tout entier $k$, l'entier*

$$m_k \;=\; (q-1)\sum_{i=0}^{k-1}\left(p^i\, q^{(k-i-1)}\right) - (q^k - 1) \;.$$

*est tel que pour tout entier $n \geqslant m_k$, $\frac{n}{q^k} \in V_{\frac{p}{q}}$*

Démonstration. Par récurrence. On peut vérifier que $m_0 = 0$ et $m_1 = 0$ ; et il est vrai que $\mathbb{N} \subseteq V_{\frac{p}{q}}$ et que $\frac{\mathbb{N}}{q} \subseteq V_{\frac{p}{q}}$ puisque le mot '1' est évalué à $\frac{1}{q}$ et que $V_{\frac{p}{q}}$ est stable par addition (propriété 5.12).

Soient deux entiers $k > 0$ et $n \geqslant m_{(k+1)}$. On réécrit cette deuxième condition comme

$$\begin{array}{lcccc} n \geqslant & (q-1)\sum_{i=0}^{k}(p^i q^{(k-i)}) & - & (q^{k+1}-1) & \\ n \geqslant & (q-1)\left[p^k + q\sum_{i=0}^{k-1}(p^i q^{(k-i-1)})\right] & - & q(q^k-1) & - \quad (q-1) \\ n \geqslant & (q-1)\,p^k \quad + \quad q\left[(q-1)\sum_{i=0}^{k-1}(p^i q^{(k-i-1)}) - (q^k-1)\right] & - & (q-1) & \end{array}$$

et, enfin, comme

$$n \;\geqslant\; (q-1)\,p^k + q\,m_k - (q-1)\;. \qquad (*)$$

Puisque $p^k$ et $q$ sont premiers entre eux, il existe un entier $j \in \mathbb{Z}/q\mathbb{Z}$ tel que $(n - j\,p^k)$ est divisible par $q$.

**Assertion 5.15.1** – $(n - j\,p^k) \;\geqslant\; q\,m_k$ .

*Démonstration de l'assertion.* On note $n' = ((q-1)\,p^k + q\,m_k)$.

Si $n \geqslant n'$, alors (puisque $j \in \mathbb{Z}/q\mathbb{Z}$ et donc $j \leqslant (q-1)$)

$$n - j\,p^k \;\geqslant\; n - (q-1)\,p^k \;\geqslant\; n' - (q-1)\,p^k \;=\; q\,m_k\;.$$

Si au contraire, $n < n'$, alors d'après l'équation $(*)$, $n \geqslant [n' - (q-1)]$. La différence entre $n$ et $n'$ appartient donc à $\{1, 2, \ldots, q-1\}$, donc leurs classes respectives de congruence modulo $q$ sont distinctes ; or $n' \equiv (q-1)\,p^k \;[q]$ donc l'entier $j \in \mathbb{Z}/q\mathbb{Z}$ (défini par $n \equiv j\,p^k \;[q]$) est nécessairement différent de $(q-1)$, donc $j \leqslant (q-2)$. Si bien que :

$$\begin{array}{rcll} j\,p^k & \leqslant & (q-1)\,p^k - p^k & \\ & & (q-1)\,p^k - p^k \;<\; (q-1)\,p^k - (q-1) & (\text{car } (q-1) < p^k) \\ & & (q-1)\,p^k - (q-1) \;\leqslant\; n - q\,m_k & (\text{eq. } (*)) \end{array}$$

Il s'ensuit que $(n - j\,p^k) > q\,m_k$.

On note $i = \frac{n - j\,p^k}{q}$ qui est, par définition de $j$, un entier. Il découle de l'assertion précédente que $i \geqslant m_k$, donc que l'on peut lui appliquer l'hypothèse de récurrence : le nombre $\frac{i}{q^k}$ est représentable. On note $u \in [\![p]\!]^*$ sa représentation, si bien que

$$\pi_{\frac{p}{q}}(u) \;=\; \frac{i}{q^k} \;=\; \frac{i\,q}{q^{k+1}}\;.$$

D'après le cas de base $k = 1$ de la récurrence, il existe également un mot $v \in [\![p]\!]^*$ dont la valeur est $\frac{j}{q}$ ; si bien que le mot $v\,0^k$ a pour valeur :

$$\pi_{\frac{p}{q}}(v\,0^k) \;=\; \pi_{\frac{p}{q}}(v)\left(\frac{p}{q}\right)^k \;=\; \frac{j\,p^k}{q^{(k+1)}}\;.$$

Puisque $V_{\frac{p}{q}}$ est stable par addition (propriété 5.12), il existe un mot $w$ (qui est la sortie $w = \mathcal{C}(u, v\,0^k)$ de l'additionneur $\mathcal{C}$) dont la valeur est

$$\pi_{\frac{p}{q}}(w) \;=\; \pi_{\frac{p}{q}}(u) + \pi_{\frac{p}{q}}(v\,0^k) \;=\; \frac{j\,p^k}{q^{(k+1)}} + \frac{i\,q}{q^{k+1}} \;=\; \frac{n}{q^{k+1}}\;. \qquad \square$$

Le lemme 5.15, et l'existence de la valeur $m_k$ implique évidemment celle d'une borne $l_k$ tel que

$$l_k \;=\; \min\left\{ l \in \mathbb{N} \;\middle|\; \forall n \geqslant l \qquad \frac{n}{q^k} \in V_{\frac{p}{q}} \right\} ,$$

donc inférieure ou égale à $m_k$ pour tout $k$. L'existence de la borne $l_k$ peut être démontré plus simplement : les mots $10^{k-1}$ et 1 s'évaluent respectivement à $\frac{p^{k-1}}{q^k}$ et $\frac{1}{q}$, ce qui implique que $V_{\frac{p}{q}}$ contient $\frac{1}{q^k}(p^k\mathbb{N} + q^k\mathbb{N})$ ; puisque $p^k$ et $q^k$ sont premiers entre eux, l'identité de Bezout montre que $l_k$ existe et est inférieure à $(p\,q)^{(k-1)}$.

Les calculs explicites des bornes $l_k$, reportés à la figure 1, donnent $l_k = m_k$ dans tous les cas $p$, $q$ et $k$ où ils ont été conduits. Ce sont en effet ces données qui ont guidé le choix de l'expression de $m_k$ donnée au lemme 5.15. Il reste à démontrer que cette égalité $m_k = l_k$ est toujours vrai, ce que l'on peut exprimer par la conjecture suivante.

**Conjecture 5.16** – *Soit une base $\frac{p}{q}$. Pour tout entier $k$ le nombre $\frac{m_k-1}{q^k}$ n'appartient pas à $V_{\frac{p}{q}}$.*

| $\mathbb{N}/q^k$ | $l_k(=m_k)$ |
|---|---|
| $\mathbb{N}/1$ | 0 |
| $\mathbb{N}/2$ | 0 |
| $\mathbb{N}/4$ | 2 |
| $\mathbb{N}/8$ | 12 |
| $\mathbb{N}/16$ | 50 |
| $\mathbb{N}/32$ | 180 |
| $\mathbb{N}/64$ | 602 |
| $\mathbb{N}/128$ | 1932 |

(a) base $\frac{3}{2}$

| $\mathbb{N}/q^k$ | $l_k(=m_k)$ |
|---|---|
| $\mathbb{N}/1$ | 0 |
| $\mathbb{N}/3$ | 0 |
| $\mathbb{N}/9$ | 6 |
| $\mathbb{N}/27$ | 48 |
| $\mathbb{N}/81$ | 270 |
| $\mathbb{N}/243$ | 1 320 |
| $\mathbb{N}/729$ | 6 006 |
| $\mathbb{N}/2\,187$ | 26 208 |

(b) base $\frac{4}{3}$

| $\mathbb{N}/q^k$ | $l_k(=m_k)$ |
|---|---|
| $\mathbb{N}/1$ | 0 |
| $\mathbb{N}/4$ | 0 |
| $\mathbb{N}/16$ | 18 |
| $\mathbb{N}/64$ | 216 |
| $\mathbb{N}/256$ | 1 890 |
| $\mathbb{N}/1\,024$ | 14 760 |
| $\mathbb{N}/4\,096$ | 109 458 |
| $\mathbb{N}/16\,384$ | 790 776 |

(c) base $\frac{7}{4}$

Figure 1 – Quelques bornes $l_k$ calculées empiriquement

**Exemple 5.17** – *La figure 2 est une représentation de $V_{\frac{3}{2}}$* par tranches*, c'est-à-dire que $V_{\frac{3}{2}} \bigcap \frac{\mathbb{N}}{q^k}$ est montré comme une bande (discrète), pour des petits entiers $k$. Les nombres appartenant à $V_{\frac{3}{2}}$ sont hachurés ; ainsi, on peut déduire que $\frac{\mathbb{N}}{2} \subseteq V_{\frac{3}{2}}$, et que $\frac{1}{4}$ est le seul nombre de dénominateur 4 qui n'est pas contenu dans $V_{\frac{3}{2}}$, etc.*

*Deux carreaux se trouvant sur deux bandes différentes à la même abscisses $n$ ne représentent pas le même nombre, mais des nombres différents dont le numérateur est identique ; c'est-à-dire de la forme $\frac{n}{q^k}$ pour des certains $k$. Par exemple, la colonne étiquetée par 12, montre que $V_{\frac{3}{2}}$ contient, de haut en bas les nombres $\frac{12}{16}$, $\frac{12}{8}$, $\frac{12}{4}$, $\frac{12}{2}$ et enfin 12.*

**Exemple 5.18** – *La figure 3 est une autre représentation de $V_{\frac{3}{2}}$ par tranches mais qui est cette fois renormalisée. La présence d'un certain nombre $x$ dans $V_{\frac{p}{q}} \bigcap \frac{\mathbb{N}}{q^k}$*

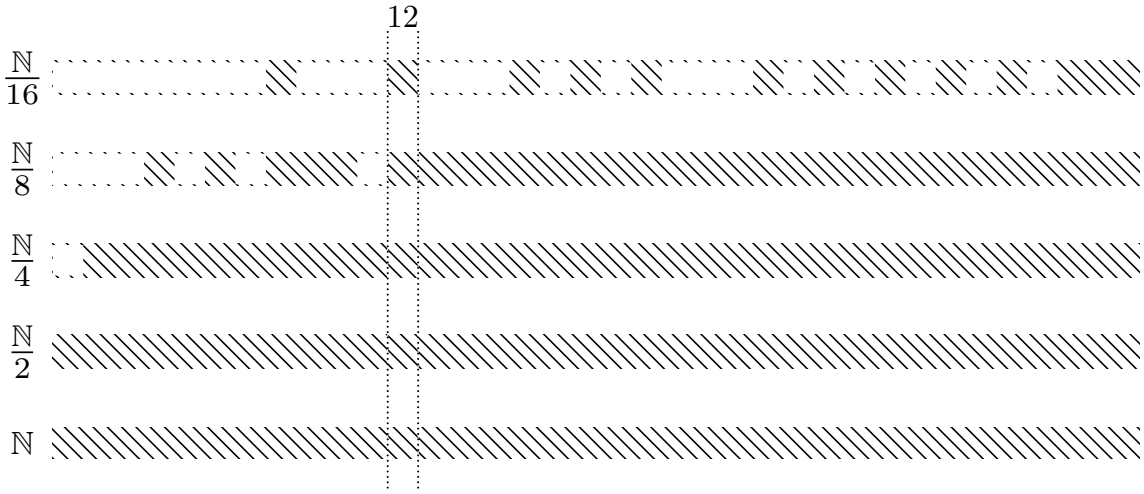



Figure 2 – Représentation de $V_{\frac{3}{2}}$ par tranche

*est représentée par une bande noire qui est toujours centrée autour de l'abscisse $x$ mais dont la largeur est de $\frac{1}{q^k}$ unités. Par exemple, la bande rouge sur chaque ligne représente toujours le nombre 10. Cette figure est donc une représentation de $V_{\frac{3}{2}}$ par raffinements successifs, et met en évidence son caractère lacunaire.*

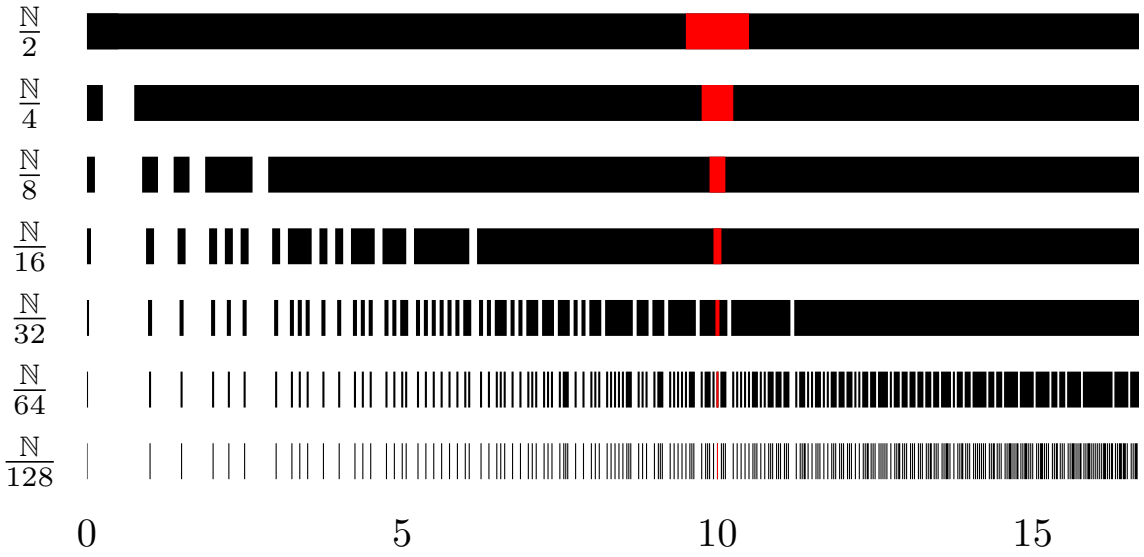



Figure 3 – Représentation de $V_{\frac{3}{2}}$ par tranche, renormalisé

## Monoïdes finiment engendrés

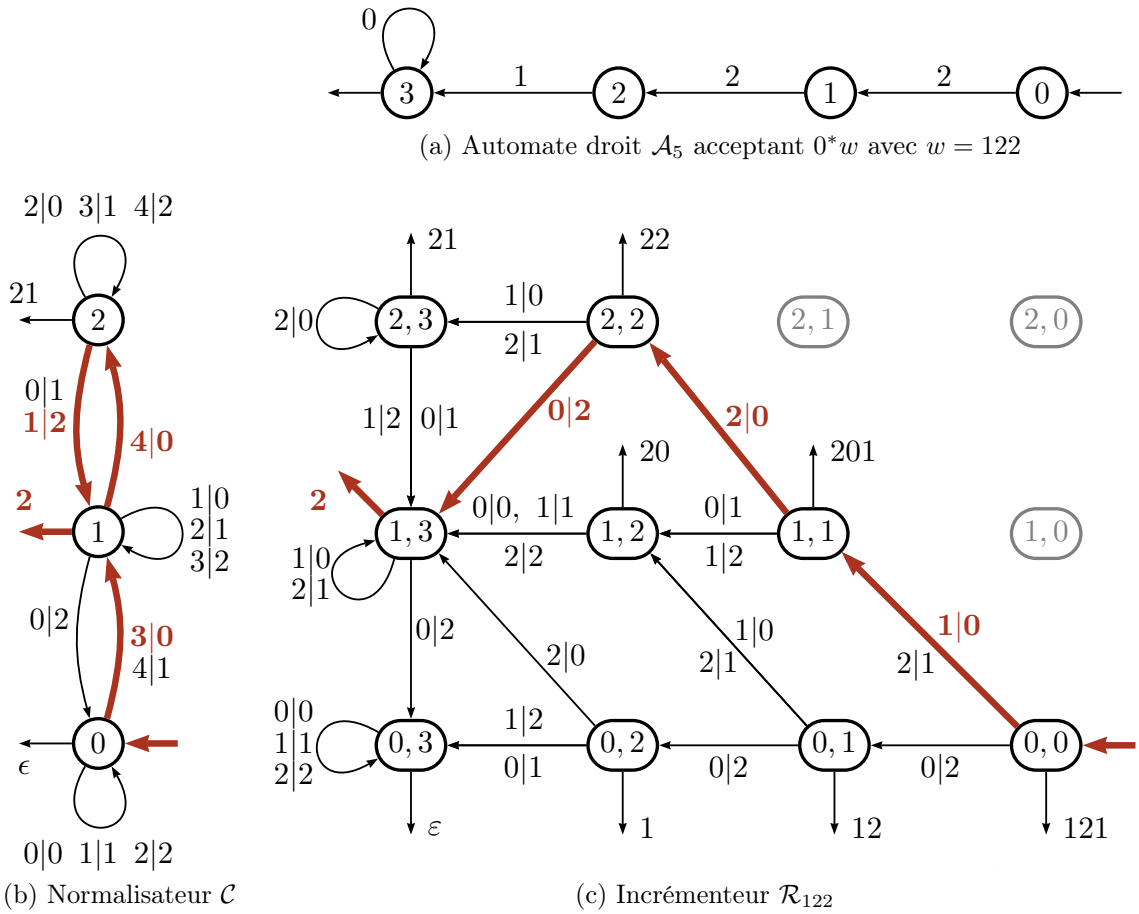
Soit $M$ un sous-monoïde additif de $V_{\frac{p}{q}}$ ; le langage $\langle M \rangle_{\frac{p}{q}}$ est formé des $\frac{p}{q}$-représentations des nombres appartenant à $M$. Dans la suite, on appellera simplement $M$ *un monoïde* et $\langle M \rangle_{\frac{p}{q}}$ la *représentation* du monoïde $M$. Une *représentation*

*de monoïde* n'a donc pas ici le sens classique, hérité de *représentation de groupe*, d'action linéaire sur un espace vectoriel.

Le but de cette section est d'établir le théorème II.

**Théorème II** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q > 1$. Tous les sous-monoïdes additifs et finiment engendrés de $V_{\frac{p}{q}}$ sont représentés en base $\frac{p}{q}$ par des langages FLIP.*

Pour démontrer ce théorème, nous définissons d'abord l'incrémenteur, un transducteur qui réalise l'addition par une constante sur les représentations de monoïdes (corollaire 5.26). On montre ensuite que l'image d'un langage FLIP par ce transducteur est FLIP, si le langage en entrée est la représentation d'un monoïde (proposition 5.28). Puis il est établi que chaque monoïde finiment engendré est inclus dans une union finie de translations de $\mathbb{N}$ (lemme 5.29). Le théorème découle enfin du fait que la propriété FLIP est stable par inclusion, union finie, et passage par l'incrémenteur.




(a) Automate droit $\mathcal{A}_5$ acceptant $0^*w$ avec $w = 122$

(b) Normalisateur $\mathcal{C}$

(c) Incrémenteur $\mathcal{R}_{122}$

Figure 4 – Construction de l'ncrémenteur par $w = 122$ en base $\frac{3}{2}$

## L'incrémenteur

Nous avons vu dans le chapitre 4 que l'addition, en base rationnelle, est réalisée par un transducteur droit, lettre-à-lettre et séquentiel, appelé additionneur (*cf.* définition 4.25, page 104). Nous allons le spécialiser en un transducteur (lui aussi droit, lettre-à-lettre et séquentiel) qui réalise l'addition par une constante.

Étant donné un paramètre $w \in [\![p]\!]^*$, *l'incrémenteur par* $w$ réalise la fonction $u \mapsto v$ où $v$ vérifie $\pi_{\frac{p}{q}}(v) = \pi_{\frac{p}{q}}(u) + \pi_{\frac{p}{q}}(w)$. Ce transducteur est construit par union de l'automate **droit** acceptant le langage $0^*w$ sur une bande d'entrée (la deuxième) de l'additionneur $\mathcal{C}$ (qui est un transducteur droit).

**Exemple 5.19** – *La figure 4 montre la construction de l'incrémenteur par le mot* $w = 122$. *La figure 4a représente* $\mathcal{A}_5$, *l'automate dont le langage est* $0^*w$ *et la figure 4b représente le normalisateur plutôt que l'additionneur pour des raisons de lisibilité (celui-ci comporte deux fois plus d'étiquettes que celui-là).*

*Une transition* $(x', y') \xleftarrow{a \mid c} (x, y)$ *de l'incrémenteur existe si les transitions suivantes existent :* $x' \xleftarrow{b} x$ *dans* $\mathcal{A}_5$ *pour un certain* $b$ *(qui est en fait unique) et* $y' \xleftarrow{(a+b) \mid c} y$ *dans* $\mathcal{C}$. *Le chemin du mot* 021 *est mis en évidence dans l'incrémenteur et correspond au chemin de* 143 *(également mis en évidence) dans le normalisateur ;* 143 *est en effet la somme chiffre-à-chiffre de* 021 *et* $w = 122$.

La définition formelle de l'incrémenteur est donnée ci-dessous.

**Définition 5.20** – *Soit un mot* $w = b_{n-1} \cdots b_1 b_0 \in [\![p]\!]^*$. *On appelle* incrémenteur par $w$ *le transducteur droit*

$$\mathcal{R}_w \;=\; \langle\, \mathbb{N} \times \{0, 1, \ldots, n\}, [\![p]\!], [\![p]\!], (0,0), E_{\mathcal{R}}, \psi \,\rangle \;,$$

*dont les alphabets d'entrée et de sortie sont tous deux* $[\![p]\!]$, *dont les transitions sont définies par*

$\forall s, s' \in \mathbb{N},\ \ \forall a, c \in [\![p]\!]$,

$$(s', n) \xleftarrow{a \mid c} (s, n) \iff q\,s + a \;=\; p\,s' + c \tag{5.1a}$$

$$\forall i < n \qquad (s', i+1) \xleftarrow{a \mid c} (s, i) \iff q\,s + (a + b_i) \;=\; p\,s' + c \tag{5.1b}$$

*et dont la fonction finale* $\psi$ *est définie par :*

$\forall s \in \mathbb{N}$,

$$\psi((s, n)) \;=\; \langle s \rangle_{\frac{p}{q}} \tag{5.1c}$$

$$\forall i < n \qquad \psi((s, i)) \;=\; \psi((s', i+1))\,c \qquad \textit{si}\ \ (s', i+1) \xleftarrow{0 \mid c} (s, i) \tag{5.1d}$$

La dernière équation (5.1d) traite le cas particulier où le mot $u$ en entrée est plus court que le paramètre $w$, la fonction finale $\psi$ simule alors la lecture de $k$ lettres 0, comme l'énonce la propriété suivant.

**Propriété 5.21** – *Soit un mot* $w \in [\![p]\!]^*$ *dont on note la longueur* $n$. *Pour tout mot* $u$ *de longueur* $i < n$, *alors*

$$\forall k < (n - i) \qquad \mathcal{R}_w(u) \;=\; \mathcal{R}_w(0^k u) \;.$$

Démonstration. Il suffit de montrer que $\mathcal{R}_w(u) = \mathcal{R}_w(0\,u)$.

On note le calcul de $0\,u$ dans $\mathcal{R}_w$ par

$$(s',\, i+1) \xleftarrow{0 \mid c} (s,i) \xleftarrow{u \mid v} (0,0)\ . \qquad (*)$$

Si bien que les images respectives de $u$ et de $0\,u$ par $\mathcal{R}_w$ sont

$$\mathcal{R}_w(0\,u) = \psi((s,\, i+1))\,c\,v \quad \text{et} \quad \mathcal{R}_w(u) = \psi((s,i))\,v\ .$$

Or, puisque la transition à gauche du chemin $(*)$ existe, il découle de la équation (5.1d) que $\psi((s,i)) = \psi((s,\, i+1))\,c$, donc que $\mathcal{R}_w(0\,u) = \mathcal{R}_w(u)$. □

On rappelle que l'additionneur (définition 4.25, page 104) est le transducteur droit $\mathcal{C} = \langle\, \mathbb{N},\, [\![p]\!] \times [\![p]\!],\, [\![p]\!]^*,\, 0,\, E_{\mathcal{C}},\, \omega \,\rangle$ dont les transitions sont définies par :

$$\forall s, s' \in \mathbb{N},\ \forall a, b, c \in [\![p]\!] \qquad s' \xleftarrow[\mathcal{C}]{(a,b) \mid c} s \iff q\,s + a + b = p\,s' + c$$

et la fonction finale par :

$$\forall s \in \mathbb{N} \qquad \omega(s) \ = \ \langle s \rangle_{\frac{p}{q}}\ .$$

Comparer la définition 5.20 à celle-ci implique immédiatement le lemme suivant.

**Lemme 5.22** – *Soient deux mots $u, w \in [\![p]\!]^*$ dont on note respectivement les longueurs $i = |u|$ et $n = |w|$.*

**a)** *Si $i < n$, on note $w'$ le préfixe de $w$ de longueur $i$ et*

$$(s,i) \xleftarrow{u \mid v} (0,0) \ \textit{existe dans } \mathcal{R}_w \iff s \xleftarrow{(u,w') \mid v} 0 \ \textit{existe dans } \mathcal{C}.$$

**b)** *Si $i \geqslant n$, on note $k = (i-n)$ et*

$$(s,n) \xleftarrow{u \mid v} (0,0) \ \textit{existe dans } \mathcal{R}_w \iff s \xleftarrow{(u,0^k w) \mid v} 0 \ \textit{existe dans } \mathcal{C}.$$

Ce lemme permet de transférer les propriétés de l'additionneur à l'incrémenteur, en particulier qu'il est correct (proposition 5.23) et que sa partie accessible est finie (la proposition 5.24) découlent du théorème 4.26.

**Proposition 5.23** – *Soit un mot $w \in [\![p]\!]^*$. Pour tout mot $u \in [\![p]\!]^*$, le mot $v = \mathcal{R}_w(u)$ vérifie l'équation $\pi_{\frac{p}{q}}(v) = \pi_{\frac{p}{q}}(u) + \pi_{\frac{p}{q}}(w)$.*

**Proposition 5.24** – *Soit un mot $w \in [\![p]\!]^*$. La partie accessible de $\mathcal{R}_w$ est finie.*

Le lemme suivant montre que l'incrémenteur envoie la représentation d'un nombre sur la représentation d'un autre nombre. Cela permettra d'avoir une équivalence entre l'addition constante sur un ensemble de nombres et l'action de l'incrémenteur sur leurs représentations, comme énoncé au corollaire 5.26.

**Lemme 5.25** – *Soit un nombre $x \in V_{\frac{p}{q}}$ dont on note la représentation $w = \langle x \rangle_{\frac{p}{q}}$. Pour tout nombre $y \in V_{\frac{p}{q}}$, $\langle x + y \rangle_{\frac{p}{q}} = \mathcal{R}_w(\langle y \rangle_{\frac{p}{q}})$.*

Démonstration. On note les longueurs de $w$ et $\langle y\rangle_{\frac{p}{q}}$ respectivement par $n$ et $k$.

Si $n < k$, on note $i = (k-n)$ et les deux égalités suivantes découlent alors respectivement du lemme 5.22b et de la proposition 4.28 (page 105) :

$$\mathcal{R}_{\langle x\rangle_{\frac{p}{q}}} \;=\; \mathcal{C}(\langle x\rangle_{\frac{p}{q}}, 0^i w) \;=\; \langle x+y\rangle_{\frac{p}{q}} \;.$$

Si $n \geqslant k$, on note $i = (n-k)$, et les trois équations suivantes découlent alors respectivement de la propriété 5.21, du lemme 5.22b et de la proposition 4.28

$$\mathcal{R}_w\left(\langle x\rangle_{\frac{p}{q}}\right) \;=\; \mathcal{R}_w\left(0^k\langle x\rangle_{\frac{p}{q}}\right) \;=\; \mathcal{C}\left(0^k\langle x\rangle_{\frac{p}{q}}, w\right) \;=\; \langle x+y\rangle_{\frac{p}{q}} \;. \qquad \square$$

**Corollaire 5.26** – *Soient un ensemble de nombres $M \subseteq V_{\frac{p}{q}}$ et un nombre $x \in V_{\frac{p}{q}}$ dont on note la représentation $w = \langle x\rangle_{\frac{p}{q}}$. Alors $\mathcal{R}_w(\langle M\rangle_{\frac{p}{q}}) = \langle M+x\rangle_{\frac{p}{q}}$.*

## Représentation de monoïde et incrémenteur

Dans la suite, on considère des langages qui ne sont pas FLIP pour démontrer certains résultats. Un langage $L$ n'est pas FLIP s'il existe deux mots $u, v$, un ensemble infini d'indices $I \subseteq \mathbb{N}$ et une famille de mots $(y_i)_{i\in I}$ tels que $v \neq \varepsilon$ et pour tout $i \in I$, $u\,v^i\,y_i \in L$. Cette propriété est conservée par l'incrémenteur, comme l'exprime (la contraposée de) la proposition suivante.

**Proposition 5.27** – *Soient un langage $L \subseteq [\![p]\!]^*$ et un mot $w \in [\![p]\!]^*$. Si $\mathcal{R}_w(L)$ est FLIP, alors $L$ est FLIP.*

Démonstration. Par contraposée. Supposons que $L$ n'est pas FLIP ; il existe donc deux mots $u \in [\![p]\!]^*$ et $v \in [\![p]\!]^+$, un ensemble infini d'indices $I \subseteq \mathbb{N}$ et une famille $(y_i)_{i\in I}$ de mots de $[\![p]\!]^*$ tels que

$$K \;=\; \{\, u\,v^i\,y_i \mid i \in I \,\} \;\subseteq\; L \;.$$

On peut de plus choisir les différentes variables $u, v, I, (y_i)_{i\in I}$ de façon à ce que les hypothèses suivantes soient satisfaites

($*$) *La longueur de chaque mot $y_i$ est supérieure à $n = |w|$.*
Il suffit de factoriser différemment les éléments de $K$. Soit $k$ un entier tel que $|v^k| > n$. Pour tout $i \in I$, $i \geqslant k$, on note $x_{(i-k)} = v^k\,y_i$.

$$u\,v^i\,y_i \;=\; u\,v^{(i-k)}v^k y_i \;=\; u\,v^{(i-k)}\,x_{(i-k)} \;.$$

Il s'ensuit que l'ensemble (infini) $J = \{(i-k) \mid i \in I \text{ et } i \geqslant k\}$ vérifie

$$\{u\,v^j\,x_j \mid j \in J\} \;\subseteq\; K \;\; (\subseteq\; L) \;.$$

($**$) *Le calcul de chaque mot $y_i$ dans $\mathcal{R}_w$ atteint le même état $(s,n)$ :*

$$\exists s \in \mathbb{N},\;\; \forall i \in I \qquad (s,n) \xleftarrow[\mathcal{R}_w]{y_i \mid y'_i} (0,0) \;.$$

Il suffit de raffiner $I$. L'incrémenteur $\mathcal{R}_w$ a un nombre fini d'états alors que $I$ est infini ; il existe donc un état $(s,k)$ et un sous-ensemble $J \subseteq I$ infini tel que pour tout $j \in J$, le calcul de $y_j$ atteint $(s,k)$. De plus, d'après l'hypothèse ($*$) tous les $y_i$ ont une longueur supérieure à $n$, donc $k = n$.

Les transitions de $\mathcal{R}_w$ sortants des états dont la seconde composante vaut $n$ sont définies par (équation (5.1b)) :

$$\forall s, s' \in \mathbb{N},\ \forall a, c \in [\![p]\!] \qquad (s', n) \xleftarrow{a \mid c} (s, n) \quad \Longleftrightarrow \quad q\,s + a \ = \ p\,s' + c\ .$$

Puisque $a < p$ et $q < p$, alors le membre de droite implique que $(p\,s + p) > s'p$, donc que $(s + 1) > s'$ et, puisque $s$ et $s'$ sont entiers, que $s \geqslant s'$. Il s'ensuit que s'il part d'un état dont la seconde composante est égale à $n$, tout chemin

$$(t_k, n) \longleftarrow \cdots \longleftarrow (t_1, n) \longleftarrow (t_0, n) \quad \text{vérifie} \quad t_k \ \leqslant \ \cdots \ \leqslant \ t_1 \ \leqslant \ t_0\ .$$

En particulier, le chemin partant de $(s, n)$ et étiqueté par $v^j$ (pour $j$ suffisamment grand) est ultimement stationnaire à un état noté $(t, n)$ (qui est donc son propre successeur par chacune des lettres de $v$).

Étudions maintenant l'image de $K$ par $\mathcal{R}_w$. Pour tout entier $i \in I$, $i \geqslant j$, le calcul de $u\,v^i\,y_i$ est alors le suivant pour certains mots $v'$, $u'$, $z$ et un certain état $(r, n)$

$$(r, n) \xleftarrow{u \mid u'} (t, n) \xleftarrow{v^{(i-j)} \mid (v')^{(i-j)}} (t, n) \xleftarrow{v^j \mid z} (s, n) \xleftarrow{y_i \mid y'_i} (0, 0)$$

Donc,

$$\forall i \in I,\ \ i \geqslant j \qquad \mathcal{R}_w(u\,v^i\,y_i) \ = \ [\langle r \rangle_{\frac{p}{q}}\,u']\ (v')^{(i-j)}\ [z\,y'_i]\ .$$

Puisque $\mathcal{R}_w$ est lettre-à-lettre, $v' \neq \varepsilon$, ce qui implique que $\mathcal{R}_w(K)$ n'est pas un langage FLIP. Il s'ensuit que le langage $\mathcal{R}_w(L)$ qui le contient n'est pas un langage FLIP non plus (contraposée du lemme 5.10c). □

Dans le cas particulier où $L$ est la représentation d'un monoïde et $w$ la représentation d'un nombre, la réciproque de la proposition précédente est vérifiée.

**Proposition 5.28** – *Soient un monoïde $M \subseteq V_{\frac{p}{q}}$ et un nombre $x \in V_{\frac{p}{q}}$. Si $\langle M \rangle_{\frac{p}{q}}$ est un langage FLIP, alors $\langle M + x \rangle_{\frac{p}{q}}$ est un langage FLIP.*

Démonstration. Puisque $M$ est sous-monoïde additif de $V_{\frac{p}{q}}$ (donc de $\mathbb{Q}$), il contient $m\mathbb{N}$ pour un certain entier $m$.

D'autre part, l'élément $x$ appartient à $V_{\frac{p}{q}}$ donc s'écrit comme $\frac{n}{q^k}$ pour certains entiers $k$ et $n$. Le lemme 5.15 donne l'expression d'une borne $m_k$ tel que pour tout entier $j \geqslant m_k$, le nombre $\frac{j}{q^k}$ appartient à $V_{\frac{p}{q}}$. En particulier,

$$\exists y \in V_{\frac{p}{q}} \quad \text{tel que} \quad y = \frac{j}{q^k} \quad \text{et} \quad (j + n) \ \equiv \ 0\ [m\,q^k]\ .$$

Il s'ensuit que $(x + y)$ appartient à $m\mathbb{N}$ donc à $M$ et, puisque $M$ est stable par addition, que $(M + x + y) \subseteq M$ donc que $\langle M + x + y \rangle_{\frac{p}{q}} \subseteq \langle M \rangle_{\frac{p}{q}}$.

Puisque $\langle M \rangle_{\frac{p}{q}}$ est FLIP par hypothèse, $\langle M + x + y \rangle_{\frac{p}{q}}$ l'est également (car la propriété FLIP est conservée par inclusion, lemme 5.10c). On note $w = \langle y \rangle_{\frac{p}{q}}$ ; il découle du corollaire 5.26 que $\mathcal{R}_w(\langle M + x \rangle_{\frac{p}{q}}) \ = \ \langle M + x + y \rangle_{\frac{p}{q}}$, un langage FLIP donc de la proposition 5.27 que $\langle M + x \rangle_{\frac{p}{q}}$ est FLIP. □

Le lemme suivant donne une approximation (par excès) de tout monoïde finiment engendré par une union finie de translations de $\mathbb{N}$.

**Lemme 5.29** – *Soit un monoïde $M \subseteq V_{\frac{p}{q}}$. Si $M$ est finiment engendré, alors il existe une famille **finie** $(g_i)_{i\in I}$ d'éléments de $V_{\frac{p}{q}}$ telle que $M$ est inclus dans $\bigcup_{i\in I}(g_i + \mathbb{N})$.*

Démonstration. On note $\left(\frac{n_i}{q^{k_i}}\right)_{i\in I}$ une famille finie engendrant $M$ et $k$ le plus grand des $k_i$. Il s'ensuit que tout élément de $M$ est un nombre rationnel dont le dénominateur divise $q^k$, donc

$$M \subseteq \left(\frac{\mathbb{N}}{q^k} \cap V_{\frac{p}{q}}\right) .$$

Or, tout nombre de $\frac{\mathbb{N}}{q^k}$ peut être écrit sous la forme $n + \frac{j}{q^k}$ pour certains entiers $n$ et $j$ tels que $0 \leqslant j < q^k$. Il s'ensuit que

$$\frac{\mathbb{N}}{q^k} \;=\; \bigcup_{j=0}^{q^k-1}\left(\mathbb{N} + \frac{j}{q^k}\right) \quad \text{et donc que} \quad M \;\subseteq\; \bigcup_{j=0}^{q^k-1}\left[V_{\frac{p}{q}} \cap \left(\mathbb{N} + \frac{j}{q^k}\right)\right] .$$

Pour tout entier $j$, $0 \leqslant j < q^k$, on note $g_j$ le plus petit nombre appartenant à $(V_{\frac{p}{q}} \cap (\mathbb{N} + \frac{j}{q^k}))$ (il est donc représentable). Si bien que

$$\forall j,\ 0 \leqslant j < q^k \qquad \left[V_{\frac{p}{q}} \cap \left(\mathbb{N} + \frac{j}{q^k}\right)\right] \;\subseteq\; (g_j + \mathbb{N}) .$$

ce qui implique enfin que

$$M \;\subseteq\; \bigcup_{j=0}^{q^k-1}(g_j + \mathbb{N}) . \qquad \square$$

Nous pouvons maintenant démontrer le théorème II, rappelé ci-dessous.

**Théorème II** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q > 1$. Tous les sous-monoïdes additifs et finiment engendrés de $V_{\frac{p}{q}}$ sont représentés en base $\frac{p}{q}$ par des langages FLIP.*

Démonstration. Soit un monoïde $M$ finiment engendré. D'après le lemme 5.29, il existe une famille finie $\{g_i\}_{i\in I}$ telle que $M \subseteq \bigcup_{i\in I}(\mathbb{N} + g_i)$ ce qui implique que

$$\langle M\rangle_{\frac{p}{q}} \;\subseteq\; \bigcup_{i\in I}\langle \mathbb{N} + g_i\rangle_{\frac{p}{q}} . \qquad (*)$$

Puisque $\langle \mathbb{N}\rangle_{\frac{p}{q}} = L_{\frac{p}{q}}$ est FLIP, il découle de la proposition 5.28 que pour tout $i \in I$, $\langle \mathbb{N} + g_i\rangle_{\frac{p}{q}}$ est FLIP. Donc, puisque la propriété FLIP est stable par union finie (lemme 5.10e) et inclusion (lemme 5.10c), il découle de l'équation $(*)$ que $\langle M\rangle_{\frac{p}{q}}$ est un langage FLIP. $\square$

# Périodicité

Un résultat central dans l'étude d'un système de numération $S$ est : *un ensemble ultimement périodique d'entiers est représenté dans le système $S$ par un langage régulier*. La question de sa réciproque amenant par exemple au théorème de Cobham en base entière.

Un tel résultat est évidemment impossible pour les bases rationnelles, puisque les sous-ensembles d'un langage FLIP sont FLIP. Dans cette section, nous allons introduire une notion de périodicité qui s'étend aux nombres de $V_{\frac{p}{q}}$ et peut, dans certains cas, donner des énoncés similaires au résultat susdit.

## Extension de la congruence aux nombres rationnels

Soit un entier $n$ fixé dans la suite. Nous allons définir une fonction partielle $\mathbb{Q} \to \mathbb{Z}/n\mathbb{Z}$ qui étend la fonction associant à chaque entier sa classe de congruence modulo $n$.

**Définition 5.30** – *Soit un nombre $z \in \mathbb{Q}$ qui s'écrit comme la fraction irréductible $z = \frac{m}{k}$.*

- *Si $k$ est premier avec $n$ alors $k$ possède un inverse dans (le groupe multiplicatif) $\mathbb{Z}/n\mathbb{Z}$, que l'on note $k^{-1}$. On note alors $(z \,\%\, n)$ l'entier*
$$z \,\%\, n \;=\; \left(\frac{m}{k}\right) \%\, n \;=\; (m\,k^{-1}) \%\, n \;.$$
- *Si $k$ n'est pas premier avec $n$ alors $(z \,\%\, n)$ n'est pas défini.*

*Dans le premier cas on appelle $(z \,\%\, n)$, par abus de langage, la* classe de congruence *de $z$ modulo $n$ ; dans le second on dit que $(z \,\%\, n)$ n'a pas de classe de congruence modulo $n$. De plus, deux nombres $z, z' \in \mathbb{Q}$ sont dit* congrus modulo $n$, *noté $z \equiv z'\ [n]$, si $(z \,\%\, n)$ et $(z' \,\%\, n)$ existent et que $z \,\%\, n = z' \,\%\, n$.*

Dans la suite, pour tout entier $l$ premier avec $n$ on utilisera la notation $l^{-1}$ dans le sens de cette définition, c'est-à-dire que $l^{-1}$ est l'entier de $\mathbb{Z}/n\mathbb{Z}$ tels que $(l^{-1}\,l) \,\%\, n = 0$. De plus, si un entier $l$ divise un entier $l'$, on note $(l'/l)$ le quotient (entier) de cette division qui vérifie donc $l\,(l'/l) = l'$.

Le lemme suivant et son corollaire montrent qu'il n'est pas nécessaire de considérer des fractions irréductibles mais simplement des fractions *suffisamment réduites*, c'est-à-dire dont le dénominateur est premier avec $n$.

**Lemme 5.31** – *Soit un nombre $z \in \mathbb{Q}$ qui s'écrit comme la fraction $z = \frac{m}{k}$. Si $k$ est premier avec $n$, alors $z \,\%\, n = (m\,k^{-1}) \,\%\, n$.*

Démonstration. On note $d$ le PGCD de $m$ et $k$ ; la fraction irréductible de $z$ est donc $\frac{(m/d)}{(k/d)}$. D'après la définition de congruence

$$z \,\%\, n \;=\; \left[(m/d)(k/d)^{-1}\right] \%\, n \;.$$

(Jusqu'à la fin de cette démonstration, on ne manipule que des congruences sur des entiers auxquels toutes les propriétés usuelles s'appliquent.) Puisque $d$ est un facteur de $k$, il est par hypothèse premier avec $n$ et possède donc dans $\mathbb{Z}/n\mathbb{Z}$ un inverse (que l'on note $d^{-1}$). Il s'ensuit que $(m/d) \equiv m\,d^{-1}\ [n]$ et que $(k/d) \equiv k\,d^{-1}\ [n]$ ; d'où $(k/d)^{-1} \equiv k^{-1}\,d\ [n]$, ce qui implique que

$$\left[(m/d)\,(k/d)^{-1}\right] \%\, n \;=\; \left[m\,d^{-1}\,k^{-1}\,d\right] \%\, n \;=\; \left[m\,k^{-1}\right] \%\, n \;. \qquad \square$$

**COROLLAIRE 5.32** – *Soient deux fractions $\frac{m}{k}$ et $\frac{m'}{k'}$. Si $k$ et $k'$ sont premiers avec $n$, alors $\frac{m}{k} \equiv \frac{m'}{k'}$ si et seulement si $m\,k' \equiv m'\,k\ [n]$.*

Cette généralisation de la congruence conserve plusieurs propriétés usuelles, dont on donne quelques exemples ci-dessous.

**PROPRIÉTÉ 5.33** – *Soient deux nombres $z, z' \in \mathbb{Q}$.*

**a)** *Si $z \equiv z'\ [n]$, alors pour tout entier $k$, $z\,k \equiv z'\,k\ [n]$.*

**b)** *Si $z \equiv z'\ [n]$, alors pour tout entier $k$ premier avec $n$, $z\,\frac{1}{k} \equiv z'\,\frac{1}{k}\ [n]$.*

**c)** *Si $(z\,\%\,n)$ existe et est premier avec $n$, alors $(\frac{1}{z}\,\%\,n)$ existe et est égal à $(z\,\%\,n)^{-1}$.*

**d)** *Soient deux nombres $x, x' \in \mathbb{Q}$. Si $z \equiv z'\ [n]$ et $x \equiv x'\ [n]$, alors*

$$z + x \ \equiv\ z' + x'\ [n] \quad \text{et} \quad z\,x \ \equiv\ z'\,x'\ [n]\ .$$

**e)** *Soit un nombre $x \in \mathbb{Q}$ tel que $(x\,\%\,n)$ existe et est premier avec $n$. Si $x\,z \equiv x\,z'\ [n]$, alors $z \equiv z'\ [n]$.*

**f)** *Soit un nombre $x \in \mathbb{Q}$ tel que $(x\,\%\,n)$ existe. Si $x + z \equiv x + z'\ [n]$, alors $z \equiv z'\ [n]$.*

DÉMONSTRATION. Les points (**a**) et (**b**) sont des cas particuliers du point (**d**).

**c)** On note $\frac{m}{k}$ la fraction irréductible de $z$, si bien que la fraction irréductible de $\frac{1}{z}$ est $\frac{k}{m}$. Puisque $(z\,\%\,n)$ existe, $k$ est premier avec $n$ et $(z\,\%\,n) = m\,k^{-1}$. Puisque $(z\,\%\,n)$ est premier avec $n$, $m\,k^{-1}$ est un élément inversible de $\mathbb{Z}/n\mathbb{Z}$ donc $m$ l'est également. Donc $z = \frac{k}{m}$ admet une congruence modulo $n$ égale à $(k\,m^{-1}) = (m\,k^{-1})^{-1} = (z\,\%\,n)^{-1}$.

**d)** On note les fractions irréductibles de $z, z', x, x'$ par

$$z \ =\ \frac{m}{k}\ ,\quad z' \ =\ \frac{m'}{k'}\ ,\quad x \ =\ \frac{i}{j} \quad \text{et} \quad x' = \frac{i'}{j'}\ .$$

Les hypothèses $\frac{m}{k} \equiv \frac{m'}{k'}\ [n]$ et $\frac{i}{j} \equiv \frac{i'}{j'}\ [n]$ sont donc équivalentes à

$$i'\,j \ \equiv\ i\,j'\ \ [n] \quad \text{et} \quad m'\,k \ \equiv\ m\,k'\ \ [n]\ . \tag{$*$}$$

Somme. Il s'ensuit que $z + x = \frac{j\,m + k\,i}{k\,j}$ (avec $k\,j$ premier avec $n$) et similairement que $z' + x' = \frac{j'\,m' + k'\,i'}{k'\,j'}$ (avec $k'\,j'$ premier avec $n$). Le but est de démontrer que

$$(j'\,m' + k'\,i')\,k\,j \ \equiv\ (j\,m + k\,i)\,k'\,j' \quad [n]\ ,$$

qui se réécrit comme

$$(j'\,j)(m'\,k) + (k\,k')(i'\,j) \ \equiv\ (j'\,j)(m\,k') + (k\,k')(i\,j') \quad [n]$$

ce qui découle de $(*)$.

Produit. Similairement $z\,x = \frac{m\,i}{k\,j}$ (avec $k\,j$ premier avec $n$) et $z'\,x' = \frac{m'\,i'}{k'\,j'}$ (avec $k'\,j'$ premier avec $n$). Il découle de $(*)$ que

$$(m\,k')\,(i\,j') \ \equiv\ (m'\,k)\,(i'\,j) \quad [n]\ ,$$

ce qui conclut la démonstration.

Les points (**e**) et (**f**) découlent de (**c**) et (**d**). □

## Langage régulier et ensemble de nombres modulo

On considère toujours la base rationnelle $\frac{p}{q}$ où $p$ et $q$ sont deux entiers premiers entre tels que $p > q > 1$. Dans cette sous-section, on s'intéresse aux sous-ensembles de $V_{\frac{p}{q}}$ qui appartiennent à certaines classes de congruence modulo un entier $n$, que l'on appelle *période* par abus de langage. On ne traite que deux cas, 1) quand la période est première avec $q$ et 2) quand la (plus petite) période est égale à $q$.

### Cas où la période $n$ est première avec $q$

Chaque nombre de $V_{\frac{p}{q}}$ peut s'écrire comme une fraction dont le dénominateur est $q^k$ pour un certain $k$ (propriété 5.13, page 120). Si bien que sous l'hypothèse que la période $n$ est première avec $q$, les fractions irréductibles de tous les nombres de $V_{\frac{p}{q}}$ ont un dénominateur premier avec $n$, donc ont une valeur modulo $n$.

On appelle *ensemble pseudo-périodique de paramètre* $(n, R)$ l'ensemble des nombres $z \in V_{\frac{p}{q}}$ dont la classe de congruence modulo $n$ appartient à $R$.

**Théorème 5.34** – *Soient deux entiers $p$ et $q$ premiers entre eux tels que $p > q > 1$. Soient une période $n \in \mathbb{N}$ et un ensemble $R \in \mathbb{Z}/n\mathbb{Z}$ de restes modulo $n$. Si $n$ et $q$ sont premiers entre eux alors il existe un automate qui accepte par valeur l'ensemble pseudo-périodique de paramètre $(n, R)$.*

Cet automate, noté $\mathcal{D}_R$ est défini par :

$$\mathcal{D}_R \;=\; \langle\, \mathbb{Z}/n\mathbb{Z}, [\![p]\!], 0, \delta, R \,\rangle$$

où la fonction de transition $\delta$ est :

$$\forall k, k' \in \mathbb{Z}/n\mathbb{Z},\ \forall a \in [\![p]\!] \qquad k \xrightarrow[\mathcal{D}_R]{a} k' \quad\Longleftrightarrow\quad p\,k + a \;\equiv\; q\,k' \ [n]\ . \tag{5.2}$$

Il faut noter que puisque $q$ et $n$ sont premiers entre eux, il existe pour toute lettre $a \in [\![p]\!]$ et tout état $k \in \mathbb{Z}/n\mathbb{Z}$, un unique état $k'$ tel que $p\,k + a \equiv q\,k'\ [n]$. L'automate $\mathcal{D}_R$ est donc déterministe et complet. Il vérifie de plus la proposition suivante.

**Proposition 5.35** – *Le calcul d'un mot $u \in [\![p]\!]^*$ dans $\mathcal{D}_R$ atteint l'état $i \in \mathbb{Z}/n\mathbb{Z}$, où $i = \pi_{\frac{p}{q}}(u) \,\%\, n$.*

Démonstration. Par récurrence sur la longueur de $u$. L'assertion est vérifiée dans le cas où $u = \varepsilon$ ; sa valeur est égale à 0, qui est bien l'état initial de $\mathcal{D}_R$.

Soit un mot $u\,a \in [\![p]\!]^+$, dont on note le calcul dans $\mathcal{D}_R$ par

$$0 \xrightarrow{\ u\ } k \xrightarrow{\ a\ } k'\ .$$

L'hypothèse de récurrence et la définition des transitions de $\mathcal{D}_R$, impliquent respectivement les deux équations suivantes :

$$k \;=\; \pi_{\frac{p}{q}}(u) \,\%\, q \quad \text{donc} \quad k \;\equiv\; \pi_{\frac{p}{q}}(u) \quad [n] \tag{$*$}$$

$$p\,k + a \;\equiv\; q\,k' \ [n] \quad \text{donc} \quad \frac{p}{q}\,k + \frac{a}{q} \;\equiv\; k' \quad [n] \tag{$**$}$$

Or, d'après la définition de $\pi_{\frac{p}{q}}$ (équation (2.2a))

$$\pi_{\frac{p}{q}}(u\,a) \;=\; \frac{p}{q}\,\pi_{\frac{p}{q}}(u) + \frac{1}{q}\,a\ .$$

Ce qui implique, en reportant $(*)$ puis en utilisant $(**)$, que

$$\pi_{\frac{p}{q}}(u\,a) \;\equiv\; \frac{p}{q}\,k + \frac{1}{q} \;\equiv\; k' \quad [n]\ .$$

Puisque $k'$ est un état de $\mathcal{D}_R$, il appartient à $\mathbb{Z}/n\mathbb{Z}$ donc $\pi_{\frac{p}{q}}(u\,a)\,\%\,n = k'$. □

Démonstration du théorème 5.34. Soit un mot $u \in [\![p]\!]^*$ ; on note $i \in \mathbb{Z}/n\mathbb{Z}$ l'entier tel que $i = \pi_{\frac{p}{q}}(u)\,\%\,n$. Il découle de la proposition 5.35 précédente que le calcul de $u$ atteint dans $\mathcal{D}_R$ l'état $i$. Le mot $u$ est donc accepté par $\mathcal{D}_R$ si et seulement si $i$ est final. Un mot $u \in [\![p]\!]^*$ est donc accepté par $\mathcal{D}_R$ si et seulement si $\pi_{\frac{p}{q}}(u)\,\%\,n \in R$. □

Nous allons dans la suite brièvement comparer les ensembles pseudo-périodiques en base rationnelle avec les ensembles périodiques en base entière. Il sera donc fait quelques références aux chapitres 2 et 3 ; néanmoins, nous considérons ici que nous lisons, en base entière comme en base rationnelle, le chiffre le plus significatif en premier, c'est-à-dire au contraire des chapitres 2 et 3.

On définit un deuxième automate

$$\mathcal{E}_R \;=\; \langle\, \mathbb{Z}/n\mathbb{Z}, [\![p]\!]^*, 0, \delta', R \,\rangle$$

dont la fonction de transition $\delta'$ est définie par

$$\forall k, k' \in \mathbb{Z}/n\mathbb{Z},\ \forall a \in [\![p]\!] \qquad k \xrightarrow[\mathcal{E}_R]{a} k' \quad\Longleftrightarrow\quad p\,k + a \;\equiv\; k' \ \ [n]\ . \tag{5.3}$$

Il s'agit de l'automate usuel qui accepte les mots $u$ dont la valeur en **base entière** $p$ est congrue à un certain $r \in R$ :

$$L(\mathcal{E}_R) \;=\; \left\{\, u \in [\![p]\!]^* \;\middle|\; (\pi_p(u)\,\%\,n) \in R \,\right\}\ .$$

Les définition de $\mathcal{E}_R$ et $\mathcal{D}_R$ ne diffèrent que par un facteur $q$ (présent dans de membre droit de (5.2), absent dans celui de (5.3)) et leurs langages sont le pendant l'un de l'autre dans leurs bases $p$ et $\frac{p}{q}$ respectives.

L'extension de la congruence modulo $n$ à $V_{\frac{p}{q}}$ prend alors tout son sens.

Dans le cas où la période $n$ est première avec $p$ (en plus d'être première avec $q$), on pose $k = (\frac{p}{q}\,\%\,n) = (p\,q^{-1}\,\%\,n)$. Une récurrence montre alors que

$$\forall u \in [\![p]\!]^* \qquad \pi_k(u) \;\equiv\; q\,\pi_{\frac{p}{q}}(u) \quad [n]\ .$$

On note $S = \{r \mid (rq)\,\%\,n \in R\}$. L'automate de Pascal $\mathcal{P}_n^S$ en base $k$ (voir section 3.1) est alors un automate **droit** qui accepte un mot $u \in [\![k]\!]^*$ si et seulement si sa valeur en base $k$ est dans $S$ c'est-à-dire si et seulement si sa valeur en base $\frac{p}{q}$ est dans $R$. Il est vrai que l'alphabet de cet automate est $[\![k]\!]$ et non $[\![p]\!]$ mais nous avons vu précédemment (lemme 3.15 page 51) que tous les chiffres autres que 0 et 1 sont redondants dans un automate de Pascal.

D'autre part, pour tout entier $j$, la classe de congruence modulo $p^j$ de la valeur d'un mot $u \in [\![p]\!]^*$ est entièrement déterminée par son suffixe de longueur $j$ (lemme 4.15, page 99). Dans le cas où la période $n$ divise $p^j$ pour un certain entier $j$ (en plus d'être première avec $q$) l'automate $\mathcal{D}_R$ accepte donc tous les mots dont les suffixes respectifs de longueur $j$ appartiennent à un certain ensemble fini.

On voit bien que, dans le cas où la période $n$ est première avec $q$, les ensembles pseudo-périodiques en base $\frac{p}{q}$ offrent un paysage similaire à celui des ensembles périodiques en base entière. Cela ouvre beaucoup de questions auxquelles, à notre connaissance, personne ne s'est encore intéressé.

### Cas où la période $n$ est égale à $q$

Tout nombre de $V_{\frac{p}{q}}$ qui n'est pas un entier n'a pas de classe congruence modulo $q$ avec la définition choisie dans la sous-section 5.4.1. Ceci n'est pas un manquement due à la définition : les ensembles périodiques d'entiers (au sens classique) de période $q$ sont déjà inséparables les unes des autres, comme l'énonce le théorème suivant.

**Théorème 5.36** – *Soient deux entiers $p$ et $q$ premiers entre eux tels que $p > q > 1$. Soient $S$ un ensemble d'entiers périodique de plus petite période $q$ et $S'$ son complémentaire dans $\mathbb{N}$, respectivement représentés par $L = \langle S \rangle_{\frac{p}{q}}$ et $L' = \langle S' \rangle_{\frac{p}{q}}$. Alors, il n'existe pas de langage régulier $K \subseteq [\![p]\!]^*$ tel que $L \subseteq K$ et $L' \cap K = \emptyset$.*

Nous allons d'abord définir deux transformations d'automates basé sur l'incrémenteur (lemme 5.37 et notation 5.38), puis traiter le cas particulier ou l'ensemble $S$ est formé des entiers divisible par $q$ (proposition 5.39) et enfin démontrer que le cas général se réduit à ce cas particulier.

**Lemme 5.37** – *Soient un nombre $x \in V_{\frac{p}{q}}$ et un ensemble $M \subseteq V_{\frac{p}{q}}$ de nombres. Si $\langle M \rangle_{\frac{p}{q}}$ est langage régulier, alors les langages $\langle M + x \rangle_{\frac{p}{q}}$ et $\left\langle (M - x) \cap V_{\frac{p}{q}} \right\rangle_{\frac{p}{q}}$ sont réguliers.*

Démonstration. On note $w = \langle x \rangle_{\frac{p}{q}}$ et $\mathcal{R}_w$ l'incrémenteur associé.

Puisque $\langle M \rangle_{\frac{p}{q}}$ est régulier et que $\mathcal{R}_w$ est un transducteur (droit, lettre-à-lettre et séquentiel) alors $\mathcal{R}_w(\langle M \rangle_{\frac{p}{q}})$ et $\mathcal{R}_w^{-1}(\langle M \rangle_{\frac{p}{q}})$ sont des langages réguliers, acceptés respectivement par des automates notés $\mathcal{B}$ et $\mathcal{C}$. D'après le corollaire 5.26, l'automate $\mathcal{B}$ accepte $\langle M + x \rangle_{\frac{p}{q}}$.

En général, l'automate $\mathcal{C}$ accepte des mots qui commencent par des 0. En effet, si par exemple le nombre $x$ appartient à $M$ alors l'image du mot $\varepsilon$ par $\mathcal{R}_w$ est $w$ ce qui implique que $\varepsilon$ est accepté par $\mathcal{C}$ ; il découle alors de propriété 5.21 que, pour tout entier $k < |w|$, le mots $0^k$ est accepté par $\mathcal{R}_w$ .

Montrons que les mots acceptés par $L(\mathcal{C})$ qui ne commencent pas par un 0 sont exactement les représentations de $((M - x) \cap V_{\frac{p}{q}})$. Soit $y \in ((M - x) \cap V_{\frac{p}{q}})$, qui est donc un nombre représentable. Il s'ensuit que $(x + y) \in M$ et donc $\mathcal{R}_w(\langle y \rangle_{\frac{p}{q}})$ est

égal à $\langle x+y\rangle_{\frac{p}{q}}$ (lemme 5.25) donc appartient à $\langle M\rangle_{\frac{p}{q}}$. Enfin $\langle y\rangle_{\frac{p}{q}} \in \left(\mathcal{R}_w^{-1}(\langle M\rangle_{\frac{p}{q}})\right)$ donc $\langle y\rangle_{\frac{p}{q}} \in L(\mathcal{C})$.

Soit $u$ un mot accepté par $\mathcal{C}$ qui ne commence pas par un 0 ; il existe donc $y \in V_{\frac{p}{q}}$ tel $u = \langle y\rangle_{\frac{p}{q}}$. Par définition de $\mathcal{C}$, $\mathcal{R}_w(u) \in \langle M\rangle_{\frac{p}{q}}$ donc en particulier $\pi_{\frac{p}{q}}(\mathcal{R}_w(u)) \in M$. D'après la proposition 5.23, $\pi_{\frac{p}{q}}(u) + \pi_{\frac{p}{q}}(w) = \pi_{\frac{p}{q}}(\mathcal{R}_w(u))$ donc $(x+y) \in M$ donc $y \in (M-x)$. Le mot $u$ est la représentation de $y$ qui est à la fois dans $(M-x)$ et dans $V_{\frac{p}{q}}$, donc $u \in \left\langle (M-x)\cap V_{\frac{p}{q}}\right\rangle_{\frac{p}{q}}$.

