Trees and Languages with Periodic Signature

Victor Marsault[a], Jacques Sakarovitch[b]

[a] *IRIF, Université Denis Diderot, 8 place Aurélie Nemours, 75013 Paris, France.*
[b] *LTCI, CNRS / Telecom-ParisTech, 46 rue Barrault, 75013 Paris, France.*

Abstract

The *signature* of a labelled tree (and hence of its prefix-closed branch language) is the sequence of the degrees of the nodes of the tree in the *breadth-first traversal*. In a previous work, we have characterised the signatures of the regular languages. Here, the trees and languages that have the simplest possible signatures, namely the periodic ones, are characterised as the sets of representations of the integers in rational base numeration systems.

For any pair of co-prime integers p and q, $p > q > 1$, the language $L_{\frac{p}{q}}$ of representations of the integers in base $\frac{p}{q}$ looks chaotic and does not fit in the classical Chomsky hierarchy of formal languages. On the other hand, the most basic example given by $L_{\frac{3}{2}}$, the set of representations in base $\frac{3}{2}$, exhibits a remarkable regularity: its signature is the infinite *periodic* sequence: $2, 1, 2, 1, 2, 1, \ldots$

We first show that $L_{\frac{p}{q}}$ has a periodic signature and the period (a sequence of q integers whose sum is p) is directly derived from the *Christoffel word* of slope $\frac{p}{q}$. Conversely, we give a canonical way to label a tree generated by any periodic signature; its branch language then proves to be the set of representations of the integers in a rational base (determined by the period) and written with a non-canonical alphabet of digits. This language is very much of the same kind as a $L_{\frac{p}{q}}$ since rational base numeration systems have the key property that, even though $L_{\frac{p}{q}}$ is not regular, normalisation is realised by a finite letter-to-letter transducer.

Keywords: Rational base numeration system, Breath-first signature, Abstract numeration system

1. Introduction

This work was motivated by the study of *rational base numeration systems* which were first defined in [1]. The introduction of these systems allowed to make some progress in a number theoretic problem, by means of automata theory and combinatorics of words. At the same time, they raised the problem

Email address: `Victor.Marsault@liafa.univ-paris-diderot.fr` (Victor Marsault)
URL: `http://victor.marsault.xyz` (Victor Marsault)

of understanding the structure of the sets of representations of the integers in these systems from the point of view of formal language theory.

At first sight, these sets look rather chaotic and do not fit at all in the classical Chomsky hierarchy of languages. They all enjoy a property that makes them defeat, so to speak, any kind of iteration lemma. On the other hand, the most common example given by the set of representations in the base $\frac{3}{2}$ exhibits a remarkable regularity. The set $L_{\frac{3}{2}}$ of representations, which are words written with the three digits $\{0, 1, 2\}$, is prefix-closed and thus naturally represented as a subtree of the full ternary tree which is shown in Figure 1. It is then easily observed that the *breadth-first* traversal of that tree yields an infinite *periodic* sequence of degrees: $2, 1, 2, 1, 2, 1, \ldots = (2\,1)^\omega$. Moreover, the sequence of labels of the arcs in the same breadth-first search is also a purely periodic sequence $0, 2, 1, 0, 2, 1, \ldots = (0\,2\,1)^\omega$.[1]

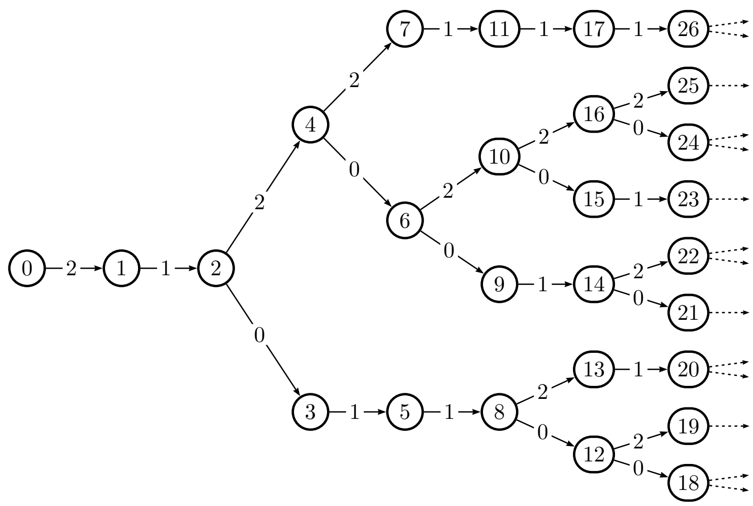


Figure 1: The tree $\mathcal{T}_{\frac{3}{2}}$, representation of the language $L_{\frac{3}{2}}$

Let us call *signature* of a tree (or of the corresponding prefix-closed language) the sequence of degrees in a breadth-first traversal of the tree. With this example, we face a situation where a regular process, a periodic signature, give birth to the highly non regular language, $L_{\frac{3}{2}}$. This paradox was the incentive to look at the breadth-first traversal description of languages in general. We have shown in [13] that regular languages are characterised by signatures belonging to a special class of morphic words. The purpose of this paper is to establish that a periodic signature is *characteristic* of the languages of representations of the integers in rational base numeration systems (roughly speaking and up to

[1] The sequence of degrees observed on the tree in the figure begins indeed with a 1 instead of a 2, the sequence of labels begins at the second term. These discrepancies will be explained later.

simple and rational transformations).