Le langage $\left\langle (M-x)\cap V_{\frac{p}{q}}\right\rangle_{\frac{p}{q}}$ est donc l'intersection des deux langages réguliers $L(\mathcal{C})$ et $[\![p]\!]^* \setminus (0[\![p]\!]^*)$ donc est un langage régulier. □

**Notation 5.38** – *Soient un nombre $x \in V_{\frac{p}{q}}$, un ensemble $M \subseteq V_{\frac{p}{q}}$ de nombres et $\mathcal{A}$ un automate qui accepte $0^*\langle M\rangle_{\frac{p}{q}}$.*

*On note $(\mathcal{A}\oplus x)$ un automate complet qui accepte $0^*\langle M+x\rangle_{\frac{p}{q}}$ et $(\mathcal{A}\ominus x)$ un automate complet qui accepte $0^*\left\langle (M-x)\cap V_{\frac{p}{q}}\right\rangle_{\frac{p}{q}}$*

Nous pouvons maintenant démontrer la proposition suivante qui est une reformulation du théorème 5.36 dans le cas où $S$ est l'ensemble des multiples de $q$.

**Proposition 5.39** – *Il n'existe pas d'automate $\mathcal{A}$ tel que, pour tout entier $m$, sa représentation $\langle m\rangle_{\frac{p}{q}}$ est accepté par $\mathcal{A}$ si et seulement $m$ est un multiple de $q$.*

Démonstration. Par l'absurde. Supposons au contraire qu'il existe un automate $\mathcal{A}_0$ qui vérifie cela. Sans perdre la généralité, on peut supposer 1) que $\mathcal{A}_0$ est complet et 2) que $L(\mathcal{A}_0) = 0^*\langle M\rangle_{\frac{p}{q}}$ pour un certain ensemble $M \in V_{\frac{p}{q}}$ ; en effet le langage $0^*\left(L(\mathcal{A}_0)\cap([\![p]\!]^*\setminus(0\,[\![p]\!]^*))\right)$ est régulier et vérifie les conditions de la proposition.. D'ailleurs, celles-ci impliquent que $M\cap\mathbb{N} = q\,\mathbb{N}$.

Pour tout $i \in \mathbb{Z}/q\mathbb{Z}$, (donc en particulier $i \in V_{\frac{p}{q}}$) on note

$$\mathcal{A}_i \;=\; \langle\, Q_i, [\![p]\!]^*, i_i, \delta_i, F_i\,\rangle \;=\; (\mathcal{A}_0\oplus i)\ .$$

Il s'agit par définition d'un automate complet qui accepte le langage $0^*\langle M+i\rangle_{\frac{p}{q}}$. L'équation suivante en découle directement :

$$\forall i\in\mathbb{Z}/q\mathbb{Z}\qquad L(\mathcal{A}_i) \;=\; \left\{\, u\in[\![p]\!]^* \;\middle|\; \pi_{\frac{p}{q}}(u)\in(M+i)\,\right\}\ .$$

Puisque $M\cap\mathbb{N} = q\,\mathbb{N}$, il s'ensuit que $(M+i)\cap\mathbb{N} = (q\mathbb{N}+i)$ et donc que

$$\forall i\in\mathbb{Z}/q\mathbb{Z},\ \forall u\in[\![p]\!]^*\quad \pi_{\frac{p}{q}}(u)\in\mathbb{N}\ \Longrightarrow\ \Big[u\in L(\mathcal{A}_i)\iff \pi_{\frac{p}{q}}(u)\,\%\,q = i\Big]\ .\ (*)$$

En d'autres termes, chaque mot dont la valeur est un entier $n$ est accepté par exactement un automate $\mathcal{A}_k$, celui dont l'indice $k = (n\,\%\,q)$ est la classe d'équivalence de $n$ modulo $q$.

L'automate $\mathcal{B}$, défini plus loin, est une variante du produit $\mathcal{A}_0\times\mathcal{A}_1\times\cdots\times\mathcal{A}_{(q-1)}$ dans laquelle 1) tous les états sont finals et 2) certaines transitions ont été supprimées. Étant donné un état $(s_0, s_1, \ldots, s_{(q-1)})$ du produit certaines de ses composantes sont des états finals dans leurs automates respectifs et d'autres non. Si par

exemple $s_i \in F_i$ alors certaines lettres sont **autorisées** à apparaître sur des transitions sortantes de $(s_0, s_1, \ldots, s_{(q-1)})$, les lettres $a \in [\![p]\!]$ tel que $(i\,p + a) \,\%\, q = 0$.

L'intuition derrière cette construction est la suivante : si un mot $u \in [\![p]\!]^*$ est de valeur entière, chacun des $\mathcal{A}_i$ est correct sur $u$, c'est-à-dire que $u$ est accepté par $\mathcal{A}_k$ où $k = \pi_{\frac{p}{q}}(u) \,\%\, q$ et n'est pas accepté par les $\mathcal{A}_i$, $i \neq k$. L'état $(s_0, s_1, \ldots, s_{(q-1)})$ atteint par le calcul de $u$ dans $\mathcal{B}$ vérifie (s'il existe) que pour tout $i$, $s_i$ est l'état atteint par le calcul de $u$ dans $\mathcal{A}_i$, donc $s_k \in F_k$ et chaque autre $i \in \mathbb{Z}/q\mathbb{Z}$ vérifie $s_i \notin F_i$. Les seules transitions sortants de l'état $(s_0, s_1, \ldots, s_{(q-1)})$ sont donc étiquetées par les lettres vérifiant $(k\,p + a) \,\%\, q = 0$, c'est-à-dire les lettres tels que $\pi_{\frac{p}{q}}(u\,a)$ est un entier.

Formellement,

$$\mathcal{B} \;=\; \langle\, Q_{\mathcal{B}},\, [\![p]\!],\, \delta_{\mathcal{B}},\, i_{\mathcal{B}},\, F_{\mathcal{B}} \,\rangle$$

où $Q_{\mathcal{B}} = Q_0 \times Q_1 \times \cdots \times Q_{(q-1)}$ est le produit cartésien des états des $\mathcal{A}_i$, où $i_{\mathcal{B}} = (i_0, i_1, \ldots, i_{q-1})$, où tous les états sont finals ($F_{\mathcal{B}} = Q_{\mathcal{B}}$) et où $\delta_{\mathcal{B}}$ est définie par :

$$(s_0, s_1, \ldots, s_{q-1}) \xrightarrow[\mathcal{B}]{a} (t_0, t_1, \ldots, t_{q-1})$$

$$\iff \begin{cases} s_0 \xrightarrow{a} t_0 & \text{dans } \mathcal{A}_0 \\ s_1 \xrightarrow{a} t_1 & \text{dans } \mathcal{A}_1 \\ \vdots \quad \vdots \quad \vdots & \vdots \\ s_{(q-1)} \xrightarrow{a} t_{(q-1)} & \text{dans } \mathcal{A}_{(q-1)} \\ s_i \in F_i, \text{ où } i \text{ est l'entier vérifiant } (i\,p + a) \,\%\, q = 0 \end{cases}$$

La ligne la plus importante de la définition de $\delta_{\mathcal{B}}$ est la dernière, les autres ne font qu'assurer que si $(s_0, s_1, \ldots, s_{q-1}) \xrightarrow[\mathcal{B}]{u} (t_0, t_1, \ldots, t_{q-1})$ est un chemin dans $\mathcal{B}$ alors, pour tout $i \in \mathbb{Z}/q\mathbb{Z}$, $s_i \xrightarrow{u} t_i$ est un chemin dans $\mathcal{A}_i$.

**Assertion 5.39.1** – *Soit un mot $u \in [\![p]\!]^*$. Si l'automate $\mathcal{B}$ existe, alors $\mathcal{B}$ admet un calcul pour $u$ si et seulement si $\pi_{\frac{p}{q}}(u) \in \mathbb{N}$.*

*Démonstration de l'assertion.* Par récurrence sur la longueur de $u$ ; l'assertion est vérifiée dans le cas où $u = \varepsilon$, son calcul atteint l'état initial $0 = \pi_{\frac{p}{q}}(\varepsilon) \,\%\, n$.

Soit un mot $u\,a \in [\![p]\!]^+$. L'hypothèse de récurrence se réécrit donc comme

$$\mathcal{B} \text{ admet un calcul pour } u \quad\iff\quad \pi_{\frac{p}{q}}(u) \in \mathbb{N}\ .$$

Or, l'hypothèse du sens direct ($ua$ possède un calcul dans $\mathcal{B}$) implique le membre gauche de cette équivalence et celle du sens réciproque ($\pi_{\frac{p}{q}}(ua) \in \mathbb{N}$) implique son membre droit (car $0^* L_{\frac{p}{q}}$ est clos par préfixe) ; les deux membres de cette équivalence sont donc vérifiés dans les deux sens. On note donc le calcul de $u$ dans $\mathcal{B}$

$$i_{\mathcal{B}} \xrightarrow[\mathcal{B}]{u} (s_0, s_1, \ldots, s_{q-1})\ .$$

Puisque $\pi_{\frac{p}{q}}(u) \in \mathbb{N}$, l'équation $(*)$ implique que $u$ est accepté par exactement un $\mathcal{A}_k$, celui indexé par

$$k \;=\; \pi_{\frac{p}{q}}(u) \,\%\, q\,; \qquad (**)$$

il s'ensuit que $s_k \in F_k$ et que pour tout $i \in \mathbb{Z}/q\mathbb{Z}$, $i \neq k$, $s_i \notin F_i$.

L'existence d'un calcul du mot $u\,a$ dans $\mathcal{B}$ est équivalente à l'existence de la transition

$$(s_0, s_1, \ldots, s_{q-1}) \xrightarrow[\mathcal{B}]{a} (t_0, t_1, \ldots, t_{q-1}) \quad \text{pour certains } t_0, t_1, \ldots t_{(q-1)} \ .$$

Puisque tous les $\mathcal{A}_i$ sont complets, celle-ci existe si et seulement si $(k\,p + a) \,\%\, q = 0$ (définition des transitions de $\mathcal{B}$) ; cette dernière équation, d'après $(**)$, se réécrit donc en

$$(p\,\pi_{\frac{p}{q}}(u) + a) \,\%\, q \ = \ 0 \quad \text{et est donc équivalente à} \quad \frac{1}{q}\left(p\,\pi_{\frac{p}{q}}(u) + a\right) \in \mathbb{N} \ ,$$

c'est-à-dire l'expression de $\pi_{\frac{p}{q}}(u\,a)$ (équation (4.5b)).

Puisque tous les états de $\mathcal{B}$ sont finals, il découle de l'assertion précédente que $L(\mathcal{B}) = 0^* L_{\frac{p}{q}}$ donc que $L_{\frac{p}{q}}$ est un langage régulier, ce qui contredit le théorème 4.10. Il n'existe donc pas d'automate $\mathcal{A}_0$ qui vérifie les conditions de la proposition. □

Nous allons, grâce au lemme suivant, montrer que le cas général du théorème 5.36 se réduit au cas particulier traité précédemment.

**Lemme 5.40** – *Soit un ensemble $R \subseteq \mathbb{Z}/q\mathbb{Z}$ de restes modulo $q$ tel que $(q, R)$ est un paramètre canonique*[2]*. Un entier $i$ est divisible par $q$ si et seulement si, pour tout $r \in R$, $(i + r) \,\%\, q \in R$*

Démonstration. Le sens direct est immédiat.

Sens réciproque. Soit un entier $i$ tel que pour tout $r \in R$, $(i + r) \,\%\, n \in R$. Démontrons que pour tout entier $j$, $(j \,\%\, q) \in R \iff (i + j) \,\%\, q \in R$, ce qui implique que $i$ est une période de l'ensemble périodique de paramètre $(q, R)$ (qui est canonique par hypothèse) donc que $i$ est un multiple de $q$.

Si $j \,\%\, q = r$ pour un certain $r \in R$, alors

$$(i + j) \ \equiv \ (i + r) \ \ [q]$$

dont la classe de congruence modulo $q$ appartient à $R$ par hypothèse, donc $(i + j) \,\%\, q \in R$.

Si au contraire $(i + j) \,\%\, q \in R$, alors appliquer $(q - 1)$ le cas précédent conclut la démonstration :

$$(j + i) \,\%\, q \in R \implies (j + 2\,i) \,\%\, q \in R \implies \cdots \implies (j + q\,i) \,\%\, q = j \,\%\, q \in R \ . \ \square$$

Démonstration du théorème 5.36. Soit $R = \{r_0, r_1, \ldots, r_k\} \subseteq \mathbb{Z}/q\mathbb{Z}$ un ensemble de restes modulo $q$ tel que $(q, R)$ est un paramètre canonique.

Par l'absurde. Supposons qu'il existe un automate $\mathcal{A}$ qui accepte un mot $u$ dont la valeur est entière si et seulement si $u \,\%\, q \in R$.

D'après le lemme 5.40 précédent, un entier $n$ est divisible par $q$ si et seulement si

$$[(n + r_0) \,\%\, q \in R] \ \wedge \ [(n + r_1) \,\%\, q \in R] \ \wedge \ \cdots \ \wedge \ [(n + r_k) \,\%\, q \in R] \ .$$

2. On rappelle qu'un *paramètre canonique* est un couple $(q, R)$ tel que $q$ est la plus petite période de $(q\mathbb{N} + R)$.

Soit un mot $u \in 0^* L_{\frac{p}{q}}$ tel que $\pi_{\frac{p}{q}}(u) \geqslant q$. Ce mot vérifie donc $\pi_{\frac{p}{q}}(u) \,\%\, q = 0$ si et seulement si

$$u \in L(\mathcal{A} \ominus r_0) \ \wedge \ u \in L(\mathcal{A} \ominus r_1) \ \wedge \ \cdots \ \wedge \ u \in L(\mathcal{A} \ominus r_k) \ .$$

Le produit d'automates $\mathcal{B}' = [(\mathcal{A} \ominus r_0) \times (\mathcal{A} \ominus r_1) \times \cdots \times (\mathcal{A} \ominus r_k)]$ accepte donc les mots dont la valeur est entière (et supérieure ou égale à $q$) si et seulement si elle divise $q$. Le langage $L(\mathcal{B}') \cup \{0\}$ est donc régulier, ce qui contredit la proposition 5.39. □

# Approximation de $L_{\frac{p}{q}}$

Le but de cette section est d'étudier les (sous-)monoïdes (additifs de $V_{\frac{p}{q}}$) contenant $\mathbb{N}$ et leurs représentations en base $\frac{p}{q}$, que nous appelons des approximations de $L_{\frac{p}{q}}$.

**Définition 5.41** – *On dit qu'un langage $L$ est une* approximation *de $L_{\frac{p}{q}}$, s'il satisfait les conditions suivantes :*

**a)** $L = \langle M \rangle_{\frac{p}{q}}$ *où $M$ est un monoïde additif;*

**b)** $L_{\frac{p}{q}} \subseteq L$.

## Approximation FLIP de $L_{\frac{p}{q}}$

Dans la suite, on appelle *diviseur unitaire de $q$* tout entier $d$ qui divise $q$ et qui est premier avec $(q/d)$. Par exemple si $q = 360 = 2^3 \times 3^2 \times 5$ ses diviseurs unitaires sont

$$1, \ 5, \ 2^3, \ 3^2, \ (2^3 \times 5), \ (3^2 \times 5), \ (2^3 \times 3^2) \quad \text{et} \quad (2^3 \times 3^2) \times 5 \ .$$

En particulier, si $q$ est un nombre premier, ou une puissance d'un nombre premier, il n'a que deux diviseurs unitaires, 1 et lui-même.

Dans la suite $d$ désigne un diviseur unitaire de $q$ *non-trivial*, c'est-à-dire que $d \neq 1$ et $d \neq q$.

**Définition 5.42** – *On note $M_d$ l'ensemble de nombre*

$$M_d \ = \ \left\{ \frac{m}{k} \in V_{\frac{p}{q}} \ \middle| \ \frac{m}{k} \text{ est irréductible et } k \text{ est premier avec } d \right\}$$

*et $H_d$ sa représentation*

$$H_d \ = \ \langle M_d \rangle_{\frac{p}{q}} \ .$$

L'ensemble $M_d$ est formé des nombres qui ont une classe de congruence modulo $d$ (selon la définition 5.30) ; il s'ensuit que $M_d$ est stable par addition (propriété 5.33d) et qu'il contient tous les entiers, ce qui est dit autrement par le lemme suivant.

**Lemme 5.43** – *Le langage $H_d$ est une approximation de $L_{\frac{p}{q}}$.*

Le langage $H_d$ est très semblable à $L_{\frac{p}{q}}$, comme le montre le lemme suivant [3].

3. Un mot $u\,a \in [\![p]\!]^+$ appartient à $L_{\frac{p}{q}}$ si et seulement si $p\,\pi_{\frac{p}{q}}(u) + a \equiv 0 \ \ [q]$

**Lemme 5.44** – *Soit un mot $u \in [\![p]\!]^*$ et une lettre $a \in [\![p]\!]$. Alors l'équivalence suivante est vérifiée :*

$$(u\,a) \in (0^*H_d) \quad \Longleftrightarrow \quad p\,\pi_{\frac{p}{q}}(u) + a \equiv 0 \ \ [d]$$

Démonstration. L'évaluation de $\pi_{\frac{p}{q}}(u\,a)$ est (d'après équation (2.2a))

$$\pi_{\frac{p}{q}}(u\,a) \ = \ \frac{p\,\pi_{\frac{p}{q}}(u) + a}{q} \ = \ \frac{p\,\pi_{\frac{p}{q}}(u) + a}{d\,(q/d)} \ .$$

Le mot $u\,a$ appartient à $0^*H_d$ si et seulement si $\pi_{\frac{p}{q}}(u\,a)$ appartient à $M_d$ donc si et seulement si le nombre

$$\pi_{\frac{p}{q}}(u\,a) \ = \ \frac{p\,\pi_{\frac{p}{q}}(u) + a}{d\,(q/d)}$$

admet une classe de congruence modulo $d$, c'est-à-dire si et seulement si la fraction irréductible de $(p\,\pi_{\frac{p}{q}}(u) + a)$ est $\frac{d\,m}{k}$ pour certains $m$ et $k$ (avec $k$ premier avec $d$). c'est-à-dire si et seulement si $p\,\pi_{\frac{p}{q}}(u) + a \equiv 0 \ \ [d]$. □

**Corollaire 5.45** – *Le langage $H_d$ est clos par préfixe et prolongeable.*

Démonstration. Soit un mot $u \in [\![p]\!]^*$. Pour toute lettre $a \in [\![p]\!]$, le nombre $(p\,\pi_{\frac{p}{q}}(u) + a)$ admet une classe de congruence modulo $d$ si et seulement si $\pi_{\frac{p}{q}}(u)$ en admet une. En effet, l'addition, la multiplication et division (par un entier $p$ premier avec $q$ donc avec $d$) conserve l'existence d'une classe de congruence modulo $d$ (propriétés 5.33).

Si $(u\,a) \in (0^*H_d)$, il suit du lemme précédent que $p\,\pi_{\frac{p}{q}}(u) + a \equiv 0 \ \ [d]$, donc, d'après le paragraphe précédent que $\pi_{\frac{p}{q}}(u)$ admet une classe de congruence modulo $d$ donc $\pi_{\frac{p}{q}}(u) \in M_d$ et enfin que $u \in (0^*H_d)$.

Si, au contraire, $u \in (0^*H_d)$ alors le nombre $(p\,\pi_{\frac{p}{q}}(u))$ admet une classe de congruence modulo $d \leqslant q < p$ donc il existe une lettre $b \in \{1, 2, \ldots, q\} \subsetneq [\![p]\!]$ telle que $(p\,\pi_{\frac{p}{q}}(u) + b) \ \equiv \ 0\ [d]$ donc, d'après le lemme précédent, que $(u\,b) \in (0^*H_d)$.

On a donc démontré que $(0^*H_d)$ est clos par préfixe et prolongeable à droite par des lettres non nulles ; il s'ensuit que $H_d$ est clos par préfixe et prolongeable à droite. □

Un raisonnement analogue à celui du lemme 5.44 donne le lemme suivant, plus général.

**Lemme 5.46** – *Soit un mot $u \in [\![p]\!]^*$. Pour tout mot $w \in [\![p]\!]^*$ dont on note la longueur $i$, l'équivalence suivante est vérifiée :*

$$(u\,w) \in (0^*H_d) \quad \Longleftrightarrow \quad p^i\,\pi_{\frac{p}{q}}(u) + q^i\,\pi_{\frac{p}{q}}(w) \ \equiv \ 0 \ \ [d^i] \ .$$

Démonstration. Soit un mot $w \in [\![p]\!]^*$ dont on note la longueur $i$. Le mot $(u\,w)$ appartient $0^*H_d$ si et seulement si le nombre

$$\pi_{\frac{p}{q}}(u\,w) \ = \ \pi_{\frac{p}{q}}(u)\left(\frac{p}{q}\right)^i + \pi_{\frac{p}{q}}(w) \ = \ \frac{1}{q^i}(\underbrace{\pi_{\frac{p}{q}}(u)p^i + \pi_{\frac{p}{q}}(w)q^i}_{x})$$

admet une classe de congruence modulo $d$. Le nombre $\pi_{\frac{p}{q}}(u\,w)$ a une classe de congruence modulo $d$ si et seulement si le dénominateur de sa fraction irréductible ne contient pas le facteur $d$, c'est-à-dire si et seulement si le nombre $x$ a une fraction irréductible de la forme

$$x \;=\; \pi_{\frac{p}{q}}(u)\,p^i + \pi_{\frac{p}{q}}(w)\,q^i \;=\; \frac{d^i\,m}{k}\;.$$

pour certains entiers $m$ et $k$ tel que $k$ est premier avec $d$. Cette dernière équation est enfin équivalente à

$$\pi_{\frac{p}{q}}(u)\,p^i + \pi_{\frac{p}{q}}(w)\,q^i \;\equiv\; 0 \quad [d^i]\;. \qquad \square$$

**Corollaire 5.47** – *Soient deux mots $u, v \in [\![p]\!]^*$ et un entier $k$*

**a)** *Si $\pi_{\frac{p}{q}}(u) \equiv \pi_{\frac{p}{q}}(v)\ [d^i]$, alors pour tout mot $w$ de longueur $i$, $u\,w$ appartient à $0^*H_d$ si et seulement si $v\,w$ appartient à $0^*H_d$.*

**b)** *Si il existe un mot $w$ de longueur $i$ tel que $u\,w$ et $v\,w$ appartiennent à $0^*H_d$, alors $\pi_{\frac{p}{q}}(u) \equiv \pi_{\frac{p}{q}}(v)\ [d^i]$.*

La proposition suivante est une conséquence de ce résultat, tout comme la proposition 4.18.

**Proposition 5.48** – *Le langage $H_d$ est FLIP.*

Démonstration. Par l'absurde. Supposons qu'il n'est pas FLIP ; il existe donc deux mots $u \in [\![p]\!]^*$ et $v \in [\![p]\!]^+$ tels que $u\,v^* \subseteq H_d$. Il découle donc du corollaire précédent que $\pi_{\frac{p}{q}}(u)$ et $\pi_{\frac{p}{q}}(u\,v)$ sont congrus modulo $d^{i\,|v|}$ pour tout $i$. Si l'on note leurs fractions irréductibles respectives par $\frac{m}{k}$ et $\frac{m'}{k'}$, il découle du corollaire 5.32 que les entiers $m\,k'$ et $m'\,k$ sont congrues modulo $d^{i\,|v|}$ pour tout $i$. ce qui implique que $m\,k' = m'\,k$ donc que $\pi_{\frac{p}{q}}(u) = \pi_{\frac{p}{q}}(u\,v)$. Ceci implique que $u = \varepsilon$ et $v = 0$ et donc que $0 \in H_d$ ce qui contredit sa définition. $\square$

**Remarque 5.49** – *Contrairement à ce qu'il peut sembler au premier abord, le langage $H_d$ n'est pas (en général) égal au langage $L_{\frac{p}{d}}$ (où le dénominateur de la base est $d$). En effet, le premier est défini récursivement par*

$$(u\,a) \in (0^*H_d) \quad\Longleftrightarrow\quad p\,\pi_{\boxed{\frac{p}{q}}}(u) + a \equiv 0 \quad [d]\;,$$

*alors que le second est défini par*

$$(u\,a) \in (0^*L_{\frac{p}{d}}) \quad\Longleftrightarrow\quad p\,\pi_{\boxed{\frac{p}{d}}}(u) + a \equiv 0 \quad [d]\;.$$

*Ce n'est pas la même fonction d'évaluation qui est utilisée ; ces deux conditions ne sont équivalentes que dans le cas où $(q/d) \equiv 1\ [d]$.*

*Par exemple, soit $\frac{p}{q} = \frac{7}{6}$ et $d = 3$ ; les évaluations du mot $31$ dans les bases $\frac{7}{6}$ et $\frac{7}{3}$ sont respectivement*

$$\pi_{\frac{7}{6}}(31) \;=\; \frac{7\times 3}{6^2} + \frac{1}{6} \;=\; \frac{21+6}{4\times 9} \;=\; \frac{3}{4} \;\in M_3$$

$$\pi_{\frac{7}{3}}(31) \;=\; \frac{7\times 3}{3^2} + \frac{1}{3} \;=\; \frac{21+3}{9} \;=\; \frac{8}{3} \;\notin \mathbb{N}$$

*Le mot* 31 *appartient donc à* $H_3$ *mais pas à* $L_{\frac{7}{3}}$ *(alors que par exemple, le mot* 3 *appartient aux deux).*

## Approximations régulières de $L_{\frac{p}{q}}$

On dit qu'une approximation $K$ de $L_{\frac{p}{q}}$ est *régulière* si $K$ est un langage régulier. Par exemple, le langage $\left\langle V_{\frac{p}{q}} \right\rangle_{\frac{p}{q}} = [\![p]\!]^* \setminus (0^*[\![p]\!]^*)$ est une approximation rationnelle de $L_{\frac{p}{q}}$. Nous verrons dans la suite que pour tout entier $n$, le langage $K_n$, défini ci-dessous, est une approximation régulière de $L_{\frac{p}{q}}$

**Définition 5.50** – *Pour tout entier* $n > 0$*, on note* $K_n$ *le langage des mots dont chaque préfixe de longueur inférieure à* $n$ *est dans* $L_{\frac{p}{q}}$*.*

$$K_n \;=\; \{\, a_0a_1\cdots a_{(k-1)} \in [\![p]\!]^* \mid \forall i < \min(n,k) \qquad (a_0a_1\cdots a_{(i-1)}) \in L_{\frac{p}{q}} \,\}\ .$$

*De plus on note* $K_0 = [\![p]\!]^* \setminus (0\,[\![p]\!]^*)$*.*

Puisque $L_{\frac{p}{q}}$ est clos par préfixe, les deux équations suivantes donnent une définition équivalente de $K_n$ :

$$\forall u \in [\![p]\!]^* \qquad |u| < n \quad\Longrightarrow\quad \Big[u \in K_n \quad\Longleftrightarrow\quad u \in L_{\frac{p}{q}}\Big]\ ; \tag{5.4a}$$

$$\forall u, v \in [\![p]\!]^* \qquad |u| = n \quad\Longrightarrow\quad \Big[(u\,v) \in K_n \quad\Longleftrightarrow\quad u \in L_{\frac{p}{q}}\Big]\ . \tag{5.4b}$$

Soient un entier $n$ et un mot $w \in K_n$ tel que $|w| \geqslant n$. Pour tout mot $v \in [\![p]\!]^*$ le mot $(w\,v) \in K_n$ ; il s'ensuit que tous les mots de $K_n \cap [\![p]\!]^{\geqslant n}$ sont Nérode-équivalents ; puisque les autres mots de $K_n$ sont en nombre fini, le lemme suivant est vérifié.

**Lemme 5.51** – *Pour tout* $n \in \mathbb{N}$*,* $K_n$ *est un langage régulier.*

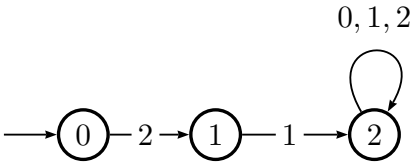



Figure 5 – Automate acceptant le langage $K_2$ en base $\frac{3}{2}$

Un automate acceptant $K_n$ peut être obtenu en coupant à la profondeur $n$ l'arbre étiqueté représentant le langage $L_{\frac{p}{q}}$ et en rajoutant une boucle pour chaque lettre sur chaque feuille ainsi créée. Par exemple, les figures 5 et 6 représentent des automates acceptant respectivement $K_2$ et $K_5$ en base $\frac{3}{2}$.

D'autre part, il découle de l'équation (5.4) que les langages $K_n$ forment une hiérarchie (qui est évidemment stricte).

$$L_{\frac{p}{q}} \;\subsetneq\; \cdots \;\subsetneq\; K_n \;\subsetneq\; \cdots \;\subsetneq\; K_1 \;\subsetneq\; K_0 \;=\; [\![p]\!]^* \setminus (0\,[\![p]\!]^*) \;=\; \left\langle V_{\frac{p}{q}} \right\rangle_{\frac{p}{q}}\ .$$

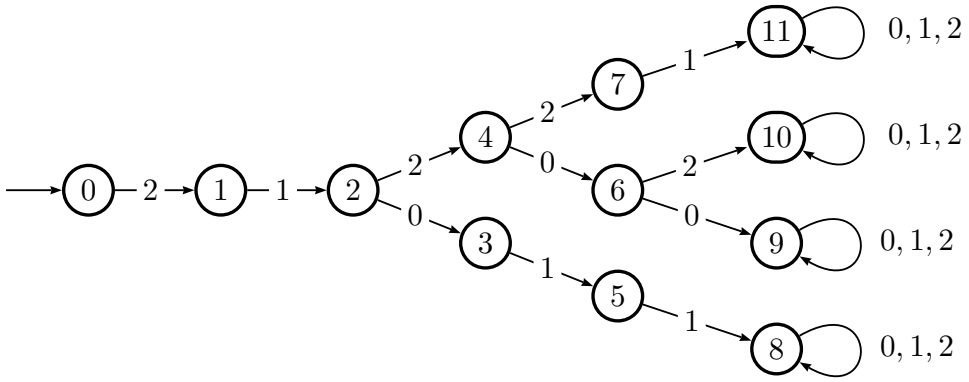



Figure 6 – Automate acceptant le langage $K_5$ en base $\frac{3}{2}$

Tous les $K_i$ sont des représentations de monoïdes. En effet, l'image de deux mots de $K_i$ par l'additionneur $\mathcal{C}$, est un mot de $K_i$, comme énoncé par la proposition suivante.

**Proposition 5.52** – *Soient un entier $n$ et deux mots $u, v \in K_n$ tels que $|v| \geqslant |u|$ ; on note la différence de leurs longueurs $k = |v| - |u|$. Alors $\mathcal{C}(0^k u, v)$ appartient à $K_n$.*

Démonstration. Si $u$ et $v$ sont tous les deux des longueurs inférieures à $n$, ils appartiennent à $L_{\frac{p}{q}}$ et la proposition est vérifiée, on suppose dans la suite que $|v| > n$.

On note $w = \mathcal{C}(0^k u, v)$. Puisque $u$ et $v$ appartiennent à $K_n$, ils ne commencent pas par des 0, il découle donc de la proposition 4.28 que $w$ ne commence pas par un 0 non plus.

On factorise $v = v'v''$ et $0^k u = u'u''$ tel que $|v'| = |u'| = n$ (ce qui implique que $|v''| = |u''|$). Le mot $v'$ est le préfixe de $v$ de longueur $n$ donc (d'après l'équation (5.4b)) il appartient à $L_{\frac{p}{q}}$. Le mot $u'$ est de la forme $0^i w$, où $w$ est un préfixe de $u$ de longueur inférieure à $n$ donc appartenant à $L_{\frac{p}{q}}$ (définition 5.50). Les mots $u'$ et $v$' appartiennent donc tous les deux à $O^* L_{\frac{p}{q}}$ ou, autrement dit,

$$\pi_{\frac{p}{q}}(u') \in \mathbb{N} \quad \text{et} \quad \pi_{\frac{p}{q}}(v') \in \mathbb{N} \; . \tag{$*$}$$

On note le calcul de $(u'u'', v'v'')$ dans $\mathcal{C}_{\frac{p}{q}}$ par

$$\xleftarrow{\langle s' \rangle \, \frac{p}{q}} \; s' \; \xleftarrow{(u', v') \mid w'} \; s \; \xleftarrow{(u'', v'') \mid w''} \; 0 \; .$$

On note de plus $m = (\pi_{\frac{p}{q}}(u') + \pi_{\frac{p}{q}}(v') + s)$. D'après le lemme 4.27 (page 105),

$$m \;=\; \pi_{\frac{p}{q}}(u') + \pi_{\frac{p}{q}}(v') + s \;=\; \pi_{\frac{p}{q}}(w') + s' \left(\frac{p}{q}\right)^{|w'|} \quad \text{et} \tag{$**$}$$

$$|w'| = |u'| = |v'| = n \; . \tag{$\overset{*}{**}$}$$

Puisque d'une part $s$ est un entier (car un état de $\mathcal{C}$) et d'autre part $\pi_{\frac{p}{q}}(u')$ et $\pi_{\frac{p}{q}}(v')$ sont des entiers (d'après $(*)$), il découle de $(**)$ que $m$ est un entier.

Puisque $w(=\langle s\rangle_{\frac{p}{q}} w'w'')$ ne commence pas par un 0, $\langle s\rangle_{\frac{p}{q}} w$ non plus, donc :

$$\langle s\rangle_{\frac{p}{q}} w' \;=\; \left\langle \pi_{\frac{p}{q}}(\langle s\rangle_{\frac{p}{q}} w')\right\rangle_{\frac{p}{q}} \;=\; \left\langle \pi_{\frac{p}{q}}(w') + s'\left(\frac{p}{q}\right)^{|w'|}\right\rangle_{\frac{p}{q}} \;=\; \langle m\rangle_{\frac{p}{q}}\ .$$

Puisque, d'après l'équation $(\overset{*}{**})$, $|w'| = n$, le préfixe de longueur $n$ de $w(=\langle s'\rangle_{\frac{p}{q}} w'w'')$ est un préfixe de $\langle m\rangle_{\frac{p}{q}}(=\langle s'\rangle_{\frac{p}{q}} w')$. Il découle enfin de l'équation (5.4b) que $w = \mathcal{C}_{\frac{p}{q}}(u,v)$ appartient à $K_n$. □

Le théorème suivant est alors simplement la compilation de la proposition 5.52 et du lemme 5.51.

**Théorème 5.53** – *Pour tout entier $n$, $K_n$ est une approximation régulière de $L_{\frac{p}{q}}$.*

Ce sont néanmoins des approximations grossières, dans le sens où chacune ne conserve qu'une partie finie de la structure de $L_{\frac{p}{q}}$. Ceci étant dit, nous conjecturons qu'on ne puisse pas faire mieux.

**Conjecture 5.54** – *Si $K$ est une approximation régulière de $L_{\frac{p}{q}}$ ; alors $K_n$ est inclus dans $K$ pour un certain $n$.*

Notez que si $K, L$ sont deux approximations de $L_{\frac{p}{q}}$ telles que $K \subseteq L$, alors $K$ est *plus fine*, car elle contient moins d'éléments superflus.

Défense de la conjecture. Soit $K$ une approximation régulière de $L_{\frac{p}{q}}$ accepté par un automate $\mathcal{A}$. Si $K$ ne contient aucun $K_n$, alors il existe deux familles infinies mots $(u_i)_{i\in\mathbb{N}}$ et $(v_i)_{i\in\mathbb{N}}$ tels que pour tout entier $i$,

- $|u_i| = i$ et $u_i$ appartient à $L_{\frac{p}{q}}$ (donc à $K$) :
- $u_i\, v_i$ n'appartient ni à $L_{\frac{p}{q}}$ ni à $K$.

Le mot $u_i\, v_i$ appartient à $K_i$, donc est le témoin du fait que $K_i \not\subseteq L$.

Puisque $\mathcal{A}$ est fini, il existe donc un ensemble $I \subseteq \mathbb{N}$ tel que pour tout $i \in I$, les calculs de tous les $u_i$ atteignent le même état $s$ et donc qu'il existe $v$ tel que pour tout $i \in I$, $(u_i v) \notin K$ (donc $\notin L_{\frac{p}{q}}$). Tous les mots $u \in K$ qui atteignent cet état $s$ vérifient donc

- soit $\pi_{\frac{p}{q}}(u) \notin L_{\frac{p}{q}}$ ;
- soit $\pi_{\frac{p}{q}}(u) \in L_{\frac{p}{q}}$ et donc $\pi_{\frac{p}{q}}(u)\,\%\, q^{|v|} \neq r$, pour un certain $r$.

C'est une condition (certes bien plus forte) de ce type qui mène à la contradiction de la démonstration du théorème 5.36.

Ceci est sans compter que l'ensemble $M = \pi_{\frac{p}{q}}(K)$ doit être un monoïde, et que celui-ci ne peut pas être finiment engendré (sinon $K$ serait FLIP d'après le théorème II).

## Chapitre 6

# Mots minimaux et envergures

Dans ce chapitre, on s'intéresse particulièrement aux mots minimaux (voir section 4.4, page 106 ) ; ce sont les mots infinis de $[\![q]\!]^\omega$ (et non $[\![p]\!]^\omega$) qui étiquettent des branches (infinies) de l'automate infini $\mathcal{T}_{\frac{p}{q}}$ qui, rappelons le, accepte le langage $0^*L_{\frac{p}{q}}$ et admet $\mathbb{N}$ comme ensemble d'états. Il existe une unique branche partant de chaque état/entier $n$ qui est étiquetée par un mot de $[\![q]\!]^\omega$, ce mot est donc naturellement associé à l'entier $n$.

Dans la section 6.1, nous étudions la fonction $\xi$ qui associe, pour tout entier $n$, le mot minimal de $n$ au mot minimal de $(n+1)$. Le préfixe de longueur $i$ du mot minimal de $n$ est caractérisé par la congruence de $n$ modulo $q^i$. Si bien que le préfixe de longueur $i$ de $\xi(w)$ est caractérisé par celui de $w$. Il s'ensuit que la fonction $\xi$ peut-être réalisée par un transducteur **infini** lettre-à-lettre et séquentiel.

On s'emploie donc à construire ce transducteur, noté $\mathcal{D}_{\frac{p}{q}}$. En base $\frac{3}{2}$, ou plus généralement dans les bases $\frac{p}{q}$ vérifiant $p = 2q-1$, il est obtenu en remplaçant chaque étiquette de $\mathcal{T}_{\frac{p}{q}}$ par un ensemble de couples de lettres (qui ne dépend que de cette étiquette). En d'autres termes, dans ce cas particulier, $\mathcal{D}_{\frac{p}{q}}$ et $\mathcal{T}_{\frac{p}{q}}$ ont le même graphe sous-jacent, et le premier est obtenu à partir du second par une simple substitution de $[\![p]\!]$ vers $\mathbb{P}([\![q]\!] \times [\![q]\!])$.

Le cas général nécessite une étape supplémentaire qui consiste à modifier préalablement l'alphabet de $\mathcal{T}_{\frac{p}{q}}$. Son nouvel alphabet, noté $B_{\frac{p}{q}}$, est l'intervalle (entier) de cardinal $(2q-1)$ dont le plus grand élément est $(p-1)$. A partir de $\mathcal{T}_{\frac{p}{q}}$ est donc construit un autre automate infini $\mathcal{T}'_{\frac{p}{q}}$ ; soit en supprimant les transitions de $\mathcal{T}_{\frac{p}{q}}$ dont les étiquettes ne sont pas dans $B_{\frac{p}{q}}$ ; soit en ajoutant de nouvelles transitions étiquetées par les nouvelles lettres (et qui respectent la définition des transitions de $\mathcal{T}_{\frac{p}{q}}$, étendue à $B_{\frac{p}{q}}$). Le transducteur dérivé $\mathcal{D}_{\frac{p}{q}}$ est ensuite obtenu à partir de $\mathcal{T}'_{\frac{p}{q}}$ grâce à une simple substitution, comme dans le cas particulier décrit plus haut.

On démontre enfin que le transducteur construit de cette manière a le comportement attendu :

**Théorème III** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q > 1$. Le transducteur $\mathcal{D}_{\frac{p}{q}}$ réalise la fonction $\xi$.*

Dans la section 6.2, on s'intéresse à toutes les branches de $\mathcal{T}_{\frac{p}{q}}$ qui partent d'un état donné $n$ ; elles sont étiquetées par des mots (infinis) dont les évaluations (après

Les résultats présentés dans ce chapitre ont été publiés dans les actes de WORDS 2013, voir [3].

la virgule) forment un intervalle. On appelle *envergure*[1] *de* $n$ la longueur de cet intervalle, c'est-à-dire la différence des valeurs après la virgule du mot maximal et du mot minimal de $n$. L'envergure de $n$ est donc égal à la valeur d'un *mot-témoin* $w$ qui résulte de la soustraction chiffre-à-chiffre[2] du mot minimal de $n$ au mot maximal de $n$.

La relation triviale entre le mot minimal de $(n+1)$ et le mot maximal de $n$ (voir lemme 4.34, page 108) indique une relation entre la fonction $\xi$ et les mots-témoins : chaque mot-témoin est essentiellement la sortie moins l'entrée d'un couple accepté par $\xi$.

Il s'avère que les mots-témoins sont acceptés par $\mathcal{T}'_{\frac{p}{q}}$, et même que chaque mot infini accepté par $\mathcal{T}'_{\frac{p}{q}}$ appartient à l'adhérence des mots témoins. Si bien que l'adhérence de $\mathsf{S}_{\frac{p}{q}}$ est formé des évaluations des mots acceptés par $\mathcal{T}'_{\frac{p}{q}}$. Les propriétés topologiques de cet ensemble dépendent donc du signe de $(p-(2\,q-1))$.

**Théorème IV** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q > 1$.*

**a)** *Si $p \leqslant (2\,q-1)$, alors l'adhérence de $\mathsf{S}_{\frac{p}{q}}$ est un intervalle.*

**b)** *Si $p > (2\,q-1)$, alors l'adhérence de $\mathsf{S}_{\frac{p}{q}}$ est un ensemble de Cantor.*

Un *ensemble de Cantor* est ensemble fermé, borné, d'intérieur vide et qui ne contient aucun point isolé. La démonstration du théorème IV repose sur l'intuition suivante. Dans les *petites bases* ($p \leqslant (2q-1)$), des transitions sont *ajoutées* à $\mathcal{T}'_{\frac{p}{q}}$ (par rapport à $\mathcal{T}_{\frac{p}{q}}$), et l'on verra que les mots supplémentaires acceptés par $\mathcal{T}'_{\frac{p}{q}}$ ne contribuent pas à $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$. Au contraire, dans les *grandes bases* ($p > (2q-1)$), certaines transitions de $\mathcal{T}_{\frac{p}{q}}$ ont été *supprimées* pour construire $\mathcal{T}'_{\frac{p}{q}}$ et chacune retire un intervalle de $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$.

## La fonction successeur sur les mots minimaux

Dans toute la suite, $p$ et $q$ désignent deux entiers premiers entre eux tels que $p > q > 1$. Ils définissent le système de numération à base $\frac{p}{q}$ ; voir chapitre 4, particulièrement la section 4.4 (page 106) qui fixe les notations utilisées dans la suite.

On note $\xi$ la fonction partielle de $[\![q]\!]^\omega$ dans lui même définie par, pour tout entier $n$, $\xi(w_n^-) = w_{(n+1)}^-$. Bien qu'elle prenne en argument un mot infini (donc pris dans un ensemble indénombrable) le domaine de définition est un ensemble dénombrable : $\mathsf{Dom}(\xi) = \Omega_{\frac{p}{q}}$.

**Lemme 6.1** – *Soient deux entiers $n$ et $m$. Pour tout entier $i$, les préfixes de longueur $i$ de $w_n^-$ et $w_m^-$ sont égaux si et seulement si les préfixes de longueur $i$ de $\xi(w_n^-)$ et $\xi(w_m^-)$ sont égaux.*

1. La notion d'envergure utilisée dans ce chapitre (traduction du terme anglais *span* utilisé dans [3]) est normalisée, voir remarque 6.21, page 156.
2. La soustraction chiffre-à-chiffre est similaire à l'addition chiffre-à-chiffre décrite dans sous-section 4.3.

Démonstration. On note respectivement $u, v \in [\![q]\!]^*$ les préfixes de longueurs $i$ de $w_n^-, w_m^-$ et $u', v' \in [\![q]\!]^*$ ceux de $w_{(n+1)}^-, w_{(m+1)}^-$.

Si $u = v$, alors $(n \cdot u)$ et $(m \cdot u)$ existent dans $\mathcal{T}_{\frac{p}{q}}$ donc $n \equiv m\ [q^i]$ (lemme 4.14b, page 98). Il s'ensuit que $(n+1) \equiv (m+1)\ [q^i]$. Puisque par définition, $((n+1) \cdot u')$ existe, il en découle que $((m+1) \cdot u')$ existe (lemme 4.14a, page 98). Puisque $u'$ est sur l'alphabet minimal, $u'$ est préfixe de $w_{(m+1)}^-$, donc $u' = v'$.

Un raisonnement identique montre que $u' = v'$ implique $u = v$. □

Le lemme précédent implique que la lecture des $i$ premières lettres de $w_n^-$ suffit à déterminer les $i$ premières lettres de $\xi(w_n^-)$. Puisque la clôture topologique des mots minimaux est $[\![q]\!]^\omega$ tout entier (propriété 4.33c, page 108), $\xi$ est étendue par continuité en une fonction totale de $[\![q]\!]^\omega$ dans lui-même ; elle est toujours noté $\xi$.

Il découle ainsi du lemme 6.1 précédent que pour tout mot $w \in [\![q]\!]^\omega$ la connaissance des $i$ premières lettres de $w$ détermine les $i$ premières lettres de $\xi(w)$. La fonction $\xi$ est donc réalisée par transducteur infini, lettre-à-lettre et séquentiel de $[\![q]\!]^*$ vers $[\![q]\!]^*$, appelé *transducteur dérivé* et noté $\mathcal{D}_{\frac{p}{q}}$.

Ce transducteur est construit par deux transformations successives de $\mathcal{T}_{\frac{p}{q}}$ ; elles sont décrites par les deux prochaines sous-sections 6.1.1 et 6.1.2.

## Changement d'alphabet

On note $B_{\frac{p}{q}}$ l'alphabet $\{p-(2\,q-1), \ldots, (p-1)\}$. Il s'agit de l'intervalle d'entiers dont le cardinal est $(2\,q-1)$ et dont le plus grand élément est $(p-1)$. Les propriétés suivantes découlent immédiatement de cette définition.

**Propriété 6.2** –

**a)** *L'alphabet maximal* $\{(p-q), (p-q+1), \ldots, (p-1)\}$ *est inclus dans* $B_{\frac{p}{q}}$.
**b)** *Si* $p = (2\,q-1)$, *alors* $B_{\frac{p}{q}} = [\![p]\!]$.
**c)** *Si* $p > (2\,q-1)$, *alors* $B_{\frac{p}{q}} \subsetneq [\![p]\!]$ ; *plus précisément,* $B_{\frac{p}{q}}$ *est l'ensemble des* $(2\,q-1)$ *plus grandes lettres de* $[\![p]\!]$.
**d)** *Si* $p < (2\,q-1)$, *alors* $B_{\frac{p}{q}} \supsetneq [\![p]\!]$ *et contient* $(2\,q-1-p)$ *chiffres* ***négatifs***.

De plus, l'alphabet $B_{\frac{p}{q}}$ est un intervalle entier de cardinal $(2\,q-1)$, un nombre impair, il admet donc un *centre*. Il s'agit de la lettre $(p-q)$ qui aura un rôle lors de l'étape de remplacement des étiquettes.

L'automate infini $\mathcal{T}'_{\frac{p}{q}}$ est défini (voir ci-dessous) de la même manière que $\mathcal{T}_{\frac{p}{q}}$ à l'exception de l'alphabet : le premier est sur $B_{\frac{p}{q}}$ alors que le second est sur $[\![p]\!]$. Suivant le signe de $(2\,q-1)$, l'un de ces deux alphabets est inclus dans l'autre ; $\mathcal{T}'_{\frac{p}{q}}$ résulte donc soit de l'ajout dans $\mathcal{T}_{\frac{p}{q}}$ de transitions étiquetés par des chiffres négatifs soit de la suppression des transitions de $\mathcal{T}'_{\frac{p}{q}}$ dont les étiquettes sont trop petites.

**Définition 6.3** – *On note* $\mathcal{T}'_{\frac{p}{q}}$ *l'automate suivant :*

$$\mathcal{T}'_{\frac{p}{q}} \;=\; \langle\ \mathbb{N},\ B_{\frac{p}{q}},\ \delta,\ 0,\ \mathbb{N}\ \rangle$$

*où $\delta$ est la fonction de transitions de $\mathcal{T}_{\frac{p}{q}}$ restreinte ou étendue à $B_{\frac{p}{q}}$, c'est-à-dire :*

$$\forall n, m \in \mathbb{N},\ \forall a \in B_{\frac{p}{q}} \qquad n \xrightarrow[\mathcal{T}'_{\frac{p}{q}}]{a} m \iff q\,m = p\,n + a\ . \tag{6.1}$$

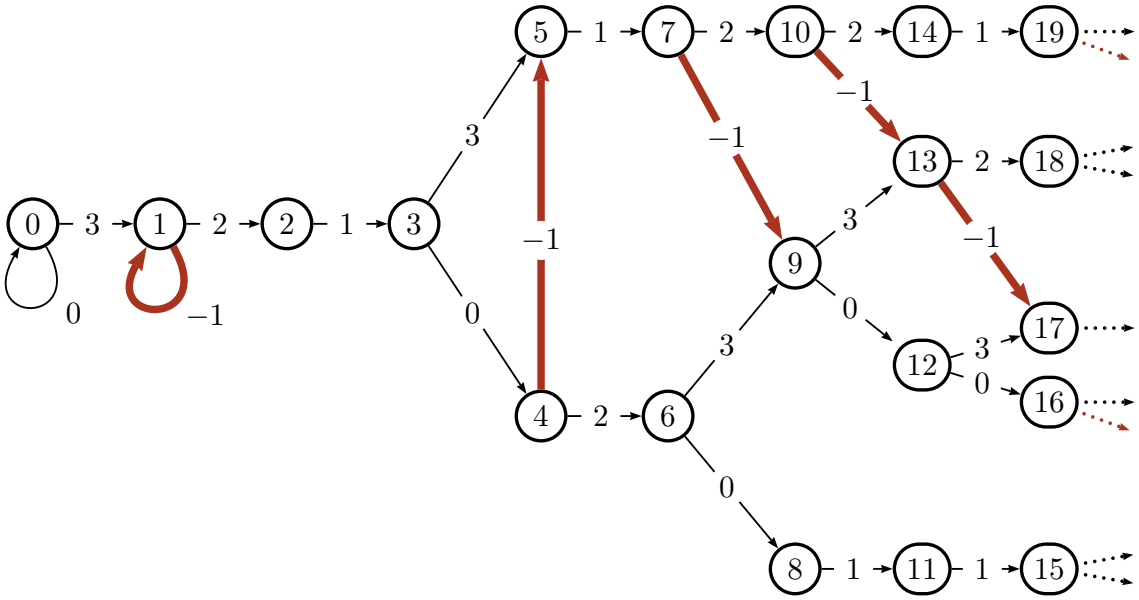



Figure 1 – Transformation de $\mathcal{T}_{\frac{4}{3}}$ en $\mathcal{T}'_{\frac{4}{3}}$

**Exemple 6.4** – *La base $\frac{3}{2}$ correspond au cas où $p = (2q-1)$ donc où $B_{\frac{p}{q}} = [\![p]\!]$ ; les automates $\mathcal{T}'_{\frac{3}{2}}$ et $\mathcal{T}_{\frac{3}{2}}$ sont donc égaux.*

**Exemple 6.5** – *La figure 1 représente l'automate $\mathcal{T}'_{\frac{4}{3}}$. La base $\frac{4}{3}$ illustre le cas où $p < (2q-1)$, c'est-à-dire que $B_{\frac{p}{q}}$ contient $[\![p]\!]$ et des chiffres négatifs (ici un seul, –1). Sur la figure, les transitions de $\mathcal{T}'_{\frac{4}{3}}$ qui ont été ajoutées à celles de $\mathcal{T}_{\frac{4}{3}}$ sont dessinées en gras. Le graphe (orienté) sous-jacent de $\mathcal{T}'_{\frac{4}{3}}$ n'est pas un arbre, mais un DAG.*

**Exemple 6.6** – *La figure 2 représente l'automate $\mathcal{T}'_{\frac{7}{3}}$. La base $\frac{7}{3}$ illustre le cas où $p > (2q-1)$, c'est-à-dire où $B_{\frac{p}{q}}$ est strictement inclus dans $[\![p]\!]$. Les transitions de $\mathcal{T}_{\frac{p}{q}}$ étiquetées par les $2 (= p - (2q-1))$ plus petites lettres de $[\![p]\!]$ n'existent pas dans $\mathcal{T}'_{\frac{p}{q}}$ ; ces transitions sont dessinées en pointillé sur la figure. Le graphe sous-jacent de $\mathcal{T}'_{\frac{7}{3}}$ est un ensemble (infini) d'arbres ; sa partie accessible reste un arbre.*

Une simple vérification établit le lemme suivant.

**Lemme 6.7** – *Soit $n$ un entier.*

**a)** *Si $(n-1) \mid q$, alors l'état $n$ admet exactement une transition sortante dans $\mathcal{T}'_{\frac{p}{q}}$.*

**b)** *Sinon, l'état $n$ admet exactement deux transitions sortantes dans $\mathcal{T}'_{\frac{p}{q}}$.*

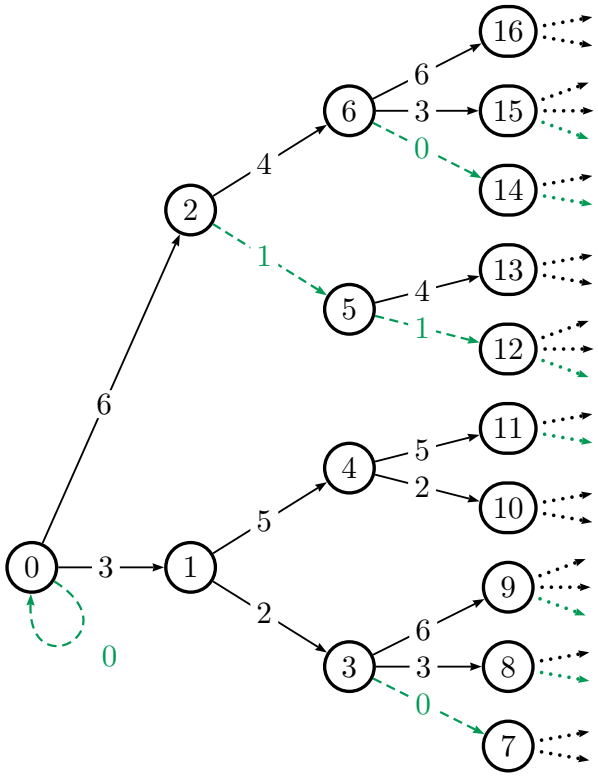



Figure 2 – Transformation de $\mathcal{T}_{\frac{7}{3}}$ en $\mathcal{T}'_{\frac{7}{3}}$

**Lemme 6.8** – *Soit un mot $u \in B_{\frac{p}{q}}^*$ accepté par $\mathcal{T}'_{\frac{p}{q}}$. Alors $\pi_{\frac{p}{q}}(u)$ est un entier et le calcul de $u$ atteint l'état $(\pi_{\frac{p}{q}}(u))$.*

Démonstration. Par récurrence sur la longueur de $u$ ; le lemme est évidemment vérifié si $u = \varepsilon$.

Soit un mot $u\,a \in B_{\frac{p}{q}}^+$ accepté par $\mathcal{T}'_{\frac{p}{q}}$. Par hypothèse de récurrence, l'état atteint par le calcul de $u$ est $\pi_{\frac{p}{q}}(u) \in \mathbb{N}$ ; on note alors le calcul de $u\,a$ dans $\mathcal{T}'_{\frac{p}{q}}$ par :

$$0 \xrightarrow{\ u\ } \pi_{\frac{p}{q}}(u) \xrightarrow{\ a\ } m\ .$$

La transition de droite ci-dessus implique que $p\,\pi_{\frac{p}{q}}(u) + a = q\,m$ (équation (6.1)). Si bien que

$$\pi_{\frac{p}{q}}(u\,a) \ = \ \pi_{\frac{p}{q}}(u)\,\frac{p}{q} + \frac{a}{q} \ = \ m\ ,$$

ce qui conclut la récurrence □

## Remplacement des étiquettes

La *fonction de remplacement* $\psi$ envoie une lettre $a \in B_{\frac{p}{q}}$ sur l'ensemble des couples $b \mid c \in [\![q]\!] \times [\![q]\!]$ dont la différence (sortie moins entrée) est égale à la différence de $a$ au centre de $B_{\frac{p}{q}}$ : $(c - b) = a - (p - q)$. Plus formellement, $\psi$ est définie par :

$$\begin{array}{rccl} \psi : & B_{\frac{p}{q}} & \to & \mathbb{P}([\![q]\!] \times [\![q]\!]) \\ & a & \mapsto & \{\,(b \mid c) \ \mid\ 0 \leqslant b, c < q \quad \text{et} \quad (c - b) \ = \ \overline{\psi}(a)\,\} \end{array} \tag{6.2}$$

où $\overline{\psi}$ est la fonction qui calcule la distance au centre de $B_{\frac{p}{q}}$ :

$$\forall a \in B_{\frac{p}{q}} \qquad \overline{\psi}(a) \;=\; a - (p-q) \tag{6.3}$$

Le transducteur dérivé $\mathcal{D}_{\frac{p}{q}}$ résulte du remplacement dans $\mathcal{T}_{\frac{p}{q}}$ de chaque étiquette $a$ par $\psi(a)$. Une définition formelle est donnée après quelques exemples.

**Exemple 6.9** – *En base $\frac{3}{2}$, le centre de $B_{\frac{p}{q}}$ est $1(= p-q)$ ; les fonctions $\psi$ et $\overline{\psi}$ sont données ci-dessous.*

$$\begin{array}{llll|lll} \overline{\psi}\colon & 0 & \longmapsto & -1 & \psi\colon & 0 & \longmapsto & \{\ 1|0\ \} \\ \overline{\psi}\colon & 1 & \longmapsto & 0 & \psi\colon & 1 & \longmapsto & \{\ 1|1,\ 0|0\ \} \\ \overline{\psi}\colon & 2 & \longmapsto & 1 & \psi\colon & 2 & \longmapsto & \{\ 0|1\ \} \end{array}$$

*Le transducteur $\mathcal{D}_{\frac{3}{2}}$ est représenté par la figure 3.*

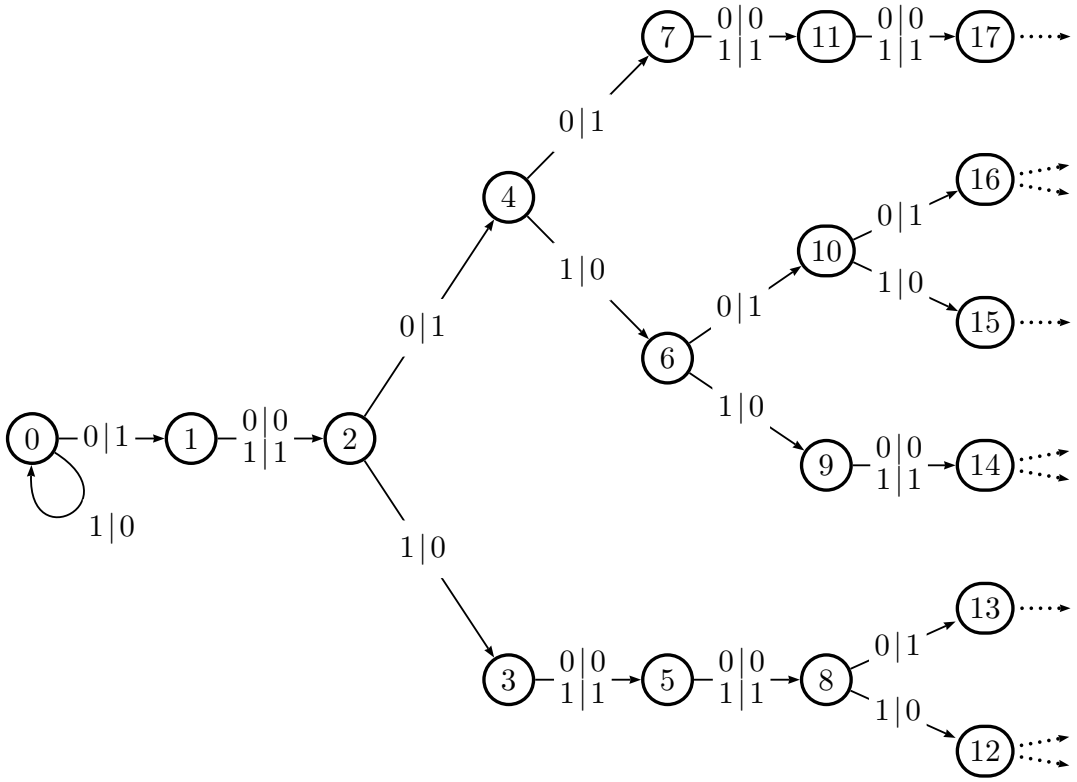



Figure 3 – Le transducteur dérivée en base $\frac{3}{2}$

**Exemple 6.10** – *En base $\frac{4}{3}$, le centre de $B_{\frac{p}{q}}$ est $1(= p-q)$ ; les fonctions $\psi$ et $\overline{\psi}$ sont données ci-dessous.*

$$\begin{array}{llll|lll} \overline{\psi}\colon & -1 & \longmapsto & -2 & \psi\colon & -1 & \longmapsto & \{\ 2|0\ \} \\ \overline{\psi}\colon & 0 & \longmapsto & -1 & \psi\colon & 0 & \longmapsto & \{\ 2|1,\ 1|0\ \} \\ \overline{\psi}\colon & 1 & \longmapsto & 0 & \psi\colon & 1 & \longmapsto & \{\ 2|2,\ 1|1,\ 0|0\ \} \\ \overline{\psi}\colon & 2 & \longmapsto & 1 & \psi\colon & 2 & \longmapsto & \{\ 1|2,\ 0|1\ \} \\ \overline{\psi}\colon & 3 & \longmapsto & 2 & \psi\colon & 3 & \longmapsto & \{\ 0|2\ \} \end{array}$$

*Le transducteur $\mathcal{D}_{\frac{4}{3}}$ est représenté par la figure 4.*

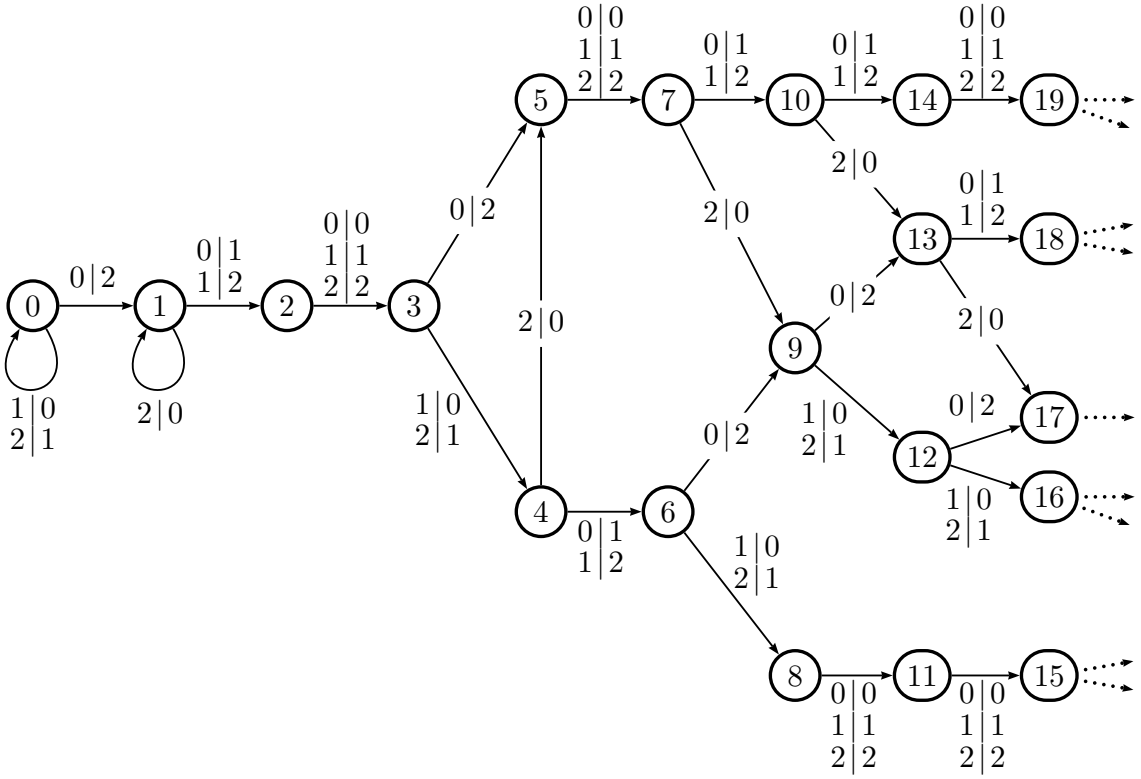



Figure 4 – Le transducteur dérivée en base $\frac{4}{3}$

**Exemple 6.11** – *En base $\frac{7}{3}$, le centre de $B_{\frac{p}{q}}$ est $4 (= p - q)$ ; les fonctions $\psi$ et $\overline{\psi}$ sont données ci-dessous.*

$$\begin{array}{llll|lll} \overline{\psi}\colon & 2 & \longmapsto & -2 & \psi\colon & 2 & \longmapsto & \{\ 2|0\ \} \\ \overline{\psi}\colon & 3 & \longmapsto & -1 & \psi\colon & 3 & \longmapsto & \{\ 2|1,\ 1|0\ \} \\ \overline{\psi}\colon & 4 & \longmapsto & 0 & \psi\colon & 4 & \longmapsto & \{\ 2|2,\ 1|1,\ 0|0\ \} \\ \overline{\psi}\colon & 5 & \longmapsto & 1 & \psi\colon & 5 & \longmapsto & \{\ 1|2,\ 0|1\ \} \\ \overline{\psi}\colon & 6 & \longmapsto & 2 & \psi\colon & 6 & \longmapsto & \{\ 0|2\ \} \end{array}$$

*Le transducteur $\mathcal{D}_{\frac{7}{3}}$ est représenté par la figure 5.*

**Remarque 6.12** – *Les fonctions $\psi$ des exemples 6.10 et 6.11 sont identiques à un décalage sur l'entrée près. Ceci toujours est vérifié dans le cas où deux bases $\frac{p}{q}$ et $\frac{p'}{q}$ ont le même dénominateur.*

**Définition 6.13** – *On appelle* transducteur dérivé, *le transducteur*

$$\mathcal{D}_{\frac{p}{q}} \;=\; \langle\ \mathbb{N},\ [\![q]\!],\ [\![q]\!],\ E_{\mathcal{D}},\ 0,\ \mathbb{N}\ \rangle$$

*dont les transitions sont définies par :*

$$\forall n, m \in \mathbb{N},\ \ \forall b, c \in [\![q]\!] \qquad n \xrightarrow[\mathcal{D}_{\frac{p}{q}}]{b\,|\,c} m \iff q\,m \;=\; p\,n + (c - b) + (p - q) \quad (6.4)$$

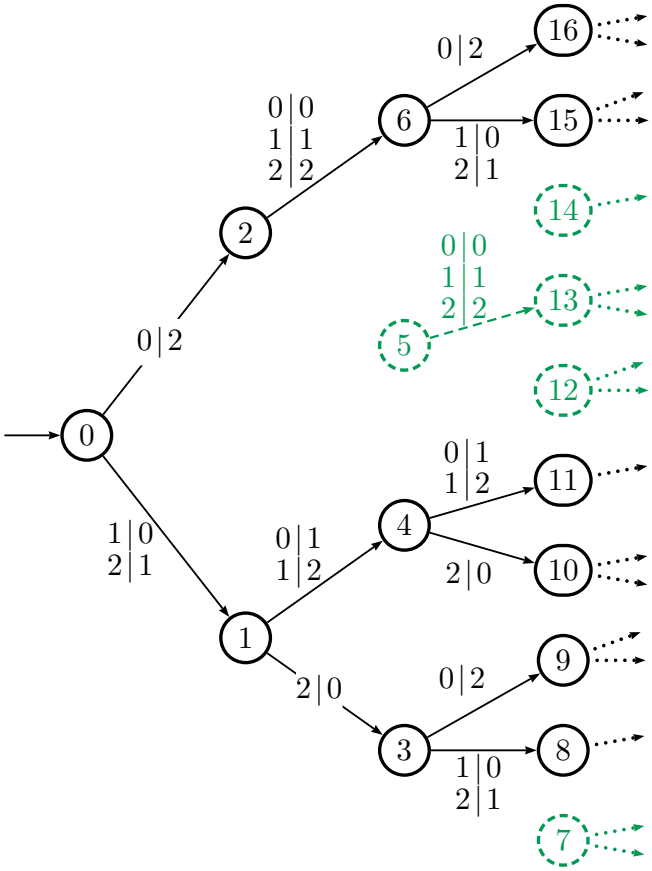



Figure 5 – Le transducteur dérivée en base $\frac{7}{3}$

Localement, ce transducteur réalise une permutation des lettres de $[\![q]\!]$, comme l'exprime le lemme suivant.

**Lemme 6.14** – *Soient un état $n$ de $\mathcal{D}_{\frac{p}{q}}$ et $x$ une lettre de $[\![q]\!]$.*

**a)** *Il existe une unique lettre $b \in [\![q]\!]$ et un unique entier $m$ tels que $n \xrightarrow{b \mid x} m$.*

**b)** *Il existe une unique lettre $c \in [\![q]\!]$ et un unique entier $m$ tels que $n \xrightarrow{x \mid c} m$.*

Démonstration. L'équation à droite de l'équation (6.4) correspond exactement à la division euclidienne de $(pn + c + (p - q))$ par $q$ : le reste est $b$ et le quotient est $m$. Si l'on fixe l'entier $n$ et la lettre $c \in [\![q]\!]$, il existe donc un unique entier $m$ et une unique lettre $b \in [\![q]\!]$ satisfaisant cette équation. Il s'ensuit que chaque lettre $c$ de $[\![q]\!]$ apparaît en entrée sur exactement une transition sortante de chaque état $n$ ; un raisonnement similaire montre le résultat symétrique pour les lettres en sortie. □

**Corollaire 6.15** – *Soient un état $n$ de $\mathcal{D}_{\frac{p}{q}}$ et un mot infini $w \in [\![q]\!]^\omega$.*

**a)** *Il existe un unique mot infini $w' \in [\![q]\!]^\omega$ tel que $n \xrightarrow{w \mid w'} \cdots$*

**b)** *Il existe un unique mot infini $w'' \in [\![q]\!]^\omega$ tel que $n \xrightarrow{w'' \mid w} \cdots$*

**Corollaire 6.16** – *Le transducteur dérivé $\mathcal{D}_{\frac{p}{q}}$ réalise une bijection de $[\![q]\!]^\omega$ dans lui-même.*

## Fonction réalisée par le transducteur dérivé

Le but de cette sous-section 6.1.3 est d'établir le théorème III, rappelé ci-dessous.

**Théorème III** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q > 1$. Le transducteur $\mathcal{D}_{\frac{p}{q}}$ réalise la fonction $\xi$.*

La démonstration de ce théorème requiert quelques résultats intermédiaires. Le lemme 6.14b a précédemment établi que chaque état possède une unique transition sortante dont l'entrée est une lettre donnée ; le lemme suivant donne l'expression de la lettre sortante et l'état d'arrivée de cette transition.

**Lemme 6.17** – *Soient un entier $n$ et une lettre $b \in [\![q]\!]$. La transition $n \xrightarrow{b \mid c} m$ existe dans $\mathcal{D}_{\frac{p}{q}}$, où*

$$c \;=\; (b - p\,(n+1)) \,\%\, q \qquad \textit{et} \qquad m \;=\; \left\lceil \frac{p\,(n+1) - b}{q} - 1 \right\rceil \;.$$

Démonstration. D'après le lemme 6.14b, il existe $c \in [\![q]\!]$ et $m \in \mathbb{N}$ tels que $n \xrightarrow{b \mid c} m$, donc qui satisfont (d'après l'équation (6.4))

$$q\,m \;=\; p\,n + (c - b) + (p - q) \;. \qquad (*)$$

Cette équation se réécrit en

$$c \;=\; b - p\,(n+1) + (q+1)\,m \;\equiv\; (b - p\,(n+1)) \quad [q] \;.$$

De plus, $c$ appartient à $[\![q]\!]$, ce qui implique que $c \,\%\, q = c$, donc

$$c \;=\; (b - (n+1)\,p) \,\%\, q \;.$$

D'autre part, l'équation $(*)$ peut également se réecrire en

$$m \;=\; \frac{p\,n + (c - b) + (p - q)}{q} \;=\; \frac{(p\,(n+1) - b)}{q} + \frac{c}{q} - 1 \;,$$

or, puisque $c < q$ et que $m$ est un entier, $m = \left\lceil \frac{p\,(n+1) - b}{q} - 1 \right\rceil$. □

La proposition suivante explicite la fonction réalisée par $\mathcal{D}_{\frac{p}{q}}$ sur les mots finis à partir de chaque état.

**Proposition 6.18** – *Soient deux mots sur $u, v \in [\![q]\!]^*$, et quatre entiers $i, j, m$ et $n$. L'implication suivante est alors vérifiée :*

$$\left.\begin{array}{ll} n \xrightarrow{u} m & \textit{dans } \mathcal{T}_{\frac{p}{q}} \\ i \xrightarrow{u \mid v} j & \textit{dans } \mathcal{D}_{\frac{p}{q}} \end{array}\right\} \;\Longrightarrow\; (n+i+1) \xrightarrow{v} (m+j+1) \quad \textit{dans } \mathcal{T}_{\frac{p}{q}} \;.$$

Démonstration. On envisage d'abord le cas où $u$ est réduit à une lettre $a \in [\![q]\!]$, et donc $v$ à une lettre $b \in [\![q]\!]$. Les deux hypothèses sont alors et impliquent

(l'équation (4.6), page 98 et lemme 6.17) :

$$n \xrightarrow{a} m \text{ dans } \mathcal{T}_{\frac{p}{q}} \quad \text{ce qui implique que} \quad \begin{cases} p\,n - a \equiv 0 \quad [q] \\ m \;=\; \frac{p\,n+a}{q} \end{cases}$$

$$i \xrightarrow{a \mid b} j \text{ dans } \mathcal{D}_{\frac{p}{q}} \quad \text{ce qui implique que} \quad \begin{cases} b \;=\; (a - p\,(i+1)) \,\%\, q \\ j \;=\; \left\lceil \frac{p\,(i+1)-a}{q} - 1 \right\rceil \end{cases}$$

Il s'ensuit que

$$p\,(n+i+1) - b \;=\; p\,(n+i+1) - (a - p\,(i+1)) \,\%\, q \;\equiv\; p\,n - a \;\equiv\; 0 \quad [q]\ ,$$

ou autrement dit que $b$ est une étiquette sortante de $(n+i+1)$ dans $\mathcal{T}_{\frac{p}{q}}$. Le successeur de $(n+i+1)$ par cette lettre est (équation (4.6))

$$\begin{aligned} \frac{p\,(n+i+1) + (a - p\,(i+1)) \,\%\, q}{q} &= \frac{p\,n}{q} + \frac{p\,(i+1)}{q} + \frac{(a - p\,(i+1)) \,\%\, q}{q} \\ &= m + \frac{p\,(i+1) - a}{q} + \frac{(a - p\,(i+1)) \,\%\, q}{q} \\ &= m + \left\lceil \frac{p\,(i+1) - a}{q} \right\rceil \\ &= m + j + 1 \end{aligned}$$

Le cas général s'obtient par induction sur la longueur de $u$ ; si $u = \varepsilon$ alors $v = \varepsilon$ et le résultat est immédiat. Soient deux mots $u\,a \in [\![q]\!]^+$ et $v\,b \in [\![q]\!]^+$ tels que

$$n \xrightarrow{u} l \xrightarrow{a} m \text{ dans } \mathcal{T}_{\frac{p}{q}} \quad \text{et} \quad i \xrightarrow{u \mid v} k \xrightarrow{a \mid b} j \text{ dans } \mathcal{D}_{\frac{p}{q}}\ .$$

L'hypothèse de récurrence d'une part et le cas traité dans le paragraphe précédent d'autre part impliquent les existences respectives du chemin et de la transition suivants :

$$(n+i+1) \xrightarrow{v} (l+k+1) \xrightarrow{b} (m+j+1)\ . \qquad \square$$

Le résultat suivant transpose le résultat précédent aux mots infinis.

**Proposition 6.19** – *Pour tous entiers $n$ et $i$, il existe un unique mot $w \in [\![q]\!]^\omega$ tel que $i \xrightarrow[\mathcal{D}_{\frac{p}{q}}]{w_n^- \mid w'} \cdots$ de $\mathcal{D}_{\frac{p}{q}}$ ; ce mot est égal à $w' = w^-_{(n+i+1)}$.*

Démonstration. L'existence et l'unicité de la branche découlent du corollaire 6.15. Soit $u$ un préfixe fini de $w_n^-$ et $v$ le préfixe fini de $w'$ de même longueur : $|u| = |v|$. On note $j$ et $m$ les états tels que

$$i \xrightarrow[\mathcal{D}_{\frac{p}{q}}]{u \mid v} j \quad \text{et} \quad n \xrightarrow[\mathcal{T}_{\frac{p}{q}}]{u} m\ .$$

La proposition 6.18, implique que $(n+i+1) \xrightarrow[\mathcal{T}_{\frac{p}{q}}]{v} (m+j+1)$. Puisque $v$ est sur l'alphabet minimal $[\![q]\!]$, $v$ est donc un préfixe de $w^-_{(n+i+1)}$.

Tout préfixe de $w'$ est un préfixe de $w^-_{(n+i+1)}$ donc $w' = w^-_{(n+i+1)}$. $\square$

Le théorème III est le cas particulier de la proposition précédente appliquée à $i = 0$. Celle-ci montre de plus que changer l'état initial de $\mathcal{D}_{\frac{p}{q}}$ en $(i-1)$ donne un transducteur qui réalise la fonction $w_n^- \mapsto w_{n+i}^-$, pour tout entier $n$ (ou plus précisément la fonction totale qui étend celle-ci par continuité).

# Envergures

Cette section utilise les notations et les résultats de la section 4.4 page 106. En particulier, la fonction d'évaluation après la virgule $\rho_{\frac{p}{q}}$ est définie par :

$$\begin{array}{rccc} \rho_{\frac{p}{q}} : & [\![p]\!]^\omega & \longrightarrow & \mathbb{R} \\ & a_0\, a_1\, a_2 \cdots a_k \cdots & \longmapsto & \displaystyle\sum_{i \geqslant 0} \frac{a_i}{q} \left(\frac{p}{q}\right)^{-(i+1)} \end{array}$$

On rappelle que cette fonction est continue (lemme 4.30), qu'elle préserve l'ordre de $(W_{\frac{p}{q}}, \leqslant_{\mathrm{rad}})$ dans $(\mathbb{R}, \leqslant)$ (lemme 4.31) et que l'image de $W_{\frac{p}{q}}$ par $\rho_{\frac{p}{q}}$ est l'intervalle (théorème 4.36)

$$\rho_{\frac{p}{q}}(W_{\frac{p}{q}}) \;=\; \left[0,\ \rho_{\frac{p}{q}}(w_0^+)\right] \ .$$

On note $V_n \subseteq [\![q]\!]^\omega$ l'ensemble des mots infinis qui étiquettent les branches infinies de $\mathcal{T}_{\frac{p}{q}}$ partant de $n$ ; ou plus formellement

$$\forall n \in \mathbb{N} \qquad V_n \;=\; \{\, w \in [\![p]\!]^\omega \mid \forall u \in [\![q]\!]^\omega, \text{ préfixe de } w,\ (n \cdot u) \text{ existe} \,\} \ . \tag{6.5}$$

En particulier, $V_0 = W_{\frac{p}{q}}$. Une démonstration analogue à celle du théorème 4.36 (qui énonce à la page 109 que $\rho_{\frac{p}{q}}(W_{\frac{p}{q}})$ est un intervalle) permet de montrer le lemme suivant.

**Lemme 6.20** – *Pour tout entier $n$, l'ensemble $\rho_{\frac{p}{q}}(V_n)$ est l'intervalle*

$$\rho_{\frac{p}{q}}(V_n) \;=\; \left[\rho_{\frac{p}{q}}(w_n^-),\ \rho_{\frac{p}{q}}(w_n^+)\right] \ .$$

On appelle *envergure de $n$*, noté $\mathsf{span}(n)$, la longueur de l'intervalle $\rho_{\frac{p}{q}}(V_n)$, c'est-à-dire

$$\mathsf{span}(n) \;=\; \rho_{\frac{p}{q}}(w_n^+) - \rho_{\frac{p}{q}}(w_n^-) \ .$$

On note $\mathsf{S}_{\frac{p}{q}} \subseteq \mathbb{R}$ l'ensemble (dénombrable) $\mathsf{S}_{\frac{p}{q}} = \{\mathsf{span}(n) \mid n \in \mathbb{N}\}$. Le but de cette section 6.2 est de montrer que les propriétés topologiques de $\mathsf{S}_{\frac{p}{q}}$ dépendent du signe de $(p - 2\,q - 1)$, comme énoncé ci-dessous.

**Théorème IV** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q > 1$.*

**a)** *Si $p \leqslant (2\,q-1)$, alors l'adhérence de $\mathsf{S}_{\frac{p}{q}}$ est un intervalle.*

**b)** *Si $p > (2\,q-1)$, alors l'adhérence de $\mathsf{S}_{\frac{p}{q}}$ est un ensemble de Cantor.*

Le point (**a**) est démontré dans la sous-section 6.2.2 et le (**b**) dans la sous-section 6.2.3. Ces deux démonstrations reposent sur le lien entre $\mathsf{S}_{\frac{p}{q}}$ et l'automate $\mathcal{T}'_{\frac{p}{q}}$, établi dans une première sous-section 6.2.1.

**Remarque 6.21** – *La figure 6 est une représentation fractale de $W_{\frac{3}{2}}$ ; chaque mot $u$ qui est un préfixe d'un mot de $W_{\frac{3}{2}}$ atteint sur cette représentation un nœud dont l'ordonnée vaut $\rho_{\frac{3}{2}}(u\,0^\omega)$ et dont l'étiquette est $n$ si le calcul de $u$ dans $\mathcal{T}_{\frac{3}{2}}$ atteint l'état $n$. En particulier, chaque entier étiquette une infinité de nœuds et les intervalles respectifs couverts par les mots partants de deux nœuds (distincts) avec la même étiquette ne sont pas de la même longueur.*

*Au contraire de la fonction* span *qui est une fonction de* $\mathbb{N}$ *dans* $\mathbb{R}$*, on peut trouver naturel de définir l'envergure autrement : il s'agirait de la fonction, noté* span*, des nœuds de cette fractale dans* $\mathbb{R}$ *qui associe à un nœud la longueur de l'intervalle couvert par les mots partant de ce nœud.*

*Pour tout nœud $x$,* span$(x)$ *est donc la taille de l'intervalle des valeurs des mots qui* ***passent par*** *le nœud $x$ (et partent de la racine), alors que* span$(n)$ *est celle des mots qui* ***partent de*** *l'état $n$. Si $x$ est un nœud de profondeur $k$ étiqueté par $n$ alors* span$(x)\left(\frac{p}{q}\right)^k =$ span$(n)$.

*Le* span *est donc* normalisé *en fonction de la profondeur du nœud. Cela rend l'étude de cette fonction beaucoup plus intéressante. En particulier, cela évite d'avoir une fonction qui tend évidemment vers 0 quand la profondeur tend vers l'infini.*

## Mots-témoins et adhérence des envergures

Le but de cette sous-section est de démontrer le théorème suivant, qui donne une relation entre l'automate $\mathcal{T}'_{\frac{p}{q}}$ et l'ensemble $\mathsf{S}_{\frac{p}{q}}$ des envergures. On rappelle que si $\mathcal{A}$ est un automate (fini ou infini), on note $\mathfrak{L}(\mathcal{A})$ l'ensemble des mots infinis acceptés par $\mathcal{A}$.

**Théorème 6.22** – *Le langage de mots infinis accepté par $\mathcal{T}'_{\frac{p}{q}}$ a pour évaluation après la virgule l'adhérence des envergures :*

$$\rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})) \;=\; \mathsf{adh}(S_{\frac{p}{q}})\ .$$

Soit $a$ une lettre de l'alphabet minimal $[\![q]\!]$, et $b$ une lettre de l'alphabet maximal $\{(p-q),\cdots,(p-1)\}$. L'entier $(b-a)$ appartient nécessairement à $\{p-(2\,q-1),\cdots,(p-1)\}$, c'est-à-dire est une lettre de $B_{\frac{p}{q}}$.

Ainsi, la soustraction d'un mot minimal à un mot maximal effectuée chiffre-à-chiffre donne un mot (infini) sur l'alphabet $B_{\frac{p}{q}}$. On note cette opération par $\ominus$, et pour chaque entier $n$, le mot $(w_n^+ \ominus w_n^-)$ est appelé *le mot-témoin* (de l'envergure) de $n$. Cette appellation provient de la formule suivante qu'une simple vérification permet d'établir :

$$\forall n \in \mathbb{N} \qquad \mathsf{span}(n) \;=\; \rho_{\frac{p}{q}}(w_n^+ \ominus w_n^-)\ . \tag{6.6}$$

On note $\mu$ le morphisme de mots lettre-à-lettre des mots sur l'alphabet minimal $[\![q]\!]$ dans les mots sur l'alphabet maximal $\{(p-q),(p-q+1),\ldots,(p-1)\}$ qui est défini par son action sur chaque lettre :

$$\forall a \in [\![q]\!] \qquad \mu(a) \;=\; (a+(p-q))$$

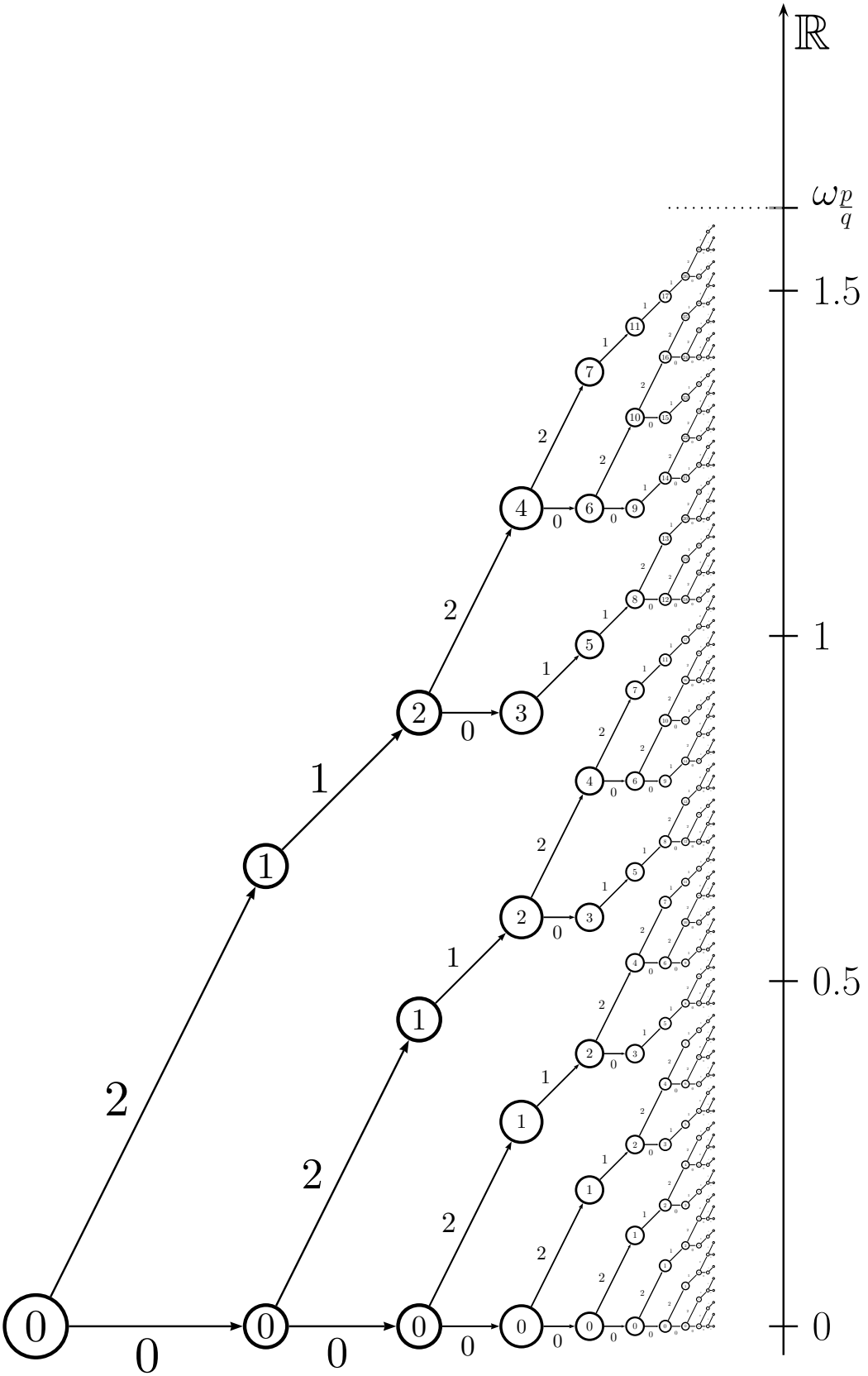



Figure 6 – Représentation fractale des évaluations après la virgule de $W_{\frac{p}{q}}$

Le lemme 4.34, page 108, se réecrit donc en :

$$\forall n \in \mathbb{N} \qquad \mu(w^-_{(n+1)}) \;=\; w^+_n \;. \tag{6.7}$$

L'existence de la fonction $\mu$ indique une certaine proximité entre les transformations $\xi : w^-_n \mapsto w^-_{(n+1)}$ et $w^-_n \mapsto w^+_n$. Cette idée est matérialisée par la proposition suivante, qui relie $\mathcal{T}'_{\frac{p}{q}}$ (l'étape intermédiaire de la construction du transducteur $\mathcal{D}_{\frac{p}{q}}$ qui réalise $\xi$) et les mots témoins.

**Proposition 6.23** – *Un mot fini est accepté par $\mathcal{T}'_{\frac{p}{q}}$ si et seulement s'il est le préfixe d'un mot-témoin.*

Démonstration. Le transducteur $\mathcal{D}_{\frac{p}{q}}$ réalise la fonction $\xi$ (théorème III) qui est l'extension par continuité de la fonction $w^-_n \mapsto w^-_{(n+1)}$, pour tout entier $n$. Il s'ensuit que pour tous mots finis $u, v \in [\![q]\!]^*$, $u \mid v$ est accepté par $\mathcal{D}_{\frac{p}{q}}$ si et seulement si $u$ et $v$ sont les préfixes respectifs de $w^-_n$ et $w^-_{n+1}$, pour un certain entier $n$.

D'autre part, un mot témoin est un mot de la forme $w^+_n \ominus w^-_n$ pour un certain entier $n$, ce qui implique que

$$\forall n \in \mathbb{N} \qquad w^+_n \ominus w^-_n \;=\; \mu(w^-_{(n+1)}) \ominus w^-_n \;. \tag{$*$}$$

Il découle donc du paragraphe précédent qu'un mot fini $x \in B^*_{\frac{p}{q}}$ est le préfixe d'un mot témoin si et seulement si il est de la forme $x = \mu(u) \ominus v$ pour certains $u, v \in [\![q]\!]^*$ tels que $u \mid v$ est accepté par $\mathcal{D}_{\frac{p}{q}}$.

Il reste maintenant à démontrer que pour tous mots finis $u, v \in [\![q]\!]^*$ de même longueur $k$, le mot $\mu(u) \ominus v$ est accepté par $\mathcal{T}_{\frac{p}{q}}$ si et seulement si $u \mid v$ est accepté par $\mathcal{D}_{\frac{p}{q}}$. Ceci se montre par récurrence sur $k$ ; le cas où $u = v = \varepsilon$ est immédiat. Le cas général requiert l'assertion suivante (qui donne un résultat plus fort dans le cas où $u$ et $v$ sont réduits à des lettres).

**Assertion 6.23.1** – *Pour toutes lettres $a, b \in [\![q]\!]$ et tous entiers $n, m$, les deux énoncés suivants sont équivalents :*

**a)** $n \xrightarrow{a \mid b} m$ *est une transition de* $\mathcal{D}_{\frac{p}{q}}$

**b)** $n \xrightarrow{\mu(b)-a} m$ *est une transition de* $\mathcal{T}'_{\frac{p}{q}}$

*Démonstration de l'assertion.* D'après la définition de $\mathcal{D}_{\frac{p}{q}}$ (équation (6.4)), l'énoncé (**a**) est équivalent à

$$\begin{aligned} q\,m &= p\,n + (b - a) + (p - q) \\ &= p\,n + (\mu(b) - a) \;. \end{aligned} \tag{$*$}$$

Or, $(\mu(b) - a)$ est le résultat de la soustraction d'une lettre minimale à une lettre maximale, donc est une lettre de $B_{\frac{p}{q}}$. L'équation $(*)$ est donc équivalente au fait que la transition $n \xrightarrow{\mu(b)-a} m$ existe dans $\mathcal{T}'_{\frac{p}{q}}$ (d'après l'équation (6.1)), c'est-à-dire la condition (**b**).

Soient deux mots $u = u'\,a \in [\![q]\!]^+$ et $v = v'\,b \in [\![q]\!]^+$ non-vides. Le mot $\mu(u) \ominus v$ est accepté par $\mathcal{T}'_{\frac{p}{q}}$ si et seulement si il existe deux entiers $n$ et $m$ tels que

$$0 \xrightarrow{\mu(u')\ominus v'} n \xrightarrow{\mu(a)\ominus b} m \quad \text{dans } \mathcal{T}'_{\frac{p}{q}} \;.$$

D'une part, ce chemin (à gauche) est équivalent au chemin (à gauche) suivant par hypothèse de récurrence, et d'autre part cette transition (à droite) est équivalente à la transition suivante (à droite) d'après l'assertion 6.23.1 :

$$0 \xrightarrow{u' \mid v'} n \xrightarrow{a \mid b} m \quad \text{dans } \mathcal{D}_{\frac{p}{q}} \ ,$$

ce qui conclut la récurrence. □

**Corollaire 6.24** – *Le langage de mots infinis accepté par $\mathcal{T}'_{\frac{p}{q}}$ est la clôture topologique de l'ensemble des mots-témoins :*

$$\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}}) \ = \ \mathsf{adh}\left(\left\{ w_n^+ \ominus w_n^- \ \middle|\ n \in \mathbb{N} \right\}\right) \ .$$

Puisque la fonction $\rho_{\frac{p}{q}}$ d'évaluation après la virgule est continue (lemme 4.30 page 106), le théorème 6.22 découle du corollaire 6.24.

## Adhérence des envergures dans une petite base

Le but de cette sous-section est d'établir le point (**a**) du théorème IV. Nous commençons par démontrer un lemme préliminaire puis la proposition 6.26 dont le théorème est un corollaire.

**Lemme 6.25** – *On suppose que $p \leqslant (2q-1)$. Si un mot $u \in B_{\frac{p}{q}}^*$ est accepté par $\mathcal{T}'_{\frac{p}{q}}$, alors $|u| > \left|\langle m \rangle_{\frac{p}{q}}\right|$, où $m = \pi_{\frac{p}{q}}(u)$.*

Démonstration. Dans le cas où $p = (2q-1)$, $B_{\frac{p}{q}}$ et $[\![p]\!]$ sont égaux donc $\mathcal{T}_{\frac{p}{q}}$ et $\mathcal{T}'_{\frac{p}{q}}$ sont égaux, si bien que le mot $u$ est accepté par $\mathcal{T}_{\frac{p}{q}}$ donc appartient à $0^* L_{\frac{p}{q}}$ et satisfait donc la condition requise. On suppose dans la suite que $p < (2q-1)$, ce qui implique en particulier que $[\![p]\!] \subsetneq B_{\frac{p}{q}}$.

Par récurrence ; le cas $u = \varepsilon$ est évidemment vérifié. Soit un mot $u = u'a \in B_{\frac{p}{q}}^+$ non-vide dont on note le calcul dans $\mathcal{T}'_{\frac{p}{q}}$ par :

$$0 \xrightarrow[\mathcal{T}'_{\frac{p}{q}}]{u'} n \xrightarrow[\mathcal{T}'_{\frac{p}{q}}]{a} m \ .$$

Il découle donc du lemme 6.8 que $\pi_{\frac{p}{q}}(u') = n$ et $\pi_{\frac{p}{q}}(u) = m$ et donc, par hypothèse de récurrence que

$$|u'| \ \geqslant \ |\langle n \rangle_{\frac{p}{q}}| \ . \qquad (*)$$

Si $a$ appartient à $[\![p]\!]$ alors $\langle m \rangle_{\frac{p}{q}} = \langle n \rangle_{\frac{p}{q}} a$ ce qui implique d'après l'équation précédente que

$$|u| \ = \ |u'a| \ \geqslant \ \left|\langle n \rangle_{\frac{p}{q}} a\right| \ = \ \left|\langle m \rangle_{\frac{p}{q}}\right|$$

On suppose dans la suite que $a \in (B_{\frac{p}{q}} \setminus [\![p]\!])$, on note $b \in [\![p]\!]$ l'unique lettre et $n'$ l'unique entier tel que

$$\langle m \rangle_{\frac{p}{q}} \ = \ \langle n' \rangle_{\frac{p}{q}} b \ . \qquad (**)$$

Il s'ensuit que $n' \xrightarrow{b} m$ est une transition de $\mathcal{T}_{\frac{p}{q}}$ et donc de $\mathcal{T}'_{\frac{p}{q}}$ (puisque nous sommes dans le cas $p \leqslant (2q-1)$). D'autre part, $[\![p]\!]$ est formé des $p$ plus grandes

lettres de $B_{\frac{p}{q}}$, toute lettre de $[\![p]\!]$ est donc strictement supérieure à toute lettre de $(B_{\frac{p}{q}} \setminus [\![p]\!])$ ; en particulier, $b > a$. Puisque de plus,

$$n' \xrightarrow{b} m \quad \text{et} \quad n \xrightarrow{a} m$$

sont deux transitions de $\mathcal{T}'_{\frac{p}{q}}$, il découle de l'équation (6.1) que $n' < n$ et donc que

$$|\langle n' \rangle_{\frac{p}{q}}| \leqslant |\langle n \rangle_{\frac{p}{q}}| \,. \tag{$\overset{*}{**}$}$$

Combiner l'équation $(**)$ avec les inéquations $(\overset{*}{**})$ et $(**)$ conclut la récurrence :

$$\left|\langle m \rangle_{\frac{p}{q}}\right| = \left|\langle n' \rangle_{\frac{p}{q}}\right| + 1 \leqslant \left|\langle n \rangle_{\frac{p}{q}}\right| + 1 \leqslant |u'| + 1 = |u| \,. \qquad \square$$

**Proposition 6.26** – *Si $p \leqslant (2q-1)$, alors les langages de mots infinis acceptés par $\mathcal{T}_{\frac{p}{q}}$ et $\mathcal{T}'_{\frac{p}{q}}$ ont des évaluations égales :*

$$\rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})) = \rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}_{\frac{p}{q}})) \,.$$

Démonstration. Puisque $p \leqslant (2q-1)$, $[\![p]\!] \subseteq B_{\frac{p}{q}}$ donc toutes les transitions de $\mathcal{T}_{\frac{p}{q}}$ sont des transitions de $\mathcal{T}'_{\frac{p}{q}}$ ; si bien que tout mot accepté par $\mathcal{T}_{\frac{p}{q}}$ est accepté par $\mathcal{T}'_{\frac{p}{q}}$ donc

$$\rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}_{\frac{p}{q}})) \subseteq \rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})) \,.$$

On note la branche maximale de $\mathcal{T}_{\frac{p}{q}}$ partant de l'état initial par

$$0 = n_0 \xrightarrow{a_0} n_1 \xrightarrow{a_1} n_2 \xrightarrow{a_2} \cdots \xrightarrow{a_{k-1}} n_k \xrightarrow{a_k} \cdots \,,$$

ce qui implique que $w_0^+ = a_0\, a_1\, a_2 \cdots a_k \cdots$ et que

$$\forall k, m \in \mathbb{N} \qquad |\langle m \rangle_{\frac{p}{q}}| \leqslant k \implies m < n_k \,. \tag{$*$}$$

Soit un mot infini $w = (b_0\, b_1\, b_2 \cdots b_k \cdots) \in B_{\frac{p}{q}}^{\omega}$ accepté par $\mathcal{T}'_{\frac{p}{q}}$ dont on note le calcul par :

$$0 = m_0 \xrightarrow{b_0} m_1 \xrightarrow{b_1} m_2 \xrightarrow{b_2} \cdots \xrightarrow{b_{k-1}} m_k \xrightarrow{b_k} \cdots \,,$$

ce qui implique que (lemme 6.8)

$$\forall k \in \mathbb{N} \qquad \pi_{\frac{p}{q}}(b_0\, b_1\, b_2 \cdots b_{k-1}) = m_k \geqslant 0 \,. \tag{$**$}$$

Pour tout entier $k$, il découle lemme 6.25 que

$$k = |b_0\, b_1\, b_2 \cdots b_{k-1}| \geqslant |\langle m_k \rangle_{\frac{p}{q}}| \qquad \Big(\text{car } m_k = \pi_{\frac{p}{q}}(b_0\, b_1\, b_2 \cdots b_{k-1})\Big) \,.$$

Donc $\langle m_k \rangle_{\frac{p}{q}} \leqslant k$, ce qui implique d'après $(*)$ que

$$\forall k \in \mathbb{N} \qquad \pi_{\frac{p}{q}}(b_0\, b_1\, b_2 \cdots b_{k-1}) = m_k \leqslant n_k = \pi_{\frac{p}{q}}(a_0\, a_1\, a_2 \cdots a_{k-1}) \,. \tag{$\overset{*}{**}$}$$

Or pour tout mot (fini) $u$,

$$\rho_{\frac{p}{q}}(u\, 0^{\omega}) = \pi_{\frac{p}{q}}(u) \left(\frac{p}{q}\right)^{-|u|} \,,$$

les inéquations $(**)$ et $(\overset{*}{**})$ impliquent donc que

$$\forall k \in \mathbb{N} \qquad 0 \leqslant \rho_{\frac{p}{q}}(b_0\, b_1\, b_2 \cdots b_{k-1}\, 0^{\omega}) \leqslant \rho_{\frac{p}{q}}(a_0\, a_1\, a_2 \cdots a_{k-1}\, 0^{\omega}) \,.$$

donc, puisque $\rho_{\frac{p}{q}}$ est une fonction continue, que

$$0 \leqslant \rho_{\frac{p}{q}}(w) \leqslant \rho_{\frac{p}{q}}(w_0^+) .$$

La valeur du mot $w$ est donc dans

$$\rho_{\frac{p}{q}}(w) \in [0, \rho_{\frac{p}{q}}(w_0^+)] = \rho_{\frac{p}{q}}(W_{\frac{p}{q}}) = \rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}_{\frac{p}{q}})) .$$

(La première égalité découle du théorème 4.36 page 109 et la seconde de l'équation (4.11) page 106 définissant l'ensemble $W_{\frac{p}{q}}$.)

En d'autres termes,

$$\rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})) \subseteq \rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}_{\frac{p}{q}})) . \qquad \square$$

**Théorème IVa** – *Si $p \leqslant (2\,q-1)$, alors l'adhérence de $\mathsf{S}_{\frac{p}{q}}$ est un intervalle.*

Démonstration. Il découle du théorème 6.22 que

$$\mathsf{adh}(\mathsf{S}_{\frac{p}{q}}) = \rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})) ,$$

de la proposition 6.26 précédente que

$$\rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})) = \rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}_{\frac{p}{q}})) ,$$

de la définition de $W_{\frac{p}{q}}$ (équation (4.11) page 106) que

$$\rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}_{\frac{p}{q}})) = \rho_{\frac{p}{q}}(W_{\frac{p}{q}})$$

et enfin du théorème 4.36 (page 109) que

$$\rho_{\frac{p}{q}}(W_{\frac{p}{q}}) = \left[0,\ \rho_{\frac{p}{q}}(w_0^+)\right] . \qquad \square$$

## Adhérence des envergures dans une grande base

**Définition 6.27** – *On appelle* ensemble de Cantor *tout ensemble*

**a)** *fermé,*
**b)** *borné,*
**c)** *nulle part dense,*
**d)** *et qui ne possède aucun point isolé.*

**Exemple 6.28** (Ensemble triadique de Cantor [18]) – *L'ensemble triadique de Cantor, noté $\mathcal{K}_3$, est un sous-ensemble de $[0,1]$ défini récursivement par soustraction d'intervalles (ouverts). On définit $\chi$ comme l'opération qui retire le tiers central (ouvert) d'un intervalle fermé donné ; plus formellement $\chi$ associe à tout intervalle fermé $[x,y]$ l'union d'intervalle fermé :*

$$\chi([x,y]) = \left[x,\ x+\frac{y-x}{3}\right] \bigcup \left[x+\frac{2\,(y-x)}{3},\ y\right] .$$

[18] Georg Cantor, 1884, *De la puissance des ensembles parfaits de points.*.

*Par exemple* $\chi([0,1]) = [0,\frac{1}{3}] \cup [\frac{2}{3},1]$. *On note* $S_0 = [0,1]$ *; et pour tout* $i \in \mathbb{N}$ *si* $S_i$ *est une union finie d'intervalles fermée* $S_i = \bigcup_{i\in I} F_i$ *alors* $S_{(i+1)} = \bigcup_{i\in I} \chi(F_i)$. *L'ensemble triadique de Cantor est alors l'intersection de tous les* $S_i$ *:*

$$\mathcal{K}_3 \;=\; \bigcap_{i\in\mathbb{N}} S_i \;,$$

*c'est-à-dire la limite de ce processus, représenté à la figure 7.*


Figure 7 – L'ensemble triadique de Cantor

*Une définition alternative peut être donnée à partir de l'évaluation (après la virgule) en base entière 3 définie par :*

$$\forall a_0\, a_1 \cdots a_k \cdots \in [\![3]\!]^\omega \qquad \rho_3(a_0\, a_1 \cdots a_k \cdots) \;=\; \sum_{i=0}^{\infty} a_i\, 3^{-(i+1)} \;.$$

*L'ensemble triadique de Cantor est l'ensemble des évaluations des mots infinis dans lesquelles le chiffre '1' n'apparaît pas :* $\mathcal{K}_3 = \{\, \rho_3(w) \mid w \in \{0,2\}^\omega \,\}$.

Dans le cas $p > (2\,q-1)$ considéré dans cette sous-section 6.2.3, les mots infinis acceptés par $\mathcal{T}'_{\frac{p}{q}}$ sont les mots accepté par $\mathcal{T}_{\frac{p}{q}}$ qui ne contiennent pas certaines lettres (à savoir les lettres de $([\![p]\!] \setminus B_{\frac{p}{q}}) = \{0,1,\cdots(p-(2\,q))\}$). L'ensemble $\rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}}))$ peut être décrit de façon similaire à la définition alternative de $\mathcal{K}_3$. Le but de cette section est de démontrer le théorème IVb, rappelé ci-dessous.

**Théorème IVb** – *Si* $p > (2q-1)$*, alors l'adhérence de* $\mathsf{S}_{\frac{p}{q}}$ *est un espace de Cantor.*

La démonstration est découpée de la façon suivante. La proposition 6.29 montre que $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$ est fermé et borné, la proposition 6.32 qu'il est d'intérieur vide et la proposition 6.34 qu'il ne contient aucun point isolé.

**Proposition 6.29** – *L'ensemble* $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$ *est fermé et borné.*

Démonstration. Il est fermé, car égal à l'adhérence d'un autre ensemble, et est borné car égal à $\rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}}))$ (théorème 6.22) lui-même inclus dans $\rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}_{\frac{p}{q}})) = [0,\omega_{\frac{p}{q}}]$ (théorème 4.36). □

Les démonstrations des autres propositions demandent plus de travail. Nous donnons d'abord deux propriétés valables dans les grandes bases.

**PROPRIÉTÉ 6.30** – *On suppose $p > (2q-1)$.*

**a)** *Tous les chiffres de $B_{\frac{p}{q}}$ sont strictement positifs.*

**b)** *Pour tout entier $n$, $\mathsf{span}(n) > 0$.*

DÉMONSTRATION. **a)** Par définition, $B_{\frac{p}{q}}$ est l'intervalle entier de cardinal $(2q-1)$ dont le plus grand élément est $p$. Sous l'hypothèse $p > (2q-1)$, le plus petit élément de $B_{\frac{p}{q}}$, $(p-2q-1)$, est donc strictement supérieur à 0.

**b)** Tous les mots témoins sont acceptés par $\mathcal{T}'_{\frac{p}{q}}$ (découle du corollaire 6.24) donc sont des mots de $B^{\omega}_{\frac{p}{q}}$ ; cet alphabet ne contient que des lettres strictement positives d'après le point (**a**) donc, pour tout entier $n$, la valeur $\rho_{\frac{p}{q}}(w_n^+ \ominus w_n^-)$ du mot-témoin de $n$ est strictement supérieure à 0.

Or, l'envergure d'un entier $n$ est égale à la valeur après la virgule du mot-témoin de $n$ (équation (6.6), rappelée ci-dessous)

$$\forall n \in \mathbb{N} \qquad \mathsf{span}(n) \;=\; \rho_{\frac{p}{q}}(w_n^+ \ominus w_n^-) \;,$$

donc, pour tout entier $n$, $\mathsf{span}(n) > 0$. □

Le lemme suivant donne l'idée de la démonstration de la proposition 6.32 qui le suit : *étant donné un mot $w \in \mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})$ on peut trouver un mot $w' \in \mathfrak{L}(\mathcal{T}_{\frac{p}{q}})$ arbitrairement proche (au sens topologique) de $w$, mais dans lequel apparaît la lettre 0 qui n'est pas dans $B_{\frac{p}{q}}$.*

**LEMME 6.31** – *On suppose que $p > (2q-1)$. A partir de chaque état $n$, il existe un état $m$ qui est accessible dans $\mathcal{T}_{\frac{p}{q}}$ mais pas dans $\mathcal{T}'_{\frac{p}{q}}$.*

DÉMONSTRATION. Soit $n$ un entier. On note $S_i$, pour tout entier $i$, l'ensemble des successeurs d'ordre $i$ de $n$ :

$$\forall i \in \mathbb{N} \qquad S_i \;=\; \left\{ m \;\middle|\; \exists u \in [\![p]\!]^* \text{ tel que } n \xrightarrow{u} m \text{ existe dans } \mathcal{T}_{\frac{p}{q}} \text{ et } |u| = i \right\} .$$

Puisque $p > (2q-1)$, chaque état admet, dans $\mathcal{T}_{\frac{p}{q}}$, au moins deux transitions sortantes donc, pour tout entier $i$, $\mathsf{Card}(S_i) \geqslant 2^i$ ; il existe donc un entier $j$ tel que $\mathsf{Card}(S_j) \geqslant p$. L'ensemble des successeurs d'ordre $j$ de $n$ forment un intervalle entiers donc il existe dans $S_j$ un entier $m$ divisible par $p$ (attention, par $p$ et non par $q$)

L'état $m$ est donc accessible dans $\mathcal{T}_{\frac{p}{q}}$ par une (unique) transition (qui est) étiquetée par 0. L'alphabet $B_{\frac{p}{q}}$ ne contient pas la lettre 0 (propriété 6.30a) donc la transition entrante de $m$ qui existait dans $\mathcal{T}_{\frac{p}{q}}$ n'existe pas dans $\mathcal{T}'_{\frac{p}{q}}$ donc $m$ n'est pas accessible dans $\mathcal{T}'_{\frac{p}{q}}$. □

**PROPOSITION 6.32** – *L'ensemble $\mathsf{adh}(S_{\frac{p}{q}})$ est d'intérieur vide.*

Avant de démontrer cette proposition, nous donnons quelques notations supplémentaires. Pour tout mot fini $u \in [\![p]\!]^*$, on note $Z_u \subseteq [\![p]\!]^{\omega}$ l'ensemble des mots infinis acceptés par $\mathcal{T}_{\frac{p}{q}}$ qui ont comme préfixe $u$ :

$$Z_u \;=\; u\left(u^{-1}\,\mathfrak{L}(\mathcal{T}_{\frac{p}{q}})\right)$$

Il s'agit d'une variante "dénormalisée" de l'ensemble $V_n$ (défini par l'équation (6.5), page 155) ; en effet l'équation suivante est vérifiée :

$$Z_u \;=\; u\,V_n \quad \text{où} \quad n = \pi_{\frac{p}{q}}(u)\ .$$

Il découle alors du lemme 6.20 que

$$\forall u \in L(\mathcal{T}_{\frac{p}{q}}) \qquad \rho_{\frac{p}{q}}(Z_u) \;=\; \Big[\rho_{\frac{p}{q}}\big(u\,w_n^-\big)\,,\ \rho_{\frac{p}{q}}\big(u\,w_n^+\big)\Big] \quad \text{où} \quad n = \pi_{\frac{p}{q}}(u)\ . \tag{6.8}$$

Démonstration de la proposition 6.32. Par l'absurde. Supposons que $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$ n'est pas d'intérieur vide, il contient donc un certain intervalle $]x, y[$, et, puisqu'il est fermé, il contient même $[x, y]$. On peut alors démontrer qu'il est dense dans un certain $\rho_{\frac{p}{q}}(Z_u)$ comme exprimé ci-dessous.

**Assertion 6.32.1** – *Si $\rho_{\frac{p}{q}}(\mathsf{S}_{\frac{p}{q}})$ contient l'intervalle $[x, y]$, alors il existe un mot $u \in B_{\frac{p}{q}}{}^*$ tel que $\mathsf{S}_{\frac{p}{q}}$ contient $\rho_{\frac{p}{q}}(Z_u)$.*

*Démonstration de l'assertion.* On note $d = \frac{y-x}{2}$ la demi longueur de $[x, y]$ et $z := x + d$ le milieu de $[x, y]$. Puisque la série $\sum_{i=0}^{\infty} \left(\frac{p}{q}\right)^{-i}$ converge, pour tout réel $\varepsilon > 0$, il existe un entier $k$ tel que

$$\forall u \in [\![p]\!]^*,\ \forall w, w' \in [\![p]\!]^\omega \qquad |u| \geqslant k \quad \Longrightarrow \quad \mathsf{abs}\Big(\rho_{\frac{p}{q}}(uw) - \rho_{\frac{p}{q}}(uw')\Big) \;<\; \varepsilon\ . \quad (*)$$

On pose $\varepsilon = \frac{d}{3}$ et on note $k$ l'entier correspondant.

D'autre part, puisque $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$ contient $[x, y]$, il existe un élément de $\mathsf{S}_{\frac{p}{q}}$ arbitrairement proche de $z (\in [x, y])$, autrement dit, il existe un entier $i$ tel que

$$\mathsf{abs}\,(\mathsf{span}(i) - z) \;<\; \varepsilon\ .$$

On note $u \in B_{\frac{p}{q}}{}^*$ le préfixe du mot-témoin $(w_i^+ \ominus w_i^-)$ de longueur $k$ et on pose $n = \pi_{\frac{p}{q}}(u)$. Puisque toutes les lettres de $B_{\frac{p}{q}}$ sont strictement positives (propriété 6.30a), $n > 0$.

Il découle de l'équation (6.8) que $\rho_{\frac{p}{q}}(Z_u) = [\rho_{\frac{p}{q}}(u\,w_n^-),\ \rho_{\frac{p}{q}}(u\,w_n^+)]$ et de $(*)$ que sa longueur est inférieure à $\varepsilon$. Or nous avons choisi $u$ comme un préfixe de $(w_i^+ \ominus w_i^-)$ donc celui-ci appartient à $Z_u$, ce qui implique que

$$\mathsf{span}(i) \;=\; \rho_{\frac{p}{q}}(w_i^+ \ominus w_i^-) \ \text{ appartient à } \ \rho_{\frac{p}{q}}(Z_u)\ .$$

Or, $\mathsf{span}(i)$ est à une distance au plus $\varepsilon$ de $z$ (inéquation $(*)$), donc chaque point de $\rho_{\frac{p}{q}}(Z_u)$ est à une distance au plus $2\varepsilon = 2\,\frac{d}{3} < d$ de $z$ (le milieu de $[x, y]$), donc est un point de $[x, y]$. Il s'ensuit que $\rho_{\frac{p}{q}}(Z_u)$ est inclus dans $[x, y]$ et donc dans $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$.

On note $u \in B_{\frac{p}{q}}{}^*$ le mot dont il est fait référence dans l'assertion précédente et $n = \pi_{\frac{p}{q}}(u)$, de telle sorte que $\rho_{\frac{p}{q}}(Z_u)$ est inclus dans $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$.

**Assertion 6.32.2** – *Si $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$ contient $\rho_{\frac{p}{q}}(Z_u)$, alors il existe un mot $v \in [\![p]\!]^*$ tel que*

- **a)** $Z_{uv} \cap \mathfrak{L}(\mathcal{T}'_{\frac{p}{q}}) = \emptyset$ ,
- **b)** $\rho_{\frac{p}{q}}(Z_{uv})$ *est un intervalle non-trivial et*
- **c)** $\rho_{\frac{p}{q}}(Z_{uv}) \subseteq \rho_{\frac{p}{q}}(Z_u)$.

*Démonstration de l'assertion.* D'après le lemme 6.31, il existe un état $m \in \mathbb{N}$, inaccessible depuis $n$ dans $\mathcal{T}_{\frac{p}{q}}$ mais accessible depuis $n$ dans $\mathcal{T}'_{\frac{p}{q}}$ ; il existe donc un mot $v \in [\![p]\!]^*$ tel que $n \xrightarrow{v} m$ existe dans $\mathcal{T}_{\frac{p}{q}}$. Les mots de $Z_{uv}$ sont donc tous acceptés par $\mathcal{T}_{\frac{p}{q}}$ et tous refusés par $\mathcal{T}'_{\frac{p}{q}}$ ; la condition (**a**) est donc vérifiée.

**b)** D'après l'équation (6.8),

$$\rho_{\frac{p}{q}}(Z_{uv}) \;=\; \left[\rho_{\frac{p}{q}}(u\,v\,w^-_m),\ \rho_{\frac{p}{q}}(u\,v\,w^+_m)\right] \ .$$

Or, il découle de la propriété 6.30b que $\rho_{\frac{p}{q}}(w^-_m) < \rho_{\frac{p}{q}}(w^+_m)$ ce qui implique que $\rho_{\frac{p}{q}}(u\,v\,w^-_m) < \rho_{\frac{p}{q}}(u\,v\,w^+_m)$ et donc $\rho_{\frac{p}{q}}(Z_{uv})$ est un intervalle non-trivial.

**c)** Tous les mots de $\mathfrak{L}(\mathcal{T}_{\frac{p}{q}})$ qui commencent par $u\,v$ commencent par $u$, donc $Z_{uv} \subseteq Z_u$, ce qui implique que

$$\rho_{\frac{p}{q}}(Z_{uv}) \;\subseteq\; \rho_{\frac{p}{q}}(Z_u) \ .$$

Soit un mot $w \in \mathfrak{L}(\mathcal{T}_{\frac{p}{q}})$ dont la valeur est dans l'intérieur de $\rho_{\frac{p}{q}}(Z_{uv})$, c'est-à-dire tel que

$$\rho_{\frac{p}{q}}(u\,v\,w^-_m) \;<\; \rho_{\frac{p}{q}}(w) \;<\; \rho_{\frac{p}{q}}(u\,v\,w^+_m) \ .$$

Puisque $\rho_{\frac{p}{q}}$ conserve l'ordre (lemme 4.31, page 106), il s'ensuit que

$$u\,v\,w^-_m \;<_{\text{rad}}\; w \;<_{\text{rad}}\; u\,v\,w^+_m \ ,$$

et donc que $u\,v$ est un préfixe de $w$. Si bien que $w \in Z_{uv}$ et n'appartient donc pas à $\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})$ (d'après l'assertion précédente).

Puisque $p > (2\,q-1)$, $\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}}) \subseteq \mathfrak{L}(\mathcal{T}_{\frac{p}{q}})$, il découle donc du paragraphe précédent qu'il n'existe aucun mot de $\mathcal{T}'_{\frac{p}{q}}$ dont la valeur est dans $\mathsf{int}(\rho_{\frac{p}{q}}(Z_{uv}))$. Or, il découle de l'assertion précédente que $\mathsf{int}(\rho_{\frac{p}{q}}(Z_{uv}))$ est un intervalle ouvert non-vide et inclus dans $Z_u$. Puisque $\rho_{\frac{p}{q}}(\mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})) = \mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$ (théorème 6.22), $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$ ne contient pas $Z_{uv}$ donc pas $Z_u$. Contradiction. □

La dernière étape consiste à démontrer que $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$ ne contient pas de point isolé. Une fois encore, le lemme suivant donne l'idée de la démonstration de la proposition 6.34 qui le suit : *étant donné un mot $w \in \mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})$, il existe un mot $w' \in \mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})$ arbitrairement proche (au sens topologique) de $w$, mais de valeur différente.*

**Lemme 6.33** – *On suppose que $p > (2\,q-1)$. Partant de chaque état $n \in \mathbb{N}$, il existent deux branches de $\mathcal{T}'_{\frac{p}{q}}$ dont les étiquettes $w, w' \in B^{\omega}_{\frac{p}{q}}$ ont des valeurs différentes : $\rho_{\frac{p}{q}}(w) \neq \rho_{\frac{p}{q}}(w')$.*

Démonstration. Soit $n$ un entier. L'alphabet $B_{\frac{p}{q}}$ contient l'alphabet maximal $\{(p-q), (p-q+1), \ldots, (p-1)\}$ (propriété 6.2a), donc le mot maximal de $n$ étiquette une branche partant de $n$ présente à la fois dans $\mathcal{T}_{\frac{p}{q}}$ et $\mathcal{T}'_{\frac{p}{q}}$.

Supposons que $w^+_n = (p-q)^\omega$. Puisque l'automate $\mathcal{T}_{\frac{p}{q}}$ accepte le langage $0^*L_{\frac{p}{q}}$ et que $(p-q) \neq 0$, il en découle que $L_{\frac{p}{q}}$ contient le langage régulier infini $\langle n \rangle_{\frac{p}{q}}(p-q)^*$, ce qui contredit le fait que $L_{\frac{p}{q}}$ est FLIP (proposition 4.18, page 99). Donc, $w^+_n \neq (p-q)^\omega$.

Il existe donc une lettre $a \in B_{\frac{p}{q}}$, $a \neq (p-q)$, un mot $u \in B_{\frac{p}{q}}^*$ et deux états $n', m \in \mathbb{N}$ tels que $u\,a$ est un préfixe de $w_n^+$ et que

$$n \xrightarrow{u} n' \xrightarrow{a} m \ . \qquad (*)$$

La lettre $a$ est différente de $(p-q)$ et appartient à un mot maximal, donc $a > (p-q)$, si bien que la lettre $b = (a-q) > (p-2q)$ donc appartient à $B_{\frac{p}{q}}$. L'existence de la transition droite de $(*)$ implique (équation (6.1), page 148) que

$$n'\,p + a \ = \ q\,m$$

ce qui implique que

$$n'\,p + b \ = \ n'\,p + a - q \ = \ q\,(m-1) \ ,$$

elle même impliquant l'existence dans $\mathcal{T}'_{\frac{p}{q}}$ de la transition (équation (6.1))

$$n' \xrightarrow{b} (m-1) \ .$$

Pour les mêmes raisons que précédemment, le mot maximal de $(m-1)$ étiquette une branche partant de $(m-1)$ dans $\mathcal{T}'_{\frac{p}{q}}$ donc les deux mots suivants étiquettes deux branches distinctes de $\mathcal{T}'_{\frac{p}{q}}$ partant de $n$ :

$$w_n^+ = u\,a\,w_m^+ \quad \text{et} \quad u\,b\,w_{(m-1)}^+ \ .$$

Or, il découle de la proposition 4.35 (page 109) que

$$\rho_{\frac{p}{q}}\Big(b\,w_{(m-1)}^+\Big) \ = \ \rho_{\frac{p}{q}}\big(a\,w_m^-\big) \quad \text{donc que} \quad \rho_{\frac{p}{q}}\Big(u\,b\,w_{(m-1)}^+\Big) \ = \ \rho_{\frac{p}{q}}\big(u\,a\,w_m^-\big)$$

et de la propriété 6.30b que

$$\rho_{\frac{p}{q}}\big(w_m^+\big) \ > \ \rho_{\frac{p}{q}}\big(w_m^-\big) \quad \text{donc que} \quad \rho_{\frac{p}{q}}\big(u\,a\,w_m^+\big) \ > \ \rho_{\frac{p}{q}}\big(u\,a\,w_m^-\big) \ .$$

Les mots $w = (u\,b\,w_{(m-1)}^+)$ et $w' = (\rho_{\frac{p}{q}}(u\,a\,w_m^+))$ ont donc des valeurs différentes et étiquettent deux branches de $\mathcal{T}'_{\frac{p}{q}}$ partant de $n$. □

**Proposition 6.34** – *L'ensemble* $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$ *ne contient aucun point isolé.*

Démonstration. Soit un réel $x \in \mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$. D'après le théorème 6.22, il existe donc un mot infini $v = (a_0\,a_1 \cdots a_k \cdots) \in B_{\frac{p}{q}}^{\omega}$ accepté par $\mathcal{T}'_{\frac{p}{q}}$ tel que $\rho_{\frac{p}{q}}(v) = x$. On note son calcul dans $\mathcal{T}'_{\frac{p}{q}}$ par

$$0 = n_0 \xrightarrow{a_0} n_1 \xrightarrow{a_1} n_2 \xrightarrow{a_2} \cdots \xrightarrow{a_{k-1}} n_k \xrightarrow{a_k} \cdots \qquad (*)$$

Soit un entier $k$. On applique le lemme 6.33 à $n_k$, si bien qu'il existe deux mots $w, w' \in B_{\frac{p}{q}}^{\omega}$ tels que

$$n_k \xrightarrow{w} \cdots \ , \quad n_k \xrightarrow{w'} \cdots \quad \text{et} \quad \rho_{\frac{p}{q}}(w') \neq \rho_{\frac{p}{q}}(w) \ .$$

Si bien que les mots infinis

$$a_0\,a_1 \cdots a_{k-1}\,w \quad \text{et} \quad a_0\,a_1 \cdots a_{k-1}\,w'$$

ont des valeurs différentes, et, puisque le calcul de $(a_0\,a_1 \cdots a_{k-1})$ atteint dans $\mathcal{T}'_{\frac{p}{q}}$ l'état $n_k$ (d'après $(*)$), qu'ils sont acceptés par $\mathcal{T}'_{\frac{p}{q}}$. Un de ces deux mots (que l'on

note $w_k$) à donc une valeur différente de $\rho_{\frac{p}{q}}(v)$ ; il vérifie donc

$$w_k \in \mathfrak{L}(\mathcal{T}'_{\frac{p}{q}}) \qquad (**)$$

$$\rho_{\frac{p}{q}}(w_k) \neq \rho_{\frac{p}{q}}(v) \qquad (\overset{*}{**})$$

$$w_k \text{ admet le même préfixe de longueur } k \text{ que } v\ . \qquad (\overset{**}{**})$$

On note de plus $y_k = \rho_{\frac{p}{q}}(w_k)$. Puisque $w_k \in \mathfrak{L}(\mathcal{T}'_{\frac{p}{q}})$ (d'après $(**)$), $y_k$ appartient à $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$ (théorème 6.22) ; il découle donc de $(\overset{*}{**})$ que

$$y_k \in \Big(\mathsf{adh}(\mathsf{S}_{\frac{p}{q}}) \setminus \{x\}\Big)\ .$$

La suite $(w_k)_{k\in\mathbb{N}}$ converge vers $v$ (d'après $(\overset{**}{**})$), donc, puisque $\rho_{\frac{p}{q}}$ est continue (lemme 4.30), la suite $(y_k)_{k\in\mathbb{N}}$ converge vers $\rho_{\frac{p}{q}}(v) = x$. En définitive, $(y_k)_{k\in\mathbb{N}}$ est une suite à valeurs dans $\Big(\mathsf{adh}(\mathsf{S}_{\frac{p}{q}}) \setminus \{x\}\Big)$ qui converge vers $x$ ; celui-ci n'est donc pas un point isolé dans $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$. □

Une autre propriété de l'ensemble triadique de Cantor (défini dans l'exemple 6.28) est qu'il est de mesure nulle. Cette propriété n'est, dans le cas général, pas vérifié par tous les ensembles de Cantor, mais nous conjecturons que c'est le cas pour $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$.

**Conjecture 6.35** – *Si $p > (2\,q - 1)$, alors $\mathsf{adh}(\mathsf{S}_{\frac{p}{q}})$ est de mesure nulle.*

# Conclusion de la deuxième partie

Dans cette partie, nous avons poursuivi l'étude des systèmes de numération à base rationnelle commencée dans [2], et notamment celle de $L_{\frac{p}{q}}$, le langage des représentations des entiers en base $\frac{p}{q}$. La complexité de ce langage est déjà établie dans [2] mais la propriété (que nous avons appelée *itération préfixe finie (FLIP)*) qui la caractérise n'y est pas mise en avant. Dire que $L_{\frac{p}{q}}$ n'est pas un langage algébrique est en deçà de la réalité.

Le chapitre 5 tire profit de l'existence d'un additionneur en base $\frac{p}{q}$ (établie elle aussi dans [2]). On s'intéresse d'abord aux monoïdes (additifs et $\frac{p}{q}$-représentables) de nombres et aux langages qui les représentent (en base $\frac{p}{q}$). En particulier, il y est montré que si un monoïde est finiment engendré, alors le langage qui le représente est FLIP.

D'autre part, dans tous les systèmes de numération, l'étude des représentations des ensembles périodiques est un passage obligé ; nous l'avons initiée ici pour la base $\frac{p}{q}$. Quand la période est première avec $q$, la situation est similaire à celle de la base entière : la représentation de chaque ensemble périodique est un langage régulier. Au contraire, un ensemble d'entiers périodique de période $q$ est représenté par un langage inclus dans $L_{\frac{p}{q}}$ ; ce langage est inséparable de son complémentaire (dans $L_{\frac{p}{q}}$) par un langage régulier. Nous conjecturons que ce résultat s'étend à toutes les périodes qui ne sont pas premières avec $q$.

Dans le chapitre 6 est étudiée la fonction qui associe le mot minimal de $n$ au mot minimal de $(n+1)$. Nous avons montré que cette fonction est réalisée par un transducteur $\mathcal{D}_{\frac{p}{q}}$ dont la structure est essentiellement celle du langage $L_{\frac{p}{q}}$ lui-même. Ceci a permis de distinguer deux classes de systèmes de numération à base rationnelle suivant les propriétés topologiques de l'ensemble des envergures : l'adhérence de celui-ci est dans un cas un intervalle et dans l'autre un ensemble de Cantor.

La proximité des structures de $\mathcal{D}_{\frac{p}{q}}$ et de $L_{\frac{p}{q}}$ est un fait remarquable mais relativement abstrait. Nous pensons qu'il est possible d'utiliser ce résultat pour faire émerger un peu d'ordre au sein de $L_{\frac{p}{q}}$. Nos tentatives en ce sens n'ont pour l'instant pas abouti.

Les systèmes de numération à base rationnelle restent largement inexplorés et de nombreuses questions à leurs propos sont toujours ouvertes. Par exemple, peu d'éléments sont connus sur les différents facteurs d'un mot minimal, sans même parler des fréquences de leurs apparitions ; ces questions sont généralement liées à des problèmes de théorie des nombres ouverts depuis longtemps.

À notre connaissance, il n'y a que deux résultats sur la complexité factorielle (*cf.* [20]) en base rationnelle. Dans [60], les auteurs considèrent l'ensemble des représentations des nombres de 1 à $n$ et une longueur $l$ ; ils montrent que la fréquence

[2] Shigeki Akiyama, Christiane Frougny et Jacques Sakarovitch, 2008, *Powers of rationals modulo 1 and rational base number systems.*

[20] Julien Cassaigne et François Nicolas, 2010, *Factor Complexity.*

[60] Johannes F. Morgenbesser, Wolfgang Steiner et Jörg M. Thuswaldner, 2013, *Patterns in rational base number systems.*

d'apparitions de chaque motif de longueur $l$ au sein de toutes ces représentations tend vers la moyenne quand $n$ tend vers l'infini. Une second article [29] donne l'expression d'une borne inférieure de la complexité factorielle des mots minimaux ; il ne concerne néanmoins pas toutes les bases rationnelles et n'atteint pas les attentes que suggèrent les observations empiriques. En effet, les études statistiques de mots minimaux (voir par exemple nos travaux [52] non-repris dans ce mémoire) suggèrent que tous les motifs y apparaissent très tôt avec des fréquences d'apparitions qui tendent rapidement vers la moyenne ; à vrai dire, chaque mot minimal semble même se comporter comme une suite pseudo-aléatoire uniforme.

La grande généralité de ces observations sont à comparer avec la question ouverte suivante, qui illustre bien les lacunes des connaissances théoriques actuelles : *en base $\frac{3}{2}$, le motif 11 apparaît-il au moins une fois au sein de chaque mot minimal* ?

[29] Artūras Dubickas, 2009, *On integer sequences generated by linear maps.*

[52] Victor Marsault, 2016, *A few statistical experiments on minimal words in rational base numeration systems.*

# Troisième partie

# Signature et étiquetage

Chapitre 7

# Sérialisation d'arbres et de langages infinis

Ce chapitre introduit la notion de *signature* d'un arbre (ou d'un langage) ; c'est un mot infini qui caractérise l'arbre (ou le langage). Suivant la direction considérée (d'arbre vers mot, ou de mot vers arbre), on parle de *sérialisation* de l'arbre ou de *génération* de l'arbre par sa signature.

La signature d'un arbre (enraciné, ordonné, de degré fini) est une suite d'entiers, la suite des degrés des nœuds de l'arbre parcouru en largeur. Puisque l'arbre est ordonné (c'est-à-dire que l'ensemble des successeurs de chaque nœud est muni d'un ordre total), un parcours en largeur est *canonique*, celui qui visite les successeurs dans l'ordre. C'est ce parcours en largeur que l'on utilise pour la sérialisation, une procédure qui est donc entièrement déterministe. Nous verrons que deux arbres distincts ont de plus toujours des signatures différentes, si bien que la signature est caractéristique de l'arbre. Nous ne considérerons que des arbres de degré borné, dont la signature est donc une suite d'entiers bornée, autrement dit, un mot infini sur un alphabet fini de chiffres.

Si de plus les arcs de l'arbre sont étiquetés, on appelle *étiquetage* la suite des étiquettes de ses arcs visités par le même parcours en largeur canonique. Le couple signature/étiquetage est caractéristique de l'arbre étiqueté. Puisqu'un langage clos par préfixe (et sur un alphabet ordonné) est essentiellement un arbre étiqueté ; on utilise pour celui-là toutes les notions définies pour celui-ci ; on parle donc de signature, d'étiquetage, de sérialisation et de génération d'un langage clos par préfixe.

Le parcours en largeur d'un langage $L$ (clos par préfixe) visite les nœuds dans le même ordre que l'énumération des mots de $L$ dans l'ordre radiciel ; le calcul de la signature $L$ considère donc implicitement $L$ comme un SNA (Système de Numération Abstrait, voir section 1.7 page 28). C'est pourquoi ce formalisme est particulièrement adapté à la description des langages des représentations d'entiers dans des systèmes de numération. Par exemple, dans le cas des bases entières, la signature du langage des représentations en base $p$ est simplement $p^\omega$. C'est aussi

Une version courte de chapitre a été publié dans les actes de DLT 2014, voir [55]. Nous avons appris entre temps que le théorème principal de cet article (appelé théorème V dans ce mémoire) découle de résultats connus ; le formalisme de signature/étiquetage utilisé est néanmoins nouveau.

le cas pour certains systèmes de numération non-standard, comme le système de Fibonacci, dont la signature est une version du mot de Fibonacci. Nous verrons aussi dans le chapitre 8 que la signature du langage des représentations en base $\frac{p}{q}$ est périodique, et que sa période est liée au mot de Christoffel de pente $\frac{p}{q}$.

La section 7.1 introduit en détail les notions mentionnées précédemment (signature, étiquetages, *etc.*). Dans la section 7.2, nous montrons que les signatures et étiquetages des langages réguliers forment une sous-classe des mots morphiques (théorème V) ; ce résultat est une conséquence d'un résultat de Rigo et Maes [71].

**Théorème V** – *Un langage clos par préfixe $L$ est régulier si et seulement si la signature étiquetée de $L$ est s-morphique.*

Enfin, dans la section 7.3, nous utilisons le formalisme signature/étiquetage pour donner une définition des *systèmes de numération morphiques*, originellement dus à Dumont et Thomas [30].

# Arbre, signature, étiquetage

## Arbre et i-arbre

Dorénavant, nous appellerons *arbre* un graphe orienté connexe qui possède les propriétés suivantes.

- Il existe un nœud distingué, appelé *racine*, à partir duquel tous les autres nœuds sont accessibles ; sur les figures, la racine est toujours le nœud le plus à gauche.
- Tout nœud, excepté la racine, possède exactement un arc entrant et un nombre fini d'arcs sortants.
- Les successeurs de chaque nœud sont ordonnés ; cet ordre est représenté sur les figures par la règle : un plus petit successeur est en dessous d'un plus grand successeur.

Les conventions établies ci-dessus impliquent donc que les deux arbres de la figure 1 sont distincts. Leurs racines respectives sont dans les deux cas le nœud étiqueté par $r$ mais, le plus petit successeur de la racine est, à gauche, une feuille étiquetée par $x$ et, à droite, un nœud interne étiqueté par $y$.


Figure 1 – Deux arbres différents

[71] Michel Rigo et Arnaud Maes, 2002, *More on Generalized Automatic Sequences.*

[30] Jean-Marie Dumont et Alain Thomas, 1989, *Systèmes de Numération et Fonctions Fractales Relatifs aux Substitutions.*

On appelle *degré* $\mathsf{d}(x)$ *d'un nœud* $x$ le nombre de successeurs de $x$. On ne considère que des arbres infinis, mais puisque le degré de chaque nœud est fini, un parcours en largeur visite à terme chaque nœud. Notre but est plus tard de décrire des langages (sur des alphabets finis) comme des arbres étiquetés, qui sont donc de degré borné ; c'est pourquoi on ne considérera dans la suite que des arbres de degré borné.

Puisque les arbres sont ordonnés, il y a un *parcours en largeur canonique*, celui qui visite les successeurs d'un nœud par ordre croissant : sur les figures les nœuds sont visités par ce parcours de gauche à droite, puis de bas en haut pour les nœuds de la même colonne (c'est-à-dire à la même profondeur). Ce parcours associe à chaque nœud un entier, la position à laquelle il a été visité ; on peut donc considérer que l'ensemble des nœuds d'un arbre est toujours $\mathbb{N}$ : 0 est la racine et le nœud $i \in \mathbb{N}$ est le $(i+1)$-ème nœud visité par le parcours en largeur (canonique) de l'arbre. La définition formelle d'un arbre est la suivante.

**Définition 7.1** – *Un* arbre $T$ *est un sous-ensemble de* $\mathbb{N} \times \mathbb{N}$ *vérifiant les conditions suivantes.*

**a)** *Si* $(n, m)$ *et* $(n', m')$ *sont deux arcs de* $T$*, alors* $n < n'$ *implique* $m < m'$*.*

**b)** *Pour tout entier* $m > 0$*, il existe un unique entier* $n < m$ *tel que* $(n, m) \in E$*.*

**c)** $(0, 0) \notin T$*.*

On note $n \xrightarrow[T]{} m$ si $(n, m) \in T$. L'ordre des successeurs du nœud $n$ est l'ordre usuel sur les entiers, restreint à l'ensemble des successeurs de $n$. Enfin, $\mathsf{d}(n)$ est le cardinal de l'ensemble $\{m \mid (n, m) \in T\}$ et on appelle *profondeur de* $n$ la longueur du chemin allant de la racine à $n$ (l'existence de ce chemin est une conséquence de la condition 7.1b).

Le prochain résultat regroupe des propriétés essentielles sur les arbres ; en particulier la dernière montre que la définition 7.1 respecte les spécifications précédentes.

**Propriété 7.2** – *Soit un arbre* $T$*.*

**a)** *Si* $n \xrightarrow[T]{} m$ *et* $n' \xrightarrow[T]{} m'$ *sont deux arcs de* $T$ *tels que* $m < m'$*, alors* $n \leqslant n'$*.*

**b)** *Si deux entiers* $m \leqslant m'$ *sont les successeurs d'un entier* $n$ *dans* $T$*, alors tout entier* $i$*,* $m \leqslant i \leqslant m'$*, est également un successeur de* $n$*.*

**c)** *Soit un entier* $k$ *Si deux entiers* $m \leqslant m'$ *sont les successeurs d'ordre* $k$ *d'un entier* $n$ *dans* $T$*, alors tout entier* $i$*,* $m \leqslant i \leqslant m'$*, est également un successeur d'ordre* $k$ *de* $n$*.*

**d)** *Si* $m < m'$*, alors* $m$ *est moins (ou aussi) profond que* $m'$ *dans* $T$*.*

**e)** *Le nœud* $n$ *est le* $(n+1)$*-ème nœud visité par le parcours en largeur canonique.*

Démonstration. **a)** Si l'on supposait au contraire que $n > n'$, alors la condition 7.1a impliquerait que $m > m'$ ce qui contredirait les hypothèses.

**b)** Si $i = m$ ou $i = m'$, alors le point découle de l'unicité du prédécesseur (condition 7.1b). Autrement, puisque $m$ a un prédécesseur, alors $m > 0$ donc $i > 0$ qui a donc également un prédécesseur (condition 7.1b) que l'on note $j$. Puisque par hypothèse $m < i < m'$, il découle du (**a**) que $n \leqslant i \leqslant n$ donc $i = n$

**c)** Par récurrence sur $k$ ; l'unique successeur d'ordre 0 de $n$ est $n$ lui même. Si $k > 1$, on note $x$, $j$ et $j'$ les prédécesseurs respectifs de $i$, $m$ et $m'$ ; il s'ensuit en particulier

que $j$ et $j'$ sont des successeurs d'ordre $(k-1)$ de $n$. Et les inéquations $m \leqslant i \leqslant m'$ impliquent (d'après (**a**)) que $j \leqslant x \leqslant j'$, donc par hypothèse de récurrence le nœud $x$ (le prédécesseur de $i$), est un successeur de $n$ d'ordre $(k-1)$.

**d)** Soit $k$ et $k'$ les profondeurs respectives de $m$ et $m'$. Par l'absurde. Supposons que $k' < k$ donc le prédécesseur d'ordre $k'$ de $m$ n'est pas la racine ; c'est donc un nœud $n > 0$ alors que le prédécesseur d'ordre $k'$ de $m'$ est la racine 0. Appliquer $k'$ fois le point (**b**) implique que $n \leqslant 0$, contradiction.

**e)** Il découle de la condition 7.1b que tous les nœuds sont accessibles depuis la racine donc il suffit de prouver que pour tout entier $m < m'$, $m$ est visité avant $m'$ par le parcours en largeur canonique. Il découle du point (**c**) que $m$ ne peut être strictement plus profond que $m'$. Si $m$ est strictement moins profond que $m'$, alors il sera nécessairement visité avant $m'$ par le parcours en largeur ; on suppose donc dans la suite que $m$ et $m'$ sont à la même profondeur dans $T$.

On note $n$ leur plus grand ancêtre commun et les chemins respectifs de $n$ à $m$ et $m'$ par :

$$\begin{aligned} n = m_0 &\longrightarrow m_1 \longrightarrow m_2 \longrightarrow \cdots \longrightarrow m_k = m \\ n = m'_0 &\longrightarrow m'_1 \longrightarrow m'_2 \longrightarrow \cdots \longrightarrow m'_k = m' \end{aligned}$$

ce qui implique que pour tout entier $i$, $0 < i \leqslant k$, $m_i \neq m'_i$. Utiliser itérativement le point (**a**) montre que pour tout entier $i \leqslant k$, $m_i < m'_i$. En particulier $m_1 < m'_1$ donc $m_1$ est visité avant $m'_1$ par le parcours en largeur canonique de $T$ car ils sont tous les deux successeurs du même nœud $n$. Il s'ensuit itérativement que pour tout $i \leqslant k$, $m_i$ est visité avant $m'_i$, donc en particulier que $m$ est visité avant $m'$. □

Il sera parfois utile de considérer que la racine est aussi son propre successeur, c'est-à-dire qu'elle présente une boucle. Cela rend la fonction prédécesseur totale et permettra plus tard un parallèle plus naturel avec les systèmes de numération. Cette structure *d'arbre avec une boucle sur la racine* est appelée *i-arbre* et est définie comme un arbre, si ce n'est que la condition 7.1c est inversée.

**Définition 7.3** – *Un* i-arbre $T$ *est un sous-ensemble de* $\mathbb{N} \times \mathbb{N}$ *vérifiant les conditions suivantes.*

**a)** *Si* $(n, m)$ *et* $(n', m')$ *sont deux arcs de* $T$*, alors* $n < n'$ *implique* $m < m'$*.*
**b)** *Pour tout entier* $m > 0$*, il existe un unique entier* $n < m$ *tel que* $(n, m) \in E$*.*
**c)** $(0, 0) \in T$*.*

Les notions d'arbre et d'i-arbre sont tellement proches qu'on passe d'un arbre à l'i-arbre correspondant sans plus de cérémonie.

## Signature

**Définition 7.4** – *On appelle* signature *une suite infinie* $\mathbf{s} = s_0\, s_1\, s_2 \cdots$ *d'entiers bornée ; il s'agit donc d'un mot infini sur un alphabet (implicite) fini de chiffres.*

*On note alors $S_j$, la somme partielle des $j$ premiers termes de* $\mathbf{s}$ *:*

$$\forall j \in \mathbb{N} \qquad S_j \;=\; \sum_{i=0}^{j-1} s_i \;. \tag{7.1}$$

*En particulier, si $j > 0$, $S_j = S_{(j-1)} + s_j$.*

*La signature* $\mathbf{s} = s_0\, s_1\, s_2 \cdots$ *est dite* valide *si elle satisfait la condition suivante :*

$$\forall j \in \mathbb{N} \qquad S_{(j+1)} \;>\; (j+1) \;. \tag{7.2}$$

**Définition 7.5** – *La signature* $\mathbf{s}$ *d'un i-arbre $T$ est la suite des degrés des nœuds de $T$ dans l'ordre du parcours en largeur :*

$$\mathbf{s} \;=\; s_0\, s_1\, s_2 \cdots \qquad \textit{où} \qquad \forall n \in \mathbb{N} \qquad s_n \;=\; \mathsf{d}(n) \;.$$

*Par convention, la signature d'un arbre $T$ est celle de l'i-arbre correspondant.* [1]

Par exemple la figure 2 représente un i-arbre dont la signature commence par :

$$\begin{array}{cccccccccc} & 2 & 1 & 1 & 2 & 1 & 2 & 2 & 1 & \cdots \\ = & \mathsf{d}(0) & \mathsf{d}(1) & \mathsf{d}(2) & \mathsf{d}(3) & \mathsf{d}(4) & \mathsf{d}(5) & \mathsf{d}(6) & \mathsf{d}(7) & \cdots \end{array}$$

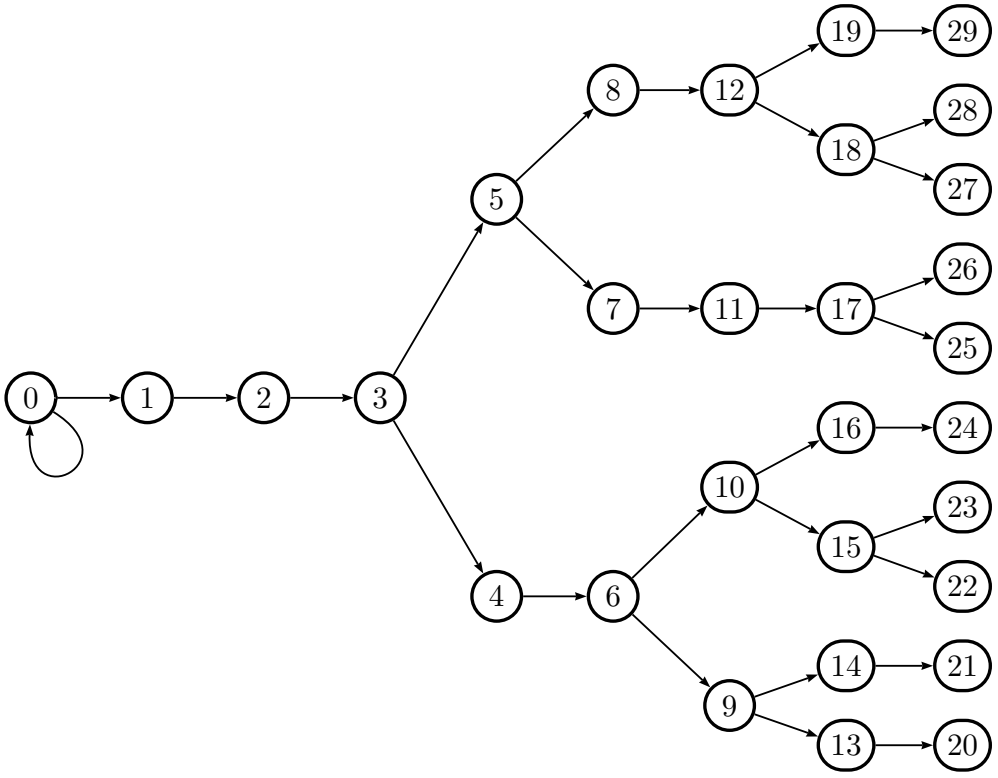



Figure 2 – Un arbre dont la signature commence par $2\,1\,1\,2\,1\,2\,2\,1\,1\,2\,1\,2\,1\,1\,2\cdots$

1. En pratique, si $T$ est un arbre, la première composante de sa signature, $s_0$, est égale à $(\mathsf{d}_T(0)+1)$ car il faut prendre en compte la boucle qui apparaît sur la racine lors du passage à l'i-arbre.

**Lemme 7.6** – *Soient un arbre $T$ dont on note la signature $\mathbf{s} = s_0 s_1 s_2 \cdots$ et deux nœuds $j, m \in \mathbb{N}$. Si $m$ est le plus petit successeur de $j$, alors $m = S_j$.*

Démonstration. On note $m' = S_j$ et $j'$ le prédécesseur de $m'$. Les nœuds compris entre 0 et $(j-1)$ ont en tout $S_j = m'$ successeurs qui sont les $m'$ plus petits entiers (allant 0 à $(m'-1)$) donc $j' \geqslant j$.

D'autre part, puisque chaque nœud n'a qu'un prédécesseur, aucun nœud strictement plus petit que $m'$ n'est successeur de $j$ donc $m \geqslant m'$. Puisque la fonction prédécesseur est croissante (propriété 7.2a) il s'ensuit que $j \geqslant j'$, donc, d'après le paragraphe précédent que $j = j'$.

Le nœud $m'$ est donc un successeur de $j$ et, puisque $m$ est le plus petit successeur de $j$, il s'ensuit $m \leqslant m'$, donc que $m = m'$. □

La réciproque de ce lemme est fausse dans le cas général. En effet, même si l'équation $m = S_n$ est vérifiée, $m$ n'est pas nécessairement le successeur de $n$, car celui-ci peut n'avoir aucun successeur. Dans ce cas $s_n = 0$, ce qui implique que l'équation $m = (S_n + s_n) = S_{(n+1)}$ est également vérifiée, le lemme suivant s'ensuit.

**Lemme 7.7** – *Soient un arbre $T$ dont on note la signature $\mathbf{s} = s_0 s_1 s_2 \cdots$ et un entier $j > 0$. Le nœud $S_j$ est le plus petit successeur du nœud $n$, où $n$ est le plus petit entier tel que $s_n \neq 0$ et $n \geqslant j$.*

Les deux propositions suivantes, réciproque l'une de l'autre, énoncent la bijection entre signatures valides et i-arbres.

**Proposition 7.8** – *La signature d'un (i-)arbre est valide.*

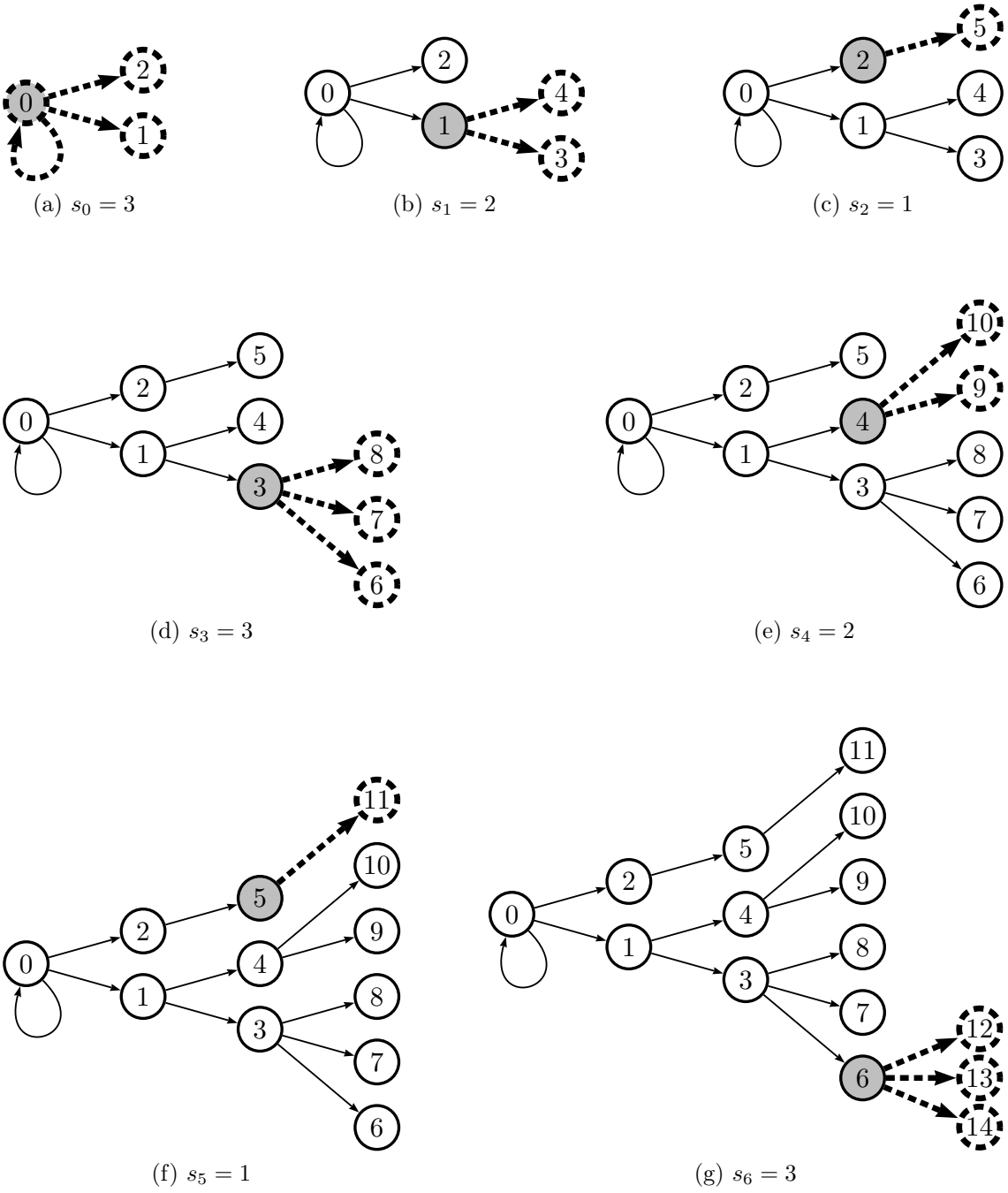
Démonstration. Soit un arbre $T$ dont on note la signature $\mathbf{s} = s_0\, s_1 \cdots s_i \cdots$. Soit un entier $j$. D'après le lemme 7.7 précédent, $S_{(j+1)}$ est le plus petit successeur de $n$, où $n$ est le plus petit entier tel que $s_n \neq 0$ et $n \geqslant (j+1)$. En particulier, $n \longrightarrow S_j$ donc (d'après la condition 7.1b,) $n < S_{j+1}$, ce qui implique $S_{j+1} > n \geqslant (j+1)$. □

**Proposition 7.9** – *Une signature valide $\mathbf{s}$ définit de façon unique un arbre $T_{\mathbf{s}}$ dont la signature est $\mathbf{s}$.*

La démonstration de la proposition 7.9 consiste en une procédure qui engendre un i-arbre à partir d'une signature valide $\mathbf{s} = s_0\, s_1\, s_2 \cdots$ donnée. Cette procédure utilise deux entiers : le point de départ $n$ et le point d'arrivée $m$ de l'arc courant, qui sont tous les deux initialisés à $n = m = 0$.

A l'étape $(n+1)$ de la procédure, $s_n$ nouveaux nœuds sont créés ; il s'agit des nœuds $m,\ m+1, \ldots, (m+s_n-1)$. Puis $s_n$ transitions sont créées, toutes partant de $n$ et chacune arrivant à l'un de ces nouveaux nœuds. Enfin, $n$ est incrémenté de 1 et $m$ de $s_n$

La validité de $\mathbf{s}$ assure qu'à chaque étape de la procédure $n < m$, sauf à la toute première étape, quand $n = m = 0$. Il s'ensuit que chaque nœud est strictement plus grand que son prédécesseur, à l'exception de la racine qui est son propre prédécesseur. La figure 3 montre les sept premières étapes de la génération de l'arbre $T_{(321)^\omega}$.




(a) $s_0 = 3$ (b) $s_1 = 2$ (c) $s_2 = 1$

(d) $s_3 = 3$ (e) $s_4 = 2$

(f) $s_5 = 1$ (g) $s_6 = 3$

Figure 3 – Les sept premières étapes de la génération de $T_{(321)^\omega}$

## Étiquetage

Un arbre étiqueté $\mathcal{T}$ est un arbre dont les arcs sont étiquetés sur un alphabet ordonné $A$. Les ordres de $\mathcal{T}$ et de $A$ doivent alors être cohérents, c'est-à-dire qu'un arc vers un plus petit successeur doit être étiqueté par une plus petite lettre.

**Définition 7.10** – *Un* (i-)arbre étiqueté $\mathcal{T}$ *sur un alphabet ordonné $A$ est un sous-ensemble de $\mathbb{N} \times A \times \mathbb{N}$ vérifiant les deux conditions suivantes.*

**a)** *La projection $T = \{(n, m) \mid (n, a, m) \in \mathcal{T}\}$ sur $\mathbb{N} \times \mathbb{N}$ de $\mathcal{T}$ est un (i-)arbre.*

**b)** *Pour tous triplets $(n, a, m)$ et $(n, b, m')$ appartenant à $\mathcal{T}$, si $m < m'$ alors $a < b$.*

*On note $n \xrightarrow[\mathcal{T}]{a} m$ plutôt que $(n, a, m) \in \mathcal{T}$, auquel cas $m$ est appelé* le successeur par *$a$ de $n$.*

**Définition 7.11** – *On appelle* étiquetage *d'un i-arbre étiqueté $\mathcal{T}$ la suite des étiquettes des arêtes de $\mathcal{T}$ dans l'ordre de leurs visites par le parcours en largeur :*

$$\boldsymbol{\lambda} \;=\; \lambda_0\, \lambda_1\, \lambda_2 \cdots \lambda_i \cdots$$

*où, pour tout entier $i$, $\lambda_i$ est la* **lettre** *qui étiquette l'unique arc entrant du nœud $i$.*

On appelle *signature étiquetée* un couple $(\mathbf{s}, \boldsymbol{\lambda})$ formé d'une signature et d'un étiquetage. La *décomposition de $\boldsymbol{\lambda}$ par rapport à $\mathbf{s}$* est l'unique suite $(u_i)_{i\in\mathbb{N}}$ de **mots** finis permettant de factoriser

$$\boldsymbol{\lambda} \;=\; u_0\, u_1\, u_2 \cdots u_i \cdots \qquad \text{avec} \qquad \forall i \in \mathbb{N} \;\; |u_i| = s_i \;. \tag{7.3}$$

**Lemme 7.12** – *Soit un i-arbre étiqueté $\mathcal{T}$ dont on note la signature étiquetée $(\mathbf{s}, \boldsymbol{\lambda})$ et $(u_i)_{i\in\mathbb{N}}$ la décomposition de $\boldsymbol{\lambda}$ (par rapport à $\mathbf{s}$). Alors pour tout nœud $n \in \mathbb{N}$, $u_n$ est le mot formé des étiquettes sortantes de $n$ dans l'ordre.*

Démonstration. Soit un nœud $n \in \mathbb{N}$. Si $s_n = 0$ alors $u_n$ est de longueur 0 (donc égal à $\varepsilon$) et le lemme est trivialement vérifié. Sinon, on note $m$ le plus petit successeur de $n$ et il découle du lemme 7.6 que

$$m \;=\; S_n \;=\; \sum_{i=0}^{n-1} s_i \;.$$

La première lettre de $u_n$, noté $a$, est dans $\boldsymbol{\lambda}$ à l'indice $|u_0| + |u_1| + \cdots + |u_{n-1}| = m$, donc l'étiquette de l'arc $n \xrightarrow{a} m$. De même, pour tout $i < s_n$, la $(i+1)$-ème lettre de $u_n$ est l'étiquette de la transition $n \xrightarrow{a} (m + i)$, ce qui conclut la preuve. □

Un mot $a_0\, a_1 \cdots a_k$ (sur un alphabet ordonné) est dit *croissant* si ses lettres sont rangées par ordre croissant : $a_0 < a_1 < \cdots < a_k$. On dit qu'un étiquetage $\boldsymbol{\lambda}$ est *cohérent* avec une signature $\mathbf{s}$ si tous les facteurs $u_i$ de la décomposition de $\boldsymbol{\lambda}$ par rapport $\mathbf{s}$ sont croissants. Enfin, une signature étiquetée $(\mathbf{s}, \boldsymbol{\lambda})$ est dite *valide* si $\mathbf{s}$ est une signature valide et que $\boldsymbol{\lambda}$ est cohérent avec $\mathbf{s}$.

**Proposition 7.13** – *La signature étiquetée d'un i-arbre étiqueté est valide.*

Démonstration. Il découle du lemme 7.12 que le $(n+1)$-ème facteur de la décomposition de l'étiquetage d'un i-arbre étiqueté est la suite des étiquettes sortantes de $n$, dans l'ordre. Ce facteur est donc croissant d'après la condition 7.10b. □

**Proposition 7.14** – *Une signature étiquetée* $(\mathbf{s}, \boldsymbol{\lambda})$ *définit de façon unique un i-arbre étiqueté* $\mathcal{T}_{(\mathbf{s},\boldsymbol{\lambda})}$ *dont la signature étiquetée est* $(\mathbf{s}, \boldsymbol{\lambda})$.

La procédure engendrant l'i-arbre étiqueté $\mathcal{T}_{(\mathbf{s},\boldsymbol{\lambda})}$ est identique à celle correspondant à la proposition 7.9 si ce n'est que chaque nouvel arc $n \xrightarrow{a} m$ est étiqueté par $a = \lambda_m$. Le lemme suivant donne une caractérisation compacte de i-arbre étiqueté $\mathcal{T}_{(\mathbf{s},\boldsymbol{\lambda})}$

**Lemme 7.15** – *Soit une signature étiquetée valide* $(\mathbf{s}, \boldsymbol{\lambda})$ *qui définit l'i-arbre étiqueté* $\mathcal{T}_{(\mathbf{s},\boldsymbol{\lambda})}$. *L'équivalence suivante est alors vérifiée :*

$$\forall n, m,\ \forall a \in A \qquad n \xrightarrow[\mathcal{T}_{(\mathbf{s},\boldsymbol{\lambda})}]{a} m \quad \Longleftrightarrow \quad S_n \leqslant m < S_{(n+1)} \quad et \quad a \ = \ \lambda_m \ .$$

Comme pour les signatures, l'étiquetage d'un arbre étiqueté $\mathcal{T}$ sur un alphabet $A$ est celui de son i-arbre étiqueté associé ; il faut néanmoins décider avec quelle lettre étiqueter la boucle rajoutée sur la racine. On prend alors la convention que l'i-arbre associé est étiqueté sur l'alphabet $A \uplus \{\#\}$ où $\#$ est une nouvelle lettre ($\# \notin A$) plus petite que toutes les lettres de $A$, et qui n'apparaît que sur la boucle $0 \xrightarrow{\#} 0$. La lettre $\#$ ne sera généralement pas utilisée car on considérera des langages fournit avec leur lettre de calage (voir sous-section suivante).

## Signature, langage calable et système de numération abstrait

La notion d'arbre étiqueté spécifiée à la définition 7.10 est alors très proche de celle de langage clos par préfixe ; on peut passer de la première à la seconde de la façon usuelle, rappelée ci-dessous.

**Définition 7.16** – *Un arbre étiqueté* $\mathcal{T} \subseteq \mathbb{N} \times A \times \mathbb{N}$ *définit le langage* $L$ *des étiquettes de ses branches :*

$$L_{\mathcal{T}} \ = \ \{\, u \in A^* \mid 0 \xrightarrow[\mathcal{T}]{u} n \text{ , pour un certain } n \in \mathbb{N} \,\} \ .$$

Soit un alphabet (ordonné) $A$ et un langage $L \subseteq E^*$ clos par préfixe. Réciproquement à la définition 7.16, on pourrait définir, à partir du langage $L$, l'arbre $\langle\, L,\, A,\, \varepsilon,\, E \,\rangle$ dont les arcs sont définis par

$$\forall u \in A^*,\ \forall a \in A \qquad u \xrightarrow{a} ua \qquad \text{si } \in L \ .$$

Néanmoins, cette transformation du langage $L$ en arbre étiqueté n'utilise pas le formalisme de la définition 7.10 ; en particulier, l'ensemble des nœuds n'est pas $\mathbb{N}$. On peut démontrer que cet arbre satisfait les propriétés nécessaires (enraciné, ordonné, *etc.*) pour que l'arbre $\mathcal{T}_L$ soit entièrement défini (implicitement) suivant notre formalisme.

Considérer le langage $L$ comme un système de numération abstrait (SNA, voir section 1.7 page 28) permet d'en donner la définition explicite suivante. En particulier, on rappelle que $\langle n \rangle_L$ désigne le $(n+1)$-ème mot de $L$ dans l'ordre radiciel.

**Définition 7.17** – *Soit un alphabet $A$ ordonné et un langage $L \subseteq E^*$ clos par préfixe. On associe à $L$ l'arbre étiqueté $\mathcal{T}_L \subseteq \mathbb{N} \times A \times \mathbb{N}$ défini par :*

$$\forall n, m \in \mathbb{N},\ \forall a \in A \qquad n \xrightarrow[\mathcal{T}_L]{a} m \qquad \text{si et seulement si} \qquad \langle n \rangle_L a = \langle m \rangle_L \ . \quad (7.4)$$

Une simple vérification montre que l'objet $\mathcal{T}_L$ dans la définition précédente est conforme à la définition 7.10 d'arbre. Il découle du lemme suivant que les transformations des définitions 7.16 et 7.17 sont l'inverse l'une de l'autre.

**Lemme 7.18** – *Soit un langage $L \subseteq A^*$ clos par préfixe. Pour tout nœud $n \in \mathbb{N}$, le chemin $0 \xrightarrow{u} n$ existe dans $\mathcal{T}_L$ si et seulement si $u = \langle n \rangle_L$.*

Démonstration. Sens direct. Par récurrence ; le mot qui étiquette le chemin $0 \xrightarrow[\mathcal{T}_L]{u} 0$ est nécessairement $\varepsilon = \langle 0 \rangle_L$.

Soient un mot $u \in A^*$ et une lettre $a \in A$ tels que

$$0 \xrightarrow[\mathcal{T}_L]{u} m \xrightarrow[\mathcal{T}_L]{a} n \ .$$

D'après l'hypothèse de récurrence, $u = \langle m \rangle_L$ et d'après la définition 7.17, le dernier arc de ce chemin existe si et seulement si $\langle m \rangle_L a = \langle n \rangle_L$. Il s'ensuit que $u\,a = \langle n \rangle_L$ ce qui conclut la récurrence.

Sens réciproque. Soit $n$ un entier. Puisque $\mathcal{T}_L$ est un arbre, le nœud $n$ est accessible depuis la racine par un certain mot $v$ ; il découle du sens direct que $v = \langle n \rangle_L$. □

Dorénavant, on identifie la notion d'arbre étiqueté avec celle de langage clos par préfixe. Si bien que l'on pourra écrire $n \xrightarrow[L]{u} m$ pour un certain langage $L$ (clos par préfixe) et noter $L_{(\mathbf{s},\boldsymbol{\lambda})}$ le langage engendré par $(\mathbf{s}, \boldsymbol{\lambda})$.

Des constructions similaires permettent de passer d'un i-arbre étiqueté à un langage calable (c'est-à-dire de la forme $z^*L$, où $z$ est une lettre strictement plus petite que toutes les lettres qui débutent des mots de $L$, voir définition 1.17 page 28). Une démonstration analogue à celle du lemme 7.18 donne le résultat suivant.

**Lemme 7.19** – *Soit un langage calable $z^*L \subseteq A^*$ clos par préfixe. Pour tout nœud $n \in \mathbb{N}$, le chemin $0 \xrightarrow[z^*L]{u} n$ existe si et seulement si $u = z^k \langle n \rangle_L$, pour un certain entier $k$.*

## Signature d'un langage régulier

Le but de cette section est d'établir une relation entre (une sous-classe des) mots morphiques et les langages réguliers. On trouve dans la littérature des termes différents (comme les adjectifs *substitutif* et *morphique*) pour désigner des objets semblables ou identiques. Nous faisons le choix d'utiliser la nomenclature basée sur le mot morphisme.

D'autre part, on utilisera $\Sigma$ pour noter l'alphabet utilisé conjointement avec un endomorphisme de mots, et des lettres grecques minuscules ($\alpha$, $\beta$, etc) pour désigner les lettres de $\Sigma$. Cette convention a pour but de bien différencier cet alphabet avec

ceux des automates (usuellement noté $A, B, D$) ce qui rendra, le moment venu, le passage des mots morphiques aux automates beaucoup plus clair.

On considère dans la suite un alphabet $\Sigma$. Soit $\sigma$ un endomorphisme de mots de $\Sigma^*$ dans lui-même. Il est dit *prolongeable en* $\alpha \in \Sigma$ si $\sigma(\alpha)$ commence par $\alpha$ et si la suite de mots $(\sigma^n(\alpha))_{n\in\mathbb{N}}$ a une longueur qui tend vers l'infini. Pour alléger la formulation, on dira dans la suite qu'un endomorphisme est *prolongeable* sans préciser *en* $\alpha$ ; la variable $\alpha$ est donc réservée à cet effet.

Quand $\sigma$ est prolongeable, il existe (par définition) un mot $u$ tel que $\sigma(\alpha) = \alpha\, u$ et la suite $(\sigma^n(\alpha))_{n\in\mathbb{N}}$ converge vers le mot infini

$$\sigma^\omega(\alpha) \;=\; \alpha\; u\; \sigma(u)\; \sigma^2(u)\; \cdots$$

Tout mot infini de cette forme, $\sigma^\omega(\alpha)$ pour un certain endomorphisme $\sigma$ prolongeable, est appelé *purement morphique.*

Un morphisme est dit *lettre-à-lettre* si l'image de chaque lettre est une lettre et il est dit *non-effaçant* s'il n'envoie aucune lettre sur le mot vide. Un mot infini est appelé *morphique* s'il est l'image d'un mot purement morphique par un second morphisme c'est-à-dire de la forme $f(\sigma^\omega(\alpha))$, pour certains morphismes $\sigma : \Sigma^* \to \Sigma^*$ prolongeable et $f : \Sigma^* \to A^*$. On peut supposer des propriétés supplémentaires sur ces deux morphismes, comme le montre le résultat suivant due à Cobham [24] (voir aussi [5]).

**Lemme 7.20** [24] – *Soit un mot $w \subseteq A^\omega$. Si $w$ est un mot morphique, alors il existe deux morphismes $\sigma : \Sigma^* \to \Sigma^*$ non-effaçant, prolongeable et $f : \Sigma^* \to A^*$ lettre-à-lettre tels que $w = f\big(\sigma^\omega(\alpha)\big)$.*

## Signature s-morphique

**Définition 7.21** – *Soit $\sigma$ un endomorphisme $\Sigma^* \to \Sigma^*$. On définit le morphisme lettre-à-lettre $f_\sigma : \Sigma^* \to D^*$ par*

$$\forall \beta \in \Sigma \qquad f_\sigma(\beta) \;=\; |\sigma(\beta)|\ ,$$

*et où $D \subseteq \mathbb{N}$ est un alphabet fini de chiffres, défini implicitement de façon à ce que $f_\sigma$ soit surjective.*

*Un mot est dit* s-morphique *s'il est de la forme $f_\sigma(\sigma^\omega(\alpha))$ pour un certain endomorphisme prolongeable $\sigma$.*

Un mot s-morphique est donc un mot morphique particulier ; de plus, il s'agit d'une suite d'entiers bornée, c'est donc une signature.

**Lemme 7.22** – *Toute signature s-morphique est valide.*

[24] Alan Cobham, 1968, *On the Hartmanis-Stearns problem for a class of tag machines.*

[5] Jean-Paul Allouche et Jeffrey Shallit, 2003, *Automatic Sequences : Theory, Applications, Generalizations.*

Démonstration. Soit un endomorphisme prolongeable $\sigma$ définissant le mot s-morphique

$$\mathbf{s} \;=\; s_0\, s_1\, s_2\, \cdots \;=\; f_\sigma(\sigma^\omega(\alpha))\ .$$

Par définition (équation (7.2)), il faut montrer que pour tout entier $j$, $S_{(j+1)} > (j+1)$.

On note $u$ le préfixe de longueur $(j+1)$ de $\sigma^\omega(\alpha)$. Le mot $f_\sigma(u)$ est donc le préfixe de longueur $(j+1)$ de $\mathbf{s}$ ce qui implique que la somme des lettres de $f_\sigma(u)$ est égale à $S_{(j+1)}$. Il découle de la définition 7.21 que la somme des lettres de $f_\sigma(u)$ est égale à la longueur de $\sigma(u)$, les deux inéquations suivantes sont donc égales membre-à-membre :

$$|\sigma(u)| > |u| \quad \text{et} \quad S_{(j+1)} > (j+1)\ .$$

Puisque $\sigma$ est prolongeable (et que $|u| \neq 0$), il existe un entier $n$ tel que $\sigma^n(\alpha) \sqsubseteq u \sqsubset \sigma^{(n+1)}(\alpha)$. Puisque $\sigma$ est un morphisme, l'appliquer aux trois membres donne $\sigma^{(n+1)}(\alpha) \sqsubseteq \sigma(u) \sqsubset \sigma^{(n+2)}(\alpha)$. On obtient donc $u \sqsubset \sigma^{(n+1)}(\alpha) \sqsubseteq \sigma(u)$.

Puisque $u$ est un préfixe strict de $\sigma(u)$ est en particulier strictement plus court, ce qui conclut la démonstration. □

**Définition 7.23** – *Une signature étiquetée* $(\mathbf{s}, \boldsymbol{\lambda})$ *est dite s-morphique s'il existe deux morphismes* $\sigma : \Sigma^* \to \Sigma^*$ *et* $g : \Sigma^* \to A^*$ *tels que*

**a)** $\sigma$ *est prolongeable ;*
**b)** *pour toute lettre* $\beta \in \Sigma$, $|g(\beta)| = |\sigma(\beta)|$ $(= f_\sigma(\beta))$ *;*
**c)** $\mathbf{s} = f_\sigma(\sigma^\omega(\alpha))$ *;*
**d)** $\boldsymbol{\lambda} = g(\sigma^\omega(a))$.

*On dit alors que* $(\mathbf{s}, \boldsymbol{\lambda})$ *est* définie *par* $(\sigma, g)$.

Le lemme suivant donne une caractérisation simple des signatures étiquetées s-morphiques qui sont valides.

**Lemme 7.24** – *Soit* $(\mathbf{s}, \boldsymbol{\lambda})$ *une signature étiquetée s-morphique définie par* $(\sigma, g)$, *avec* $\sigma : \Sigma^* \to \Sigma^*$ *et* $g : \Sigma^* \to A^*$ *La signature étiquetée* $(\mathbf{s}, \boldsymbol{\lambda})$ *est valide si et seulement si* $g(\beta)$ *est un mot croissant pour toute lettre* $\beta \in \Sigma$ *qui apparaît dans* $\sigma^\omega(\alpha)$.

Démonstration. On note $\sigma^\omega(\alpha) = \beta_0\, \beta_1\, \beta_2\, \cdots$, si bien que

$$\mathbf{s} = f_\sigma(\beta_0)\, f_\sigma(\beta_1)\, f_\sigma(\beta_2)\, \cdots \quad \text{et} \quad \boldsymbol{\lambda} = g(\beta_0)\, g(\beta_1)\, g(\beta_2)\, \cdots\ .$$

La décomposition de $\boldsymbol{\lambda}$ par rapport à $\mathbf{s}$ est, par définition, $\boldsymbol{\lambda} = u_0\, u_1\, \cdots\, u_i\, \cdots$ où, pour tout entier $i$, $|u_i| = f_\sigma(\beta_1) = |\sigma(\beta_i)|$.

Puisque $\mathbf{s}$ est s-morphique, elle est valide (en tant que signature, lemme 7.22) ; la signature étiquetée $(\mathbf{s}, \boldsymbol{\lambda})$ est donc valide si et seulement si $\boldsymbol{\lambda}$ est cohérent avec $\mathbf{s}$, c'est-à-dire si et seulement si pour tout entier $i$, le mot $u_i$ est croissant.

Or, puisque $(\mathbf{s}, \boldsymbol{\lambda})$ est s-morphique, toute lettre $\beta \in \Sigma$ vérifie $|\sigma(\beta)| = |g(\beta)|$ (condition 7.23b), donc

$$\forall i \in \mathbb{N} \qquad u_i \;=\; g(\beta_i)\ .$$

La signature étiquetée $(\mathbf{s}, \boldsymbol{\lambda})$ est donc valide si et seulement si pour tout entier $i$, le mot $g(\beta_i)$ est croissant. □

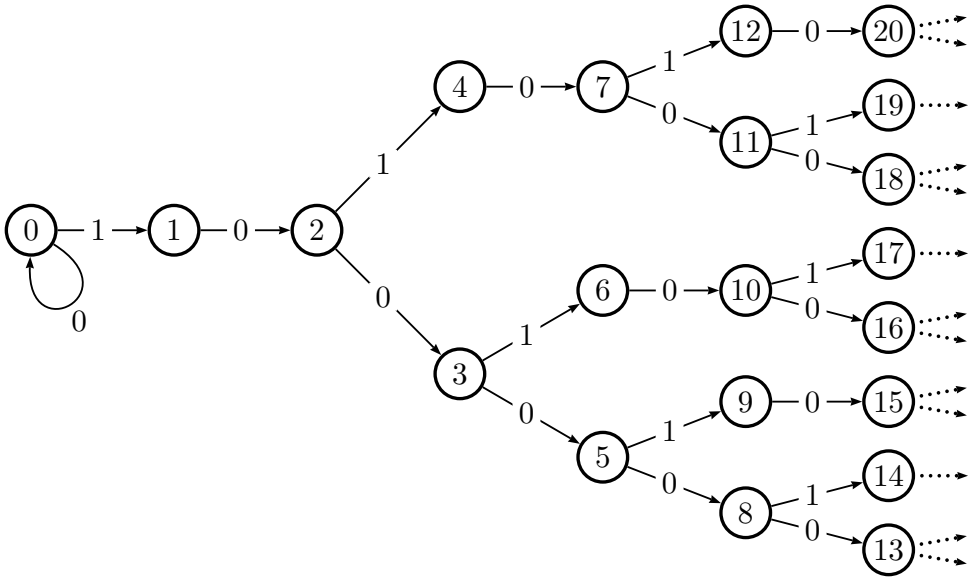



Figure 4 – Représentations des entiers dans le système de numération de Fibonacci

**Exemple 7.25** – *Le* mot de Fibonacci *est le mot purement morphique* $\sigma_1{}^\omega(\alpha)$ *où* $\sigma_1$ *est définie par* $\sigma_1(\alpha) = \alpha\,\beta$ *et* $\sigma_1(\beta) = \alpha$*, donc*

$$\sigma_1{}^\omega(\alpha) \;=\; \alpha\,\beta\,\alpha\,\alpha\,\beta\,\alpha\,\beta\,\alpha\,\alpha\,\beta\,\alpha\,\alpha\,\beta\,\alpha\,\beta\,\cdots$$

*Ce morphisme* $\sigma_1$ *induit la signature s-morphique*

$$\mathbf{s} \;=\; f_{\sigma_1}(\sigma_1^\omega(\alpha)) \;=\; 2\,1\,2\,2\,1\,2\,1\,2\,2\,1\,2\,2\,1\,2\,1\,\cdots$$

*On note de plus* $g_1$ *le morphisme* $\{\alpha,\beta\}^* \to [\![2]\!]^*$ *défini par* $g_1(\alpha) = 0\,1$ *et* $g_1(\beta) = 0$ *qui induit donc l'étiquetage*

$$\boldsymbol{\lambda} \;=\; g_1(\sigma_1^\omega(\alpha)) \;=\; 0\,1\;0\;0\,1\;0\,1\;0\;0\,1\;0\;0\,1\;0\,1\;0\;0\,1\;0\,1\;0\;0\,1\;0\;\cdots$$

*Le langage calable* $0^*L_{(\mathbf{s},\boldsymbol{\lambda})}$ *est représenté par la figure 4. Il s'agit du langage (calable) des représentations des entiers dans le système de numération de Fibonnacci, c'est-à-dire le langage régulier des mots qui ne contiennent pas le facteur* 11 *:*

$$0^*L_{(\mathbf{s},\boldsymbol{\lambda})} \;=\; [\![1]\!]^* \setminus ([\![1]\!]^*\;11\;[\![1]\!]^*)\;.$$

## Signatures s-morphiques et langages réguliers

Le lien entre les mots (purement) morphiques et les automates remonte aux travaux de Cobham [26] dont le résultat (théorème 7.26) est énoncé après quelques définitions.

Un *automate avec sortie* est un automate déterministe qui est muni d'une fonction donnant une valeur à chaque état final, de sorte qu'un mot n'est pas simplement accepté, mais qu'il lui est attribué une lettre dans un nouvel alphabet que l'on pourra appeler *couleur*. Un tel automate est formellement représenté par

[26] Alan Cobham, 1972, *Uniform Tag Sequences*.

un 6-tuple $\mathcal{A} = \langle Q, A, \delta, i, F, B, c \rangle$ où $\langle Q, A, \delta, i, F \rangle$ est un automate déterministe, où $B$ est un alphabet fini, et où $f$ est une fonction totale $F \to B$.

Soit un entier $p > 1$ considéré comme une base entière et un alphabet fini $B$. Un mot infini $\mathbf{s} = s_0 s_1 s_2 \cdots \subseteq B^\omega$ est dit $p$-automatique s'il existe un automate avec sortie $\mathcal{A}$ qui attribue, pour tout entier $i$, la couleur $s_i$ au mot $\langle i \rangle_p$. Un endomorphisme de mots $\sigma : \Sigma^* \to \Sigma^*$ est dit $p$-uniforme s'il envoie chaque lettre sur un mot de longueur $p$. De plus, un mot infini est dit *$p$-uniformément morphique* s'il s'agit d'un mot morphique $f(\sigma^\omega(\alpha))$ où $\sigma$ est un morphisme $p$-uniforme.

**Théorème 7.26** [26] – *Soit un entier $p > 1$. Un mot infini $\mathbf{s}$ est $p$-automatique si et seulement si $\mathbf{s}$ est $p$-uniformément morphique.*

Ce théorème a été adapté par Rigo et Maes aux systèmes de numération abstraits réguliers (SNAR). Soit un langage régulier $L$ considéré comme un SNAR et un alphabet $B$. Un mot infini $\mathbf{s} = s_0 s_1 s_2 \cdots \subseteq B^\omega$ est dit $L$-automatique s'il existe un automate avec sortie $\mathcal{A}$ qui attribue, pour tout entier $i$, la couleur $s_i$ au mot $\langle i \rangle_L$.

**Théorème 7.27** [71] – *Une suite $\mathbf{s}$ est $L$-automatique pour un certain SNAR $L$, si et seulement si $\mathbf{s}$ est un mot morphique.*

Le but de cette sous-section 7.2.2 est de démontrer le théorème V, qui est apparenté au résultat précédent.

**Théorème V** – *Un langage clos par préfixe $L$ est régulier si et seulement si la signature étiquetée de $L$ est s-morphique.*

Notez que le langage $L$ est régulier si et seulement si un langage calable $z^*L$ l'est (pour une certaine lettre $z$) ; dans ce cas, leurs signatures respectives sont égales et leurs étiquetages respectifs ne diffèrent que par la première lettre (# pour $L$ et $z$ pour $z^*L$). On peut reformuler le théorème V par :

**Théorème V′** – *Un langage calable $z^*L$ clos par préfixe est régulier si et seulement si la signature étiquetée de $z^*L$ est s-morphique.*

Ce théorème est proche du théorème 7.27 mais n'en est pas une conséquence immédiate. En effet, on peut appliquer celui-ci pour montrer que la signature et l'étiquetage sont des mots morphiques mais rien n'assure (a priori) que leurs générations utilisent le même endomorphisme ($\sigma$). Un problème analogue se pose dans le sens réciproque. Nous verrons que le théorème V′ est en fait une conséquence de la proposition 7.33, un des résultats intermédiaires de [71].

Les définitions 7.28 et 7.31 donnent les deux sens d'une correspondance automate $\leftrightarrow$ signature qui sont l'inverse l'une de l'autre (lemme 7.32).

**Définition 7.28** – *Soit une signature $(\mathbf{s}, \boldsymbol{\lambda})$ s-morphique (et valide) définie par $(\sigma, g)$ avec $\sigma : \Sigma^* \to \Sigma^*$ et $g : \Sigma^* \to A^*$.*

*L'*automate (déterministe) induit *par cette signature est $\mathcal{A}_{(\sigma,g)} = \langle \Sigma, A, \delta, \alpha, \Sigma \rangle$. L'ensemble des **états** est l'alphabet $\Sigma$, l'état initial est $\alpha$ (la lettre sur laquelle $\sigma$ est*

[71] Michel Rigo et Arnaud Maes, 2002, *More on Generalized Automatic Sequences.*

*prolongeable) et tous les états sont finals ; l'alphabet est $A$ ; la fonction de transition $\delta$ est définie dans le paragraphe suivant.*

*Soit une lettre de $\beta \in \Sigma$, donc un état de $\mathcal{A}_{(\sigma,g)}$. Les images de $\beta$ respectivement par $\sigma$ et $g$ sont des mots d'une même longueur $k$ (condition 7.23b) ; on note ces images par*

$$\sigma(\beta) \;=\; \gamma_0\,\gamma_1\,\cdots\,\gamma_{(k-1)} \qquad et \qquad g(\beta) \;=\; a_0\,a_1\,\cdots\,a_{(k-1)}\ .$$

*Il y a alors $k$ transitions sortantes de $\beta$ dans $\mathcal{A}_{(\sigma,g)}$ définies par*

$$\forall i < k \qquad \beta \xrightarrow[\mathcal{A}_{(\sigma,g)}]{a_i} \gamma_i\ .$$

*L'état $\beta$ a donc exactement $k$ transitions sortantes, une étiquetée par chaque lettre qui apparaît dans $g(\beta)$. Puisque $(s,\lambda)$ est valide et induite par $(\sigma,g)$, le mot $g(\beta)$ est un mot croissant (lemme 7.24) donc toutes ses lettres sont distinctes ; l'automate $\mathcal{A}$ est donc bel et bien déterministe.*

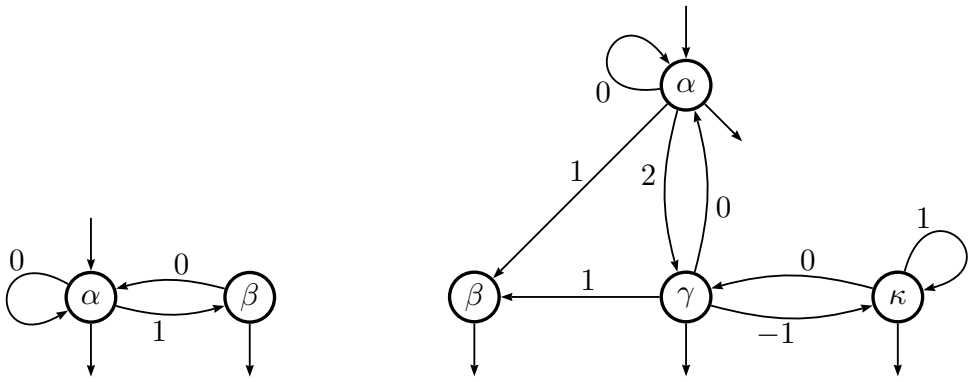
**Exemple 7.29** (Suite de l'exemple 7.25) – *On rappelle que $\sigma_1$ est le morphisme de Fibonacci de $\{\alpha,\beta\}^*$ dans lui-même, défini par*

$$\sigma_1(\alpha) \;=\; \alpha\,\beta \qquad et \qquad \sigma_1(\beta) \;=\; \beta$$

*et $g_1$ est le morphisme lettre-à-lettre de $\{\alpha,\beta\}^*$ dans $\{0,1\}^*$ défini par*

$$g_1(\alpha) \;=\; 0\,1 \qquad et \qquad g_1(\beta) \;=\; 0\ .$$

*L'automate $\mathcal{A}_{(\sigma_1,g_1)}$ est représenté à la figure 5a.*




(a) L'automate $\mathcal{A}_{(\sigma_1,g_1)}$ (b) L'automate $\mathcal{A}_{(\sigma_2,g_2)}$

Figure 5 – Deux transformations de signatures s-morphiques en automates

**Exemple 7.30** – *Soit l'endomorphisme $\sigma_2$ de $\{\alpha,\beta,\gamma,\kappa\}^*$ défini par*

$$\sigma_2(\alpha) \;=\; \alpha\,\beta\,\gamma\ ,\quad \sigma_2(\beta) \;=\; \varepsilon\ ,\quad \sigma_2(\gamma) \;=\; \kappa\,\alpha\,\beta \quad et \quad \sigma_2(\kappa) \;=\; \gamma\,\kappa \quad ;$$

*et le morphisme $g_2$ de $\{\alpha,\beta,\gamma,\kappa\}^*$ dans $\{-1,\,0,\,1,\,2\}^*$ défini par*

$$g_2(\alpha) \;=\; 0\,1\,2\ ,\quad g_2(\beta) \;=\; \varepsilon\ ,\quad g_2(\gamma) \;=\; (-1)\,0\,1 \quad et \quad g_2(\kappa) \;=\; 0\,1\ .$$

*Ces deux morphismes vérifient bien les conditions de la définition 7.23 et définissent la signature et l'étiquetage (s-morphique) suivants :*

$$\mathbf{s} \;=\; (3032)^\omega \quad \textit{et} \quad \boldsymbol{\lambda} \;=\; (012\;\;(-1)01\;01)^\omega;$$

*L'automate $\mathcal{A}_{(\sigma_2,g_2)}$ est représenté à la figure 5b. Nous verrons dans le chapitre 8 suivant que le langage $0^*L$ de cet automate est une* représentation non-canonique des entiers en base 2 *: la valeur en base 2 du $(i+1)$-ème mot de $L$ dans l'ordre radiciel est égale à $n$.*

**Définition 7.31** – *Soit $\mathcal{A} = \langle\, Q, A, \delta, i, F \,\rangle$ un automate (déterministe). Nous allons définir deux morphismes $\sigma_\mathcal{A} : Q^* \to Q^*$ et $g_\mathcal{A} : Q^* \to A^*$. Soit $s$ un état de $\mathcal{A}$ et l'on note ses transitions sortantes par $s \xrightarrow{a_0} t_0$, $s \xrightarrow{a_1} t_1, \ldots$, $s \xrightarrow{a_k} t_k$. On suppose de plus, sans perdre la généralité, que $a_0 < a_1 < \cdots < a_k$ ; les images de $s$ par $\sigma_\mathcal{A}$ et $g_\mathcal{A}$ sont alors respectivement définies par*

$$\sigma_\mathcal{A}(s) \;=\; t_0\, t_1 \cdots t_k \quad \textit{et} \quad g_\mathcal{A}(s) \;=\; a_0\, a_1 \cdots a_k \;.$$

Une simple vérification montre que les définitions 7.28 et 7.31 spécifient deux transformations inverses l'une de l'autre, comme exprimé par le lemme suivant.

**Lemme 7.32** –

**a)** *Si $\mathcal{B}$ est un automate dont tous les états sont final, alors $\mathcal{A}_{(\sigma_\mathcal{B}, g_\mathcal{B})} = \mathcal{B}$.*

**b)** *Si $(\mathbf{s}, \boldsymbol{\lambda})$ est une signature étiquetée s-morphique induite par $(\tau, h)$, alors*

$$\sigma_{\mathcal{A}_{\tau,h}} \;=\; \tau \quad \textit{et} \quad g_{\mathcal{A}_{\tau,h}} \;=\; h \;.$$

La proposition suivante constitue le point central de la démonstration du théorème V.

**Proposition 7.33** [71] – *Soit un langage $z^*L$ calable, régulier et clos par préfixe dont on note $\mathcal{M}$ l'automate minimal et émondé. On note de plus $\sigma^\omega_\mathcal{M}(\alpha) = \beta_0\, \beta_1\, \beta_2 \cdots$.*

*Pour tout entier $n$, le calcul de $\langle n \rangle_L$ atteint dans $\mathcal{M}$ l'état $\beta_n$.*

Démonstration. On note $\mathcal{M} = \langle\, \Sigma, A, \delta, \alpha, \Sigma \,\rangle$ l'automate (minimal émondé) acceptant $z^*L$ de telle sorte que $\sigma_\mathcal{M}$ est un endormophisme $\Sigma^* \to \Sigma^*$.

On considère le langage $z^*L$ comme un i-arbre étiqueté, et on définit une fonction $\varphi : \mathbb{N} \to \Sigma$ qui associe à chaque nœud $n$ de cet i-arbre un état de l'automate $\mathcal{M}$, celui atteint par $\langle n \rangle_L$. Il s'ensuit que [2]

$$\forall n, m \in \mathbb{N},\;\; \forall a \in A \qquad n \xrightarrow[z^*L]{a} m \quad \Longleftrightarrow \quad \varphi(n) \xrightarrow[\mathcal{M}]{a} \varphi(m) \;.$$

Puisque $z$ est accepté par $\mathcal{M}$, la transition $\alpha \xrightarrow{z} \beta$ existe et puisque $\mathcal{M}$ est minimal elle vérifie $\beta = \alpha$ (les mots $\varepsilon$ et $z$ ont le même langage de suffixe : $z^*L$ tout entier) ; elle est associée à la boucle $0 \xrightarrow{z} 0$ sur la racine de l'i-arbre par l'équivalence précédente. Cette transition implique également que $\sigma_\mathcal{M}(\alpha)$ commence par $\alpha$ ; en effet, Puisque $z$ est la lettre de calage, elle est strictement plus petites que les premières lettres des mots de $L$, donc c'est la plus petite lettre parmi les

2. Si on considère l'i-arbre $z^*L$ comme un automate infini, la fonction $\varphi$ est l'équivalent d'un morphisme d'automates.

étiquettes des transitions sortantes de l'état initial $\alpha$ (car $\mathcal{M}$ est émondé). Il découle donc de la définition 7.31 que $\sigma_{\mathcal{M}}(\alpha)$ commence par $\alpha$.

Soit un nœud $n \in \mathbb{N}$ ; on note ses successeurs par $m, (m+1), \ldots, m+k$. Si bien qu'il existe des lettres $a_0 < a_1 \cdots < a_k$ tels que

$$\forall i \leqslant k \qquad n \xrightarrow{a_i} (m+i) \qquad \text{donc que} \qquad \forall i \leqslant k \qquad \varphi(n) \xrightarrow{a_i} \varphi(m+i)$$

ce qui implique, d'après la définition 7.31, que

$$\sigma_{\mathcal{M}}(\varphi(n)) \;=\; \varphi(m)\;\varphi(m+1)\;\cdots\;\varphi(m+k)\;.$$

On applique le même raisonnement à chacun des nœuds $m$ à $(m+k)$. Leurs successeurs sont donc les successeurs d'ordre 2 de $n$ et forment un intervalle (propriété 7.2c) ; on les note par $m',\ (m'+1),\ \ldots,\ m'+k'$ pour certains $m'$ et $k'$ ; ils vérifient donc

$$\begin{aligned} \sigma^2_{\mathcal{M}}(\varphi(n)) \;&=\; \sigma_{\mathcal{M}}(\varphi(m))\;\sigma_{\mathcal{M}}(\varphi(m+1))\;\cdots\;\sigma_{\mathcal{M}}(\varphi(m+k)) \\ &=\; \varphi(m')\;\varphi(m'+1)\;\cdots\;\varphi(m'+k')\;. \end{aligned}$$

En itérant ce procédé, il s'ensuit que

$$\forall i \in \mathbb{N} \qquad \sigma^i_{\mathcal{M}}(\varphi(n)) \;=\; \varphi(x)\;\varphi(x+1)\;\cdots\;\varphi(x+y)$$
$$\text{où} \quad x, (x+1), \ldots, (x+y) \text{ sont les successeurs d'ordre } i \text{ de n . } (*)$$

Pour tout entier $i$, le plus petit successeur d'ordre $i$ de la racine est elle-même. Appliquer l'équation précédente à $n = 0$ donne donc :

$$\forall i \in \mathbb{N} \qquad \sigma^i_{\mathcal{M}}(\varphi(0)) \;=\; \varphi(0)\;\varphi(1)\;\cdots\;\varphi(m)$$
$$\text{où} \quad m \text{ est le plus grand nœud à la profondeur } i\;.$$

Puisque $\varphi(0)$ est l'état atteint par $\langle 0 \rangle_L = \varepsilon$, il s'agit de l'état initial : $\varphi(0) = \alpha$. Il s'ensuit que $\sigma^{\omega}_{\mathcal{M}}(\alpha)$ est $\varphi(0)\;\varphi(1)\;\varphi(2)\cdots$. □

Le théorème V′ est essentiellement une conséquence de cette proposition.

Démonstration du théorème V′. Sens direct. Soit un langage calable $z^*L$ clos par préfixe et régulier. On note $\mathcal{M} = \langle\, \Sigma, A, \delta, \alpha, \Sigma \,\rangle$ l'automate minimal émondé acceptant $z^*L$. Il découle de la proposition précédente que $\sigma^{\omega}_{\mathcal{M}}(\alpha)$ est la suite des états atteints dans $\mathcal{M}$ par les mots de $L$ pris dans l'ordre radiciel.

Soit un entier $n$ ; on note $\beta \in \sigma$ l'état atteint par $\langle n \rangle_L$ dans $\mathcal{M}$. Il découle de la définition 7.21 que $f_{\sigma_{\mathcal{M}}}(\beta) = |\sigma_{\mathcal{M}}(\beta)|$, donc en particulier que $f_{\sigma_{\mathcal{M}}}(\beta)$ est le nombre de transitions sortantes de l'état $\beta$ dans $\mathcal{M}$. Le nœud $n$ admet donc $f_{\sigma_{\mathcal{M}}}(\beta)$ arcs sortants dans $z^*L$.

Puisque d'après la proposition 7.33 $\sigma^{\omega}_{\mathcal{M}}(\alpha)$ est la suite des états atteints par $L$ dans l'ordre radiciel ; il s'ensuit que $f_{\sigma_{\mathcal{M}}}(\sigma^{\omega}_{\mathcal{M}}(\alpha))$ est bien la signature de $z^*L$. Un raisonnement analogue montre que les étiquettes des arcs sortants de $n$ forment le mot $g_{\sigma_{\mathcal{M}}}(\beta)$, ce qui implique que l'étiquetage de $z^*L$ est bien $g_{\sigma_{\mathcal{M}}}(\sigma^{\omega}_{\mathcal{M}}(\alpha))$.

Sens réciproque. Soit une signature étiquetée $(\mathbf{s}, \boldsymbol{\lambda})$ s-morphique définie par $(\tau, h)$. Le langage calable accepté par $\mathcal{A}_{(\tau,h)}$ est régulier donc d'après le sens réciproque il admet une signature étiqueté s-morphique définie par $(\sigma_{\mathcal{A}_{(\tau,h)}}, g_{\mathcal{A}_{(\tau,h)}})$ qui est égale à $(\tau, h)$ d'après le lemme 7.32. Puisque la signature étiquetée est caractéristique du langage calable, $(\mathbf{s}, \boldsymbol{\lambda})$ engendre le langage calable accepté par $\mathcal{M}_{(\tau,h)}$. □

## Systèmes de numération morphiques

Dans cette section, nous utilisons le formalisme des signatures pour redéfinir les *systèmes de numération morphiques* définis dans la série d'articles [30, 31, 32] ; ils sont aussi appelés *systèmes de numération de Dumont-Thomas* en hommage à leurs auteurs.

Dans la suite, on manipule un alphabet $B_\sigma$ dont les *lettres sont des mots* appartenant à autre monoïde libre $\Sigma^*$. Pour éviter les confusions, si un mot de $\Sigma^*$ est noté $u$ ou $\alpha_0\,\alpha_1\,\cdots\,\alpha_k$, la lettre correspondante de $B_\sigma$ sera noté $[w]$ ou $[\alpha_0\,\alpha_1\,\cdots\,\alpha_k]$. Par exemple, les mots $[\varepsilon]$ et $[\varepsilon]\,[\varepsilon]$ sont deux mots distincts de ${B_\sigma}^*$ ; aucun des deux n'est égal au mot vide de ${B_\sigma}^*$ (que l'on note toujours $\varepsilon$).

Soit un endomorphisme $\sigma : \Sigma^* \to \Sigma^*$ prolongeable en $\alpha$. On note $B_\sigma$ l'alphabet de mots composé de tous les préfixes stricts des images par $\sigma$ des lettres de $\Sigma$ :

$$B_\sigma \;=\; \{\; [w] \mid u \sqsubset \sigma(\beta) \quad \text{avec} \quad \beta \in \Sigma \;\} \tag{7.5}$$

On définit de plus le morphisme $g_\sigma : \Sigma^* \to {B_\sigma}^*$ qui, intuitivement, imite $\sigma$ en bégayant. Soit une lettre $\beta \in \Sigma$ ; on note $\sigma(\beta) = \gamma_0\,\gamma_1\,\cdots\,\gamma_k$. Alors $g_\sigma(\beta)$ est le mot de longueur $k$ dont la $(i+1)$-ème lettre est $[\gamma_0\,\gamma_1\,\cdots\,\gamma_{(i-1)}]$, le préfixe de $\sigma(\beta)$ de longueur $i$ :

$$\forall \beta \in \Sigma \qquad g_\sigma(\beta) \;=\; [\varepsilon]\,[\gamma_0]\,[\gamma_0\,\gamma_1]\,\cdots\,[\gamma_0\,\gamma_1\,\cdots\,\gamma_{k-1}] \quad \text{si } \sigma(\beta) = \gamma_0\,\gamma_1\,\cdots\,\gamma_k\ . \tag{7.6}$$

Notez que le mot $g_\sigma(\beta)$ ne contient pas la lettre $[\gamma_0\,\gamma_1\,\cdots\,\gamma_k] = [\sigma(\beta)]$ (qui, a priori, n'appartient même pas à $B_\sigma$) ; en particulier, si $\sigma(\beta)$ est le mot vide alors $g_\sigma(\beta)$ est aussi égal au mot vide $\varepsilon$ (et n'est pas égal au mot non-vide $[\varepsilon]$)

L'automate $\mathcal{A}_{\sigma,g_\sigma}$ (*cf.* définition 7.28), noté plus simplement $\mathcal{A}_\sigma$ dans la suite, est alors appelé *l'automate des préfixes (associé à $\sigma$)*. Cet automate est implicitement présent dès l'article original [30] à travers la notion de *suite admissible* qui simule le calcul d'un automate, puis formalisé comme automate dans un article ultérieur [32]. La caractérisation suivante des transitions de $\mathcal{A}_\sigma$ correspond à cette définition originelle.

**Lemme 7.34** – *Soit un endomorphisme $\sigma : \Sigma^* \to \Sigma^*$ prolongeable auquel est associé l'automate des préfixes $\mathcal{A}_\sigma$. La transition $\beta \xrightarrow{[w]} \gamma$ existe dans $\mathcal{A}_\sigma$ si et seulement si $w\,\gamma$ est un préfixe de $\sigma(\beta)$.*

Démonstration. Il découle de la définition 7.28 que la transition $\beta \xrightarrow{[w]} \gamma$ existe dans $\mathcal{A}_\sigma$ si et seulement si il existe un indice $i$ tel que $[w]$ est la $(i+1)$-ème lettre de $g_\sigma(\beta)$ et $\gamma$ est la $(i+1)$-ème lettre de $\sigma(b)$.

Sens direct. Puisque $[w]$ est la $(i+1)$-ème lettre de $g_\sigma(\beta)$, $w$ est le préfixe de longueur $i$ de $\sigma(b)$. Donc puisque $\gamma$ est la $(i+1)$-ème lettre de $\sigma(b)$, $w\,\gamma$ est le préfixe de longueur $(i+1)$ de $\sigma(b)$.

[30] Jean-Marie Dumont et Alain Thomas, 1989, *Systèmes de Numération et Fonctions Fractales Relatifs aux Substitutions.*

[31] Jean-Marie Dumont et Alain Thomas, 1991, *Digital sum problems and substitutions on a finite alphabet.*

[32] Jean-Marie Dumont et Alain Thomas, 1993, *Digital sum moments and substitutions.*

Sens réciproque. Soit $w\,\gamma$ un préfixe de $\sigma(\beta)$. On note $i = |w|$ ce qui implique déjà que $\gamma$ est la $(i+1)$-ème lettre de $\sigma(\beta)$. De plus $w$ est alors un préfixe strict de $\sigma(b)$ donc $[w]$ apparaît dans $g_\sigma(\beta)$ et puisque $i = |w|$, il en est la $(i+1)$-ème lettre. D'après la définition 7.28, la transition $\beta \xrightarrow{[w]} \gamma$ existe donc dans $\mathcal{A}_\sigma = (\mathcal{A}_{(\sigma,g_\sigma)})$. □

On note $\rho_\sigma$ la fonction ${B_\sigma}^* \to \Sigma^*$ définie par :

$$\rho_\sigma\left([w_k]\cdots[w_2]\,[w_1]\,[w_0]\right) \;=\; \sigma^k(w_k)\cdots\sigma^2(w_2)\,\sigma(w_1)\,w_0\;. \tag{7.7}$$

Alternativement, cette fonction $\rho_\sigma$ peut être définie récursivement par la formule

$$\forall u \in {B_\sigma}^*,\;\; \forall w \in \Sigma^* \qquad \rho_\sigma(u\,[w]) \;=\; \sigma(\rho_\sigma(u))\,w\;. \tag{7.8}$$

Le théorème suivant est l'un des résultats principaux de [30] et définit les systèmes de numération morphiques.

**Théorème 7.35** [30] – *Soit un endomorphisme $\sigma : \Sigma^* \to \Sigma^*$ prolongeable et un entier $n$. Il existe un unique mot $u \in {B_\sigma}^*$ qui vérifie les trois conditions suivantes :*

**a)** *$u$ ne commence pas par la lettre $[\varepsilon]$,*
**b)** *$u$ est accepté par $\mathcal{A}_\sigma$*
**c)** *et $\rho_\sigma(u)$ est un mot de longueur $n$.*

Le mot $u$ associé à l'entier $n$ dont il est question dans le théorème 7.35 est alors la représentation de $n$ dans le système de numération morphique $\sigma$ : il est noté $\langle n\rangle_\sigma = u$. La proposition suivante est démontrée dans le même article.

**Proposition 7.36** [30] – *Soit un endomorphisme $\sigma : \Sigma^* \to \Sigma^*$ prolongeable en $\alpha$ et un entier $n$. Pour tout entier $n$, $\rho_\sigma(\langle n\rangle_\sigma)$ est le préfixe de longueur $n$ de $\sigma^\omega(\alpha)$.*

Le théorème 7.35 implique que le langage $L_\sigma$ des représentations des entiers dans le système de numération morphique $\sigma$ est régulier ; l'automate des préfixe $\mathcal{A}_\sigma$ acceptant le langage calable $[\varepsilon]^*L_\sigma$. En revanche, ce théorème ne suffit pas à démontrer que ce système de numération est un système de numération abstrait. En effet rien n'implique, a priori, que $\langle n\rangle_\sigma <_{\text{rad}} \langle m\rangle_\sigma$ pour tous entiers $n$ et $m\mathbb{N}$ tels que $n < m$. Ce résultat, énoncé ci-dessou, a été démontré par Berthé et Rigo dans [14].

**Théorème 7.37** [14] – *Les systèmes de numération morphiques sont des systèmes de numération abstrait réguliers.*

Nous allons dans la suite redémontrer les théorèmes 7.35 et 7.37 en prenant le problème à l'envers. Nous considérons le langage calable $[\varepsilon]^*L_\sigma$, accepté par l'automate $\mathcal{A}_\sigma$, comme un SNAR. La proposition suivante montre qu'il satisfait la condition correspondante à la proposition 7.36, exprimé à la proposition 7.38.

**Proposition 7.38** – *Soit un entier $m$. Alors $\rho_\sigma\left(\langle m\rangle_{L_\sigma}\right)$ est le préfixe de longueur $m$ de $\sigma^\omega(\alpha)$.*

[14] Valérie Berthé et Michel Rigo, 2007, *Odometers on Regular Languages.*

Démonstration. On note $\sigma^\omega(\alpha) = \gamma_0\gamma_1\gamma_2\cdots$ et il découle de la proposition 7.33 que pour tout entier $i$, $\gamma_i$ est l'état atteint par $\langle i\rangle_{L_\sigma}$ dans $\mathcal{A}_\sigma$.

Démontrons par récurrence que $\rho_\sigma(\langle m\rangle_{L_\sigma}) = \gamma_0\gamma_1\cdots\gamma_{m-1}$ ; c'est le cas pour $\rho_\sigma(\langle 0\rangle_{L_\sigma}) = \rho_\sigma(\varepsilon) = \varepsilon$. On suppose dans la suite que $m > 0$ et que la proposition est vérifiée pour tout entier $j < m$.

*Cas 1* : la dernière lettre de $\langle m\rangle_{L_\sigma}$ est $[\varepsilon]$. On note $n, m, k$ les entiers et les mots $u, v \in L_\sigma$, $w \in \Sigma^*$ tels que

$$0 \xrightarrow[L_\sigma]{u} n \xrightarrow[L_\sigma]{[\varepsilon]} m \quad \text{et} \quad 0 \xrightarrow[L_\sigma]{v} (n-k) \xrightarrow[L_\sigma]{[w]} (m-1)\ . \tag{$*$}$$

Puisque $(m-1)$ et $m$ sont deux entiers consécutifs,

- $(n-k) \xrightarrow[L_\sigma]{[w]} (m-1)$ est le plus grand arc sortant de $(n-k)$ donc $[w]$ est la plus grande étiquette sortante du nœud $(n-k)$ ; en d'autres termes $w$ est le plus grand préfixe strict de $\sigma(\gamma_{n-k})$, il découle donc du lemme 7.34
$$\sigma(\gamma_{n-k}) \;=\; w\,\gamma_{(m-1)}\,; \tag{$**$}$$
- pour tout entier $i$, $(n-k) < i < n$, l'état $\gamma_i$ n'a pas de transitions sortantes dans $\mathcal{A}_\sigma$, ce qui implique que
$$\forall\ i \in \mathbb{N},\ \ (n-k) < i < n, \qquad \sigma(\gamma_i) \;=\; \varepsilon\ . \tag{$\overset{*}{**}$}$$

Il découle de $(*)$ que $\langle n-k\rangle_{L_\sigma} = v$, que $\langle m-1\rangle_{L_\sigma} = v\,[w]$ et que $\langle m\rangle_{L_\sigma} = u\,[\varepsilon]$. Appliquer l'hypothèse de récurrence successivement à $n$ et $(n-k)$ et $(m-1)$ donne la suite de calcul suivante qui conclut le cas 1 :

$$\begin{array}{rcll}
\rho_\sigma(\langle m\rangle_{L_\sigma}) & = & \rho_\sigma(u\,[\varepsilon]) \;=\; \sigma(\rho_\sigma(u)) \;=\; \sigma(\rho_\sigma(\langle n\rangle_{L_\sigma})) & \\
& = & \sigma(\gamma_0\gamma_1\cdots\gamma_{(n-1)}) & \text{(HR)} \\
& = & \sigma(\underbrace{\gamma_0\gamma_1\cdots\gamma_{(n-k-1)}}_{=\rho_\sigma(\langle n-k\rangle_{L_\sigma})})\,\underbrace{\sigma(\gamma_{(n-k)})}_{=w\,\gamma_{(m-1)}}\,\underbrace{\sigma(\gamma_{(n-k+1)}\cdots\gamma_{(n-1)})}_{=\varepsilon} & \text{(HR), $(**)$, $(\overset{*}{**})$} \\
& = & \sigma(\rho_\sigma(\langle n-k\rangle_{L_\sigma}))\,w\,\gamma_{(m-1)} & \\
& = & \sigma(\rho_\sigma(v))\,w\,\gamma_{(m-1)} & \text{(eq. (7.8))} \\
& = & \rho_\sigma(v\,[w])\ \ \gamma_{(m-1)} & \\
& = & \rho_\sigma(\langle m-1\rangle_{L_\sigma})\,\gamma_{(m-1)} & \text{(HR)} \\
& = & \gamma_0\gamma_1\cdots\gamma_{(m-2)}\,\gamma_{(m-1)} &
\end{array}$$

*Cas 2* : la dernière lettre de $\langle m\rangle_{L_\sigma}$ est de la forme $[w\,\beta]$, pour certains $w \in \Sigma^*$ et $\beta \in \Sigma$.

On note $n \in \mathbb{N}$ le nœud et $u \in B_\sigma^*$ le mot tels que

$$0 \xrightarrow[L_\sigma]{u} n \xrightarrow[L_\sigma]{[w\,\beta]} m \quad \text{donc que} \quad \gamma_n \xrightarrow[\mathcal{A}_\sigma]{[w\,\beta]} \gamma_m\ .$$

Il découle du lemme 7.34 que $(w\,\beta\,\gamma_m)$ est un préfixe de $\sigma(\gamma_n)$ ; il s'ensuit que $(w\,\beta)$ est aussi un préfixe $\sigma(\gamma_n)$, donc (d'après le même lemme) que $\mathcal{A}_\sigma$ possède la transition

$$\gamma_n \xrightarrow[\mathcal{A}_\sigma]{[w]} \beta \quad \text{, ce qui implique que} \quad 0 \xrightarrow[L_\sigma]{u} n \xrightarrow[L_\sigma]{[w]} m'$$

pour un certain $m'$. Or les mots $(u\,[w])$ et $(u\,[w\,\beta])$ sont deux mots consécutifs dans l'ordre radiciel donc $m' = (m-1)$ et $\gamma_{(m-1)} = \beta$.

D'après l'hypothèse de récurrence, $\rho_\sigma(u\,[w]) = \rho_\sigma(\langle m-1\rangle_{L_\sigma}) = \gamma_0\,\gamma_1\cdots\gamma_{(m-2)}$, donc

$$\begin{aligned}\rho_\sigma(\langle m\rangle_{L_\sigma}) = \rho_\sigma(u\,[w\,\beta]) = \rho_\sigma(u\,[w\,\gamma_{(m-1)}]) &= \rho_\sigma(u\,[w])\,\gamma_{(m-1)}\\ &= \gamma_0\,\gamma_1\cdots\gamma_{(m-2)}\,\gamma_{(m-1)}\ .\end{aligned}$$ □

Les deux corollaires suivants de la proposition 7.38 sont des reformulations des théorèmes 7.35 et 7.37.

**Corollaire 7.39** – *Soit $\sigma : \Sigma \to \Sigma$ un morphisme prolongeable. Pour tout entier $n$, $\langle n\rangle_{L_\sigma}$ est l'unique mot $u \in B_\sigma^*$ tel que*

**a)** *$u$ ne commence pas par la lettre $[\varepsilon]$,*

**b)** *$u$ est accepté par $\mathcal{A}_\sigma$*

**c)** *et $\rho_\sigma(u)$ est un mot de longueur $n$.*

Démonstration. Le mot $\langle n\rangle_{L_\sigma}$ appartient à $L_\sigma$ donc vérifie les conditions (**a**) et (**b**) ; il découle de la proposition 7.38 que $\rho_\sigma(\langle n\rangle_{L_\sigma})$ est le préfixe de longueur $n$ de $\sigma^\omega(\alpha)$ donc que $\langle n\rangle_{L_\sigma}$ satisfait la condition (**c**).

Réciproquement, tout mot $u$ qui satisfait les conditions (**a**) et (**b**) appartient à $L_\sigma$ donc est égal à $u = \langle m\rangle_{L_\sigma}$ pour un certain $m$. Il découle donc de la proposition 7.38 que $\rho_\sigma(u)$ est de longueur $m$ donc la condition (**c**) implique que $m = n$, donc $u = \langle n\rangle_{L_\sigma}$. □

**Corollaire 7.40** – *Le système de numération abstrait régulier $[\varepsilon]^*L_\sigma$ et le système de numération morphique $\sigma$ sont identiques.*

Chapitre 8

# Signatures périodiques

Ce chapitre poursuit l'étude des signatures et étiquetages, il s'appuie donc sur les notations, les définitions et la terminologie de la section 7.1, page 174. Nous traitons le cas des signatures (purement) périodiques $\mathbf{s} = \mathbf{r}^\omega$ ; dans ce cas $\mathbf{r}$ est appelé *un rythme*, c'est-à-dire une suite finie d'entiers. Les rythmes seront couplés avec des étiquetages périodiques $\boldsymbol{\lambda} = \boldsymbol{\gamma}^\omega$ dont la période $\boldsymbol{\gamma}$ sera également appelée *étiquetage* par abus de langage. Le langage engendré par le rythme $\mathbf{r}$ et l'étiquetage $\boldsymbol{\gamma}$ est défini comme le langage $L_{(\mathbf{s},\boldsymbol{\lambda})}$ engendré par la signature étiquetée $(\mathbf{s}, \boldsymbol{\lambda}) = (\mathbf{r}^\omega, \boldsymbol{\gamma}^\omega)$.

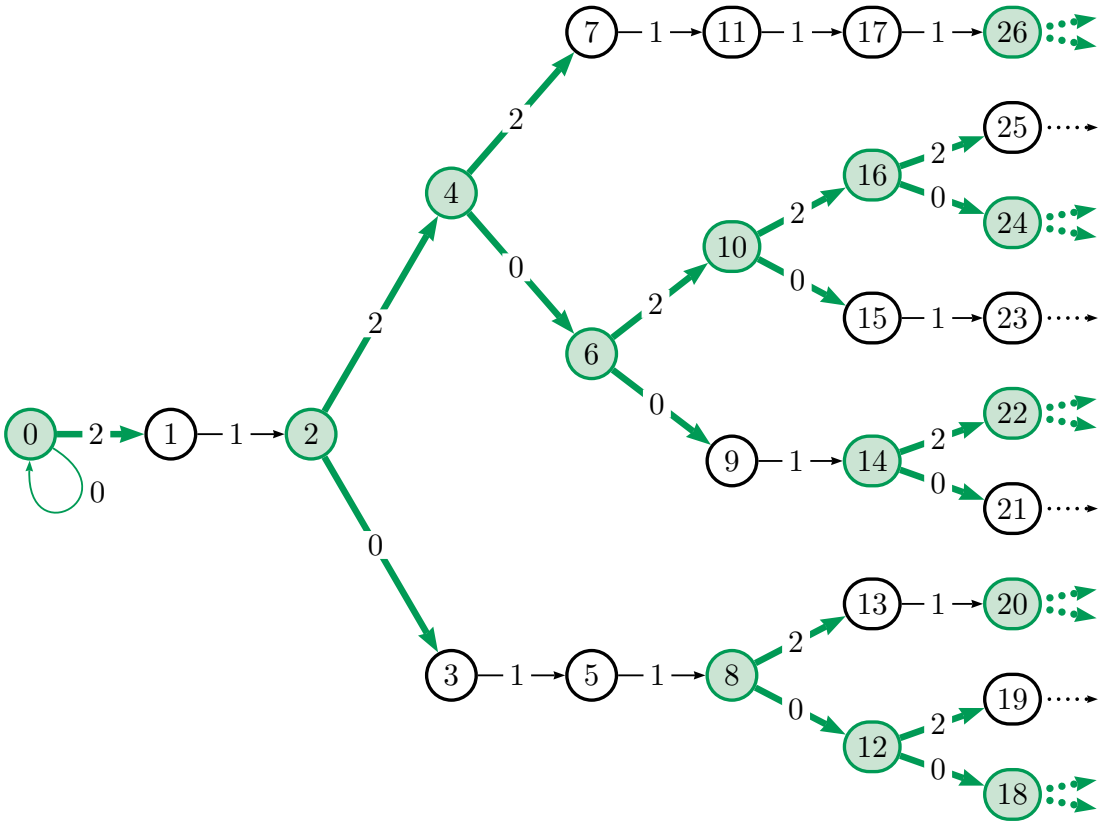



Figure 1 – Langage dont le rythme est 21 et l'étiquetage est 021

Les résultats présentés dans ce chapitre sont à paraître dans les actes de LATIN 2016, voir [56].

Considérons par exemple le langage calable $0^*L_{\frac{3}{2}}$ représenté par la figure 1. Les nœuds pairs y sont représentés en traits verts et gras ; chacun d'entre eux admet deux transitions sortantes, étiquetées respectivement par 0 et par 2. Les autres nœuds sont impairs et chacun admet une seule transition sortante étiquetée par 1. Ce langage est donc engendré par le rythme $\mathbf{r} = 21$ et l'étiquetage $\boldsymbol{\gamma} = 201$.

Puisque le langage $L_{\frac{3}{2}}$ est FLIP (Finite Left Iteration Property, voir section 5.1, page 117), les outils classiques de la théorie des langages formels ne semblent pas adaptés à le décrire ; il est donc remarquable qu'on puisse le construire de façon périodique (en utilisant la génération par signature présentée dans le chapitre précédent page 178). Nous verrons dans ce chapitre que ce n'est pas un cas isolé et que les systèmes de numération à base rationnelle sont liés aux signatures périodiques. A vrai dire, c'est la découverte de cette propriété des bases rationnelles qui nous a conduit à définir la notion plus générale de signature.

Le *paramètre directeur* d'un rythme $\mathbf{r}$ est le couple $(q, p)$ si $\mathbf{r}$ est de longueur $q$, et que la somme de ses composantes est égale à $p$. À ce rythme $\mathbf{r}$ est également associé un chemin dans le plan discret $\mathbb{Z} \times \mathbb{Z}$ allant de $(0, 0)$ à $(q, p)$. Dans le cas général, $p$ et $q$ ne sont pas nécessairement premiers entre eux ; on note $p'$ et $q'$ leurs quotients respectifs par leur PGCD, si bien que $\frac{p'}{q'}$, appelé le *taux de croissance* de $\mathbf{r}$, est la fraction irréductible égale à $\frac{p}{q}$.

Soient $p$ et $q$ deux entiers premiers entre eux tels que $p > q \geqslant 1$. Le mot de Christoffel de pente $\frac{p}{q}$ code justement un chemin dans le plan discret, le plus petit chemin allant de $(0, 0)$ à $(q, p)$ et qui reste au dessus de la droite de pente $\frac{p}{q}$ passant par l'origine. Ces mots sont étudiés en combinatoire des mots (voir par exemple [13]) pour leur nombreuses propriétés. Les rythmes associés à ces chemins sont appelés *rythmes de Christoffel* et nous montrons dans la section 8.2 qu'ils engendrent les langages de représentations des entiers dans les différentes bases rationnelles.

**Théorème VI** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q \geqslant 1$.*

**a)** *La signature du langage $L_{\frac{p}{q}}$ est $\mathbf{r}_{\frac{p}{q}}^{\omega}$, où $\mathbf{r}_{\frac{p}{q}}$ est le rythme dont le chemin dans $\mathbb{Z} \times \mathbb{Z}$ est le mot de Christoffel de pente $\frac{p}{q}$.*

**b)** *L'étiquetage du langage $L_{\frac{p}{q}}$ est $\boldsymbol{\gamma}_{\frac{p}{q}}^{\omega}$, où $\boldsymbol{\gamma}_{\frac{p}{q}}$ est la suite résultant de la génération de $\mathbb{Z}/p\mathbb{Z}$ par $q$ : $\boldsymbol{\gamma}_{\frac{p}{q}} = 0\,q\,(2\,q\,\%\,p) \cdots ((p-1)\,q\,\%\,p)$.*

Le rythme de Christoffel de pente $\frac{p}{q}$ est un rythme de paramètre directeur $(q, p)$ (puisqu'il code un chemin de l'origine à $(q, p)$) mais aussi de taux de croissance $\frac{p}{q}$ (puisqu'il n'est défini que dans le cas où $p$ et $q$ sont premiers entre eux). Le rythme qui engendre $L_{\frac{p}{q}}$ a donc un taux de croissance égal à $\frac{p}{q}$.

Soit maintenant un rythme $\mathbf{r}$ quelconque ; on note $(q, p)$ son paramètre directeur et $\frac{p'}{q'}$ son taux de croissance. On lui associe ensuite un *étiquetage* $\boldsymbol{\gamma}_{\mathbf{r}}$ qui imite l'étiquetage de $L_{\frac{p'}{q'}}$. Nous montrons dans la section 8.3 que le langage engendré par $\mathbf{r}$ et $\boldsymbol{\gamma}_{\mathbf{r}}$ est alors fortement lié au système de numération en base $\frac{p'}{q'}$.

[13] Jean Berstel, Aaron Lauve, Christophe Reutenauer et Franco Saliola, 2008, *Combinatorics on Words : Christoffel Words and Repetition in Words*.

**Théorème VII** – *Soit un rythme* $\mathbf{r}$ *de taux de croissance* $\frac{p'}{q'}$*. Le langage engendré par* $\mathbf{r}$ *et l'étiquetage spécial associé* $\boldsymbol{\gamma}_{\mathbf{r}}$ *est une représentation non-canonique des entiers en base* $\frac{p'}{q'}$*.*

Une *représentation non-canonique des entiers en base* $\frac{p'}{q'}$ est un langage sur un alphabet non-canonique qui contient exactement un mot de chaque valeur entière.

Dans un système de numération, la *normalisation* est la fonction qui associe chaque mot sur un alphabet non-canonique au mot de même valeur sur l'alphabet canonique. En base rationnelle, si l'alphabet non-canonique d'entrée est fini, la normalisation est réalisée par un transducteur droit, lettre-à-lettre et séquentiel, appelé normalisateur (voir [2] ou section 4.3, page 100). En particulier, $L_{\frac{p'}{q'}}$ est l'image par le normalisateur du langage engendré par $\mathbf{r}$ et $\boldsymbol{\gamma}_{\mathbf{r}}$. Ces deux langages sont donc aussi complexes l'un que l'autre (ou aussi simples dans le cas dégénéré où $q' = 1$ donc où $\frac{p'}{q'}$ est un entier).

Il s'avère que la démonstration du théorème VII n'utilise pas directement le fait que la signature considérée est périodique mais uniquement les propriétés de l'étiquetage spécial. Suite à cette observation, le théorème VII est généralisé aux signatures ultimement périodiques et même à des signatures dites *dirigées par* $\frac{p}{q}$, c'est-à-dire dont le chemin dans le plan $\mathbb{Z} \times \mathbb{Z}$ est confiné entre deux droites (parallèles) de pente $\frac{p}{q}$. Il s'agit du cas où la définition de l'étiquetage spécial se généralise sur un alphabet fini.

# Arbre et langage engendrés par un rythme

## Les rythmes et leurs représentations géométriques

**Définition 8.1** – *Soient deux entiers* $p$ *et* $q$*.*

**a)** *On appelle rythme de paramètre directeur* $(q, p)$*, un* $q$*-uplet* $\mathbf{r}$ *d'entiers dont la somme vaut* $p$ *:*

$$\mathbf{r} \;=\; (r_0, r_1, \ldots, r_{q-1}) \quad \textit{et} \quad \sum_{i=0}^{q-1} \;=\; p \;.$$

*On considérera généralement qu'un rythme est un mot sur un alphabet d'entiers donc noté :* $\mathbf{r} = r_0\, r_1 \cdots r_{q-1}$*.*

**b)** *Un rythme* $\mathbf{r}$ *est dit* valide *s'il satisfait l'équation suivante :*

$$\forall j,\;\; 0 \leqslant j < q, \quad \sum_{i=0}^{j} r_i \;>\; (j+1) \;.$$

**c)** *On appelle* taux de croissance *la fraction irréductible* $\frac{p'}{q'}$ *égale à* $\frac{p}{q}$*.*

[2] Shigeki Akiyama, Christiane Frougny et Jacques Sakarovitch, 2008, *Powers of rationals modulo 1 and rational base number systems.*

**Exemple 8.2** – *Voici quelques rythmes dont le taux de croissance est* $\frac{5}{3}$ *:*

$$221\,,\ 302\,,\ 122\,,\ 221221 \quad \textit{et} \quad 213004\ .$$

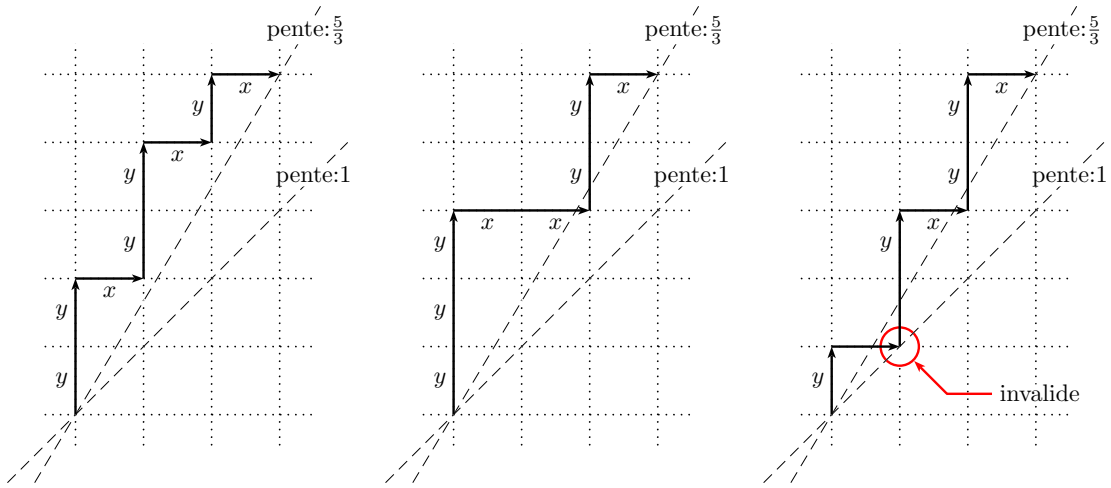
*Ils sont tous valides à l'exception du troisième* $(122)$*. Le paramètre directeur des trois premiers est* $(3,5)$ *et celui des deux derniers est* $(6,10)$*.*

On donne à chaque rythme une représentation géométrique en tant que chemin dans le plan discret $\mathbb{Z}\times\mathbb{Z}$. Ces chemins sont codés par des mots de $\{x,y\}^*$ où $x$ code un segment horizontal unitaire et $y$ un segment vertical unitaire. Chaque mot $u$ de $\{x,y\}^*$ représente donc un chemin de $(0,0)$ à $(q,p)$ où $q$ et $p$ sont respectivement le nombre de $x$ et de $y$ dans $u$.

**Définition 8.3** – *Soit un rythme* $\mathbf{r} = r_0\, r_1 \cdots r_{q-1}$ *de paramètre directeur* $(q,p)$*. On associe à* $\mathbf{r}$ *le mot* $\mathsf{path}(\mathbf{r}) \in \{x,y\}^*$ *défini par :*

$$\mathsf{path}(\mathbf{r}) \;=\; \mathsf{path}(r_0\, r_1 \cdots r_{q-1}) \;=\; y^{r_0}\, x\, y^{r_1}\, x\, y^{r_{q-1}} x\ .$$

*Ce mot contient* $q$ *fois la lettre* $x$ *et* $p$ *fois la lettre* $y$ *donc code un chemin allant de* $(0,0)$ *à* $(q,p)$ *dans le plan discret* $\mathbb{Z}\times\mathbb{Z}$*.*




(a) 221 (b) 302 (c) 122

Figure 2 – Chemins dans le plan associés à trois rythmes de paramètre directeur $(3,5)$

Par exemple, les trois rythmes de paramètre directeur $(3,5)$ de l'exemple 8.2 sont associés aux trois chemins (représentés à la figure 2) allant de $(0,0)$ à $(3,5)$ codés par les mots

$$\begin{aligned}
\mathsf{path}(221) &= y^2\, x\, y^2\, x\, y\, x \;=\; y\, y\, x\, y\, y\, x\, y\, x\ ,\\
\mathsf{path}(302) &= y^3\, x\, y^0\, x\, y^2\, x \;=\; y\, y\, y\, x\, x\, y\, y\, x\ ,\\
\mathsf{path}(122) &= y^2\, x\, y^2\, x\, y^1\, x \;=\; y\, x\, y\, y\, x\, y\, y\, x\ .
\end{aligned}$$

La validité d'un rythme $\mathbf{r}$ coïncide avec la validité de la signature $\mathbf{r}^\omega$ et admet une caractérisation géométrique donnée par le lemme suivant et l'illustré par la figure 2c.

**Lemme 8.4** – *Soit un rythme* $\mathbf{r}$*. Les trois énoncés suivants sont équivalents.*

**a)** *Le rythme* $\mathbf{r}$ *est valide.*

**b)** *La signature* $\mathbf{r}^\omega$ *est valide.*

**c)** *En dehors du point* $(0,0)$*, le chemin* $\mathsf{path}(\mathbf{r})$ *se trouve strictement au dessus de la droite de pente 1 passant par l'origine.*

Dans la suite, on considère implicitement que tous les rythmes sont valides. Ceci implique en particulier que si un rythme est de paramètre directeur $(q,p)$ alors $p > q$ et son taux de croissance $\frac{p'}{q'}$ est strictement supérieur à 1.

Si un rythme est noté $\mathbf{r} = r_0\, r_1 \cdots r_{(q-1)}$, on pourra écrire $r_i$ même si $i \geqslant q$, avec le sens implicite que $r_i = r_{(i \,\%\, q)}$. On adapte de plus la notation des signatures pour les sommes partielles : pour tout entier $i$, on note $R_i$ la somme des $i$ premiers termes de $\mathbf{r}^\omega$ :

$$\forall i \in \mathbb{N} \qquad R_i \;=\; \sum_{j=0}^{i-1} r_j \;.$$

Si de plus $i > 0$, alors $R_i = R_{(i-1)} + r_{(i-1)}$.

## L'(i-)arbre engendré par un rythme

L'i-arbre engendré par un rythme $\mathbf{r}$, noté $T_{\mathbf{r}}$, est l'i-arbre engendré par la signature $\mathbf{r}^\omega$, comme décrit dans le chapitre 7. Par exemple, la figure 3 (page 179) montre la génération de l'arbre par la signature $(3\,2\,1)^\omega$ donc par le rythme $3\,2\,1$. D'autres exemples sont donnés à la figure 3, qui représente l'i-arbre $T_{311}$ et à la figure 4 (plus loin, page 203) qui représente un langage calable dont l'i-arbre sous-jacent est $T_{302}$.

Le lemme suivant explique pourquoi $\frac{p'}{q'}$, la fraction réduite de $\frac{p}{q}$, est appelé 'taux de croissance' d'un rythme dont le paramètre directeur est $(q,p)$ : le nombre de successeurs de $k$ nœuds consécutifs est environ $k\frac{p'}{q'}$.

**Lemme 8.5** – *Soit un rythme* $\mathbf{r}$ *de paramètre directeur* $(q,p)$*. Pour tous nœuds* $n, m$ *de* $T_{\mathbf{r}}$*, l'arc* $n \longrightarrow m$ *existe si et seulement si l'arc* $(n+q) \longrightarrow (m+p)$ *existe.*

Démonstration. Sens direct. L'assertion suivante énonce un résultat plus fort dans un cas particulier, qui permet de démontrer le lemme dans le cas général.

**Assertion 8.5.1** – *Si* $m$ *est le plus petit successeur de* $n$*, alors* $(m+p)$ *est le plus petit successeur de* $(n+q)$*.*

> *Démonstration de l'assertion.* Les nœuds $n$ et $(n+q)$ ont le même nombre de successeurs donc, puisque $n$ a un plus petit successeur $m$, c'est aussi le cas de $(n+q)$. Il découle du lemme 7.6 (page 178) que $m = R_n$ et que le plus petit successeur de $(n+q)$ est $R_{(n+q)}$. Or $\mathbf{r}$ est de paramètre directeur $(q,p)$ donc toutes les composantes de $\mathbf{r}$ apparaissent dans la somme $r_n + r_{(n+1)}, \ldots, r_{(n+q-1)}$ qui vaut donc $p$. Si bien que $R_{(n+q)} = R_n + p = m + p$.

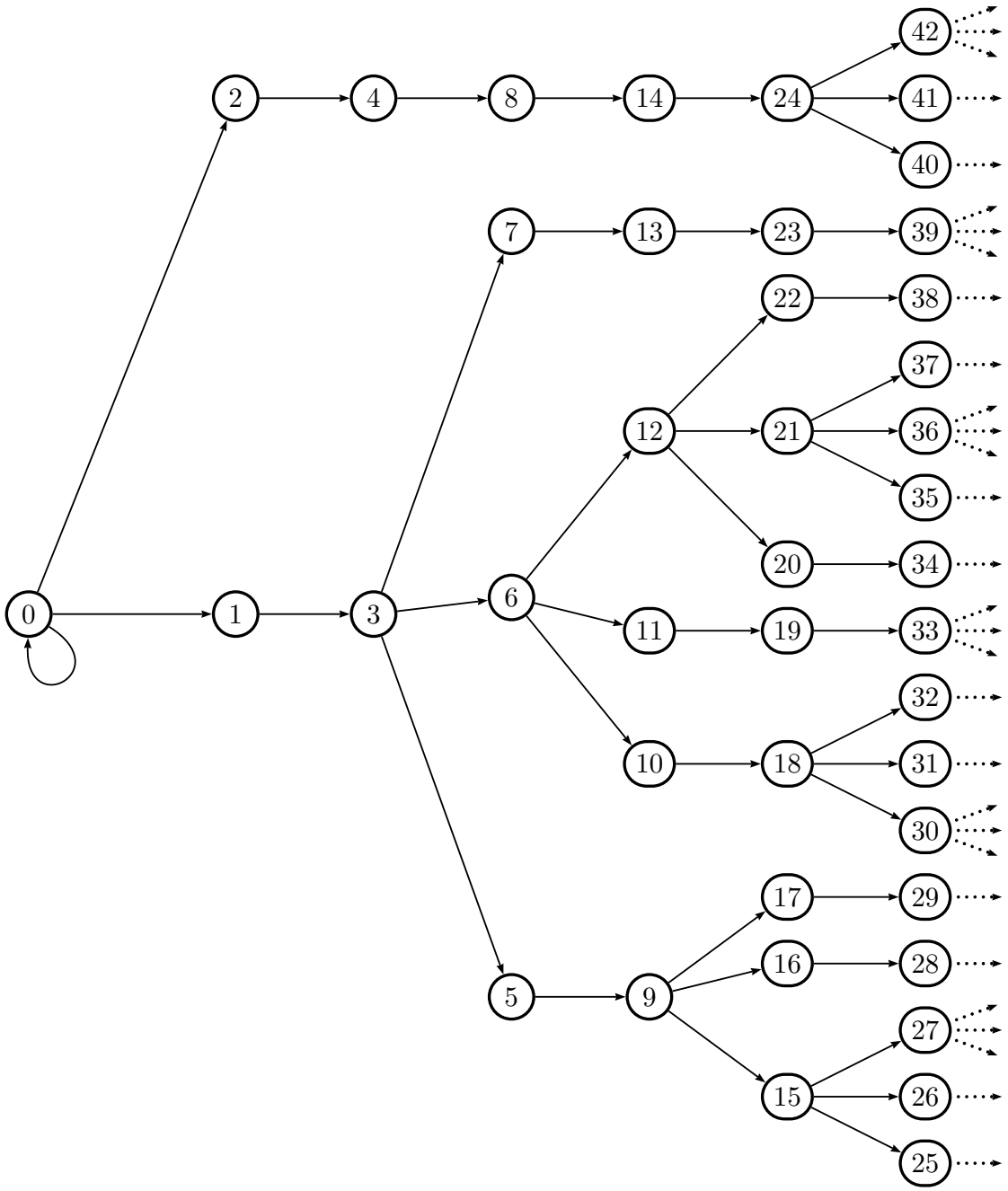



Figure 3 – L'arbre $T_{311}$

Dans le cas général, le plus petit successeur de $n$ n'est pas $m$, mais le nœud $(m-i)$, pour un certain entier $i$. Il découle donc de l'assertion précédente que $(m-i+p)$ est le plus petit successeur de $(n+q)$. Puisque $n$ et $(n+q)$ sont congrus modulo $q$, ils ont dans $T_{\mathbf{r}}$ le même nombre d'arcs sortants donc $(m+p)$ est un successeur de $(n+q)$.

Sens réciproque. On suppose que l'arc $(n+q) \longrightarrow (m+p)$ existe. On note $k$ le prédécesseur de $m$ dans $T_{\mathbf{r}}$. Appliquer le sens direct donne que $(k+q) \longrightarrow (m+p)$ est un arc de $T_{\mathbf{r}}$, or chaque nœud de $T_{\mathbf{r}}$ a exactement un prédécesseur donc $k = n$. □

## Le langage engendré par un rythme et un étiquetage

Soit un rythme $\mathbf{r} = r_0 r_1 \cdots r_{(q-1)}$ de paramètre directeur $(q, p)$ qui définit la signature $\mathbf{s} = \mathbf{r}^\omega$.

Un étiquetage cohérent avec $\mathbf{s}$ n'est pas nécessairement périodique. Prenons par exemple le rythme $\mathbf{r} = 2\,1$ et considérons les mots $\boldsymbol{\gamma}_1 = 0\,1\,2$ et $\boldsymbol{\gamma}_2 = 0\,2\,1$. N'importe quel mot de $\{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2\}^\omega$ est un étiquetage cohérent avec $\mathbf{s}$ mais n'est évidemment pas nécessairement périodique. Dans la suite, on ne va cependant considérer que des étiquetages purement périodiques $\boldsymbol{\lambda} = \boldsymbol{\gamma}^\omega$ où $\boldsymbol{\gamma} = \gamma_0\gamma_1 \cdots \gamma_{(p-1)}$ est de longueur $p$.

Dans ce cas, le couple $(\mathbf{r}^\omega, \boldsymbol{\gamma}^\omega)$ est une signature étiquetée et la décomposition de $\boldsymbol{\gamma}^\omega$ par rapport à $\mathbf{r}^\omega$ est de la forme :

$$\boldsymbol{\gamma}^\omega \;=\; \big(u_0\, u_1 \cdots u_{(q-1)}\big)^\omega \qquad (\text{avec } \;\forall i < q \qquad |u_i| = r_i)\ .$$

On dit alors que $u_0\, u_1 \cdots u_{(q-1)}$ est la *décomposition de* $\boldsymbol{\gamma}$ *(par rapport à* $\mathbf{r}$*)* et que $\boldsymbol{\gamma}$ est *cohérent* avec $\mathbf{r}$ si $\boldsymbol{\gamma}$ est de longueur $p$ et si tous les facteurs $u_i$ sont des mots croissants.

**Définition 8.6** – *Un* rythme étiqueté *est un couple* $(\mathbf{r}, \boldsymbol{\gamma})$ *où* $\mathbf{r}$ *est un rythme et* $\boldsymbol{\gamma}$ *est un mot fini. Il est dit* valide *si*

**a)** *le rythme* $\mathbf{r}$ *est valide ;*

**b)** $\boldsymbol{\gamma}$ *est cohérent avec* $\mathbf{r}$*.*

**Lemme 8.7** – *Soient un rythme* $\mathbf{r}$ *de paramètre directeur* $(q, p)$ *et* $\boldsymbol{\gamma}$ *un mot de longueur* $p$*. La signature étiquetée* $(\mathbf{r}^\omega, \boldsymbol{\gamma}^\omega)$ *est valide si et seulement si le rythme étiqueté* $(\mathbf{r}, \boldsymbol{\gamma})$ *est valide.*

Démonstration. D'après le lemme 8.4, le rythme $\mathbf{r}$ est valide si et seulement si la signature $\mathbf{r}^\omega$ est valide.

La décomposition de $\boldsymbol{\gamma}$ est $u_0\, u_1 \cdots u_{(q-1)}$ si et seulement si celle de $\boldsymbol{\gamma}^\omega$ est $(u_0\, u_1 \cdots u_{(q-1)})^\omega$. Tous les facteurs de la première sont croissants si et seulement si tous ceux de la seconde sont croissants. En d'autres termes, $\boldsymbol{\gamma}$ est cohérent avec $\mathbf{r}$ si et seulement si $\boldsymbol{\gamma}^\omega$ est cohérent avec $\mathbf{r}^\omega$. □

Dans la suite, on ne considérera que des rythmes étiquetés $(\mathbf{r}, \boldsymbol{\gamma})$ valides. On note alors $L_{(\mathbf{r},\boldsymbol{\gamma})}$ le langage engendré par la signature étiquetée $(\mathbf{r}^\omega, \boldsymbol{\gamma}^\omega)$ que l'on dira plus simplement engendrés par $(\mathbf{r}, \boldsymbol{\gamma})$. En particulier, il découle alors de l'algorithme de génération par signature/étiquetage que l'arc entrant dans un nœud $m$ de $L_{(\mathbf{r},\boldsymbol{\gamma})}$ est

étiqueté par $\gamma_i$, où $i = m \,\%\, p$. De façon analogue, le langage calable $z^*L_{(\mathbf{r},\boldsymbol{\gamma})}$ engendré par $(\mathbf{r},\boldsymbol{\gamma})$ vérifie $z = \gamma_0$.

Le lemme suivant donne une caractérisation des transitions de $z^*L_{(\mathbf{r},\boldsymbol{\gamma})}$ ; il s'agit d'une spécialisation du lemme 7.15 (page 181) au cas considéré.

**Lemme 8.8** – *Soit un rythme étiqueté (valide) $(\mathbf{r},\boldsymbol{\gamma})$ de paramètre directeur $(q,p)$. Alors le langage calable $z^*L_{(\mathbf{r},\boldsymbol{\gamma})}$ engendré par ce rythme vérifie*

$$\forall n, m,\ \forall a \in A \qquad n \xrightarrow[z^*L_{(\mathbf{r},\boldsymbol{\gamma})}]{a} m \quad \Longleftrightarrow \quad S_n \leqslant m < S_{(n+1)} \quad \text{et} \quad a = \gamma_{m \,\%\, p}\ .$$

**Corollaire 8.9** – *Pour tous nœuds $n$ et $m$ de $z^*L_{(\mathbf{r},\boldsymbol{\gamma})}$. Alors, $n \xrightarrow{a} m$ si et seulement si $(n+q) \xrightarrow{a} (m+p)$.*

## Quelques étiquetages

Quand il n'y a pas ambiguïté, on appelle désormais le mot $\boldsymbol{\gamma}$ *étiquetage* plutôt que *période d'étiquetage*.

### Étiquetage naïf

L'étiquetage le plus simple, que l'on appelle *étiquetage naïf* est le mot formé des entiers de 0 à $(p-1)$ par ordre croissant :

$$\forall p \in \mathbb{N} \qquad \boldsymbol{\nu}_p = 0\,1\cdots(p-1)\ . \tag{8.1}$$

On note $K_{\mathbf{r}}$ le langage engendré par un rythme $\mathbf{r}$ et l'étiquetage naïf correspondant (donc $K_{\mathbf{r}} = L_{(\mathbf{r},\boldsymbol{\nu}_p)}$) ; par exemple, la figure 4 représente le langage $K_{302}$. Cet étiquetage présente deux propriétés :

- il est cohérent avec chaque rythme dont le paramètre directeur a pour seconde composante $p$) car tous les facteurs de $\boldsymbol{\nu}_p$ sont des mots croissants ;
- il distingue toutes les transitions construites lors d'un cycle complet du rythme, ce qui implique que $K_{\mathbf{r}}$ est au moins aussi complexe que $L_{(\mathbf{r},\boldsymbol{\gamma})}$, quelque soit $\boldsymbol{\gamma}$ (proposition 8.10, ci-dessous).

**Proposition 8.10** – *Soit un rythme étiqueté $(\mathbf{r},\boldsymbol{\gamma})$ qui engendre le langage $L_{(\mathbf{r},\boldsymbol{\gamma})} \subseteq B^*$. Il existe un morphisme lettre-à-lettre $\varphi : [\![p]\!]^* \to B^*$ tel que $\varphi(K_{\mathbf{r}}) = L_{(\mathbf{r},\boldsymbol{\gamma})}$.*

Démonstration. On note $\boldsymbol{\gamma} = \gamma_0\,\gamma_1\cdots\gamma_{p-1}$ et $\varphi^* : [\![p]\!] \to B^*$ le morphisme défini par :

$$\forall a \in [\![p]\!] \qquad \varphi(a) = \gamma_a\ .$$

Le résultat découle alors du fait qu'un arc $n \xrightarrow{a} m$ existe dans $K_{\mathbf{r}}$ si et seulement si $n \xrightarrow{\gamma_a} m$ existe dans $L_{(\mathbf{r},\boldsymbol{\gamma})}$ (lemme 8.8). □

### Étiquetage spécial

L'étiquetage spécial est une généralisation de celui de $L_{\frac{p}{q}}$, comme nous le verrons dans la section 8.2 suivante. Il sera ensuite utilisé pour montrer le théorème principal de la section 8.3.

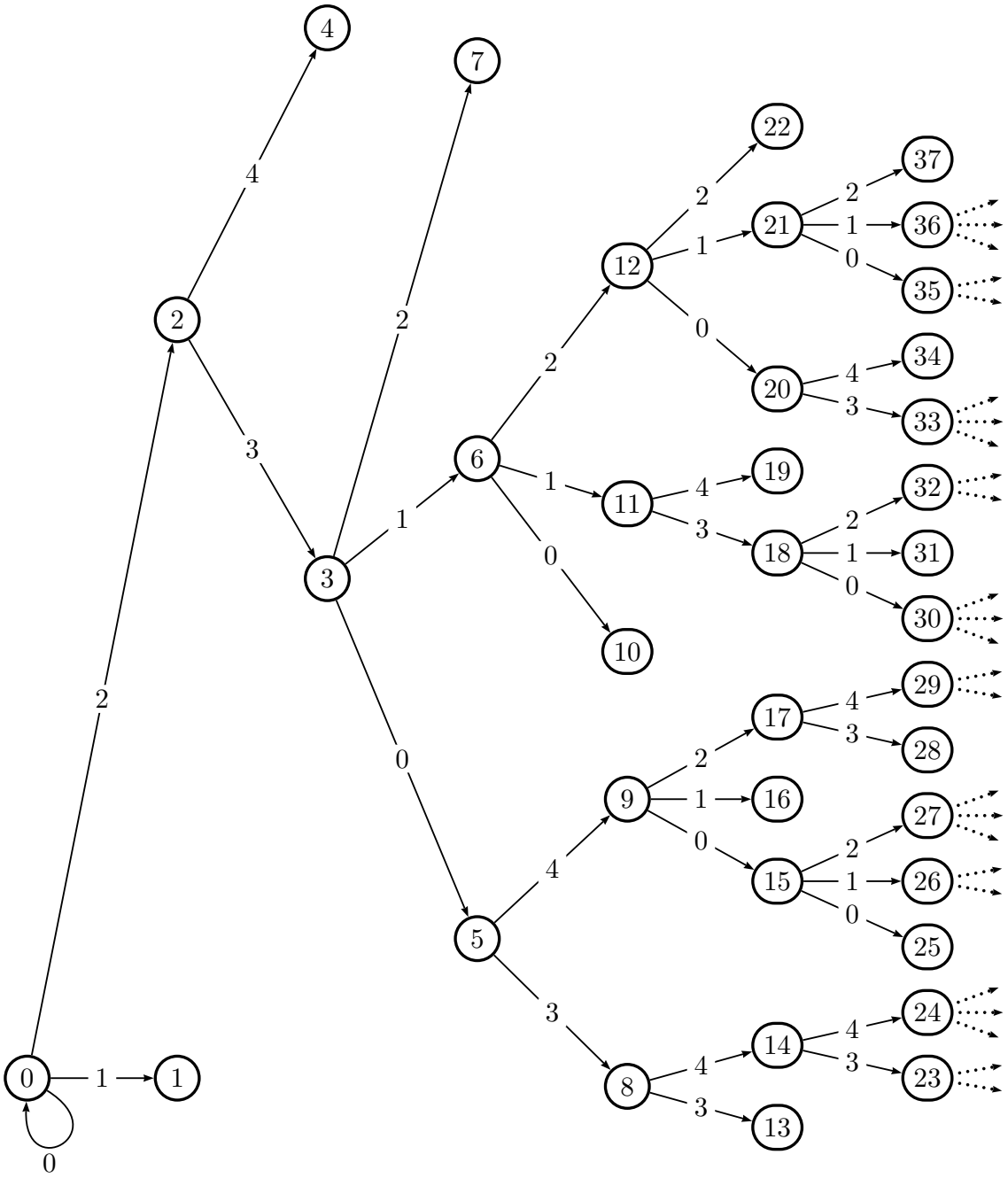



Figure 4 – Le langage $K_{302}$

**Définition 8.11** – *Soit* $\mathbf{r} = r_0 r_1 \cdots r_{q-1}$ *un rythme de paramètre directeur* $(q,p)$ *et de taux de croissance* $\frac{p'}{q'}$. *On appelle* étiquetage spécial *(associé à* $\mathbf{r}$*), noté* $\boldsymbol{\gamma}_{\mathbf{r}}$, *le mot* $\boldsymbol{\gamma}_{\mathbf{r}} = \gamma_0\gamma_i\cdots\gamma_{(p-1)}$ *dont les lettres sont définies par*

$$\forall i < p \qquad \gamma_i \;=\; q'\,i - p'\,j \qquad \text{où } j \text{ est l'entier vérifiant} \qquad R_j \leqslant i < R_{(j+1)}\ .$$

Dans le cas où $\mathbf{r}$ ne contient pas de lettre 0, une autre définition "dynamique" de l'étiquetage spécial peut être donnée. On commence par décomposer $\boldsymbol{\gamma}_{\mathbf{r}} = \gamma_0\gamma_i\cdots\gamma_{(p-1)}$ (par rapport à $\mathbf{r}$) :

$$\boldsymbol{\gamma}_{\mathbf{r}} \;=\; u_0\,u_1\cdots u_{(q-1)} \qquad \text{tel que} \qquad \forall i \in [\![q]\!] \qquad |u_i| \;=\; r_i\ .$$

On pose $\gamma_0 = 0$ et pour tout entier $i < q$

- $\gamma_{(i+1)} = \gamma_i + q'$ si $\gamma_i$ et $\gamma_{(i+1)}$ appartiennent au même facteur $u_j$ ;
- $\gamma_{(i+1)} = \gamma_i + q' - p'$ sinon (auquel cas $\gamma_i$ et $\gamma_{(i+1)}$ sont dans deux facteurs consécutifs $u_j$ et $u_{(j+1)}$ de la décomposition de $\boldsymbol{\gamma}_{\mathbf{r}}$).

Cette définition alternative est moins élégante quand $\mathbf{r}$ contient des 0. Dans le deuxième cas ci-dessus, les lettres $\gamma_i$ et $\gamma_{(i+1)}$ peuvent appartenir à des facteurs de la décomposition qui ne sont pas consécutifs, les facteurs intermédiaires étant tous réduits au mot vide. Dans ce cas $\gamma_{(i+1)} = \gamma_i + q' - k\,p'$ où $k$ est la différence d'indices entre les facteurs $u_j$ et $u_{(j+k)}$ auxquels appartiennent respectivement $\gamma_i$ et $\gamma_{(i+1)}$

**Exemple 8.12** – *Soit* $\mathbf{r} = 3133$ *; son paramètre directeur est* $(4,10)$, $p' = 5$, $q' = 2$ *et le calcul donne :*

$$\begin{aligned} \mathbf{r} \;&=\; \overset{u_0}{3} \quad \overset{u_1}{1} \quad \overset{u_2}{3} \quad \overset{u_3}{3} \\ \boldsymbol{\gamma}_{\mathbf{r}} \;&=\; \overbrace{0\,2\,4}^{u_0} \quad \overbrace{1}^{u_1} \quad \overbrace{-2\,0\,2}^{u_2} \quad \overbrace{-1\,1\,3}^{u_3}\ . \end{aligned}$$

*A l'intérieur d'un même facteur* $u_i$, *la différence entre deux chiffres successifs est* $+2\,(=+q')$ *et celle entre le dernier chiffre d'un facteur et le premier du facteur suivant est* $-3\,(=+q'-p')$.

*On pourra vérifier de manière analogue que* $\boldsymbol{\gamma}_{31} = 0121$, *et* $\boldsymbol{\gamma}_{221} = 03142$. *Le langage calable* $0^*L_{221}$, *qui est aussi le langage des* $\frac{5}{3}$*-représentations, est représenté par la figure 5 ; le langage calable* $0^*L_{31}$ *est représenté plus loin (page 217) par la figure 10.*

*Soit maintenant* $\mathbf{r} = 5005$ *; les entiers* $p$, $q$, $p'$ *et* $q'$ *sont inchangées et le calcul donne :*

$$\begin{aligned} \mathbf{r} \;&=\; \overset{u_0}{5} \quad \overset{u_1}{0} \quad \overset{u_2}{0} \quad \overset{u_3}{5} \\ \boldsymbol{\gamma}_{\mathbf{r}} \;&=\; \overbrace{0\,2\,4\,6\,8}^{u_0} \quad \overbrace{\varepsilon}^{u_1} \quad \overbrace{\varepsilon}^{u_2} \quad \overbrace{-5\,-3\,-1\,1\,3}^{u_3}\ . \end{aligned}$$

*A l'intérieur d'un même facteur* $u_i$, *la différence entre deux chiffres successifs est encore une fois* $+2\,(=+q')$ *mais pour passer de la cinquième à la sixième lettre la différence est* $-13\,(=+q'-k\,p')$, *car on passe du premier au quatrième facteur et la différence des indices vaut donc* $k = 3$.

### Étiquetage réduit

L'étiquetage réduit est le plus compact : il utilise le nombre minimal de lettres ; de plus, il produit en général les langages les plus simples. Il est principalement

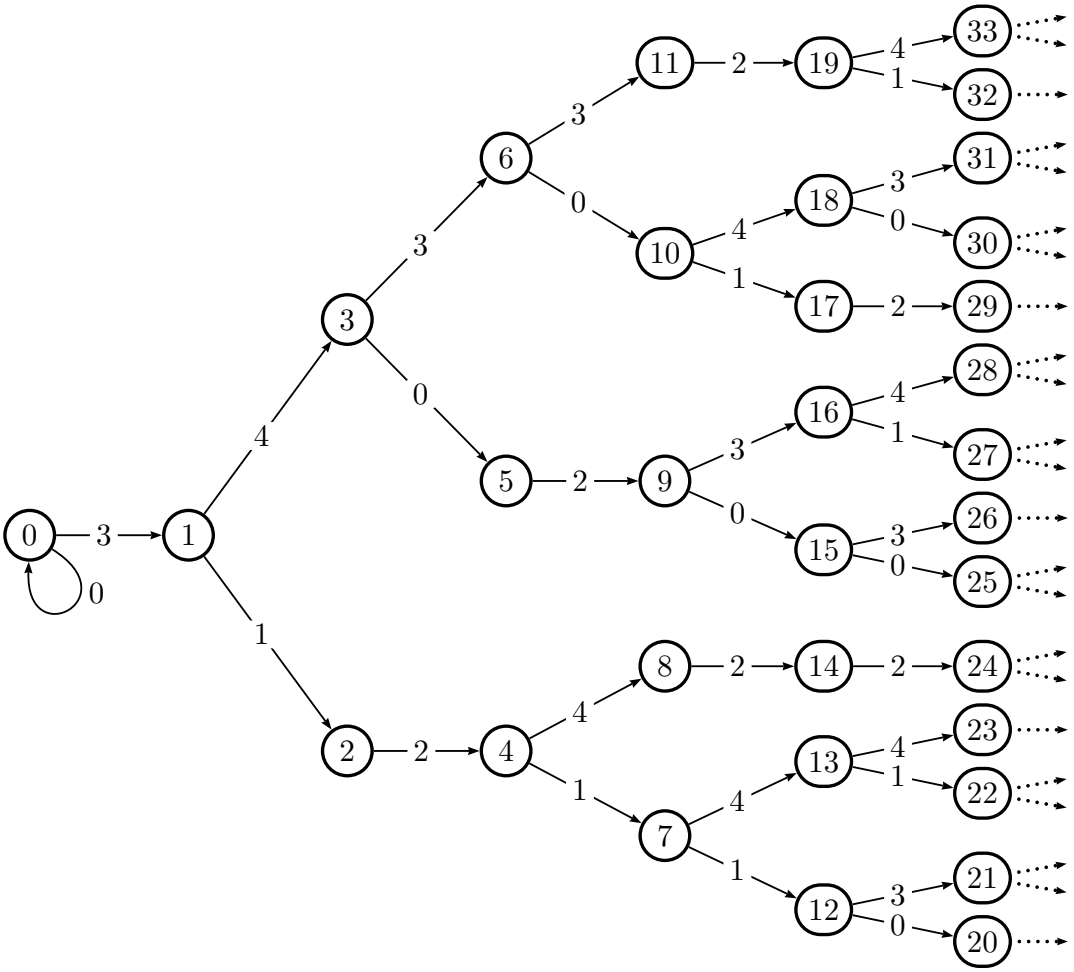



Figure 5 – Le langage $L_{2\,2\,1}(= L_{\frac{5}{3}})$

mentionné car il est à l'origine de la *surminimisation* (ou plus exactement de la *réduction d'étiquetage*) décrite dans chapitre 9.

**Définition 8.13** – *Soit* $\mathbf{r} = r_0\, r_1 \cdots r_{q-1}$. *L'*étiquetage réduit*, noté* $\boldsymbol{\rho}_{\mathbf{r}}$*, est l'étiquetage tel que chaque lettre est le plus petit chiffre positif possible tout en restant cohérent avec* $\mathbf{r}$*. Il est défini à partir de sa décomposition :*

$$\boldsymbol{\rho}_{\mathbf{r}} \;=\; v_0\, v_1 \cdots v_{(q-1)} \qquad \textit{tel que} \qquad \forall i \in [\![q]\!] \qquad |v_i| \;=\; r_i\,;$$

*et pour tout entier* $i < q$, $v_i = 0\,1 \cdots (r_i - 1)$. *(Il faut en effet que chaque facteur* $v_i$ *soit un mot croissant pour que* $\boldsymbol{\rho}_{\mathbf{r}}$ *soit cohérent avec* $\mathbf{r}$*.)*

L'étiquetage réduit est le plus compact : il utilise le nombre minimal de lettres. De plus, il produit en général les langages les plus simples. Il est principalement mentionné parce qu'il est à l'origine de la *surminimisation* (ou plus exactement de la *réduction d'étiquetage*) décrite dans chapitre 9.

# Rythme associé à une base rationnelle

Dans toute cette section, $p$ et $q$ sont deux entiers premiers entre eux tels que $p > q \geqslant 1$ ; ils définissent donc soit une base entière (si $q = 1$) soit une base rationnelle (si $q > 1$). Notre objectif est de démontrer que le langage $L_{\frac{p}{q}}$ est engendré par le rythme $\mathbf{r}_{\frac{p}{q}}$ et l'étiquetage $\boldsymbol{\gamma}_{\frac{p}{q}}$ où

- $\mathbf{r}_{\frac{p}{q}}$ est un rythme canonique de paramètre directeur $(q, p)$ (donc de taux de croissance $\frac{p}{q}$) que l'on appelle *rythme de Christoffel* car il est lié à la notion classique de mot Christoffel ;
- $\boldsymbol{\gamma}_{\frac{p}{q}}$ est l'étiquetage spécial $\boldsymbol{\gamma}_{\mathbf{r}_{\frac{p}{q}}}$ associé mais qui, dans ce cas particulier, peut être exprimé beaucoup plus simplement, il s'agit de la permutation correspondant à la génération de $\mathbb{Z}/p\mathbb{Z}$ par $q$.

**Remarque 8.14** – *L'automate infini $\mathcal{T}_{\frac{p}{q}}$ a été précédemment (page 98) défini par :*

$$\mathcal{T}_{\frac{p}{q}} \;=\; \langle\, \mathbb{N},\, [\![p]\!], \delta, 0, \mathbb{N} \,\rangle$$

*où la fonction de transition $\delta$ est la suivante :*

$$\forall n, m \in \mathbb{N},\ \ \forall a \in [\![p]\!] \qquad n \xrightarrow[\mathcal{T}_{\frac{p}{q}}]{a} m \qquad \textit{si et seulement si} \qquad q\,m \;=\; p\,n + a\ .$$

*L'ensemble d'états de cet automate est $\mathbb{N}$ et l'on peut vérifier facilement que sa fonction de transition $\delta$ est un sous-ensemble de $\mathbb{N}\times[\![p]\!]\times\mathbb{N}$ qui satisfait les conditions de la définition 7.10 d'i-arbre étiqueté (page 180). Tous les états de cet automate sont finals et il accepte le langage calable $0^*L_{\frac{p}{q}}$. Suivant notre formalisme, l'i-arbre étiqueté $\delta$ et le langage calable $0^*L_{\frac{p}{q}}$ sont donc* identifiés.

*En définitive, le langage calable $0^*L_{\frac{p}{q}}$ est caractérisé par la définition suivante de ses arcs :*

$$\forall n, m \in \mathbb{N},\ \ \forall a \in [\![p]\!] \qquad n \xrightarrow[0^*L_{\frac{p}{q}}]{a} m \qquad \textit{si et seulement si} \qquad q\,m \;=\; p\,n + a\ . \tag{8.2}$$

## Mot de Christoffel, rythme de Christoffel

Les mots de Christoffel codent les chemins donnant la *meilleure* approximation par excès ou par défaut d'un segment du plan discret $\mathbb{Z} \times \mathbb{Z}$. Ils ont été étudié en détails dans le domaine de la combinatoire des mots (voir par exemple [13]). On ne considère ici que les mots de Christoffel *par excès* que l'on appellera simplement mots de Christoffel.

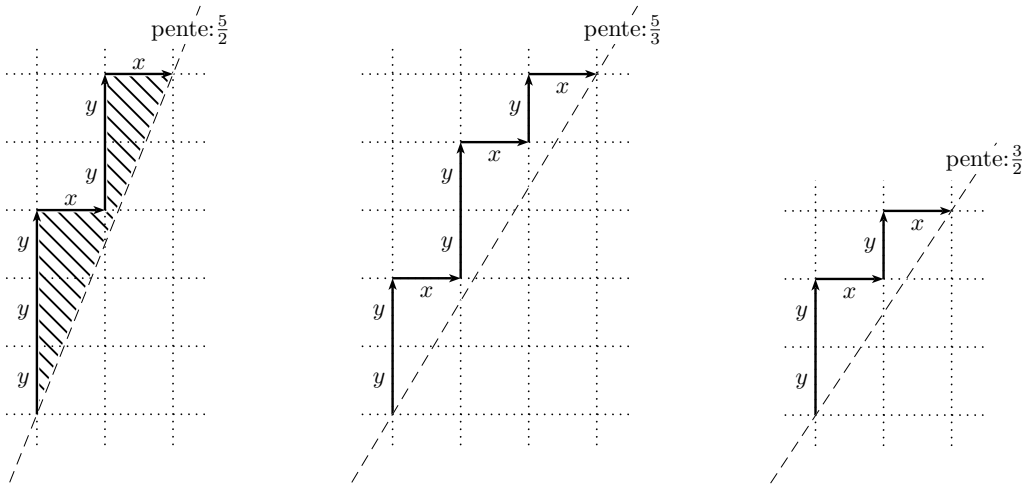
**Définition 8.15** – *Le mot de Christoffel de pente $\frac{p}{q}$, noté $\mathbf{w}_{\frac{p}{q}}$, est le mot de $\{x, y\}^*$ qui code l'unique chemin $C$ allant de $(0, 0)$ à $(q, p)$ dans le plan discret $\mathbb{Z} \times \mathbb{Z}$ tel que*

**a)** *le chemin $C$ est au dessus de la droite de pente $\frac{p}{q}$ passant par l'origine ;*

**b)** *la région (ouverte) du plan confinée entre cette droite et le chemin $C$ ne contient aucun point de coordonnées entières.*

[13] Jean Berstel, Aaron Lauve, Christophe Reutenauer et Franco Saliola, 2008, *Combinatorics on Words : Christoffel Words and Repetition in Words*.

Par exemple, la figure 6 indique les mots de Christoffel de différentes pentes et représente les chemins qu'ils codent. La figure 6a met en évidence la région du plan dont il est question dans la condition 8.15b et qui ne contient donc aucun point à coordonnée entière.




(a) $\mathbf{w}_{\frac{5}{2}} = yyyxyyx$ (b) $\mathbf{w}_{\frac{5}{3}} = yyxyyxyx$ (c) $\mathbf{w}_{\frac{3}{2}} = yyxyx$

Figure 6 – Quelques mots de Christoffel

Ces mots de Christoffel définissent des rythmes que l'on appelle rythmes de Christoffel.

**Définition 8.16** – *Le* rythme de Christoffel *associé à* $\frac{p}{q}$*, noté* $\mathbf{r}_{\frac{p}{q}}$*, est le rythme dont le chemin est codé par le mot Christoffel de pente* $\frac{p}{q}$ *:* $\mathsf{path}(\mathbf{r}_{\frac{p}{q}}) = \mathbf{w}_{\frac{p}{q}}$.

Le mot de Christoffel de pente $\frac{5}{2}$ est $yyyxyyx$ (figure 6a) donc le rythme de Christoffel associé à $\frac{5}{2}$ est $\mathbf{r}_{\frac{5}{2}} = 32$. Le rythme de Christoffel associé à $\frac{5}{3}$ est $\mathbf{r}_{\frac{5}{3}} = 221$ et celui associé à $\frac{3}{2}$ est $\mathbf{r}_{\frac{3}{2}} = 21$. La figure 7 donne un exemple pour des entiers $p$ et $q$ plus grands.

Les mots de Christoffel possèdent de nombreuses propriétés dont la plupart se transpose aux rythmes de Christoffel ; on en donne plusieurs exemples dans la suite. On rappelle qu'un mot $u$ est dit *primitif* s'il n'est la puissance d'aucun autre mot : $\forall v \quad v^k = u \implies v = u \text{ et } k = 1$.

**Lemme 8.17** – *Le mot de Christoffel* $\mathbf{w}_{\frac{p}{q}}$ *est primitif.*

Démonstration. Supposons qu'il existe un mot $u$ et un entier $k$ tel que $\mathbf{w}_{\frac{p}{q}} = u^k$. On note $i$ le nombre de $x$ dans $u$ et $j$ son nombre de $y$ ; si bien que $\mathbf{w}_{\frac{p}{q}}$ contient $q = k\,i$ fois la lettre $x$ et $p = k\,j$ fois la lettre $y$. Puisque $p$ et $q$ sont premiers entre eux, $k = 1$. □

**Corollaire 8.18** – *Le rythme de Christoffel* $\mathbf{r}_{\frac{p}{q}}$ *est primitif.*

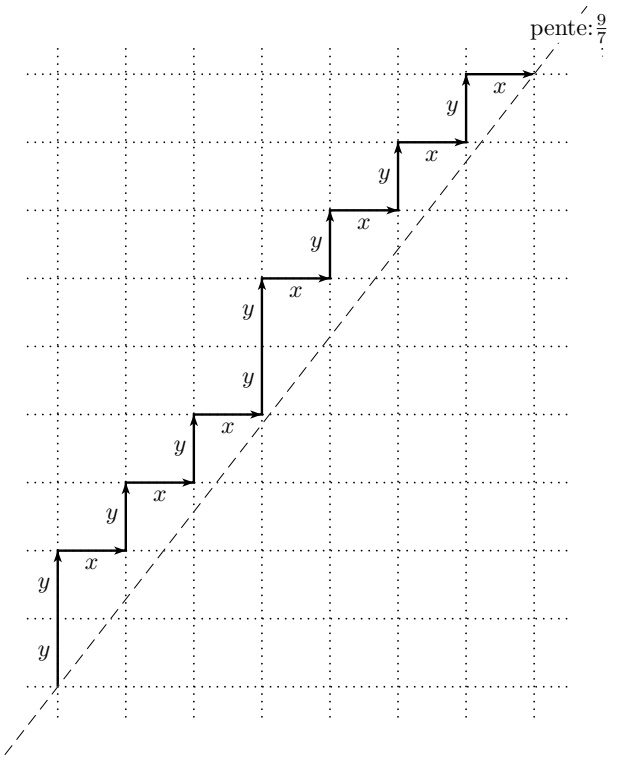



Figure 7 – Le mot $\mathbf{w}_{\frac{9}{7}} = yyxyxyxyyxyxyxyx$ et le rythme $\mathbf{r}_{\frac{9}{7}} = 2112111$ de Christoffel de pente $\frac{9}{7}$

La proposition 8.20 donne une manière de calculer directement les rythmes de Christoffel. C'est une conséquence directe du lemme suivant, qui décrit une propriété élémentaire des mots de Christoffel.

**Lemme 8.19** – *Si $ux$ est un préfixe de $\mathbf{w}_{\frac{p}{q}}$, alors il code un chemin allant de $(0,0)$ à $(k, \left\lceil k\frac{p}{q}\right\rceil)$ dans le plan discret $\mathbb{Z}\times\mathbb{Z}$, où $k$ est le nombre de $x$ dans $ux$.*

Démonstration. D'après la définition 8.15, il n'existe pas de point entier entre le chemin codé par $\mathbf{w}_{\frac{p}{q}}$ et la droite de pente $\frac{p}{q}$ (passant par l'origine).

Soit un entier $k$, $0 < k \leqslant q$. Puisque le point $(k, k\frac{p}{q})$ appartient à cette droite, le chemin codé par $\mathbf{w}_{\frac{p}{q}}$ passe par le point $(k, \lceil k\frac{p}{q}\rceil)$. Le préfixe de $\mathbf{w}_{\frac{p}{q}}$ qui atteint ce point se termine nécessairement par la lettre $x$, car dans le cas contraire, le point$(k, \lceil k\frac{p}{q}\rceil - 1)$ appartiendrait au chemin, alors qu'il est en dessous de la droite de pente $\frac{p}{q}$ (ce qui contredirait la condition 8.15a).

Les points de la forme $(k, \lceil k\frac{p}{q}\rceil)$ (tels que $0 < k \leqslant q$) sont au nombre de $q$, c'est-à-dire le nombre de $x$ dans $\mathbf{w}_{\frac{p}{q}}$ ; donc tout préfixe $ux$ de $\mathbf{w}_{\frac{p}{q}}$ atteint un de ces points $(k, \left\lceil k\frac{p}{q}\right\rceil)$, et $k$ est le nombre de $x$ dans $ux$. □

**Proposition 8.20** – *Pour tout entier $k \leqslant q$, la somme partielle $R_k$ des $k$*

*premières composantes de* $\mathbf{r}_{\frac{p}{q}}$ *est égale à la partie entière supérieure de* $k\,\frac{p}{q}$ *:*

$$\forall k \leqslant q \qquad R_k \;=\; \left\lceil k\;\frac{p}{q}\right\rceil \;.$$

La proposition suivante donne un résultat symétrique à la proposition précédente, et découle de celle-ci en utilisant l'implication :

$$\forall x, y \in \mathbb{Q} \qquad (x+y) \in \mathbb{N} \quad\Longrightarrow\quad \lceil x\rceil + \lfloor y\rfloor \;=\; x+y\;.$$

**Proposition 8.21** – *Pour tout entier* $k \leqslant q$*, la somme partielle* $r_{(q-k)} + \cdots + r_{(q-2)} + r_{(q-1)}$ *des* $k$ *dernières composantes de* $\mathbf{r}_{\frac{p}{q}}$ *est égale à la partie entière inférieure de* $k\,\frac{p}{q}$ *:*

$$\forall k \leqslant q \qquad \sum_{i=q-k}^{q-1} r_i \;=\; \left\lfloor k\;\frac{p}{q}\right\rfloor \;.$$

On veut maintenant déterminer plus finement les relations entre les différents termes de $\mathbf{r}_{\frac{p}{q}}$. On définit pour cela la suite d'entiers $e_0, e_1, \ldots, e_{(q-1)}$ par

$$\forall\; k \in [\![q]\!] \qquad e_k \;=\; q\,R_k - k\,p\;. \tag{8.3}$$

La figure 8 montre, en prenant comme exemple la base $\frac{5}{3}$, une manière plus schématique pour caractériser les rythmes de Christoffel. On considère $p$ 'petits' segments de longueur $q$ (en haut) et $q$ 'grands' segments de longueur $p$ (en bas) : un petit segment est associé à un grand si l'extrémité gauche du premier est à une abscisse appartenant au second ; ceci est représenté par une accolade décentrée sur la figure.

- L'entier $r_k$ est le nombre de petits segments associé au $(k+1)$-ème grand segment.
- L'entier $e_k$ est la différence de longueur entre les $k$ premiers grands segments (de longueur $p\,k$) et tous les petits segments qui leurs sont associés (de longueur $q\,R_k$).

Cette interprétation est une version aplatie de la *machine de Pascal gonflée (boosted Pascal machine)* présentée dans [2].

La figure 9 montre pourquoi cette interprétation est équivalente. Il faut pour cela subdiviser en $q$ les unités du plan discret $\mathbb{Z}\times\mathbb{Z}$. Notez que, sur la figure 9, les segments de longueur $q$ (en haut sur la figure 8) sont à gauche et que ceux de longueur $p$ (en bas sur la figure 8) sont à droite.

Ci-dessous sont rassemblées des propriétés arithmétiques des $e_k$ et des $r_k$ ; elles sont obtenues grâce à de simples calculs à partir des différentes définitions. On rappelle que l'on autorise les indices à dépasser $q$ : pour tout entier $k$, $e_k = e_{k\,\%\,q}$ et $r_k = r_{k\,\%\,q}$.

**Propriété 8.22** –

**a)** *Pour tout entier* $k$*,* $e_k + q\,r_k = p + e_{k+1}$.

[2] Shigeki Akiyama, Christiane Frougny et Jacques Sakarovitch, 2008, *Powers of rationals modulo 1 and rational base number systems.*

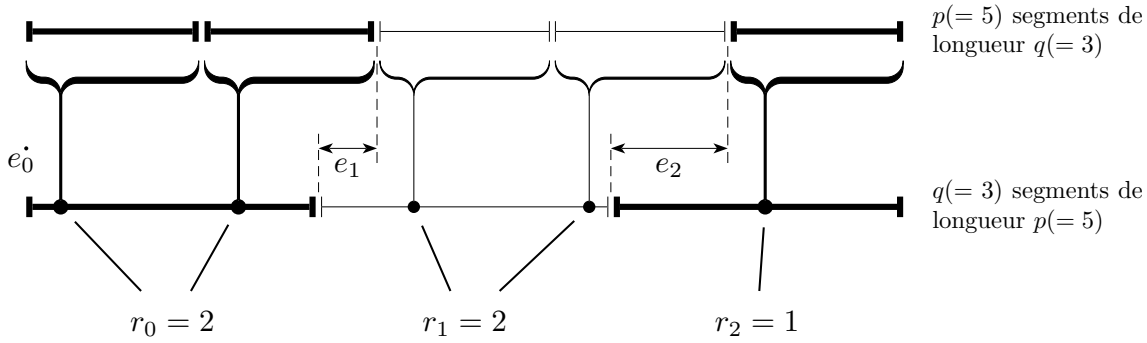



Figure 8 – Interprétation du rythme 221 de la base $\frac{5}{3}$

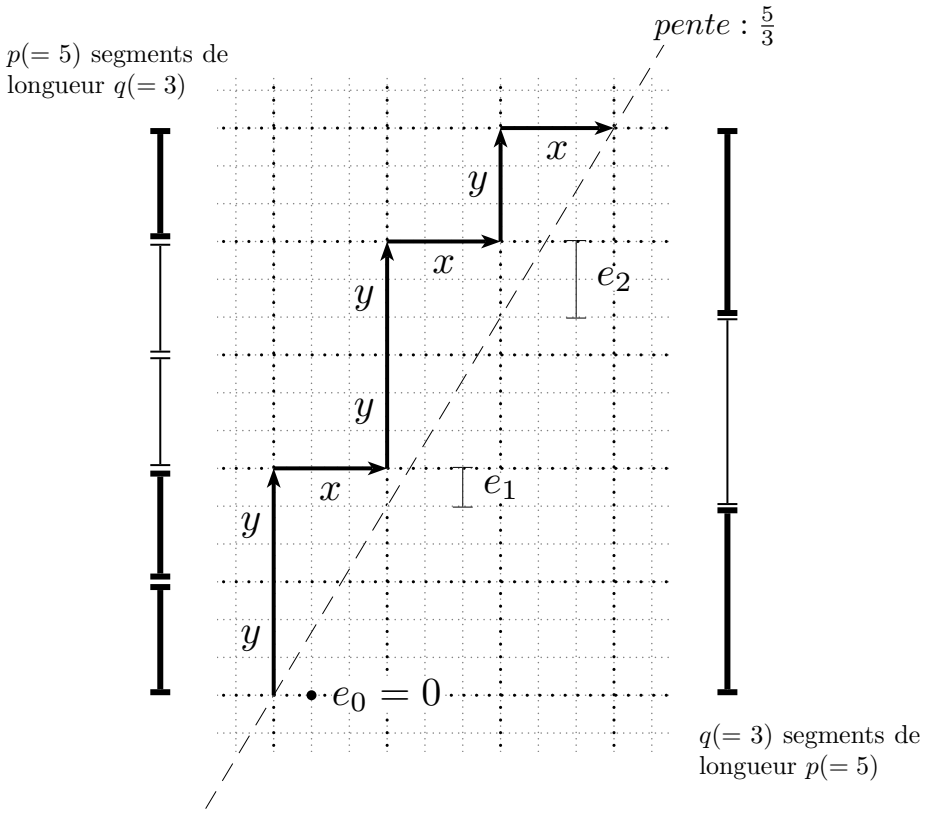



Figure 9 – Même interprétation prise de la perspective du chemin dans $\mathbb{Z} \times \mathbb{Z}$

**b)** *Pour tout entier $k$, $r_k$ est le plus petit entier tel que $(q\,r_k) + e_k \geqslant p$.*

**c)** *Pour tout entier $k$, $0 \leqslant e_k < q$.*

**d)** *Pour tous entiers $k, j$,* $e_k + q \sum_{i=0}^{j-1} r_{k+i} \;=\; j\,p + e_{(k+j)}$

**e)** *Pour tous entiers $k, j$,* $\sum_{i=0}^{j-1} r_{k+i} \;=\; \left\lceil \frac{j\,p - e_k}{q} \right\rceil \;=\; \left\lfloor \frac{j\,p + e_{(k+j)}}{q} \right\rfloor$

**f)** *Pour tous entiers $k, j < q$, si $k \neq j$, alors $e_j \neq e_k$.*

Démonstration. Le point (**a**) découle immédiatement de la définition des $e_k$

(équation (8.3)) et le (**b**) est une réécriture de la proposition 8.20. Le point (**c**) découle du (**b**) et le (**d**) est obtenu en itérant le (**a**).

**e)** On peut réécrire l'équation du (**d**) comme

$$\sum_{i=0}^{j-1} r_{k+i} \;=\; \frac{j\,p}{q} + \frac{e_{(k+j)}}{q} - \frac{e_k}{q}\;. \qquad (*)$$

Il découle du (**a**) que les deux fractions $\frac{e_{(k+j)}}{q}$ et $\frac{e_k}{q}$ sont plus petites que 1. Puisque le membre gauche de l'équation $(*)$ est un entier, ceci conclut la démonstration.

**f)** Soient deux entiers $k, j \in [\![q]\!]$ tels que $e_k = e_j$. Il découle du (**b**) que $r_k = r_j$, puis du (**c**) que $e_{i+1} = e_{j+1}$. Itérer ce processus implique que $\mathbf{r}_{\frac{p}{q}}$ est de période $\mathsf{abs}(k-j)$, donc $k = j$ (d'après le corollaire 8.18). □

## Génération de $L_{\frac{p}{q}}$ par rythme étiqueté

Puisque $p$ et $q$ sont premiers entre eux, $q$ est un générateur du groupe additif $\mathbb{Z}/p\mathbb{Z}$. On note $\boldsymbol{\gamma}_{\frac{p}{q}}$ le mot (de $[\![p]\!]^*$) induit par cette génération :

$$\boldsymbol{\gamma}_{\frac{p}{q}} \;=\; (0\,\%\,p)\,(q\,\%\,p)\,(2q\,\%\,p)\cdots((p-1)q\,\%\,p)\;. \qquad (8.4)$$

La proposition suivante résulte de propriétés des mots de Christoffel.

**Proposition 8.23** – *L'étiquetage $\boldsymbol{\gamma}_{\frac{p}{q}}$ coïncide avec l'étiquetage spécial $\boldsymbol{\gamma}_{\mathbf{r}_{\frac{p}{q}}}$ associé au rythme de Christoffel $\mathbf{r}_{\frac{p}{q}}$.*

Démonstration. On note l'étiquetage spécial $\boldsymbol{\gamma}_{\mathbf{r}_{\frac{p}{q}}} = \gamma_0\,\gamma_1\cdots\gamma_{p-1}$, c'est-à-dire tel que (définition 8.11, page 204) :

$$\forall i < p \qquad \gamma_i \;=\; q\,i - p\,j \quad \text{où } j \text{ est l'entier vérifiant} \quad R_j \;\leqslant\; i \;<\; R_{j+1}\;. \qquad (*)$$

(Dans le cas considéré ici, $p$ et $q$ sont premiers entre eux donc $p = p'$ et $q = q'$.)

Soit un entier $i < p$. Il découle de l'équation précédente que $\gamma_i \equiv i\,q\ [p]$, il est donc suffisant de démontrer que $0 \leqslant \gamma_i < p$.

On note $j$ l'entier tel que $R_j \leqslant i < R_{j+1}$ (qui vérifie nécessairement $0 \leqslant j < q-1$) et $M_k = (R_{j+1}-1)$. Il s'ensuit que chaque $k \in \{R_j,\, i,\, (M_j-1)\}$ vérifie l'inéquation $R_j \leqslant k < R_{(j+1)}$. Il découle donc de la définition (rappelée en $(*)$) de l'étiquetage spécial que

$$\begin{aligned} \gamma_{R_l} &= (q\,R_j) - (p\,j)\,; \\ \gamma_i &= (q\,i) - (p\,j)\,; \\ \gamma_{M_j} &= (q\,M_j) - (p\,j)\;. \end{aligned}$$

Ceci implique que $\boxed{\gamma_{R_k} \leqslant \gamma_i \leqslant \gamma_{M_K}}$.

D'après la proposition 8.20, $R_j$ est le plus petit entier supérieur à $j\,\frac{p}{q}$, donc $R_j \geqslant j\,\frac{p}{q}$ ce qui implique que $\boxed{0 \leqslant \gamma_{R_j}}$. Appliquer la même proposition à $(j+1)$ donne que $R_{(j+1)}$ est le plus petit entier supérieur à $(j+1)\,\frac{p}{q}$ donc $M_j = (R_{(j+1)} - 1) < (j+1)\,\frac{p}{q}$, ce qui implique que $\boxed{\gamma_{M_j} < p}$.

Les trois inéquations encadrées donne $0 \leqslant \gamma_{R_k} \leqslant \gamma_i \leqslant \gamma_{M_K} < p$. □

Le résultat principal de cette section 8.2, ci-dessous, énonce que le langage $L_{\frac{p}{q}}$ des $\frac{p}{q}$-représentations est engendré par $(\mathbf{r}_{\frac{p}{q}}, \boldsymbol{\gamma}_{\frac{p}{q}})$.

**Théorème VI** – *Soient deux entiers $p$ et $q$, premiers entre eux et tels que $p > q \geqslant 1$.*

**a)** *La signature du langage $L_{\frac{p}{q}}$ est $\mathbf{r}_{\frac{p}{q}}{}^{\omega}$, où $\mathbf{r}_{\frac{p}{q}}$ est le rythme dont le chemin dans $\mathbb{Z} \times \mathbb{Z}$ est le mot de Christoffel de pente $\frac{p}{q}$.*

**b)** *L'étiquetage du langage $L_{\frac{p}{q}}$ est $\boldsymbol{\gamma}_{\frac{p}{q}}{}^{\omega}$, où $\boldsymbol{\gamma}_{\frac{p}{q}}$ est la suite résultant de la génération de $\mathbb{Z}/p\mathbb{Z}$ par $q$ : $\boldsymbol{\gamma}_{\frac{p}{q}} = 0\,q\,(2\,q\,\%\,p) \cdots ((p-1)\,q\,\%\,p)$.*

Le langage $L_{\frac{3}{2}}$ (représenté page 195) est engendré par $(21, 021)$ et le langage $L_{\frac{5}{3}}$ (représenté page 205) est engendré par $(221, 0314)$. La démonstration du théorème VI requiert un lemme préliminaire.

**Lemme 8.24** – *Soit un entier $n$ dont on note $k$ la classe d'équivalence modulo $q$. Alors $(n \cdot e_k)$ existe dans $0^* L_{\frac{p}{q}}$.*

Démonstration. D'après l'équation (8.2) régissant les arcs de $0^* L_{\frac{p}{q}}$, $(n \cdot e_k)$ existe si et seulement si $n\,p + e_k$ est un multiple de $q$ ou, de façon équivalente, si $k\,p + e_k$ est un multiple de $q$. Or, la définition de $e_k$ (équation (8.3)) se réécrit comme

$$k\,p + e_k \;=\; q\,R_k \;,$$

ce qui implique en particulier que $k\,p + e_k$ est un multiple de $q$. □

Démonstration du théorème VI. Soient deux entiers $p$ et $q$ premiers entre eux tels que $p > q \geqslant 1$ et $\mathbf{r}_{\frac{p}{q}} = r_0\,r_1 \cdots r_{(q-1)}$ le rythme de Christoffel associé.

Soit un nœud $n$ de $L_{\frac{p}{q}}$ dont on note $k$ la classe de congruence modulo $q$. D'après la lemme 8.24, la lettre $e_k$ étiquette un arc sortant du nœud $n$, donc les étiquettes sortantes de $n$ forment l'ensemble

$$X \;=\; \{\; b \in [\![p]\!] \mid b \equiv e_k \;[q] \;\} \;.$$

Puisque de plus $e_k < q$ (propriété 8.22c), $e_k$ est l'étiquette minimale sortante de $n$, donc

$$X \;=\; \{\; (q\,i + e_k) \mid (q\,i + e_k) < p \;\text{ et }\; i \in \mathbb{N} \;\} \;.$$

Or $r_k$ est justement le plus petit entier tel que $q\,r_k + e_k \geqslant p$ (propriété 8.22b). Le nœud $n$ admet donc $r_k$ transitions sortantes, étiquetées par

$$e_k,\; (q + e_k),\; (2\,q + e_k),\; \ldots,\; ((r_k - 1)\,q + e_k) \;.$$

La $(n+1)$-ème lettre de la signature de $L_{\frac{p}{q}}$ est donc $r_{n\,\%\,q}$.

En résumé, la signature de $L_{\frac{p}{q}}$ est $(r_0\,r_1 \cdots r_{(q-1)})^{\omega} = \mathbf{r}_{\frac{p}{q}}{}^{\omega}$.

L'étiquetage de $L_{\frac{p}{q}}$ est la suite $\lambda_0\,\lambda_1\,\lambda_2 \cdots$ des lettres tels que, pour tout entier $m$, $\lambda_m$ étiquette l'arc entrant du nœud $m$ (lemme 8.8). Or l'équation (8.2) régissant ses arcs est la suivante :

$$\forall n, m \in \mathbb{N},\;\; \forall a \in [\![p]\!] \qquad n \xrightarrow[0^* L_{\frac{p}{q}}]{a} m \qquad \text{si et seulement si} \qquad q\,m \;=\; p\,n + a \;.$$

Il s'ensuit que l'arc entrant du nœud $m$ est étiqueté par $(q\,m)\,\%\,p$, donc que l'étiquetage de $L_{\frac{p}{q}}$ est $(0\ q\ (2\,q)\ \cdots\ ((p-1)\,q))^\omega = \boldsymbol{\gamma}_{\frac{p}{q}}^{\omega}$. □

## Rythme et étiquetage dans la variante FK

Dans la section 4.5 (page 111) nous avons brièvement présenté la variante FK des systèmes de numération à base rationnelle originellement définie dans [35]. On en rappelle ici les bases pour pouvoir étudier les signatures et étiquetages qui y sont associés.

Soit deux entiers $p$ et $q$ premiers entre eux tels que $p > q \geqslant 1$.[1] La fonction d'évaluation $\theta_{\frac{p}{q}}$ dans la variante FK de la base $\frac{p}{q}$, associe à tout mot $u = a_k \cdots a_1\, a_0$ sur un alphabet de chiffres, le nombre

$$\theta_{\frac{p}{q}}(u) \;=\; \theta_{\frac{p}{q}}(a_k \cdots a_1\, a_0) \;=\; \sum_{i=0}^{k} a_i \left(\frac{p}{q}\right)^i \;=\; q\,\pi_{\frac{p}{q}}(u)\ .$$

Chaque entier est représenté par un mot de $[\![p]\!]^*$ noté $\langle n\rangle_{\frac{p}{q}}^{\mathrm{FK}}$ et le langage des représentations des entiers forment un langage noté $\Theta_{\frac{p}{q}}$. Ce langage est clos par préfixe mais n'est pas prolongeable et ses arcs sont régis par :

$$\forall n, m \in \mathbb{N},\ \ \forall a \in [\![p]\!] \qquad n \xrightarrow[0^*\Theta_{\frac{p}{q}}]{a} m \quad\Longleftrightarrow\quad q\,m \;=\; p\,n + q\,a\ .$$

Soit un nœud $n$. Si $n$ n'est pas un multiple de $q$, alors l'équation du membre droit ci-dessus n'a pas de solution pour $a$ et $m$ entiers (car $p$ est premier avec $q$) ; si au contraire $n$ est un multiple de $q$, alors chaque lettre $a \in [\![p]\!]$ définit un unique entier $m$ qui satisfait cette équation donc qui est successeur de $n$ par $a$. La proposition suivante est une reformulation de cette remarque.

**Proposition 8.25** – *Soient deux entiers $p$ et $q$ premiers entre eux tels que $p > q \geqslant 1$. Le langage $\Theta_{\frac{p}{q}}$ est engendré par le rythme $\mathbf{r} = p\,0^{(q-1)}$ et l'étiquetage réduit $\boldsymbol{\rho}_{\mathbf{r}} = 0\,1\,\cdots\,(p-1)$.*

En reprenant la notation de cette proposition, on note $\mathbf{r} = p\,0^{(q-1)}$ le rythme de $\Theta_{\frac{p}{q}}$. Dans ce cas l'étiquetage spécial associé est

$$\boldsymbol{\gamma}_{\mathbf{r}} \;=\; \gamma_0\gamma_1\cdots\gamma_{(p-1)} \;=\; 0\ q\ (2\,q)\ \cdots\ ((p-1)q)$$

alors que l'étiquetage réduit est

$$\boldsymbol{\rho}_{\mathbf{r}} \;=\; \rho_0\,\rho_1\,\cdots\,\rho_{(p-1)} \;=\; 0\ 1\ 2\ \cdots\ (p-1).$$

La $(i+1)$-ème lettre du premier est le produit de $q$ par la $(i+1)$-ème lettre du second : $\forall i < p\ \ \gamma_i = q\,\rho_i$. On note $L$ le langage engendré par $\mathbf{r}$ et $\boldsymbol{\gamma}_{\mathbf{r}}$. Il s'ensuit que $\langle n\rangle_L$ est également le produit chiffre-à-chiffre de $q$ par $\langle n\rangle_{\frac{p}{q}}^{\mathrm{FK}}$. Si bien que

$$\forall n \in \mathbb{N} \qquad \pi_{\frac{p}{q}}(\langle n\rangle_L) \;=\; q\,\pi_{\frac{p}{q}}\Big(\langle n\rangle_{\frac{p}{q}}^{\mathrm{FK}}\Big) \;=\; \theta_{\frac{p}{q}}\Big(\langle n\rangle_{\frac{p}{q}}^{\mathrm{FK}}\Big) \;=\; n\ .$$

1. Si $q = 1$, alors la variante FK est identique à la définition standard (dite AFS) de la base rationnelle (et également à celle de la définition de la base entière $p$).

[35] Christiane Frougny et Karel Klouda, 2012, *Rational base number systems for p-adic numbers.*

Le théorème VII, résultat principal de la section 8.3 suivante, est la généralisation de cette observation pour n'importe quel rythme (dont le taux de croissance est $\frac{p}{q}$).

# Base rationnelle associée à un rythme

**Notation 8.26** – *On note $L_{\mathbf{r}}$ le langage engendré par un rythme donné $\mathbf{r}$ et l'étiquetage spécial associé $\boldsymbol{\gamma}_{\mathbf{r}}$.*

Le but de cette section est d'établir le théorème VII qui est, en quelque sorte, une réciproque du théorème VI. Il s'agit de montrer que pour tout rythme $\mathbf{r}$, le langage $L_{\mathbf{r}}$ est une représentation non-canonique des entiers dans la base $\frac{p'}{q'}$, le taux de croissance de $\mathbf{r}$.

Quand $\frac{p'}{q'}$ n'est pas un entier, l'existence d'un normalisateur en base $\frac{p'}{q'}$ (voir section 4.3) permet de conclure que $L_{\mathbf{r}}$ est aussi complexe que $L_{\frac{p'}{q'}}$ : c'est un langage FLIP (corollaire 8.32). Cette complexité s'étend au langage $L_{(\mathbf{r},\boldsymbol{\gamma})}$ si $\boldsymbol{\gamma}$ a suffisamment de lettres différentes, ce qui est par exemple le cas pour l'étiquetage naïf ; $K_{\mathbf{r}}$ est donc également un langage FLIP. Au contraire, quand $\frac{p'}{q'}$ est un entier, une preuve directe permet de montrer que $L_{(\mathbf{r},\boldsymbol{\gamma})}$ est toujours un langage régulier, quel que soit l'étiquetage $\boldsymbol{\gamma}$ (théorème 8.29).

Dans cette section, on considère $p$ et $q$ deux entiers $p > q \geqslant 1$ qui ne sont **pas nécessairement premiers entre eux** et un rythme $\mathbf{r}$ dont le paramètre directeur est $(q, p)$. On note $\frac{p'}{q'}$ le taux de croissance de $\mathbf{r}$, c'est-à-dire que $p'$ et $q'$ sont les quotients respectifs de $p$ et $q$ par leur PGCD.

## Langage engendré par un rythme et l'étiquetage spécial

Si $\mathbf{r}$ est un rythme de Christoffel, alors $L_{\mathbf{r}}$ est, d'après le théorème VI, égal à $L_{\frac{p'}{q'}}$ (qui est aussi $L_{\frac{p}{q}}$ car dans ce cas $p$ et $q$ sont premiers entre eux). Dans le cas général, $L_{\mathbf{r}}$ est également lié avec la base $\frac{p'}{q'}$ mais ne coïncide pas avec $L_{\frac{p'}{q'}}$ : $L_{\mathbf{r}}$ est une *représentation non-canonique* des entiers en base $\frac{p'}{q'}$, c'est-à-dire un langage sur un alphabet non-canonique ($\neq [\![p]\!]$) comportant exactement un mot de chaque valeur entière. Cette notion est définie plus formellement ci-dessous.

**Définition 8.27** – *Soit un langage $L$ sur un alphabet $B$ (fini) de chiffres considéré comme un système abstrait. Le langage $L$ est appelé une* représentation non-canonique des entiers en base rationnelle $\frac{p'}{q'}$ *si*

$$\forall n \in \mathbb{N} \qquad \pi_{\frac{p'}{q'}}(\langle n \rangle_L) \;=\; n \;.$$

**Théorème VII** – *Soit un rythme $\mathbf{r}$ de taux de croissance $\frac{p'}{q'}$. Le langage engendré par $\mathbf{r}$ et l'étiquetage spécial associé $\boldsymbol{\gamma}_{\mathbf{r}}$ est une représentation non-canonique des entiers en base $\frac{p'}{q'}$.*

On note $A_{\mathbf{r}}$ l'alphabet des lettres (différentes) de $\boldsymbol{\gamma}_{\mathbf{r}}$. La démonstration du théorème VII requiert le lemme préliminaire suivant ; celui-ci énonce que les arcs du langage calable $0^*L_{\mathbf{r}}$ satisfont essentiellement la même condition nécessaire que ceux $0^*L_{\frac{p'}{q'}}$ (*cf.* équation (8.2)).

**Lemme 8.28** – *L'implication suivante est vérifiée :*

$$\forall n, m \in \mathbb{N},\ \forall a \in A_{\mathbf{r}} \qquad n \xrightarrow[0^*L_{\mathbf{r}}]{a} m \quad \Longrightarrow \quad a \ = \ q'\,m - p'\,n \ .$$

Démonstration. Soit un arc $n \xrightarrow{a} m$ de $0^*L_{\mathbf{r}}$. On suppose d'abord que $m < p$ donc $n < p$. Il découle du lemme 8.8 que

$$R_n \leqslant m < R_{(n+1)} \ \text{ et } \ a \ = \ \gamma_m \ ,$$

et donc de la définition 8.11 de l'étiquetage spécial que $a = \gamma_m = q'\,m - p'\,n$, ce qui conclut la preuve pour ce cas particulier.

Démontrons le cas général par récurrence sur $m$ (le paragraphe précédent étant l'initialisation). Soit $(m+p)$ un entier et $n' \xrightarrow{a} (m+p)$ un arc de $L_{\mathbf{r}}$ ; on note $n$ le prédécesseur de $m$. Il découle alors du lemme 8.9 que $n' = (n+q)$ et $n \xrightarrow{a} m$ ; donc, d'après l'hypothèse de récurrence $a \ = \ q'\,m - p'\,n$. On réécrit alors le résultat attendu comme

$$q'\,(m+p) - p'\,(n+q) \ \ = \ \ q'\,m - p'\,n + (q'\,p - p'\,q) \ \ = \ \ a + (q'\,p - p'\,q) \ .$$

Or, si on note $d$ le PGCD de $p$ et $q$, $(q'\,p - p'\,q) = (q'\,p'\,d - p'\,q'\,d) = 0$, ce qui implique que $q'\,(m+p) - p'\,(n+q) = a$ et conclut ainsi la récurrence. □

Démonstration du théorème VII. On appelle $\mathbf{r}$-*représentation* d'un entier $n$ le mot $\langle n \rangle_{L_{\mathbf{r}}}$ que l'on note plus simplement $\langle n \rangle_{\mathbf{r}}$. Le théorème VII est alors équivalent à l'équation suivante :

$$\forall n \in \mathbb{N} \qquad \pi_{\frac{p'}{q'}}(\langle n \rangle_{\mathbf{r}}) \ = \ n \ . \tag{8.5}$$

Nous allons la démontrer par récurrence ; elle est évidemment vérifiée pour $\langle 0 \rangle_{\mathbf{r}} = \varepsilon$.

Soit $m > 0$ un entier ; on note $\langle m \rangle_{\mathbf{r}} = a_k\ a_{(k-1)} \cdots a_1 a_0$. Puisque $L_{\mathbf{r}}$ est clos par préfixe, il existe un entier $n$ tel que $a_k\ a_{(k-1)} \cdots a_1 = \langle n \rangle_{\mathbf{r}}$, qui vérifie donc $n \xrightarrow[L_{\mathbf{r}}]{a_0} m$. Il découle alors du lemme 8.28 précédent que

$$a_0 \ = \ q'\,m - p'\,n \quad \text{ donc que } \quad m \ = \ \frac{a_0 + p'\,n}{q'} \ . \tag{$*$}$$

Puisque $m > 0$, son prédécesseur $n$ lui est strictement inférieur (condition (**b**) de la définition 7.3 d'i-arbre) ; on peut donc lui appliquer l'hypothèse de récurrence :

$$n \ = \ \pi_{\frac{p'}{q'}}(\langle n \rangle_{\mathbf{r}}) \ = \ \pi_{\frac{p'}{q'}}(a_k\ a_{(k-1)} \cdots a_1) \ = \ \frac{1}{q'} \sum_{i=1}^{k} a_i \left(\frac{p'}{q'}\right)^{(i-1)} \ .$$

En reportant cette équation dans $(*)$, on obtient

$$\begin{aligned} m &= \frac{a_0 + p' n}{q'} \\ &= \frac{1}{q'}\left[ a_0 + \frac{p'}{q'} \sum_{i=1}^{k} a_i \left(\frac{p'}{q'}\right)^{(i-1)} \right] \\ &= \frac{1}{q'}\left[ \sum_{i=0}^{k} a_i \left(\frac{p'}{q'}\right)^{i} \right] \\ &= \pi_{\frac{p'}{q'}}(\langle m \rangle_{\mathbf{r}}) \end{aligned}$$

ce qui conclut la récurrence. □

## Complexité des langages rythmés

Dans cette sous-section 8.3.2, on montre que le taux de croissance d'un rythme détermine la complexité du langage engendré en suivant une loi du *tout ou rien*. Si le taux de croissance est un entier, le langage engendré est régulier, quel que soit l'étiquetage choisi (théorème 8.29) ; si le taux de croissance n'est pas un entier, le langage engendré est FLIP selon certaines conditions sur l'étiquetage (proposition 8.31, corollaire 8.32).

### Cas des rythmes à taux de croissance entier

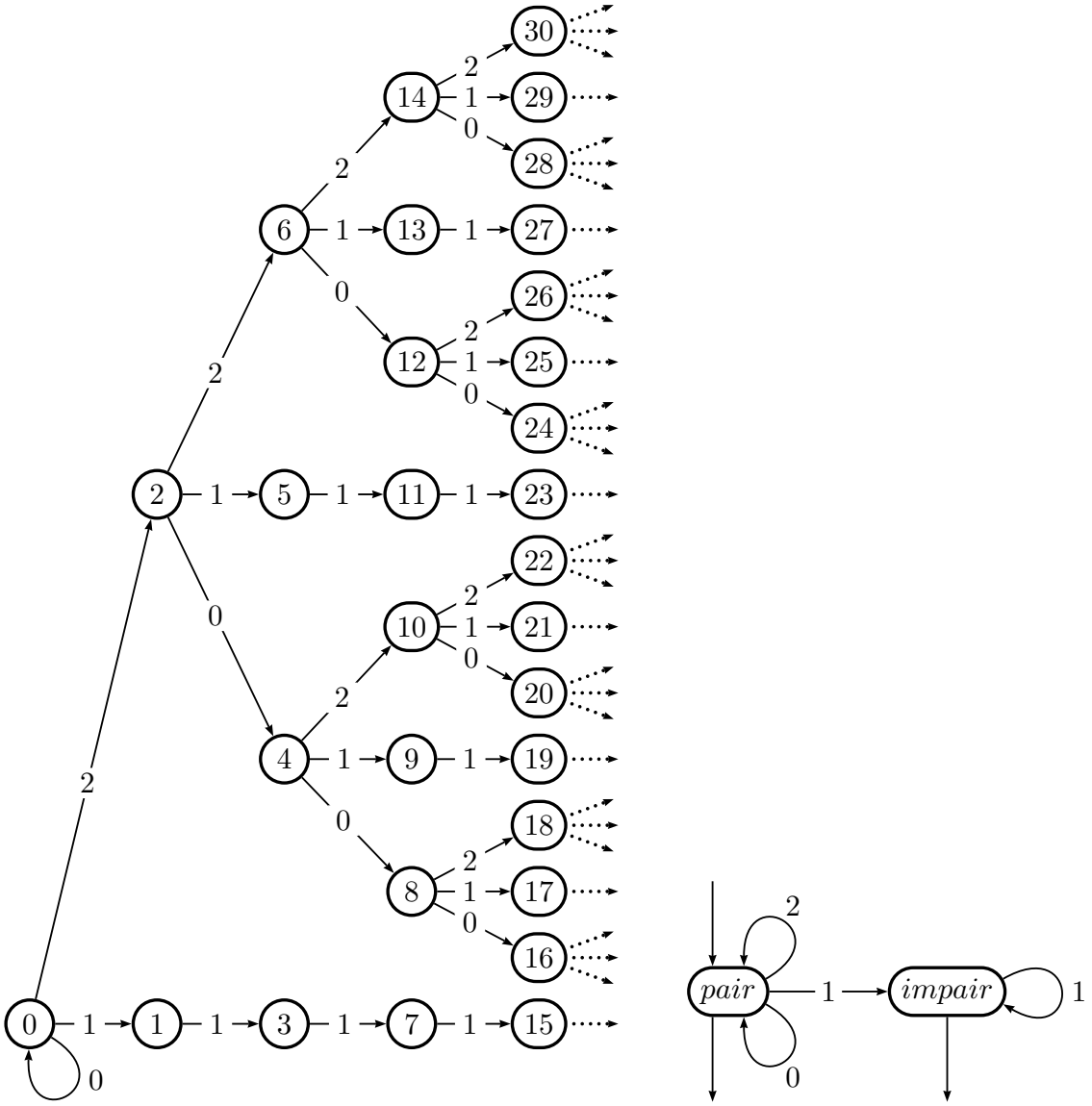
**Théorème 8.29** – *Si le taux de croissance d'un rythme* $\mathbf{r}$ *est entier, alors* $L_{(\mathbf{r},\boldsymbol{\gamma})}$ *est un langage régulier, pour tout rythme étiqueté* $(\mathbf{r}, \boldsymbol{\gamma})$ *valide.*

Démonstration. On note $\mathcal{T}_{(\mathbf{r},\boldsymbol{\gamma})}$ l'automate **infini** défini comme l'i-tree étiqueté de $0^* L_{(\mathbf{r},\boldsymbol{\gamma})}$ et dont tous les états sont finals. Si bien que $L(\mathcal{T}_{(\mathbf{r},\boldsymbol{\gamma})}) = 0^* L_{(\mathbf{r},\boldsymbol{\gamma})}$. Montrons, par récurrence sur la longueur de $u$ que si deux entiers $n$ et $m$ sont congrus modulo $q$ alors les états correspondants dans $\mathcal{T}_{(\mathbf{r},\boldsymbol{\gamma})}$ sont 'Nérode-équivalents', c'est-à-dire (puisque tous les états de $\mathcal{T}_{(\mathbf{r},\boldsymbol{\gamma})}$ sont finals) que $(n \cdot u)$ existe si et seulement si $(m \cdot u)$ existe pour tout mot $u \in [\![p]\!]^*$.

Ceci est évidemment vrai pour $u = \varepsilon$. Soit un mot $a\,u$ tel que $(n \cdot a\,u)$ existe, et soit $n'$ le nœud tel que $n \xrightarrow{a} n'$. Du lemme 8.9 découle que $m \xrightarrow{a} m'$ pour un certain $m'$ qui est congru à $n'$ modulo $p$ ; d'où, puisque $q \mid p$, que $n' \equiv m'\ [q]$. Par hypothèse de récurrence, $(m' \cdot u) = (m \cdot a \cdot u)$ existe puisque $(n' \cdot u)$ existe. Un raisonnement symétrique montre que l'existence de $(m \cdot u)$ implique celle de $(n \cdot u)$.

L'automate minimal qui accepte $L_{(\mathbf{r},\boldsymbol{\gamma})}$ possède donc au plus $q$ états. □

La figure 10 représente le langage calable $0^* L_{31}$ et l'automate qui l'accepte ; il s'agit (comme l'assure le théorème VII) d'une représentation non-canonique des entiers en base 2. D'autre part, la figure 11 donne l'automate acceptant le langage calable $0^* K_{3\,2\,1}$ dont l'arbre sous-jacent était l'exemple utilisé pour la génération par signature (figure 3, page 179).




(a) Le langage calable $0^* L_{31}$ (b) Automate acceptant $0^* L_{31}$

Figure 10 – Langage des Représentations non-canonique des entiers en base 2 engendré par le rythme 31 et l'étiquetage spécial $\gamma_{31} = 0121$

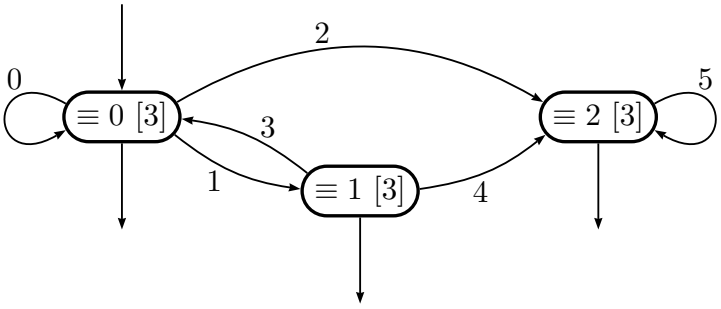



Figure 11 – Automate acceptant le langage calable $0^* K_{321}$

La combinaison des théorèmes V (page 15), qui énonce qu'un langage est régulier si et seulement si sa signature est s-morphique, et 8.29 établit le corollaire suivant. Néanmoins, passer par le théorème V est une manière compliquée pour établir le corollaire 8.30 dont la démonstration directe est simple.

**Corollaire 8.30** – *Soit un rythme* $\mathbf{r}$ *. Si le taux de croissance* $\mathbf{r}$ *est un entier, alors* $\mathbf{r}^\omega$ *est une signature s-morphique.*

Démonstration directe. Soit $\mathbf{r} = r_0\, r_1 \cdots r_{(q-1)}$ un rythme dont le paramètre directeur est $(q,\, k\,q)$ pour certains entiers $q$ et $k$.

On considère les alphabets $[\![q]\!]$ et $[\![k\,q]\!]$ et et on définit le morphisme $\varphi : [\![q]\!]^* \to [\![k\,q]\!]^*$ par :

$$\forall i \in [\![q]\!] \qquad \varphi(i) \;=\; R_i\,(R_i+1)\,(R_i+2)\cdots(R_i+r_i-1)\;.$$

Puisque $R_{(i+1)} = (R_i + r_i)$, la définition de $\varphi$ implique immédiatement que

$$\varphi\big(0\,1\,2\cdots(q-1)\big) \;=\; 0\,1\,2\cdots(k\,q-1)\;. \qquad (*)$$

On note maintenant $\sigma$ l'endomorphisme $[\![q]\!]^* \to [\![q]\!]^*$ qui projette chaque lettre du résultat de $\varphi$ sur l'alphabet $[\![q]\!]$ :

$$\forall i \in [\![q]\!] \qquad \sigma(i) \;=\; (\beta_0 \,\%\, q)\,(\beta_1 \,\%\, q)\cdots(\beta_j \,\%\, q) \qquad \text{si } \varphi(i) \;=\; \beta_0\,\beta_1\cdots\beta_j\;.$$

Il découle alors de l'équation $(*)$ que

$$\sigma\big(0\,1\,2\cdots(q-1)\big) \;=\; \big(0\,1\,2\cdots(q-1)\big)^k\;,$$

si bien que $(0\,1\,2\cdots(q-1))^\omega$ est un point fixe de $\sigma$.

Pour tout $i \in [\![q]\!]$, $|\sigma(i)| = r_i$ ce qui implique que $|\sigma(0\,1\,2\cdots(i-1))| = R_i$. La validité de $\mathbf{r}$ implique donc que $\big|\sigma\big(0\,1\,2\cdots(i-1)\big)\big| > i$, pour tout $i \in [\![q]\!]$, si bien que

$$\lim_{n\to+\infty} |\sigma^n(0)| \;=\; +\infty$$

donc que $\sigma$ est prolongeable en 0 et enfin que $\sigma^\omega(0) = (0\,1\,2\cdots(q-1))^\omega$.

La signature induite par $\sigma$ est donc

$$f_\sigma\big(\sigma^\omega(0)\big) \;=\; f_\sigma\Big(\big(0\,1\,2\cdots(q{-}1)\big)^\omega\Big) \;=\; \Big(\big|\sigma(0)\big|\,\big|\sigma(1)\big|\,\big|\sigma(2)\big|\cdots\big|\sigma(q{-}1)\big|\Big)^\omega \;=\; \mathbf{r}^\omega\;.$$

□

### Cas des rythmes à taux de croissance non-entier

A l'opposé du théorème 8.29, si un rythme a un taux de croissance qui n'est pas entier, il engendre des langages complexes. La proposition suivante traite le cas où le rythme est couplé à l'étiquetage spécial.

**Proposition 8.31** – *Soit un rythme* $\mathbf{r}$ *dont on note le taux de croissance* $\frac{p'}{q'}$*. Si* $\frac{p'}{q'}$ *n'est pas un entier, alors* $L_{\mathbf{r}}$ *est un langage FLIP.*

Démonstration. On rappelle que $A_{\mathbf{r}}$ est l'alphabet des lettres de $\boldsymbol{\gamma}_{\mathbf{r}}$ donc l'alphabet de $L_{\mathbf{r}}$. D'après l'équation (8.5) (équivalente au théorème VII), pour tout entier $n$, $\pi_{\frac{p'}{q'}}(\langle n\rangle_{\mathbf{r}}) = n$. Ceci implique que le normalisateur $\mathcal{C}_{A_{\mathbf{r}}}$ (défini à la section 4.3,

page 100) est un transducteur droit, lettre-à-lettre et séquentiel qui réalise la fonction

$$\forall n \in \mathbb{N} \qquad \langle n \rangle_{\mathbf{r}} \mapsto \langle n \rangle_{\frac{p'}{q'}} \ .$$

Il s'ensuit que $\mathcal{C}_{A_{\mathbf{r}}}(L_{\mathbf{r}}) = L_{\frac{p'}{q'}}$.

Par L'absurde. Supposons qu'il existe un langage régulier infini $K \subseteq L_{\mathbf{r}}$. Dans ce cas, $L_{\frac{p'}{q'}}$ contient le langage régulier $\mathcal{C}_{A_{\mathbf{r}}}(K)$ ; ce dernier est de plus infini, car les mots de $K$ ont tous des valeurs différentes donc $\mathcal{C}_{A_{\mathbf{r}}}$ (qui conserve la valeur) est injectif sur $K$. Il existe alors un langage régulier infini inclus dans le langage FLIP $L_{\frac{p'}{q'}}$, contradiction.

Ainsi, $L_{\mathbf{r}}$ ne contient pas de langage régulier infini ; puisqu'il est clos par préfixe, il est donc FLIP. □

Puisque $L_{\mathbf{r}}$ est l'image de $K_{\mathbf{r}}$ par un morphisme strictement alphabétique, le corollaire suivant en découle immédiatement.

**Corollaire 8.32** – *Soit un rythme* $\mathbf{r}$ *de taux de croissance* $\frac{p'}{q'}$*. Si* $\frac{p'}{q'}$ *n'est pas un entier alors* $K_r$ *est un langage FLIP.*

On ne semble pas pouvoir généraliser les résultats précédents (proposition 8.31 et corollaire 8.32) à tous les langages engendrés par des rythmes étiquetés $(\mathbf{r}, \boldsymbol{\gamma})$ tels que $\mathbf{r}$ a un taux de croissance non-entier. Par exemple l'étiquetage réduit ne produit généralement pas des langages FLIP. Il suffit en effet que la branche minimale issue d'un seul sommet soit infinie (c'est-à-dire n'aboutisse jamais sur un nœud sans arcs sortants) pour que celle-ci soit étiquetée par $0^*$ et donc que le langage ne soit pas FLIP. Une exception notable est le langage des $\frac{p}{q}$-représentations dans la variante FK (voir section 8.2.3). On peut néanmoins conjecturer que l'étiquetage réduit engendre un langage (au moins) non-algébrique, même s'il n'est pas FLIP.

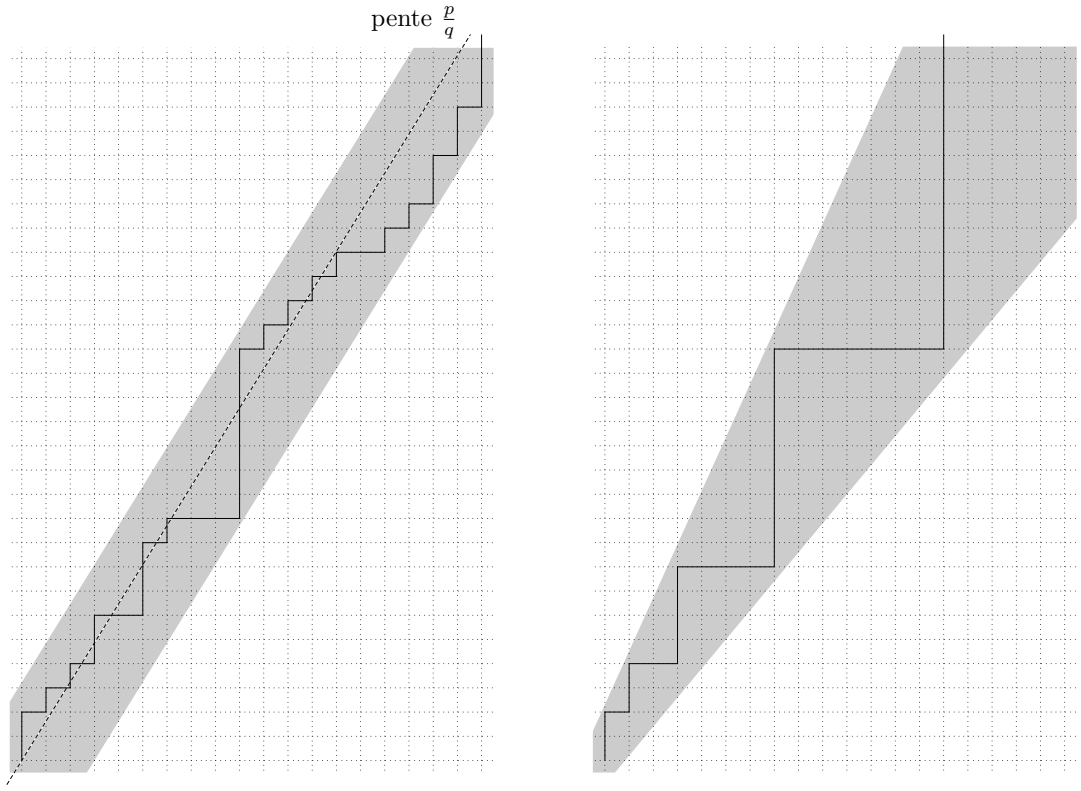
## Généralisation

La méthode utilisée précédemment dans cette section 8.3 n'utilise pas directement le fait que la signature considérée est purement périodique. En réalité, n'importe quelle signature pour laquelle on peut adapter l'étiquetage spécial permet le même raisonnement. C'est le cas par exemple pour les signatures ultimement périodiques ou, plus généralement, pour les signatures *dirigées par* $\frac{p}{q}$.

**Définition 8.33** – *Soient deux entiers* $p$ *et* $q$ *premiers entre eux tels que* $p > q \geqslant 1$*. une signature* $\mathbf{s} = s_0 s_1 \cdots s_i \cdots$ *est dite* dirigée par $\frac{p}{q}$ *si la suite* $(q\,S_i - p\,i)_{i \in \mathbb{N}}$ *est bornée.* [2]

Une définition plus graphique mais équivalente peut s'énoncer comme : *une signature est dirigée par* $\frac{p}{q}$ *si le chemin associé dans le plan discret* $\mathbb{Z} \times \mathbb{Z}$ *est confiné entre deux droites parallèles de pente* $\frac{p}{q}$. La figure 12 montre par exemple les chemins associés à deux signatures, l'une dirigée, l'autre non.

2. On rappelle que pour tout entier $i$, $S_i$ est la somme des $i$ premiers termes de $\mathbf{s}$ : $S_i = \sum_{k=0}^{i-1} s_k$.




(a) signature dirigée (b) signature qui n'est pas dirigée

Figure 12 – Direction d'une signature

**Lemme 8.34** – *Toute signature ultimement périodique* $\mathbf{s} = \mathbf{t}\,\mathbf{r}^\omega$ *est dirigée par le taux de croissance de* $\mathbf{r}$.

Démonstration. On note $(i, j)$ le paramètre directeur de $\mathbf{t}$ et $(q, p)$ celui de $\mathbf{r}$, ainsi que $\frac{p'}{q'}$ le taux de croissance de $\mathbf{r}$. Le chemin associé à $\mathbf{s}$ va d'abord de $(0, 0)$ à $(i, j)$ puis passe par les points $(i + k\,q, j + k\,p)$, pour tout entier $k$.

Il s'ensuit que ce chemin est ultimement confiné entre les droites de pente $\frac{p'}{q'}$ qui passent par les points $(i + p, j)$ et $(i, j + q)$, respectivement. Le premier point est atteint si $\mathsf{path}(\mathbf{r})$ est de la forme $y^*x^*$ (c'est-à-dire si $\mathbf{r} = p\,0^{q-1}$) et le second s'il est de la forme $x^*y^*$ (c'est-à-dire si $\mathbf{r} = 0^{q-1}\,p$). La figure 13 donne un support visuel pour ce raisonnement. □

On peut également considérer deux rythmes $\mathbf{r}$ et $\mathbf{r}'$ qui ont le même taux de croissance $\frac{p}{q}$ et les concaténer dans n'importe quel ordre ; la signature obtenue sera également dirigée par $\frac{p}{q}$, car confinée entre deux droites extrémales parmi celles qui confinent $\mathbf{r}^\omega$ et $(\mathbf{r}')^\omega$. Un raisonnement analogue démontre le lemme suivant.

**Lemme 8.35** – *Soient deux entiers* $p$ *et* $q$ *premiers entre eux tels que* $p > q \geqslant 1$. *Soit* $\mathbf{R}$ *un ensemble* ***fini*** *de rythmes dont le taux de croissance est* $\frac{p}{q}$ *et* $\mathbf{t}$ *un rythme quelconque. La concaténation de* $\mathbf{t}$ *avec une nombre infini de copies de rythmes de* $\mathbf{R}$ *produit une signature dirigée par* $\frac{p}{q}$.

Ce dernier lemme ne permet pas de construire toutes les signatures dirigées par $\frac{p}{q}$ ;

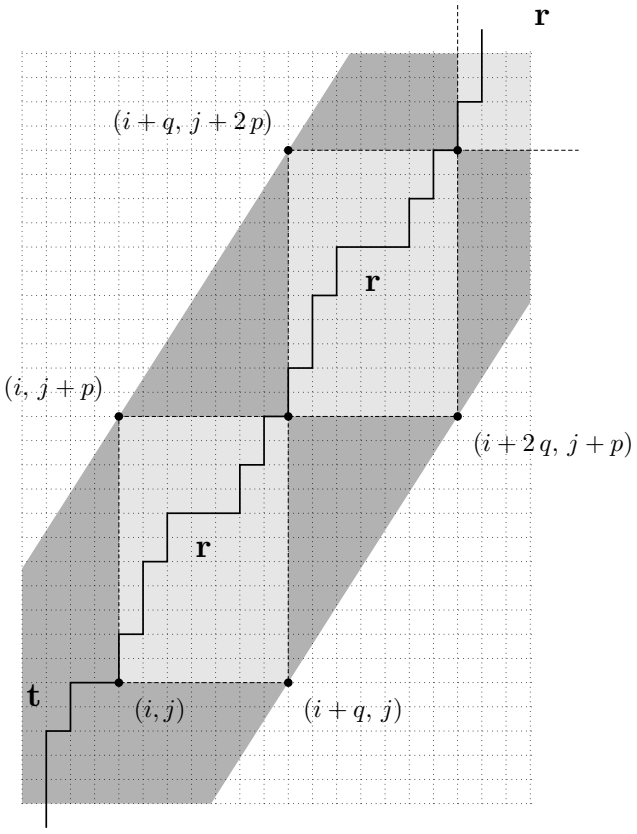



Figure 13 – Les signatures ultimement périodiques sont dirigées

par exemple, la signature suivante est dirigée par $\frac{p}{q}$ :

$$\mathbf{s} \;=\; v_0\, u_0\, v_1\, u_1 \cdots v_i\, u_i \cdots \qquad \text{où} \qquad \forall i \in \mathbb{N} \;\; u_i \;=\; (2\,(21)^i\,1) \;\text{ et }\; v_i \;=\; (1\,(21)^i\,2)\;.$$

D'autre part, on ne peut pas généraliser le lemme 8.35 au cas où $\mathbf{R}$ serait infini, comme le montre le contre-exemple suivant :

$$\mathbf{s} \;=\; w_0\, w_1 \cdots w_i \cdots \qquad \text{où} \qquad \forall i \in \mathbb{N} \quad w_i \;=\; 2^i\, 1^i\;.$$

On définit alors l'étiquetage spécial sur les signatures dirigées par $\frac{p}{q}$ en généralisant celui sur les rythmes dont le taux de croissance est $\frac{p}{q}$.

**Définition 8.36** – *Soient deux entiers $p$ et $q$ premiers entre eux tels que $p > q \geqslant 1$. Soit une signature $\mathbf{s} = s_0\, s_1 \cdots s_k \cdots$ dirigée par $\frac{p}{q}$. Pour tout entier $m$, on note $n$ est l'entier tel que $S_n \leqslant m < S_{(n+1)}$ et $\lambda_m = q\,m - p\,n$. On définit l'étiquetage spécial $\boldsymbol{\lambda}_{\mathbf{s}}$ comme $\boldsymbol{\lambda}_{\mathbf{s}} = \lambda_0\, \lambda_1 \cdots \lambda_k \cdots$*

*On note de plus $L_{\mathbf{s}}$ le langage engendré par $\mathbf{s}$ et $\boldsymbol{\lambda}_{\mathbf{s}}$.*

**Lemme 8.37** – *Soient deux entiers $p$ et $q$ premiers entre eux tels que $p > q \geqslant 1$. Soit une signature $\mathbf{s} = s_0\, s_1 \cdots s_k \cdots$ dirigée par $\frac{p}{q}$. Alors l'étiquetage spécial associé $\boldsymbol{\lambda}_{\mathbf{s}}$ est sur un alphabet fini.*

Démonstration. Soient $X = \max\{\mathsf{abs}(q\,S_i - p\,i) \mid i \in \mathbb{N}\}$ et $Y = \max\{s_i \mid i \in \mathbb{N}\}$.

Soit un entier $m$, on note $n$ l'entier tel que $S_n \leqslant m < S_{(n+1)}$ ce qui implique donc que la $(m+1)$-ème lettre de $\boldsymbol{\lambda}_{\mathbf{s}}$ est $\lambda_m = (q\,m - p\,n)$. Puisque $S_n \leqslant m < S_{(n+1)}$, l'entier $\mathsf{abs}(\lambda_m) = \mathsf{abs}(p\,n - q\,m)$ satisfait l'une des deux inéquations suivantes (suivant le signe de $\lambda_m$) :

$$\mathsf{abs}(q\,m - p\,n) \;\leqslant\; \mathsf{abs}(q\,S_n - p\,n) \;\leqslant\; X$$
$$\mathsf{abs}(q\,m - p\,n) \;\leqslant\; \mathsf{abs}(q\,S_{(n+1)} - p\,n) \;=\; \mathsf{abs}(q\,(S_n + s_n) - p\,n) \;\leqslant\; X + q\,Y\ .$$

Dans les deux cas, $\mathsf{abs}(\lambda_m)$ est donc inférieur à $(X + q\,Y)$, qui ne dépend pas de $m$. Les lettres de $\boldsymbol{\lambda}_{\mathbf{s}}$ appartiennent donc toutes à l'alphabet fini de chiffres $\{-Z, -(Z-1), \ldots, (Z-1), Z\}$, où $Z = (X + q\,Y)$. □

La démonstration du théorème VII reposait principalement sur le lien entre $a$, $n$ et $m$ quand l'arc $n \xrightarrow{\ a\ } m$ existe, qui est matérialisé par le lemme 8.28 pour les signatures périodiques. Le lemme 8.38 suivant (qui est essentiellement une reformulation de la définition de l'étiquetage spécial) en est le pendant pour les signatures dirigées ; si bien que le théorème 8.39 s'établit ensuite par une simple copie de la démonstration du théorème VII.

**Lemme 8.38** – *Soient deux entiers $p$ et $q$ premiers entre eux tels que $p > q \geqslant 1$. Soit une signature $\mathbf{s} = s_0\,s_1 \cdots s_k \cdots$ dirigée par $\frac{p}{q}$. Alors, l'implication suivante est vérifiée :*

$$\forall n, m \in \mathbb{N},\ \ \forall a \in A_{\mathbf{r}} \qquad n \xrightarrow[L_{\mathbf{s}}]{a} m \quad \Longrightarrow \quad a \;=\; q\,m - p\,n\ .$$

Démonstration. On note $\boldsymbol{\lambda}_{\mathbf{s}} = \lambda_0\,\lambda_1 \cdots \lambda_k \cdots$ l'étiquetage spécial associé à $\mathbf{s}$. Soit $n \xrightarrow[L_{\mathbf{s}}]{a} m$ un arc de $L_{\mathbf{s}}$, si bien que $a = \lambda_m$.

Il découle du lemme 7.15 (page 181) que $S_n \leqslant m < S_{(n+1)}$, ce qui implique, d'après la définition de l'étiquetage spécial que $a = \lambda_m = q\,m - p\,n$. □

**Théorème 8.39** – *Soient deux entiers $p$ et $q$ premiers entre eux tels que $p > q \geqslant 1$. Soit une signature $\mathbf{s} = s_0\,s_1 \cdots s_k \cdots$ dirigée par $\frac{p}{q}$. Le langage $L_{\mathbf{s}}$ est une représentation non-canonique des entiers en base $\frac{p}{q}$.*

**Corollaire 8.40** – *Soient deux entiers $p$ et $q$ premiers entre eux tels que $p > q \geqslant 1$. Soit une signature $\mathbf{s} = s_0\,s_1 \cdots s_k \cdots$ dirigée par $\frac{p}{q}$. Si $\frac{p}{q}$ n'est pas un entier, alors le langage $L_{\mathbf{s}}$ est FLIP.*

Bien qu'une signature dirigée par un entier $p$ engendre bel et bien une représentation canonique des entiers en base $p$ (théorème 8.39), celui-ci n'est pas nécessairement un langage régulier.

## Chapitre 9

# Surminimisation

Ce chapitre présente la notion de *surminimisation d'un automate*. Elle dérive du formalisme des signature/étiquetage de langages clos par préfixe (décrit dans le chapitre 7) mais en dépasse le cadre, car elle s'applique à des langages qui ne sont pas clos par préfixe. Conserver la généralité des résultats demande donc une adaptation de ce formalisme que nous avons jugée trop lourde. Nous donnons donc dans ce chapitre une présentation qui est presque indépendante des autres chapitres de cette troisième partie.

Nous ne considérons que des automates déterministes sur des alphabets ordonnés, si bien que les transitions sortantes de chaque état sont également ordonnées : une transition étiquetée par $a$ est inférieure aux transitions étiquetées par des lettres plus grandes que $a$.

Le processus de surminimisation d'un automate $\mathcal{A}$ s'effectue en deux étapes. On réalise d'abord un réétiquetage des transitions qui conserve (uniquement) leur ordre ; la plus petite transition est réétiquetée par la lettre 0, la deuxième par 1, etc. La seconde étape est simplement une minimisation de l'automate résultant de la première.

La surminimisation induit sur les automates une relation d'équivalence que l'on appelle *T-équivalence* : deux automates sont T-équivalents si leurs surminimisations sont isomorphes. De plus, le processus de surminimisation est idempotent, c'est-à-dire que la surminimisation est l'identité sur les automates résultants d'une première surminimisation. Ceci implique que chaque classe de T-équivalence comporte un représentant canonique : la surminimisation d'un élément quelconque de la classe.

Si deux automates émondés sont équivalents, alors leurs surminimisations respectives sont isomorphes. Ceci permet un relèvement de la T-équivalence aux langages réguliers (sur des alphabets ordonnés) : deux langages sont dit T-équivalents s'ils sont acceptés par deux automates émondés T-équivalents. De même, on appelle réduction de l'étiquetage d'un langage $L$, le langage accepté par la surminimisation de tout automate émondé acceptant $L$. Un langage est dit *irréductible* s'il est égal à sa réduction d'étiquetage.

Un système de numération abstrait régulier (SNAR, voir section 1.7 page 28) n'est rien d'autre qu'un langage régulier sur un alphabet ordonné. Nous utilisons

Les résultats présentés dans ce chapitre ont été publiés dans les actes de DLT 2015, voir [51].

donc pour les SNAR toutes les notions définies sur les langages (comme la T-équivalence, la réduction d'étiquetage, etc.). Un SNAR est essentiellement déterminé par l'ordre des mots du langage qui le définit, un ordre que la réduction d'étiquetage conserve justement, c'est pourquoi deux SNAR T-équivalents sont pratiquement égaux en tant que systèmes de numération.

**Théorème VIII** – *La fonction de conversion entre deux SNAR T-équivalents est réalisée par un transducteur fini, lettre-à-lettre et séquentiel pur.*

La réciproque de ce théorème est fausse en général ; il existe en effet des SNAR qui ne sont pas T-équivalents mais que l'on peut néanmoins convertir l'un en l'autre grâce à un transducteur fini, lettre-à-lettre et séquentiel pur. On appelle *localement croissant* un transducteur qui conserve localement l'ordre des lettres et on démontre une réciproque faible du théorème VIII.

**Théorème (9.20)** – *Soient $L$ et $K$ deux SNAR. Si la fonction de conversion de $L$ en $K$ est réalisée par un transducteur fini, lettre-à-lettre, localement croissant et séquentiel pur, alors $L$ et $K$ sont T-équivalents.*

Dans la section 9.3 on s'intéresse aux systèmes (de numération) positionnels (*cf.* [36]). Il est connu depuis longtemps que tous les systèmes positionnels sont des SNA mais ils ne sont pas toujours des SNAR (il existe des conditions nécessaires dues à Shallit [76] et Hollander [39]).

On généralise la réduction d'étiquetage aux langages quelconques (c'est-à-dire, pas nécessairement réguliers). On établit ensuite que l'algorithme glouton utilisé pour calculer les représentations des entiers dans un système positionnels $U$ assure l'irréductibilité du langage $L(U)$ des représentations des entiers, comme l'exprime le théorème suivant.

**Théorème IX** – *Pour tout système positionnel $U$, le langage calable $0^*L(U)$ est irréductible.*

# Étiquetage réduit et surminimisation

Dans la suite, tous les alphabets considérés sont **totalement ordonnés** et tous les automates considérés sont **finis et déterministes** ; de plus à partir de la sous-section 9.1.2, on ne considérera plus que les automates **émondés**.

Soit $\mathcal{A} = \langle Q_\mathcal{A}, A, \delta_\mathcal{A}, i_\mathcal{A}, F_\mathcal{A} \rangle$ un automate (fini déterministe, dont l'alphabet $A$ est ordonné). On note $\mathsf{Out}_\mathcal{A}(s)$ l'ensemble des transitions sortantes d'un

[36] Christiane Frougny et Jacques Sakarovitch, 2010, *Number representation and finite automata*.

[76] Jeffrey Shallit, 1994, *Numeration Systems, Linear Recurrences, and Regular Sets*.

[39] M. Hollander, 1998, *Greedy Numeration Systems and Regularity.*.

état $s$ et $\mathsf{d}_{\mathcal{A}}(s) = \mathsf{Card}(\mathsf{Out}_{\mathcal{A}}(s))$ le degré sortant de $s$ ; quand il n'y a pas d'ambiguïté sur l'automate, on écrit plus simplement $\mathsf{Out}(s)$ et $\mathsf{d}(s)$. De plus, on note $\mathsf{d}(\mathcal{A}) = \max\{\, \mathsf{d}_{\mathcal{A}}(s) \mid s \in Q_{\mathcal{A}} \,\}$ le degré sortant maximal parmi les états de $\mathcal{A}$.

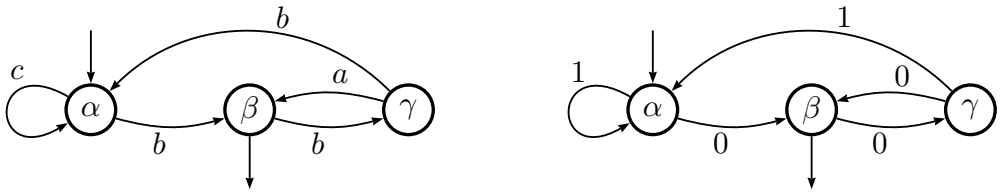
Pour tout $s$ de $Q_{\mathcal{A}}$, l'ordre de $A$ induit un ordre sur $\mathsf{Out}_{\mathcal{A}}(s)$ :

$$[\, s \xrightarrow{a} s' \,] \;<\; [\, s \xrightarrow{b} s'' \,] \quad \text{si} \;\; a < b \;.$$

On appelle *$i$-ème transition de $\mathsf{Out}(s)$* la transition ayant exactement $(i-1)$ transitions plus petites qu'elle-même dans $\mathsf{Out}(s)$.

## Réduction d'étiquetage

Tout d'abord, nous définissons un processus appelé *réduction d'étiquetage*, qui consiste à réétiqueter, pour chaque état $s$, les transitions de $\mathsf{Out}(s)$ en utilisant l'alphabet $[\![\mathsf{d}(s)]\!]$, de telle sorte que l'ordre de $\mathsf{Out}(s)$ soit conservé.




(a) Un automate $\mathcal{A}_1$ (b) $\mathsf{lred}(\mathcal{A}_1)$

Figure 1 – Réduction de l'étiquetage d'un automate

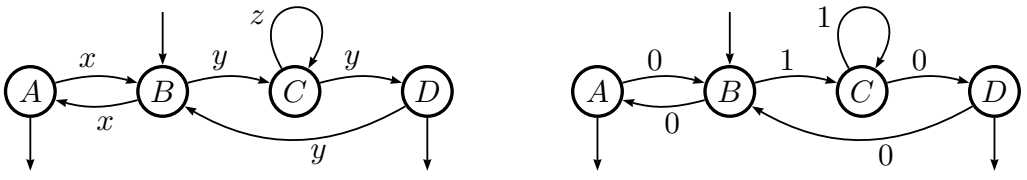
**Définition 9.1** – *Soit $\mathcal{A} = \langle\, Q_{\mathcal{A}}, A, \delta_{\mathcal{A}}, i_{\mathcal{A}}, F_{\mathcal{A}} \,\rangle$ un automate (déterministe). On appelle la réduction d'étiquetage de $\mathcal{A}$, noté $\mathsf{lred}(\mathcal{A})$, l'automate*

$$\mathsf{lred}(\mathcal{A}) \;=\; \langle\, Q_{\mathcal{A}},\, [\![\mathsf{d}(\mathcal{A})]\!],\, \delta',\, i_{\mathcal{A}},\, F_{\mathcal{A}} \,\rangle$$

*où $\delta'$ est définie par : pour tout état $s$ de $Q_{\mathcal{A}}$, si, dans $\mathcal{A}$, $s \xrightarrow{a} s_i$ est la $(i+1)$-ème transition de $\mathsf{Out}_{\mathcal{A}}(s)$, alors $s \xrightarrow{i} s_i$ existe dans $\mathsf{lred}(\mathcal{A})$.*

*Par définition, $\mathsf{lred}(\mathcal{A})$ est aussi un automate déterministe.*

Les figures 1 et 2 montrent deux exemples de réduction d'étiquetage. Ce processus commute aux morphismes d'automates, comme l'exprime le lemme suivant.




(a) Un automate $\mathcal{A}_2$ (b) $\mathsf{lred}(\mathcal{A}_2)$

Figure 2 – Réduction de l'étiquetage d'un autre automate

**Lemme 9.2** – *Soient deux automates $\mathcal{A}$ et $\mathcal{M}$. Si $\mathcal{M}$ est un quotient de $\mathcal{A}$, alors $\mathsf{lred}(\mathcal{M})$ est un quotient de $\mathsf{lred}(\mathcal{A})$.*

Démonstration. On note $\varphi : \mathcal{A} \to \mathcal{M}$ le morphisme d'automates associé au quotient. Les ensembles d'états de $\mathcal{A}$ et de $\mathsf{lred}(\mathcal{A})$ sont égaux, tout comme ceux de $\mathcal{M}$ et de $\mathsf{lred}(\mathcal{M})$, si bien que $\varphi$ est également une fonction des états de $\mathsf{lred}(\mathcal{A})$ vers les états de $\mathsf{lred}(\mathcal{M})$. Démontrons que $\varphi$ est de plus un morphisme d'automates $\mathsf{lred}(\mathcal{A}) \to \mathsf{lred}(\mathcal{M})$ ; puisque la réduction d'étiquetage ne modifie pas l'état initial ni les états finals, il suffit donc de montrer que $\varphi$ respecte les transitions (équation (1.1c)).

Soit $s$ un état de $\mathsf{lred}(\mathcal{A})$. Le morphisme $\varphi$ induit une bijection $\mathsf{Out}_{\mathcal{A}}(s) \to \mathsf{Out}_{\mathcal{M}}(s)$ (puisque $\mathcal{A}$ est déterministe) que l'on continue de noter $\varphi$ et qui conserve l'étiquette, et donc l'ordre. Si bien que $\varphi$ est également une bijection $\mathsf{Out}_{\mathsf{lred}(\mathcal{A})}(s) \to \mathsf{Out}_{\mathsf{lred}(\mathcal{M})}(s)$ qui conserve l'étiquette. □

La proposition suivante est une conséquence presque immédiate du lemme précédent.

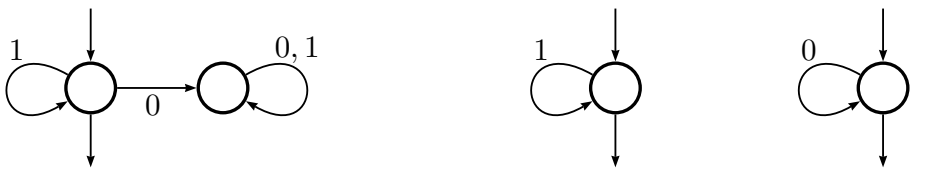
**Proposition 9.3** – *Soient deux automates émondés $\mathcal{A}$ et $\mathcal{B}$. Si $\mathcal{A}$ et $\mathcal{B}$ sont équivalents, alors $\mathsf{lred}(\mathcal{A})$ et $\mathsf{lred}(\mathcal{B})$ le sont également.*

Démonstration. On note $L$ le langage accepté par $\mathcal{A}$ et $\mathcal{B}$. L'automate émondé minimal acceptant $L$, noté $\mathcal{M}$, est donc un quotient de $\mathcal{A}$ et de $\mathcal{B}$ (puisqu'ils sont tous les deux émondés). D'après le lemme 9.2, $\mathsf{lred}(\mathcal{M})$ est donc un quotient de $\mathsf{lred}(\mathcal{A})$ et de $\mathsf{lred}(\mathcal{B})$, donc (puisque les quotients conservent le langage accepté)

$$L(\mathsf{lred}(\mathcal{B})) \;=\; L(\mathsf{lred}(\mathcal{M})) \;=\; L(\mathsf{lred}(\mathcal{A}))\ . \qquad \square$$

L'hypothèse "émondé" est indispensable dans la proposition 9.3. Par exemple, les figures 3a et 3b représentent deux automates $\mathcal{A}_3$ et $\mathcal{A}_4$ acceptant le langage $1^*$. Le premier n'est pas émondé et est égal à sa réduction d'étiquetage, alors que réduire l'étiquetage du second donne l'automate montré à la figure 3c. Si bien que

$$L(\mathsf{lred}(\mathcal{A}_3)) \;=\; 1^* \;\neq\; 0^* \;=\; L(\mathsf{lred}(\mathcal{A}_4))\ .$$




(a) $\mathcal{A}_3 = \mathsf{lred}(\mathcal{A}_3)$ (b) $\mathcal{A}_4$ (c) $\mathsf{lred}(\mathcal{A}_4)$

Figure 3 – Automates non-émondés et étiquetage réduit

Dans la suite, nous ne considérons que des automates **émondés**. La proposition 9.3 permet alors de parler de réduction de l'étiquetage pour les langages.

**Définition 9.4** – *Soit un langage $L$ (sur un alphabet ordonné) accepté par un automate (émondé) $\mathcal{A}$. On note $\mathsf{lred}(L)$ le langage accepté par $\mathsf{lred}(\mathcal{A})$.*

*Si $L = \mathsf{lred}(L)$, on dit que $L$ est* irréductible.

Par exemple, réduire l'étiquetage du langage $L_5 = ((a + b^*)\,c)^*$ produit le langage $\mathsf{lred}(L_5) = (00 + 10^* + 2)^*$.

## T-équivalence et surminimisation

**Définition 9.5** – *Deux automates $\mathcal{A}$ et $\mathcal{B}$ sont dit* T-équivalents*, noté $\mathcal{A} \overset{T}{\sim} \mathcal{B}$, si leurs réductions d'étiquetage respectives sont équivalentes :*

$$L(\mathsf{lred}(\mathcal{A})) = L(\mathsf{lred}(\mathcal{B}))\ .$$

*Deux langages sont* T-équivalents *si et seulement si leurs réductions d'étiquetage respectives sont égales.*

Par exemple, les automates $\mathcal{A}_1$ et $\mathcal{A}_2$ (figures 1a et 2a) sont T-équivalents. On peut en effet vérifier que $\mathsf{lred}(\mathcal{A}_1)$ et $\mathsf{lred}(\mathcal{A}_2)$ (figures 1b et 2b) acceptent le même langage en comparant leurs minimisations respectives : elles sont isomorphes (figure 4). Ce procédé est formalisé sous le nom de *surminimisation*, définie ci-dessous.

**Définition 9.6** – *On appelle surminimisation d'un automate $\mathcal{A}$, la minimisation de la réduction d'étiquetage de $\mathcal{A}$ : $\mathsf{surmin}(\mathcal{A}) = \mathsf{minim}(\mathsf{lred}(\mathcal{A}))$.*

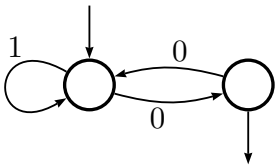



Figure 4 – Surminimisation commune de $\mathcal{A}_1$ et de $\mathcal{A}_2$

La proposition suivante découle immédiatement de la définition précédente ; elle donne à la fois une caractérisation et un algorithme efficace pour décider la T-équivalence.

**Proposition 9.7** – *Deux automates sont T-équivalents si et seulement si leurs surminimisations respectives sont isomorphes.*

**Remarque 9.8** – *Réduire l'étiquetage supprime la* signification *(éventuelle) des lettres et ne conserve que leur ordre. Par exemple, le langage $L_6 = 0^*1^*$ peut être décrit comme* des 0 suivis de 1 *alors que sa réduction d'étiquetage est $\mathsf{lred}(L_6) = 0^* + 0^*10^*$, formé des mots* avec au plus un 1. *Plus généralement, les lettres utilisées comme des balises (séparateurs, parenthèses, etc.) se fondent, lors de la réduction d'étiquetage, dans les autres symboles et ne remplissent plus leur rôle.*

*Par conséquent, la surminimisation supprime la complexité due à un choix arbitraire des étiquettes, comme le met en évidence la figure 5. Ce même exemple montre*

*aussi que la question de l'espace gagné par cette transformation est sans objet : on peut construire des automates arbitrairement grands dont la surminimisation n'a qu'un état. La taille de l'alphabet ne joue d'ailleurs aucun rôle dans ce gain : on peut par exemple, remplacer le y par un* 1 *et toutes les autres lettres par des* 0 *sur la figure 5.*

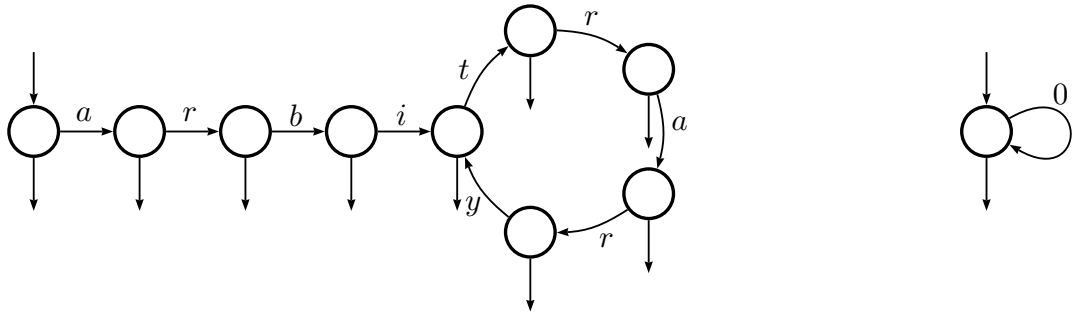



Figure 5 – Un automate dont l'étiquetage est arbitraire, et sa surminimisation

## Invariant de la T-équivalence

Les notions classiques de minimisation et d'équivalence ont un invariant évident : le langage accepté. Nous avons déjà vu que ce n'est pas le cas pour la T-équivalence ou la surminimisation ; elles ont en revanche un autre invariant : l'arbre ordonnée sous-jacent (du langage accepté).

En effet, nous avons vu dans les chapitres précédents que la donnée d'un langage $L$ clos par préfixe est équivalente à celle d'un arbre ordonné et étiqueté. Si $L$ n'est pas clos par préfixe, il faut de plus distinguer les nœuds qui correspondent aux mots de $L$, qu'on dit finals, et qu'on indique sur les figures par un double cercle.

On généralise la définition d'arbre étiqueté (définition 7.10, page 180) utilisée dans les chapitres précédents : l'ensemble des nœuds est $\mathbb{N}$, de telle sorte que $n$ soit le $(n+1)$-ème nœud visité par le parcours en largeur canonique. Un tel arbre étiqueté sur l'alphabet $A$ est donc représenté par un couple $\mathcal{T} = (E, F)$, où $E$ est un sous-ensemble de $\mathbb{N} \times A \times \mathbb{N}$ et $F$ est l'ensemble des nœuds finals, et qui vérifie les conditions suivantes

- Soient deux triplets $(n, a, m)$ et $(n', b, m')$ de $E$.
  - Si $n < n'$, alors $m < m'$.
  - Si $n = n'$ et $m < m'$ alors $a < b$.
- Pour tout entier $m > 0$, il existe un unique entier $n$ et une unique lettre $a \in A$ tels que $(n, a, m) \in E$.
- Il n'existe pas de lettre $a \in A$ telle que $(0, a, 0) \in E$.

On note plus graphiquement $n \xrightarrow{a} m$ au lieu de $(n, a, m) \in E$.

L'arbre $\mathcal{T}_L = (E_L, F_L)$ associé à $L$, défini formellement ci-dessous, est alors tel que $E_L$ est l'arbre étiqueté associé au langage clos par préfixe $\mathsf{Pre}\,(L)$ (définition 7.17,

page 182) et $F$ est constitué des nœuds $n$ tels que $0 \xrightarrow{u} n$ est étiqueté par un mot $u$ de $L$.

**Définition 9.9** – *Soit $L \subseteq A^*$ un langage dont on note $P = \mathsf{Pre}(L)$ le langage des préfixes. L'arbre étiqueté associé est $\mathcal{T}_L = (E_L, F_L)$ où $E_L \subseteq \mathbb{N} \times A \times \mathbb{N}$ est la relation définie par*

$$\forall n, m \in \mathbb{N},\ \forall a \in A \qquad n \xrightarrow{a} m \quad \text{si et seulement si} \quad \langle n \rangle_P a \ = \ \langle n \rangle_P$$

*et l'ensemble des nœuds finals est*

$$F_L \ = \ \left\{ n \in \mathbb{N} \;\middle|\; u \in L,\ \text{où } u \text{ est l'unique mot tel que } 0 \xrightarrow{u} n \right\} .$$

Dans le cas des langages réguliers, le même arbre peut être obtenu en dépliant un automate qui accepte $L$, comme l'énonce le lemme suivant.

**Lemme 9.10** – *Soient un langage régulier $L \subseteq A^*$, l'arbre étiqueté $\mathcal{T}_L = (E_L, F_L)$ associé et un automate $\mathcal{A}$ acceptant $L$. Un mot $u \in A^*$ admet un calcul dans $\mathcal{A}$ si et seulement si $0 \xrightarrow{u} n$ est une branche de $\mathcal{T}_L$, pour un certain entier $n$. De plus, ce calcul est acceptant si et seulement si $n \in F_L$.*

Démonstration. Soit un mot $u \in A^*$. Ce mot $u$ étiquette un calcul de $\mathcal{A}$ si et seulement si $u \in \mathsf{Pre}(L)$ (puisque $\mathcal{A}$ est émondé). Ceci est équivalent au fait qu'il existe un nœud $n$ tel que $0 \xrightarrow{u} n$.

D'autre part, la définition de $F$ (dans la définition 9.9), ce nœud $n$ est final si et seulement si $u \in L$, ce qui est équivalent à ce que son calcul dans $\mathcal{A}$ soit acceptant. □

Par exemple, la figure 6 montre les dépliages respectifs de trois automates T-équivalents : $\mathcal{A}_1$, $\mathcal{A}_2$ et leur surminimisation communes. On peut constater que ces trois arbres sont identiques, exceptions faites des valeurs des étiquettes. La proposition 9.11 énonce le cas général.

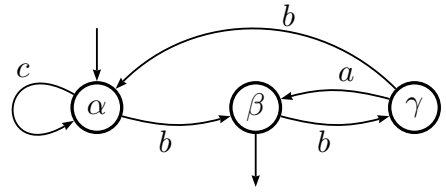
**Proposition 9.11** – *Si deux automates $\mathcal{A}$ et $\mathcal{B}$ sont T-équivalents, alors leurs dépliages respectifs ne diffèrent que par l'étiquetage.*

La proposition 9.11 résulte de la succession de deux propriétés :

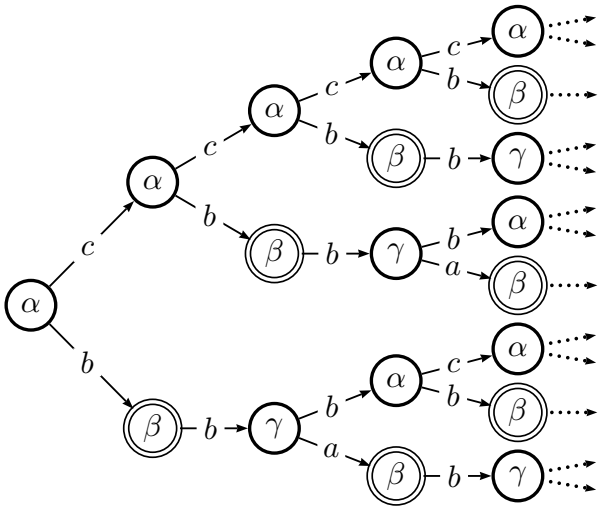
- Le dépliage d'un automate $\mathcal{A}$ et sa réduction d'étiquetage $\mathsf{Ired}(\mathcal{A})$ ne diffèrent que par l'étiquetage (lemme 9.12).
- Si un automate $\mathcal{M}$ est un quotient d'un autre automate $\mathcal{A}$, alors les dépliages de de $\mathcal{A}$ et $\mathcal{M}$ sont égaux, une conséquence du lemme 9.10.

**Lemme 9.12** – *Soit un automate $\mathcal{A}$. Alors, les dépliages respectifs de $\mathcal{A}$ et de $\mathsf{Ired}(\mathcal{A})$ ne diffèrent que par l'étiquetage.*

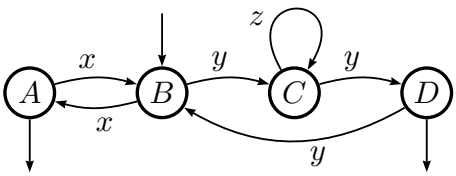
Démonstration. Pour tout état $s$ de $\mathcal{A}$ et de $\mathsf{Ired}(\mathcal{A})$, la fonction qui envoie la $(i+1)$-ème transition de $\mathsf{Out}_{\mathcal{A}}(s)$ sur la $(i+1)$-ème transition de $\mathsf{Out}_{\mathsf{Ired}(\mathcal{A})}(s)$ est une bijection d'après la définition 9.1 de la réduction d'étiquetage. Cette fonction induit une bijection entre les calculs de $\mathcal{A}$ et $\mathsf{Ired}(\mathcal{A})$ donc entre les branches de leurs dépliages (lemme 9.10). □



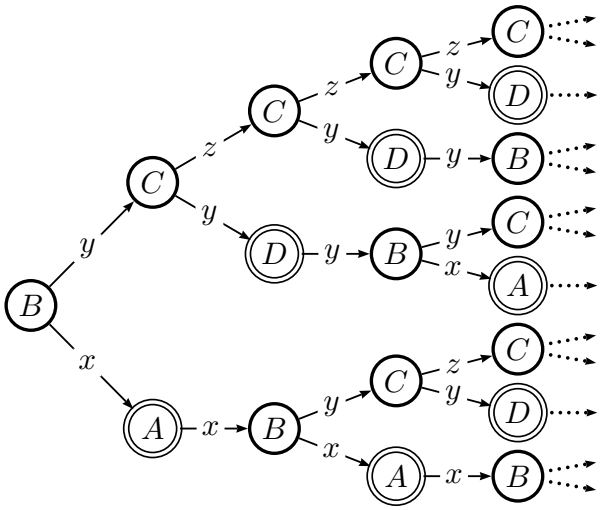

(a) L'automate $\mathcal{A}_1$



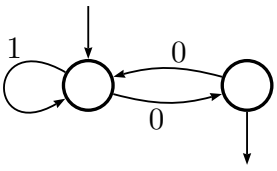

(b) ...et son dépliage



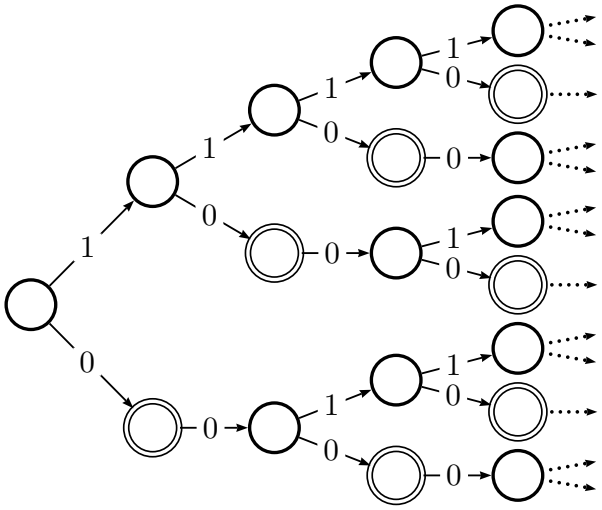

(c) L'automate $\mathcal{A}_2$




(d) ...et son dépliage




(e) Leur surminimisation




(f) ...et son dépliage

Figure 6 – Trois automates T-équivalents et leurs dépliages respectifs

# Réduction d'étiquetage et SNAR

Dans la section précédente, nous avons défini des concepts sur les langages réguliers dont les alphabets sont ordonnés ; de tel langages sont essentiellement des SNAR ; on utilise donc pour les SNAR tous les termes précédemment définis pour les langages réguliers. On s'intéresse principalement à la fonction qui permet de convertir un SNAR dans un autre qui lui est T-équivalent, c'est-à-dire la fonction qui envoie la représentation de chaque entier $n$ dans l'un vers la représentation de $n$ dans l'autre. Nous montrons que quand les deux SNAR sont T-équivalents, cette fonction est réalisée par un transducteur très simple.

## SNAR T-équivalents

**Théorème VIII** – *La fonction de conversion entre deux SNAR T-équivalents est réalisée par un transducteur fini, lettre-à-lettre et séquentiel pur.*

Cette section est dédiée à la démonstration de ce théorème. Soient deux SNAR $L$ et $K$ réalisés par deux automates $\mathcal{A}$ et $\mathcal{B}$. Nous allons d'abord définir le transducteur $\mathcal{A} \boxtimes \mathcal{B}$ qui (sous les conditions du théorème VIII) réalise la conversion de $L(\mathcal{A})$ en $L(\mathcal{B})$.

Il s'agit d'une variante du produit d'automates dans laquelle les transitions sont couplées en fonction de leur ordre dans leurs bouquets sortants respectifs plutôt que par la valeur de leur étiquette. On peut également dire que $\mathcal{A} \boxtimes \mathcal{B}$ est un réétiquetage de $\mathsf{Ired}(\mathcal{A}) \times \mathsf{Ired}(\mathcal{B})$ : chaque transition $(s,t) \xrightarrow{\ i\ } (s',t')$ de $\mathsf{Ired}(\mathcal{A}) \times \mathsf{Ired}(\mathcal{B})$ est réétiquetée par $(s,t) \xrightarrow{\ a \mid b\ } (s',t')$ dans $\mathcal{A} \boxtimes \mathcal{B}$ si

- $s \xrightarrow{\ a\ } s'$ est la $(i+1)$-ème transition de $\mathsf{Out}_{\mathcal{A}}(s)$ et
- $t \xrightarrow{\ b\ } t'$ est la $(i+1)$-ème transition de $\mathsf{Out}_{\mathcal{B}}(t)$.

**Définition 9.13** – *Soient deux automates* $\mathcal{A} = \langle\, Q_{\mathcal{A}}, A, \delta_{\mathcal{A}}, i_{\mathcal{A}}, F_{\mathcal{A}} \,\rangle$ *et* $\mathcal{B} = \langle\, Q_{\mathcal{B}}, B, \delta_{\mathcal{B}}, i_{\mathcal{B}}, F_{\mathcal{B}} \,\rangle$. *On appelle* T-produit de $\mathcal{A}$ par $\mathcal{B}$ *le transducteur*

$$\mathcal{A} \boxtimes \mathcal{B} \;=\; \langle\, Q_{\mathcal{A}} \times Q_{\mathcal{B}},\ A,\ B,\ \tau,\ (i_{\mathcal{A}}, i_{\mathcal{B}}),\ \omega \,\rangle$$

*où la fonction* $\tau$ *de transition est définie par*

$$\forall s, s' \in Q_A,\ \ \forall a_i \in A,\ \ \forall t, t' \in Q_B,\ \ \forall b_i \in B$$

$$(s,t) \xrightarrow[\mathcal{A} \boxtimes \mathcal{B}]{a_i \mid b_i} (s',t') \quad\Longleftrightarrow\quad \begin{cases} s \xrightarrow[\mathcal{A}]{a_i} s' \text{ est la } (i+1)\text{-ème transition de } \mathsf{Out}_{\mathcal{A}}(s) \\ t \xrightarrow[\mathcal{B}]{b_i} t' \text{ est la } (i+1)\text{-ème transition de } \mathsf{Out}_{\mathcal{B}}(t) \end{cases}$$

*et où la fonction finale* $\omega$ *n'est définie que sur les états* $(f,g)$ *appartenant à* $F_{\mathcal{A}} \times F_{\mathcal{B}}$ *par* $\omega(f,g) = \varepsilon$.

*De plus, on dit qu'un état* $(s,t) \in Q_{\mathcal{A}} \times Q_{\mathcal{B}}$ *est* cohérent *s'il vérifie les deux conditions suivantes :*

- *s et t ont le même statut final/non-final ;*
- *les degrés sortants de s et t sont égaux :* $\mathsf{d}_{\mathcal{A}}(s) = \mathsf{d}_{\mathcal{B}}(t)$.

*Dans le cas contraire,* $(s,t)$ *est dit* incohérent.

La figure 7 montre le transducteur $\mathcal{A}_1 \boxtimes \mathcal{A}_2$. Les états incohérents y sont dessinés en pointillés et leurs transitions sortantes sont omises ; les états inaccessibles mais cohérents sont dessinés en tiretés (ainsi que leurs transitions sortantes).

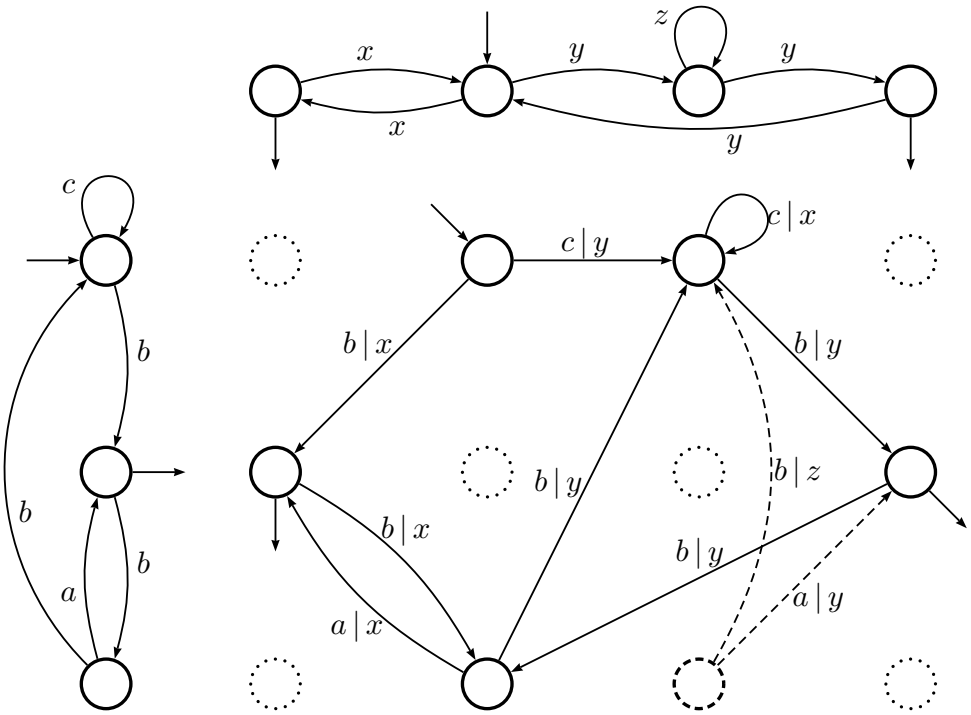



Figure 7 – Le T-produit $\mathcal{A}_1 \boxtimes \mathcal{A}_2$

Nous donnons d'abord quelques propriétés de ce transducteur qui sont des conséquences directes de la définition 9.13.

**Propriété 9.14** – *Soient $\mathcal{A}$ et $\mathcal{B}$ deux automates. Le T-produit de $\mathcal{A}$ par $\mathcal{B}$ est un transducteur fini, lettre-à-lettre et séquentiel pur.* [1]

**Propriété 9.15** – *Soient deux automates $\mathcal{A}$, $\mathcal{B}$ et quatre mots $u, u', v, v'$. Si $u \mid v$ et $u' \mid v'$ sont acceptés par $\mathcal{A} \boxtimes \mathcal{B}$, alors*

**a)** $u = u' \iff v = v'$ ;

**b)** $u <_{rad} u' \iff v <_{rad} v'$.

**Propriété 9.16** – *Soient $\mathcal{A}$ et $\mathcal{B}$ deux automates.*

**a)** *Soit $(s,t) \xrightarrow{a \mid b} (s',t')$ une transition de $\mathcal{A} \boxtimes \mathcal{B}$. Alors $s \xrightarrow{a} s'$ est une transition de $\mathcal{A}$ et $t \xrightarrow{b} t'$ est une transition de $\mathcal{B}$.*

**b)** *Soit $s \xrightarrow{a} s'$ une transition de $\mathcal{A}$ et $t$ un état de $\mathcal{B}$. Si $(s,t)$ est cohérent, alors $(s,t) \xrightarrow{a \mid b} (s',t')$ est une transition de $\mathcal{A} \boxtimes \mathcal{B}$ pour une certaine lettre $b$ et un certain état $t'$ de $\mathcal{B}$.*

1. On rappelle que l'on autorise la fonction finale à être partielle dans un transducteur séquentiel pur (*cf.* section 1.5 page 25).

**c)** *Soit $t \xrightarrow{a} t'$ une transition de $\mathcal{B}$ et $s$ un état de $\mathcal{B}$. Si $(s,t)$ est cohérent, alors $(s,t) \xrightarrow{a \mid b} (s',t')$ est une transition de $\mathcal{A} \boxtimes \mathcal{B}$ pour une certaine lettre $a$ et un certain état $s'$ de $\mathcal{A}$.*

Les états incohérents sont les témoins d'un problème dans la conversion de $L(\mathcal{A})$ en $L(\mathcal{B})$. Si aucun d'entre eux n'est accessible, le transducteur $\mathcal{A} \boxtimes \mathcal{B}$ a le comportement attendu.

**Lemme 9.17** – *Soient deux automates $\mathcal{A}$ et $\mathcal{B}$. Si tous les états accessibles de $\mathcal{A} \boxtimes \mathcal{B}$ sont cohérents, alors il réalise la fonction de conversion de $L(\mathcal{A})$ en $L(\mathcal{B})$.*

Démonstration. On note $\mathcal{A}'$ la partie émondée de l'automate d'entrée de $\mathcal{A} \boxtimes \mathcal{B}$.

**Assertion 9.17.1** – *Si tous les états accessibles de $\mathcal{A} \boxtimes \mathcal{B}$ sont cohérents, alors $\mathcal{A}$ est un quotient de $\mathcal{A}'$.*

*Démonstration de l'assertion.* On note $\Pi_1$ la projection définie par $(s,t) \mapsto s$, pour tout état $(s,t)$ de $\mathcal{A} \boxtimes \mathcal{B}$ ; montrons qu'il s'agit d'un morphisme d'automates.

*États finals.* Un état $(s,t)$ de $\mathcal{A}'$ est final si et seulement si il appartient à $F_{\mathcal{A}} \times F_{\mathcal{B}}$. Puisque $\mathcal{A}'$ est émondé, $(s,t)$ est accessible dans $\mathcal{A}$' donc dans $\mathcal{A} \boxtimes \mathcal{B}$, si bien que $(s,t)$ est cohérent par hypothèse. Les états $s$ et $t$ ont donc le même statut final/non-final donc $\Pi_1((s,t)) = s \in F_{\mathcal{A}} \iff (s,t) \in F_{\mathcal{A}} \times F_{\mathcal{B}}$.

*État initial.* L'état initial de $\mathcal{A}'$ est celui de $\mathcal{A} \boxtimes \mathcal{B}$, c'est-à-dire $(i_{\mathcal{A}}, i_{\mathcal{B}})$ ; son l'image par $\Pi_1$ est $i_{\mathcal{A}}$, l'état initial de $\mathcal{A}$.

*Transitions.* Soient $(s,t)$ un état de $\mathcal{A}'$ et $a$ une lettre. Si $(s,t) \xrightarrow{a} (s',t')$ est une transition de $\mathcal{A}'$, alors il existe $b$ tel que $(s,t) \xrightarrow{a \mid b} (s',t')$ est une transition de $\mathcal{A} \boxtimes \mathcal{B}$ ; il découle alors de la propriété 9.16a que $s \xrightarrow{a} s'$ (ou, autrement dit $\Pi_1((s,t)) \xrightarrow{a} \Pi_1((s',t'))$) est une transition de $\mathcal{A}$.

Si $s \xrightarrow{a} s'$ existe dans $\mathcal{A}$, puisque $(s,t)$ est cohérent, il existe d'après la propriété 9.16b une lettre $b$ et un état $t'$ tels que $(s,t) \xrightarrow{a \mid b} (s',t')$ est une transition de $\mathcal{A} \boxtimes \mathcal{B}$ donc $(s,t) \xrightarrow{a} (s',t')$ est une transition de $\mathcal{A}'$.

Il découle de cette assertion que le langage d'entrée de $\mathcal{A} \boxtimes \mathcal{B}$ est égal à $L(\mathcal{A})$. Un raisonnement symétrique montre que le langage de sortie de $\mathcal{A} \boxtimes \mathcal{B}$ est égal à $L(\mathcal{B})$.

Puisque $\mathcal{A} \boxtimes \mathcal{B}$ réalise une bijection (propriété 9.15a), et conserve l'ordre radiciel (propriété 9.15b), le transducteur $\mathcal{A} \boxtimes \mathcal{B}$ réalise donc la fonction $\langle n \rangle_{L(\mathcal{A})} \mapsto \langle n \rangle_{L(\mathcal{B})}$, pour tout entier $n$. □

Il s'avère que sous l'hypothèse que $\mathcal{A} \overset{T}{\sim} \mathcal{B}$, les états incohérents sont tous inaccessibles.

**Lemme 9.18** – *Soient deux automates $\mathcal{A}$ et $\mathcal{B}$. S'ils sont T-équivalents, alors tous les états accessibles de $\mathcal{A} \boxtimes \mathcal{B}$ sont cohérents.*

Démonstration. Soient $\mathcal{A} = \langle Q_{\mathcal{A}}, A, \delta_{\mathcal{A}}, i_{\mathcal{A}}, F_{\mathcal{A}} \rangle$ et $\mathcal{B} = \langle Q_{\mathcal{B}}, B, \delta_{\mathcal{B}}, i_{\mathcal{B}}, F_{\mathcal{B}} \rangle$ deux automates T-équivalents. On note $\mathcal{M} = \langle Q_{\mathcal{M}}, A, \delta_{\mathcal{M}}, i_{\mathcal{M}}, F_{\mathcal{M}} \rangle$ leur surminimisation commune et les deux morphismes associés $\varphi : \mathsf{Ired}(\mathcal{A}) \to \mathcal{M}$ et $\psi : \mathsf{Ired}(\mathcal{B}) \to \mathcal{M}$.

**Assertion 9.18.1** – *Tout état $(s,t)$ accessible satisfait l'équation $\varphi(s) = \psi(t)$.*

*Démonstration de l'assertion.* Par Induction ; l'état initial $(i_{\mathcal{A}}, i_{\mathcal{B}})$ satisfait la condition : $\varphi(i_{\mathcal{A}}) = i_{\mathcal{M}} = \varphi(i_{\mathcal{B}})$. Nous allons démontrer que si un état $(s,t)$ satisfait la condition $\varphi(s) = \psi(t)$, alors chacun de ses successeurs $(s',t')$ la satisfait également : $\varphi(s') = \psi(t')$.

Soient un état $(s,t)$ vérifiant $\varphi(s) = \psi(t)$, deux lettres $a \in A$, $b \in B$ et un état $(s',t')$ tel que

$$(s,t) \xrightarrow{a \mid b} (s',t') \ .$$

L'existence de cette transition implique (définition 9.13) que $s \xrightarrow{a} s'$ et $t \xrightarrow{b} t'$ sont respectivement les $(i+1)$-ème transitions de $\mathsf{Out}_{\mathcal{A}}(s)$ et $\mathsf{Out}_{\mathcal{B}}(t)$, pour un certain entier $i$. Il s'ensuit que $s \xrightarrow{i} s'$ et $t \xrightarrow{i} t'$ sont des transitions de $\mathsf{Ired}(\mathcal{A})$ et $\mathsf{Ired}(\mathcal{B})$, respectivement.

Puisque $\mathcal{M}$ est le quotient de $\mathsf{Ired}(\mathcal{A})$ (par $\varphi$) et de $\mathsf{Ired}(\mathcal{B})$ (par $\psi$), il existe dans $\mathcal{M}$ deux transitions $\varphi(s) \xrightarrow{i} \varphi(s')$ et $\psi(t) \xrightarrow{i} \psi(t')$. Puisque $\varphi(s) = \psi(t)$ et que $\mathcal{M}$ est déterministe, alors $\varphi(s') = \psi(t')$.

Soit $(s,t)$ un état accessible de $\mathcal{A} \boxtimes \mathcal{B}$. Il découle de l'assertion précédente que $\varphi(s) = \psi(t)$, ce qui implique que $s$ et $t$ ont d'une part le même statut final/non-final et d'autre part le même nombre de transitions sortantes. Donc $(s,t)$ est cohérent. □

Le théorème VIII découle alors essentiellement des deux lemmes (9.18 et 9.17) précédents.

Démonstration du théorème VIII. Soient $L$ et $K$ deux SNAR T-équivalents, respectivement réalisés par les automates (émondés) $\mathcal{A}$ et $\mathcal{B}$. Puisque $L$ et $K$ sont T-équivalents, tous les automates émondés qui les acceptent respectivement sont équivalents donc $\mathcal{A} \overset{T}{\sim} \mathcal{B}$. Appliquer successivement les lemmes 9.18 et 9.17 démontre que tous les état de $\mathcal{A} \boxtimes \mathcal{B}$ sont cohérents, puis que $\mathcal{A} \boxtimes \mathcal{B}$ réalise la fonction de conversion de $L(\mathcal{A}) = L$ en $L(\mathcal{B}) = K$. Puisque $\mathcal{A} \boxtimes \mathcal{B}$ est fini, lettre-à-lettre et séquentiel pur (propriété 9.14), cela conclut la démonstration. □

## Transduction lettre-à-lettre de SNAR T-inéquivalents

La réciproque du théorème VIII est fausse dans le cas général : par exemple, le transducteur représenté par la figure 8 réalise la fonction de conversion de $0 + 10^+$ en $1 + 00^+$, deux langages distincts et irréductibles, donc qui ne sont pas T-équivalents.

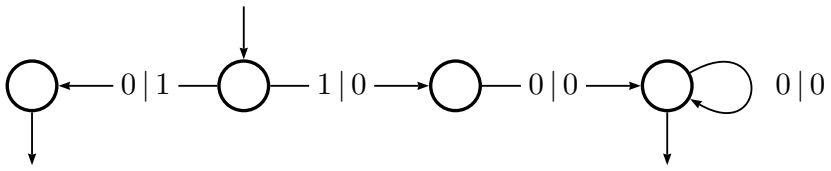



Figure 8 – Un transducteur qui n'est pas localement croissant

On dit qu'un transducteur fonctionnel $\mathcal{T}$ est *croissant* s'il conserve l'ordre radiciel, c'est-à-dire si

$$\forall u, u' \in \mathsf{Dom}(\mathcal{T}) \qquad u <_{\text{rad}} u' \iff \mathcal{T}(u) <_{\text{rad}} \mathcal{T}(v') \ . \tag{9.1}$$

Un tel transducteur réalise une certaine fonction de conversion d'un SNAR $L$ vers un autre SNAR $K$, où $L$ et $K$ sont respectivement les langages d'entrée et de sortie de $\mathcal{T}$.

On dit qu'un transducteur $\mathcal{T} = \langle\, Q_{\mathcal{T}}, A, B, E_{\mathcal{T}}, i_{\mathcal{T}}, \omega_{\mathcal{T}} \,\rangle$ lettre-à-lettre est *localement croissant* s'il conserve localement l'ordre des étiquettes ou, plus formellement, s'il satisfait la condition suivante :

$$\begin{gathered}\forall s, t, t' \in Q_{\mathcal{T}},\ \ \forall a, a' \in A,\ \ \forall b, b' \in B \quad \text{tels que} \quad s \xrightarrow[\mathcal{T}]{a \mid b} t \ \text{ et } \ s \xrightarrow[\mathcal{T}]{a' \mid b'} t' \\ a < a' \iff b < b' \ .\end{gathered} \tag{9.2}$$

Un T-produit $\mathcal{A} \boxtimes \mathcal{B}$ est toujours localement croissant (même si $\mathcal{A}$ et $\mathcal{B}$ ne sont pas T-équivalents) donc la figure 7 montre un automate localement croissant.

Le lemme suivant découle de la définition de localement-croissant.

**Lemme 9.19** – *Un transducteur lettre-à-lettre et localement croissant est croissant et séquentiel.*

En revanche, la réciproque est fausse. En effet, le transducteur représenté par la figure 8 est lettre-à-lettre et croissant mais n'est pas localement croissant (les transitions sortantes de l'état initial inversent l'ordre des lettres).

**Théorème 9.20** – *Soient $L$ et $K$ deux SNAR. Si la fonction de conversion de $L$ en $K$ est réalisée par un transducteur fini, lettre-à-lettre, localement croissant et séquentiel pur, alors $L$ et $K$ sont T-équivalents.*

Démonstration. On note $\mathcal{T}$ le transducteur réalisant la fonction de conversion, $\mathcal{A}$ l'automate d'entrée de $\mathcal{T}$, $\mathcal{B}$ son automate de sortie et $Q$ l'ensemble des états de $\mathcal{T}$ (qui est aussi celui de $\mathcal{A}$ et de $\mathcal{B}$). Puisque $\mathcal{T}$ est localement croissant, les automates $\mathcal{A}$ et $\mathcal{B}$ sont déterministes.

Soit un état $s \in Q$. Ses transitions sortantes sont énumérées par ordre croissant (à la fois en entrée et en sortie puisque $\mathcal{T}$ est localement croissant) :

$$\forall i \in [\![\mathsf{d}(s)]\!] \qquad s \xrightarrow{a_i \mid b_i} t \quad \text{avec} \quad \begin{cases} a_0 < a_1 < \cdots < a_{\mathsf{d}(s)-1} \\ b_0 < b_1 < \cdots < b_{\mathsf{d}(s)-1} \end{cases}$$

On fixe $i \in [\![\mathsf{d}(s)]\!]$. Les transitions $s \xrightarrow{a_i} t$ et $s \xrightarrow{b_i} t$ sont respectivement les (i+1)-ième transitions de $\mathsf{Out}_{\mathcal{A}}(s)$ et de $\mathsf{Out}_{\mathcal{B}}(s)$. Si bien qu'elles sont toutes deux réétiquetées par le même chiffre $i$ dans $\mathsf{Ired}(\mathcal{A})$ et $\mathsf{Ired}(\mathcal{B})$, ce qui implique que $\mathsf{Ired}(\mathcal{A})$ et $\mathsf{Ired}(\mathcal{B})$ sont isomorphes. □

# Réduction d'étiquetage et systèmes positionnels

Les systèmes de numération positionnels, ou $U$-systèmes, sont des généralisations des systèmes de numération à base entière. La *base* est alors une suite de *poids* ;

par exemple, il s'agit en base entière $p$ de la suite des puissances de $p$. Ils ont été définis dans toute leur généralité dans [34] bien que des exemples précis (comme le système de Zeckendorf/Fibonacci [78]) sont plus anciens. Nous ne donnerons qu'une présentation succincte des systèmes positionnels ; voir [15] pour plus de détails.

Dans le cadre d'un système positionnel, on prend la convention d'écrire les nombres avec le chiffre de poids fort en premier, c'est-à-dire comme c'est déjà le cas depuis le chapitre 4 et au contraire de la convention prise pour la base entière dans les chapitres qui le précèdent.

## $U$-représentations des entiers

Une *base* est une suite croissante d'entiers $U = (U_i)_{i\in\mathbb{N}}$ telle que $U_0 = 1$. Pour tout mot $w = d_k\, d_{(k-1)} \cdots d_0$ (pour l'instant sur un alphabet de chiffres quelconque) la *fonction d'évaluation en base $U$* est définie par

$$\pi_U(w) \;=\; \pi_U(d_k d_{k-1} \cdots d_0) \;=\; \sum_{i=0}^{k} d_i U_i \;. \tag{9.3}$$

Parmi tous les mots dont la valeur est un entier donné $N$, l'un est calculé grâce à l'*algorithme glouton*, est appelé *la représentation de $N$ en base $U$*, ou *$U$-représentation* et est noté $\langle N\rangle_U$. L'algorithme glouton a été décrit (page 34) précédemment dans le cas particulier de la base entière, c'est-à-dire le cas où $U = (p^i)_{i\in\mathbb{N}}$ pour un certain entier $p$.

Soit $k$ l'entier tel que $U_k \leqslant N < U_{(k+1)}$. On pose $N_k = N$ et

$$\forall i,\;\; 0 \leqslant i \leqslant k, \qquad \begin{array}{l} b_i \text{ est le plus grand entier tel que } (N_i - b_i\, U_i) \geqslant 0 \\ \text{et } N_{(i-1)} = (N_i - b_i\, U_i) \end{array} \tag{9.4}$$

La $U$-représentation de $N$ est alors $\langle N\rangle_U = b_k b_{k-1} \cdots b_0$ et une simple vérification montre que $\pi_U(\langle N\rangle_U) = N$.

Dans la suite, on supposera toujours que le rapport $\frac{U_{i+1}}{U_i}$ est borné par un entier noté $p_U$, auquel cas la $U$-représentation de chaque entier est un mot sur l'alphabet fini $[\![p_U]\!]$.

**Exemple 9.21** (Système de Fibonacci) – *Soit $F = (F_i)_{i\in\mathbb{N}}$ la suite des nombres de Fibonacci $F = 1, 2, 3, 5, \ldots$ définie par $F_0 = 1$, $F_1 = 2$ et*

$$\forall i \in \mathbb{N} \qquad F_{(i+2)} \;=\; F_{(i+1)} + F_i \;.$$

Les deux propositions suivantes sont des résultats élémentaires sur les systèmes de numération positionnels (ce sont par exemples les propositions 2.3.44 et 2.3.45 de [36]).

[34] Aviezri S. Fraenkel, 1985, *Systems of numeration.*

[78] Edouard Zeckendorf, 1972, *Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas.*

[15] Valérie Berthé et Michel Rigo, 2010, *Combinatorics, Automata and Number Theory.*

[36] Christiane Frougny et Jacques Sakarovitch, 2010, *Number representation and finite automata.*

**Proposition 9.22** – *Soit un entier $n$. Le mot $\langle n\rangle_U$ est le plus grands dans l'ordre radiciel de tous les mots qui ne commencent pas par 0 et qui s'évaluent à $n$.*

Démonstration. Soit un mot $u$ qui ne commence pas par un 0 et dont la valeur est $n$. On note $u = a_j\, a_{j-1} \cdots a_0$ et $\langle n\rangle_U = b_k\, b_{k-1} \cdots b_0$.

Il découle de la définition de la fonction d'évaluation (équation (9.3)) que $\boxed{\pi_U(u) \geqslant a_j U_j}$ et de la définition de l'algorithme glouton que $\boxed{n < U_{(k+1)}}$. Puisque $u$ ne commence pas par un zéro par hypothèse, $\boxed{a_j \geqslant 1}$.

Supposons que $j > k$ ; les inéquations encadrées impliquent alors que $\pi_U(u) > n$, ce qui contredit l'hypothèse $n = \pi_U(u)$ ; donc $j \leqslant k$. D'autre part la proposition est vérifiée dans le cas où $j < k$. On suppose donc dans la suite que $k = j$.

Soit un indice $i \geqslant 0$ tel que $a_k\, a_{k-1} \cdots a_{i+1} = b_k\, b_{k-1} \cdots b_{i+1}$. Notez que $\pi_U(b_i\, b_{i-1} \cdots b_0) = N_i$ où $N_i$ est le reste à l'étape $i$ de l'algorithme glouton ; si bien que $b_i$ est le plus grand entier tel que $(N_i - b_i U_i) \geqslant 0$. Supposons que $a_i > b_i$ ; il s'ensuit que $(a_i U_i - N_i) > 0$ donc que

$$\pi_U(u) \geqslant \pi_U(a_{k-1} \cdots a_{i+1}\, a_i\, 0^i) \;=\; (n - N_i) + a_i U_i \;>\; n\ ,$$

ce qui contredit l'hypothèse $n = \pi_U(u)$ ; donc $a_i \leqslant b_i$. Il s'ensuit que $u \leqslant_{\text{rad}} \langle n\rangle_U$. □

**Proposition 9.23** – *Soient deux entiers $n, m$. Alors $n < m$ si et seulement si $\langle n\rangle_U <_{rad} \langle m\rangle_U$.*

Démonstration. Sens direct. Soit $n$ et $m$ deux entiers tels que $n < m$. On note $\langle n\rangle_U = a_j\, a_{j-1} \cdots a_0$ et $\langle m\rangle_U = b_k\, b_{k-1} \cdots b_0$. Ceci implique que $U_j \leqslant m < U_{j+1}$ et $U_k \leqslant n < U_{k+1}$ donc que $k \leqslant j$.

Dans le cas où $k < j$ alors $\langle n\rangle_U < \langle m\rangle_U$ et la proposition est donc vérifiée. On suppose dans la suite que $k = j$. On note, pour tout entier $i \leqslant k$, $N_i$ et $M_i$ les variables intermédiaires des calculs de $\langle n\rangle_U$ et $\langle m\rangle_U$ par l'algorithme glouton.

Supposons que $\langle n\rangle_U >_{\text{rad}} \langle m\rangle_U$ ; il existe donc un entier $i$ tel que

$$a_k\, a_{(k-1)} \cdots a_{(i+1)} = b_k\, b_{(k-1)} \cdots b_{(i+1)} \quad \text{et} \quad a_i > b_i\,;$$

il s'ensuit que

$$(n - N_i) \;=\; \pi_U(a_k\, a_{(k-1)} \cdots a_{(i+1)}\, 0^{(i+1)}) \;=\; \pi_U(b_k\, b_{(k-1)} \cdots b_{(i+1)}\, 0^{(i+1)}) \;=\; (m - M_i)$$

ce qui implique que $N_i < M_i$. Les lettres $a_i$ et $b_i$ sont les plus grandes lettres telles que $(N_i - b_i U_i) \geqslant 0$ et $(M_i - a_i U_i) \geqslant 0$, respectivement. Il découle donc de $N_i < M_i$ que $a_i \leqslant b_i$. Contradiction. Donc $\langle n\rangle_U \leqslant_{\text{rad}} \langle m\rangle_U$

Les mots $\langle n\rangle_U \neq \langle m\rangle_U$ sont distincts puis qu'ayant des valeurs différentes, si bien que $\langle n\rangle_U <_{\text{rad}} \langle m\rangle_U$

Le sens réciproque découle du sens direct car $\mathbb{N}$ est totalement ordonné. □

On note $L(U)$ le langage des $U$-représentations des entiers :

$$L(U) \;=\; \{\, \langle N\rangle_U \mid N \in \mathbb{N} \,\}\ .$$

Il découle de la proposition précédente que les systèmes positionnels sont des systèmes de numération abstraits.

**Corollaire 9.24** – *Un système de numération positionnel $U$ est le SNA calable $0^* L(U)$.*

**Exemple 9.25** – *Le système de Fibonacci (défini dans l'exemple 9.21) est le SNAR calable construit à partir du langage des mots de $[\![2]\!]^*$ qui ne contiennent pas le facteur 11 :*

$$0^* L(F) \;=\; [\![2]\!]^* \setminus [\![2]\!]^* 11 [\![2]\!]^* \;=\; 0^* (1\,0^+)^* \;.$$

**Définition 9.26** – *Une suite d'entiers $(U_i)_{i\in\mathbb{N}}$ est récurrente linéaire s'il existe un entier $n$ et $n$ coefficients $\alpha_0, \alpha_1, \ldots, \alpha_{(n-1)} \in \mathbb{N}$ tels que*

$$\forall i \in \mathbb{N} \qquad U_{i+n} \;=\; \alpha_0 U_i + \alpha_1 U_{(i+1)} + \cdots + \alpha_{(n-1)} U_{(i+n-1)} \;=\; \sum_{j=0}^{n-1} \alpha_j U_{(i+j)} \;.$$

*Les $n$ premières valeurs de la suite $(U_0, U_1, \ldots, U_{n-1})$ sont alors appelées les* conditions initiales *de la suite récurrente linéaire.*

Le résultat suivant, dû à Shallit, donne une condition nécessaire pour qu'un système positionnels soit un SNAR ; une autre démonstration est donnée dans [47] utilisant un résultat classique de Chomsky et Miller (*cf.* [23]) : *la fonction génératrice d'un langage régulier est une suite récurrente linéaire.*

**Théorème 9.27** [76] – *Soit un système positionnel $U$. Si $L(U)$ est un langage régulier, alors $U$ est une suite récurrente linéaire.*

La condition donnée par ce théorème n'est pas suffisante, comme le montre l'exemple suivant, tiré du même article.

**Exemple 9.28** [76] – *Soit $V = (V_i)_{i\in\mathbb{N}}$ la suite définie par, pour tout entier $i \in \mathbb{N}$, $V_i = (i+1)^2$. Si bien que $V_0 = 1$, $V_1 = 4$, $V_2 = 9$ et*

$$\forall i \in \mathbb{N} \qquad V_{(i+3)} \;=\; 3\,V_{(i+2)} - 3\,V_{(i+1)} + V_i \;.$$

*Considérons le langage $L(V) \cap 1\,0^*\,1\,0^*$ ; il contient le mot $1\,0^n\,1\,0^m$ si et seulement si $V_{(n+m+2)} > (V_n + V_{n+m+1})$, c'est-à-dire si*

$$\begin{aligned} (n+m+3)^2 &> (n+m+2)^2 + (n+1)^2 \\ (n+m+2)^2 + 2(n+m+2) + 1 &> (n+m+2)^2 + n^2 + 2\,n + 1 \\ 2\,m + 4 &> n^2 \;. \end{aligned}$$

*Il découle alors du lemme d'itération (cf. lemme 1.14) que $L(V) \cap 1\,0^*\,1\,0^*$ n'est pas régulier, donc que $L(V)$ ne l'est pas non plus.*

## Irréductibilité des systèmes positionnels

Auparavant, la réduction d'étiquetage était définie comme une transformation d'automates, relevée sur les langages réguliers. Nous l'étendons ci-dessous aux langages quelconques par une définition directe.

[47] Nathalie Loraud, 1995, *$\beta$-shift, systèmes de numération et automates.*
[23] Noam Chomsky et George A Miller, 1958, *Finite state languages.*
[76] Jeffrey Shallit, 1994, *Numeration Systems, Linear Recurrences, and Regular Sets.*

**Définition 9.29** – *Soit un langage $L \subseteq A^*$. On note $D = [\![\mathsf{Card}(A)]\!]$ l'alphabet composé des $\mathsf{Card}(A)$ plus petits entiers, et on définit la fonction $g_L$ par*

$$\begin{array}{rccl} g_L : & A^* \times A & \longrightarrow & D \\ & (u,a) & \longmapsto & |\{u\,a \mid a < b \;\; et \;\; u\,a \in \mathsf{Pre}(L)\}| \end{array}$$

*et la fonction $f_L$ par*

$$\begin{array}{rccl} f_L : & A^* & \longrightarrow & D^* \\ & \varepsilon & \longmapsto & \varepsilon \\ & u\,a & \longmapsto & f_L(u)\,g_L(u,a) \end{array}$$

*On définit la réduction d'étiquetage de $L$ par*

$$\mathsf{lred}(L) \;=\; f_L(L) \;=\; \{\, f_L(u) \mid u \in L \,\} \;.$$

**Exemple 9.30** – *Soit le langage $L_7 = \{a^n\,b^m\,a^n\,b^m \mid n,m \in \mathbb{N}\}$ qui n'est pas algébrique. Sa réduction d'étiquetage est $\mathsf{lred}(L_7) = \{0^n\,1^m\,0^{(n+m)} \mid n,m \in \mathbb{N}\}$, un langage algébrique.*

**Exemple 9.31** – *Soit un entier $n$. On note $A$ l'alphabet $= [\![n]\!] \cup \{\overline{0}, \overline{1}, \ldots, \overline{n-1}\}$ que l'on munit de l'ordre $0 < 1 < \cdots < n-1 < \overline{0} < \overline{1} < \cdots < \overline{n-1}$. Chaque lettre $a \in [\![n]\!]$ est une* parenthèse ouvrante *d'un certain type, et $\bar{a}$ est la* parenthèse fermante *du même type. Le langage (dit de Dyck, cf.* [19]*) des mots bien parenthésés est le plus petit langage $L_8$ qui vérifie : $\varepsilon \in L_8$ et*

$$\forall u,v \in A^* \qquad u\,v \in L_8 \quad\Longrightarrow\quad \forall a \in [\![n]\!] \qquad u\,a\,\bar{a}\,v \in L_8 \;.$$

*Il s'ensuit que pour tout mot $u \in \mathsf{Pre}(L_8)$ et toute lettre $a \in [\![n]\!]$, alors $u\,a \in \mathsf{Pre}(L_8)$ ; par contre, il existe au plus une lettre $a \in [\![n]\!]$ telle que $u\,\bar{a} \in \mathsf{Pre}(L_8)$, cette lettre $a$ est la lettre la plus à droite de $u$ qui n'a pas sa parenthèse fermante. Donc,*

$$\forall u \in \mathsf{Pre}(L_8),\;\; \forall a \in [\![n]\!] \qquad g_L(u,a) = a \quad\; et \quad\; g_L(u,\overline{a}) = n \;.$$

*En d'autres termes, la fonction $f_{L_8}$ laisse inchangées toutes les parenthèses ouvrantes et envoie toutes les parenthèses fermantes sur la nouvelle lettre $n$. Le langage $\mathsf{lred}(L_8)$ est donc essentiellement le langage des mots biens parenthésés avec un seul type de parenthèses ; il en est en effet l'image inverse par un morphisme lettre-à-lettre.*

**Remarque 9.32** – *Prenons un instant la perspective signature/étiquetage (introduite dans le chapitre 7). On définit la fonction*

$$\begin{array}{rccl} \psi : & \mathbb{N} & \longrightarrow & \mathbb{N}^* \\ & N & \longmapsto & 0\,1\cdots(n-1) \end{array}$$

*En particulier la restriction de $\psi$ à tout alphabet $[\![n]\!]$ est une fonction $[\![n]\!] \to [\![n]\!]^*$ qui est étendue en un morphisme de mots $[\![n]\!]^* \to [\![n]\!]^*$*

*Soit un langage (clos par préfixe) $L$ dont la signature est $(\mathbf{s}, \boldsymbol{\lambda})$. Alors le langage $\mathsf{lred}(L)$ a pour signature $(\mathbf{s}, \boldsymbol{\mu})$ où $\boldsymbol{\mu} = \psi(\mathbf{s})$.*

*L'étiquetage $\boldsymbol{\mu}$ est le plus "petit" étiquetage cohérent avec $\mathbf{s}$ (sur un alphabet positif), dans le sens où chaque lettre est la plus petite possible. En effet, si $\mathbf{s} = s_0\,s_1\cdots s_i\cdots$ alors pour tout entier $i$, chaque mot $\psi(s_i)$ est un facteur de la décomposition de $\boldsymbol{\mu}$ par rapport à $\mathbf{s}$. Pour que $\boldsymbol{\mu}$ soit cohérent avec $\mathbf{s}$, il faut donc*

[19] Olivier Carton, 2008, *Langages formels, calculabilité et complexité*.

*que tous les $\psi(s_i)$ soient des mots croissants, et $\psi(n)$ est le plus petit mot croissant de longueur $n$.*

Suivant la définition 9.29, un langage $L$ sur un alphabet de chiffres est *irréductible* si pour tout mot $w \in L$ et toute décomposition $w = u\,a\,v$ tel que $a > 0$, alors il existe un mot $v'$ tel que $u\,(a-1)\,v'$ appartient à $L$.

**Théorème IX** – *Pour tout système positionnel $U$, le langage calable $0^*L(U)$ est irréductible.*

Démonstration. Soient deux entiers $m, i$ et une décomposition $0^i\langle m\rangle_U = u\,a\,v$ où $a$ est une lettre strictement positive ; montrons qu'il existe un mot $v'$ tel que $u\,(a-1)\,v' \in 0^*L(U)$. Puisque 0 est la lettre de calage, il suffit de traiter le cas où $u$ ne commence pas par un 0 ; c'est-à-dire où $i = 0$ et $\langle m\rangle_U = u\,a\,v$.

Si $u = \varepsilon$ et $a = 1$ alors $u\,(a-1) = 0$, auquel cas on pose $v' = \varepsilon$ ce qui implique que le mot $u\,(a-1)\,v = 0$ appartient à $0^*L(U)$ et conclut la démonstration dans ce cas particulier. On exclut donc ce cas dans la suite, ce qui implique en particulier que $u\,(a-1)$ est un mot non-vide qui ne commence pas par un 0.

On suppose de plus, sans perdre la généralité, que $m$ est choisi de telle sorte que $v$ est le plus petit mot $w$ dans l'ordre radiciel tel que $(u\,a\,w) \in L(U)$. On note $k = |v|$ et $n = \pi_U(u\,(a-1)\,v) = (m - U_k)$. Puisque $u\,(a-1)$ est non-vide ne commence pas par un 0, il en est de même pour $u\,(a-1)\,v$. Appliquer la proposition 9.22 à celui-ci puis la proposition 9.23 à $n$ et $m$ donne les deux inéquations suivantes :

$$u\,(a-1)\,v \;\leqslant_{\text{rad}}\; \langle n\rangle_U \;<_{\text{rad}}\; \langle m\rangle_U \;=\; u\,a\,v\ .$$

Or, le choix de $m$ implique que $\langle n\rangle_U$ n'est pas de la forme $u\,a\,w$ pour un certain $w <_{\text{rad}} v$. Il s'ensuit qu'il est de la forme $\langle n\rangle_U = u\,(a-1)\,v'$ pour un certain $v'(\geqslant_{\text{rad}} v)$, ce qui conclut la démonstration. □

**Remarque 9.33** – *Considérons un instant la relation de $T$-équivalence sur les systèmes de numération abstraits réguliers (SNAR). Soit $C$ une classe de $T$-équivalence, qui contient un unique SNAR $K$ irréductible ; $K$ est le représentant canonique de $C$. Tout SNAR $L$ de $C$ qui n'est pas irréductible*

- *permet de calculer $K$ facilement (en surminimisation un automate acceptant $L$) ;*
- *est accepté par un automate plus grand que celui qui accepte $K$ ;*
- *est quasiment identique à $K$ (on peut convertir l'un en l'autre avec un transducteur fini, lettre-à-lettre et séquentiel pur) ;*
- *n'est pas un système de numération positionnel (puisqu'il sont tous irréductibles).*

*Les SNAR réductibles n'apportent donc ni expressivité, ni concision, et ne capturent aucun exemple concret supplémentaire.*

# Conclusion de la troisième partie

Nous avons défini dans le chapitre 7 la signature d'un arbre infini, un mot infini qui lui est caractéristique et donc en est une sérialisation. L'ajout d'un second mot infini, appelé étiquetage, permet de sérialiser un arbre infini dont les arcs sont étiquetés, c'est-à-dire essentiellement un langage clos par préfixe. Le calcul de la signature et de l'étiquetage d'un langage $L$ (clos par préfixe) correspond à l'énumération des mots de $L$ dans l'ordre radiciel. Sérialiser $L$ est donc une façon d'étudier $L$ en tant que système de numération abstrait.

Nous avons vu ensuite que les signatures des langages réguliers forment une sous-classe des mots morphiques. Ce résultat est démontré en utilisant un parallèle classique entre morphisme de mots et système de numération abstrait régulier dont le principe remonte aux travaux de Cobham.

Il est ensuite montré dans le chapitre 8 que les signatures périodiques sont intimement liées aux systèmes de numération à base rationnelle : le taux de croissance de la période détermine (à une fonction séquentielle et rationnelle près) le langage obtenu. La généralisation de la méthode employée amène à la définition de la notion de signature dirigée par $\frac{p}{q}$, qui exprime intuitivement l'idée que l'arbre qu'elle génère croît à une vitesse qui reste toujours proche de $\frac{p}{q}$. Il semble donc sans espoir de définir une variante de la base $\frac{p}{q}$ dont le langage des représentations des entiers est simple : la signature de ce langage ne peut alors pas être dirigé par $\frac{p}{q}$.

Cette notion de direction ouvre par ailleurs de nombreuses questions. Par exemple, un système de numération lié d'une façon ou d'une autre à un nombre $\beta$ produit-il toujours un langage de représentations des entiers dont la signature est dirigée par $\beta$ ? une simple vérification montre sans surprise que le système de Fibonacci est ainsi lié au nombre d'or. Inversement, une signature dirigée par $\beta$ engendre-t-il toujours un langage lié à la $\beta$-numération ?

L'étude des signatures donne aux étiquetages un statut ambivalent. Celui-ci n'est en général pas très important mais, s'il n'est pas raisonnablement restreint, il est capable d'une grande nuisance : quelle que soit la signature considérée, on peut toujours trouver un étiquetage (par exemple, non-récursivement énumérable) qui engendre un langage arbitrairement complexe. L'idée derrière la réduction d'étiquetage, présentées dans le chapitre 9, est de donner une façon canonique pour calculer l'étiquetage.

Ces réflexions rapportées au cas des langages réguliers donnent la surminimisation, une transformation d'automates qui diminue le nombre d'états davantage que la minimisation classique. Ce gain d'espace se fait au prix de la perte du "sens des lettres", c'est-à-dire de toutes les informations qui ne modifient pas l'ordre radiciel du langage accepté.

Le fait que cette transformation fonctionne pour les langages qui ne sont pas clos par préfixe témoigne d'une certaine robustesse du formalisme. Ceci est encore amplifié par le fait que tous les systèmes positionnels, c'est-à-dire essentiellement tous les systèmes "concrets", utilisent naturellement cet étiquetage canonique.

# Bibliographie


[1] Alfred V. Aho, John E. Hopcroft et Jeffrey D. Ullman. *Data Structures and Algorithms*. Addison-Wesley, 1983.

[2] Shigeki Akiyama, Christiane Frougny et Jacques Sakarovitch. *Powers of rationals modulo 1 and rational base number systems*. In : *Israel J. Math.* 168 (2008), p. 53–91.

[3] Shigeki Akiyama, Victor Marsault et Jacques Sakarovitch. *Automatic sequences and orbits of rational numbers*. In : *Combinatorics on Words - 9th International Conference, (WORDS 2013)*. Sous la dir. de Juhani Karhumäki, Arto Lepistö et Luca Q. Zamboni. Lect. Notes Comput. Sci. 8079. Springer, 2013, p. 34–45.

[4] Jean-Paul Allouche, Narad Rampersad et Jeffrey Shallit. *Periodicity, repetitions, and orbits of an automatic sequence*. In : *Theoret. Comput. Sci* 410 (2009), p. 2795–2803.

[5] Jean-Paul Allouche et Jeffrey Shallit. *Automatic Sequences : Theory, Applications, Generalizations*. Cambridge University Press, 2003.

[6] Jorge Almeida et Marc Zeitoun. *Description and analysis of a bottom-up DFA minimization algorithm*. In : *Inf. Process. Lett.* 107.2 (2008), p. 52–59.

[7] Pierre-Yves Angrand et Jacques Sakarovitch. *Radix enumeration of rational languages*. In : *RAIRO - Theor. Inf. and Applic.* 44.1 (2010), p. 19–36.

[8] Jean-Michel Autebert, Joffroy Beauquier, Luc Boasson et Michel Latteux. *Indécidabilité de la Condition IRS*. In : *ITA* 16.2 (1982).

[9] Yehoshua Bar-Hillel, Micha A. Perles et Eli Shamir. *On formal properties of simple phrase structure grammars*. In : *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 14 (1961), p. 143–172.

[10] Èmile Barbier. *On suppose écrite la suite naturelle des nombres ; quel est le* $(10^{1000})^{\text{ième}}$ *chiffre écrit ?* In : *CRASP* 105 (1887), p. 795–798.

[11] Èmile Barbier. *On suppose écrite la suite naturelle des nombres ; quel est le* $(10^{10000})^{\text{ième}}$ *chiffre écrit ?* In : *CRASP* 105 (1887), p. 1238–1239.

[12] Jason Bell, Emilie Charlier, Aviezri S. Fraenkel et Michel Rigo. *A Decision Problem for Ultimately Periodic Sets in Nonstandard Numeration Systems*. In : *IJAC* 19.6 (2009), p. 809–839.

[13] Jean Berstel, Aaron Lauve, Christophe Reutenauer et Franco Saliola. *Combinatorics on Words : Christoffel Words and Repetition in Words*. T. 27. CRM monograph series. American Math. Soc., 2008.

[14] Valérie Berthé et Michel Rigo. *Odometers on Regular Languages*. In : *Theory Comput. Syst.* 40.1 (2007), p. 1–31.

[15] Valérie Berthé et Michel Rigo. *Combinatorics, Automata and Number Theory*. Cambridge University Press, 2010.

[16] Véronique Bruyère, Georges Hansel, Christian Michaux et Roger Villemaire. *Logic and p-recognizable sets of integers*. In : *Bull. Belg. Soc. Math.* 1 (1994), p. 191–238.

[17] John A. Brzozowski. *Canonical regular expressions and minimal state graphs for definite events*. In : *Symposium on Mathematical Theory of Automata*. 1963, p. 529–561.

[18] Georg Cantor. *De la puissance des ensembles parfaits de points.* French. In : *Acta Math.* 4 (1884), p. 381–392.

[19] Olivier Carton. *Langages formels, calculabilité et complexité*. Vuibert, 2008.

[20] Julien Cassaigne et François Nicolas. *Factor Complexity*. Chapitre 4 de *Combinatorics, Automata and Number Theory* [15]. Cambridge University Press, 2010.

[21] David G. Champernowne. *The Construction of Decimals Normal in the Scale of Ten*. In : *J. London Math. Soc.* 8 (1933), p. 254–260.

[22] Emilie Charlier, Narad Rampersad et Jeffrey Shallit. *Enumeration and Decidable Properties of Automatic Sequences*. In : *Int. J. Found. Comput. Sci.* 23.5 (2012), p. 1035–1066.

[23] Noam Chomsky et George A Miller. *Finite state languages*. In : *Information and control* 1.2 (1958), p. 91–112.

[24] Alan Cobham. *On the Hartmanis-Stearns problem for a class of tag machines*. In : *IEEE Conference Record of 9th Annual Symposium on Switching and Automata Theory*. 1968, p. 51–60.

[25] Alan Cobham. *On the Base-Dependence of Sets of Numbers Recognizable by Finite Automata*. In : *Mathematical Systems Theory* 3.2 (1969), p. 186–192.

[26] Alan Cobham. *Uniform Tag Sequences*. In : *Math. Systems Theory* 6 (1972), p. 164–192.

[27] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest et Clifford Stein. *Introduction to Algorithms (2nd ed.)* MIT Press, 2001.

[28] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest et Clifford Stein. *Introduction à l'algorithmique (2ème ed.)* Traduction française de [27] par Xavier Cazin et Georges-Louis Kocher. Dunod, 2002.

[29] Artūras Dubickas. *On integer sequences generated by linear maps*. In : *Glasgow Mathematical Journal* 51 (02 mai 2009), p. 243–252.

[30] Jean-Marie Dumont et Alain Thomas. *Systèmes de Numération et Fonctions Fractales Relatifs aux Substitutions*. In : *Theor. Comput. Sci.* 65.2 (1989), p. 153–169.

[31] Jean-Marie Dumont et Alain Thomas. *Digital sum problems and substitutions on a finite alphabet*. In : *Journal of Number Theory* 39.3 (1991), p. 351–366.

[32] Jean-Marie Dumont et Alain Thomas. *Digital sum moments and substitutions*. In : *Acta Arith* 64.3 (1993), p. 205–225.

[33] Fabien Durand. *Decidability of the HD0L ultimate periodicity problem*. In : *RAIRO - Theor. Inf. and Applic.* 47.2 (2013), p. 201–214.

[34] Aviezri S. Fraenkel. *Systems of numeration*. In : *American Mathematical Monthly* (1985), p. 105–114.

[35] Christiane Frougny et Karel Klouda. *Rational base number systems for p-adic numbers*. In : *RAIRO - Theor. Inf. and Applic.* 46.1 (2012), p. 87–106.

[36] Christiane Frougny et Jacques Sakarovitch. *Number representation and finite automata*. Chapitre 2 de *Combinatorics, Automata and Number Theory* [15]. Cambridge University Press, 2010.

[37] Seymour Ginsburg et Edwin H. Spanier. *Semigroups, Presburger formulas and languages*. In : *Pacif. J. Math.* 16 (1966), p. 285–296.

[38] Sheila A. Greibach. *One Counter Languages and the IRS Condition*. In : *J. Comput. Syst. Sci.* 10.2 (1975).

[39] M. Hollander. *Greedy Numeration Systems and Regularity*. In : *Theory Comput. Syst.* 31.2 (1998), p. 111–133.

[40] Juha Honkala. *A Decision Method for The Recognizability of Sets Defined by Number Systems*. In : *ITA* 20.4 (1986), p. 395–403.

[41] John E. Hopcroft. *An $n\log n$ algorithm for minimizing states in a finite automaton*. In : *Theory of Machines and Computations*. Sous la dir. de Z. Kohavi et Azaria Paz. Academic Press, 1971.

[42] John E. Hopcroft, Rajeev Motwani et Jeffrey D. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, 2000.

[43] Stephen C. Kleene. *Representation of events in nerve nets and finite automata*. In : *Automata Studies*. Sous la dir. de C. Shannon et John McCarthy. Princeton University Press, 1956, p. 3–41.

[44] Pierre Lecomte et Michel Rigo. *Numeration systems on a regular language*. In : *Theory Comput. Syst.* 34 (2001), p. 27–44.

[45] Pierre Lecomte et Michel Rigo. *Abstract Numeration Systems*. Chapitre 3 de *Combinatorics, Automata and Number Theory* [15]. Cambridge University Press, 2010.

[46] Jérôme Leroux. *A polynomial time Presburger criterion and synthesis for number decision diagrams*. In : *Logic in Computer Science 2005 (LICS 2005)*. IEEE Comp. Soc. Press, 2005, p. 147–156.

[47] Nathalie Loraud. *$\beta$-shift, systèmes de numération et automates*. In : *Journal de théorie des nombres de Bordeaux* 7.2 (1995), p. 473–498.

[48] M. Lothaire (collective). *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

[49] M. Lothaire (collective). *Applied Combinatorics on Words*. Cambridge University Press, 2005.

[50] Kurt Mahler. *An unsolved problem on the powers of 3/2*. In : *Journal of the Australian Mathematical Society* 8 (02 avr. 1968), p. 313–321. issn : 1446-8107.

[51] Victor Marsault. *Surminimisation of Automata*. In : *Developments in Language Theory - 19th International Conference (DLT 2015)*. Sous la dir. d'Igor Potapov. Lect. Notes Comput. Sci. 9168. Springer, 2015, p. 352–363.

[52] Victor Marsault. *A few statistical experiments on minimal words in rational base numeration systems*. Rapp. tech. 2016.

[53] Victor Marsault et Jacques Sakarovitch. *On Sets of Numbers Rationally Represented in a Rational Base Number System*. In : *Algebraic Informatics - 5th International Conference (CAI 2013)*. Sous la dir. de Traian Muntean, Dimitrios Poulakis et Robert Rolland. Lect. Notes Comput. Sci. 8080. Springer, 2013, p. 89–100.

[54] Victor Marsault et Jacques Sakarovitch. *Ultimate Periodicity of b-Recognisable Sets : A Quasilinear Procedure*. In : *Developments in Language Theory - 17th International Conference (DLT 2013)*. Sous la dir. de Marie-Pierre Béal et Olivier Carton. Lect. Notes Comput. Sci. 7907. Springer, 2013, p. 362–373.

[55] Victor Marsault et Jacques Sakarovitch. *Breadth-First Serialisation of Trees and Rational Languages*. In : *Developments in Language Theory - 18th International Conference (DLT 2014)*. Sous la dir. d'Arseny M. Shur et Mikhail V. Volkov. Lect. Notes Comput. Sci. 8633. Springer, 2014, p. 252–259.

[56] Victor Marsault et Jacques Sakarovitch. *On Sets of Numbers Rationally Represented in a Rational Base Number System*. In : *12th Latin American Theoretical Informatics Symposium (LATIN 2016)*. Lect. Notes Comput. Sci. 9644. À paraître. Version préliminaire accessible sur `arXiv:1403.5190` (2014) sous un autre titre. Springer, 2016.

[57] George H. Mealy. *A method for synthesizing sequential circuits*. In : *Bell Syst. Tech. J.* 34 (1955), p. 1045–1079.

[58] Ivan Mitrofanov. *A proof for the decidability of HD0L ultimate periodicity (in Russian)*. Preprint arXiv :1110.4780. 2011.

[59] Edward F. Moore. *Gedanken experiments on sequential machines*. In : *Automata Studies*. Sous la dir. de C. Shannon et John McCarthy. Princeton University Press, 1956, p. 129–156.

[60] Johannes F. Morgenbesser, Wolfgang Steiner et Jörg M. Thuswaldner. *Patterns in rational base number systems*. In : *J. Fourier Anal. Appl.* 19.2 (2013), p. 225–250.

[61] Andrei A. Muchnik. *The definable criterion for definability in Presburger arithmetic and its applications (in Russian)*. Preprint, Institute of New Technologies, Moscou. 1991.

[62] Andrei A. Muchnik. *The definable criterion for definability in Presburger arithmetic and its applications*. In : *Theor. Comput. Sci.* 290.3 (2003). English translation of [61], p. 1433–1444.

[63] J. Myhill. *Finite automata and the representation of events*. Rapp. tech. WADD 57-624. 1957.

[64] Anil Nerode. *Linear Automaton Transformations*. In : *American Mathematical Society* 9.4 (1958), p. 541–544.

[65] Andrew M. Odlyzko et Herbert S. Wilf. *Functional iteration and the Josephus problem*. In : *Glasgow Mathematical Journal* 33 (02 mai 1991). issn : 1469-509X.

[66] Blaise Pascal. *Œuvres complètes*. Le traité *De numeris multiplicibus*, rédigé avec les autres traités arithmétiques avant 1654, est publié chez Guillaume Desprez en 1665. Seuil, 1963.

[67] Dominique Perrin et Jean-Éric Pin. *Infinite words : automata, semigroups, logic and games*. T. 141. Academic Press, 2004.

[68] Giuseppe Pirillo. *A new characteristic property of the palindrome prefixes of a standard Sturmian word*. In : *Séminaire Lotharingien de Combinatoire [electronic only]* 43 (1999).

[69] Michael O. Rabin et Dana Scott. *Finite automata and their decision problems*. In : *IBM Journal of Research and Development* 3.2 (avr. 1959), p. 114–125.

[70] Alfréd Rényi. *Representations for real numbers and their ergodic properties*. In : *Acta Mathematica Hungarica* 8.3 (1957), p. 477–493.

[71] Michel Rigo et Arnaud Maes. *More on Generalized Automatic Sequences*. In : *Journal of Automata, Languages and Combinatorics* 7.3 (2002), p. 351–376.

[72] Jacques Sakarovitch. *Éléments de théorie des automates*. Vuibert, 2003.

[73] Jacques Sakarovitch. *Elements of Automata Theory*. Corrected English translation of [72]. Cambridge University Press, 2009.

[74] Alexei L. Semenov. *Presburgerness of predicates regular in two number systems*. In : *Siberian Mathematical Journal* 18.2 (1977). English translation of [75], p. 289–300.

[75] Alexei L. Semenov. *Presburgerness of predicates regular in two number systems (in Russian)*. In : *Sibirskii Matematicheskii Zhurnal* 18 (1977), 403–418.

[76] Jeffrey Shallit. *Numeration Systems, Linear Recurrences, and Regular Sets*. In : *Inf. Comput.* 113.2 (1994), p. 331–347.

[77] Robert E. Tarjan. *Depth-First Search and Linear Graph Algorithms*. In : *SIAM J. Comput.* 1.2 (1972), p. 146–160.

[78] Edouard Zeckendorf. *Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas*. In : *Bull. Soc. Roy. Sci. Liege* 41 (1972), p. 179–182.

# Index des définitions

## — ∗ —



## — A —

## — B —



## — C —



## — D —



## — E —



## — G —



## — I —



## — L —

## — M —



## — N —



## — O —



## — P —



## — Q —



## — R —



## — S —

## — T —



## — V —