Let us be more specific in order to state more precisely the characterisation results. An ordered tree of finite degree $\mathcal{T}$ is characterised by the infinite sequence of the degrees of its nodes visited in the order given by the breadth-first traversal, which we call the *signature* $\boldsymbol{s}$ of $\mathcal{T}$. Such a signature $\boldsymbol{s}$, together with an infinite sequence $\boldsymbol{\lambda}$ of letters taken in an ordered alphabet form a *labelled signature* $(\boldsymbol{s}, \boldsymbol{\lambda})$ and characterises then a *labelled* tree $\mathcal{T}$. The breadth-first search of $\mathcal{T}$ corresponds to the enumeration in the *radix order* of the prefix-closed language $L_{\mathcal{T}}$ of branches of $\mathcal{T}$.

We call *rhythm* of *directing parameter* (q, p) a q-tuple $\mathbf{r}$ of integers whose sum is p: $\mathbf{r} = (r_0, r_1, \ldots, r_{q-1})$ and $p = r_0 + r_1 + \cdots + r_{q-1}$. With $\mathbf{r}$, we associate sequences $\boldsymbol{\gamma}$ of p letters that meet some consistency conditions. And we consider the languages that are determined by the labelled signature $(\mathbf{r}^\omega, \boldsymbol{\gamma}^\omega)$. The characterisation announced above splits then in two parts.

We first determine (Theorem 22) the remarkable labelled signature $({\mathbf{r}_{\frac{p}{q}}}^\omega, {\boldsymbol{\gamma}_{\frac{p}{q}}}^\omega)$ of the languages $L_{\frac{p}{q}}$. We call it the *Christoffel rhythm* associated with $\frac{p}{q}$, as it can be derived from the more classical notion of Christoffel word of slope $\frac{p}{q}$ (*cf.* [2]), that is, the canonical way to approximate the line of slope $\frac{p}{q}$ on a $\mathbb{Z} \times \mathbb{Z}$ lattice. Meanwhile, the labelling $\boldsymbol{\gamma}_{\frac{p}{q}}$ is the integer sequence induced by the generation of $\mathbb{Z}/p\mathbb{Z}$ by q.

The converse is more convoluted but essentially resides in the definition of a *special labelling* $\boldsymbol{\gamma}_{\mathbf{r}}$ associated with every rhythm $\mathbf{r}$ (Definition 29). It is then established (Theorem 32) that the language $L_{\mathbf{r}}$ generated by the labelled signature $(\mathbf{r}^\omega, \boldsymbol{\gamma}_{\mathbf{r}}^\omega)$ is a *non-canonical representation* of the integers in the rational base z where $z = \frac{p}{q}$ is the growth ratio of the rhythm $\mathbf{r}$. The properties of alphabet conversion in rational base numeration systems (*cf.* [1, 7]) allow to conclude that for every rhythm $\mathbf{r}$, the language $L_{\mathbf{r}}$ is as complicated (or as simple, in the degenerate case where the growth ratio happens to be an integer) as these langages $L_{\frac{p}{q}}$.

The same techniques allow to treat the generalisation to *ultimately periodic* which raises no special difficulties and the results readily extend.

The languages with periodic labelled signature keep most of their mystery. But we have at least established that they are all alike, essentially similar to the representation languages in rational base numeration systems, and that variations in the rhythm and labelling do not really matter.

A short version of this article has been presented at the LATIN 2016 conference [14]. Earlier versions were posted on arXiv under a different title. Most of the results are also part of the PhD thesis of the first author [11]. The periodicity of the signature of the languages $L_{\frac{p}{q}}$ has been observed independently and in a slightly different setting in [5].

2. Signatures of trees and languages

For sake of completeness, we reproduce here a section from [13] describing our general framework.

We describe here a process of *serialisation* of (infinite) trees, (infinite) labelled trees, and (infinite) prefix-closed languages, that is, the representation of such objects by one, or two, (infinite) words, using the assumption of the existence of an underlying order. We also recall the related concept of abstract numeration system and introduce the one of padded language.

2.1. On trees

Classically, a tree is an undirected graph in which any two vertices are connected by exactly one path (*cf.* [4], for instance). Our point of view differs in two respects. First, a tree is a *directed* graph such that (i) there exists a *unique* vertex, called *root*, which has no incoming arc, and (ii) there is a *unique (oriented) path* from the root to every other vertex. Second, a tree is *ordered*, that is, the set of children of every node is totally ordered.

In the figures, we draw trees with the root on the left, arcs rightwards and the child order will be implicitly defined by the convention that children placed higher are greater (according to this order).

It will prove to be convenient to have a slightly different look at trees and to consider that the root of a tree is also a *child of itself*, that is, bears a loop onto itself.[2] We call such a structure an *i-tree*. It is so close to a tree that we make no real distinction between them. Nevertheless, some definitions or results are easier or more straightforward when stated for i-trees, and others when stated for trees: it is then convenient to have them both at hand. A tree will usually be denoted by $\mathcal{T}_x$ for some index x and the associated i-tree by $\mathcal{I}_x$. Figure 2 shows such a pair of a tree and the associated i-tree.

The degree of a node is the number of its children. In the sequel, we consider infinite (i-)trees of finite degree, that is, all nodes of which have finite degree. (We consider indeed infinite (i-)trees of *bounded* degree, but this restriction does not matter for the definitions to come.) The breadth-first traversal of such an ordered (i-)tree defines a total ordering of its nodes.

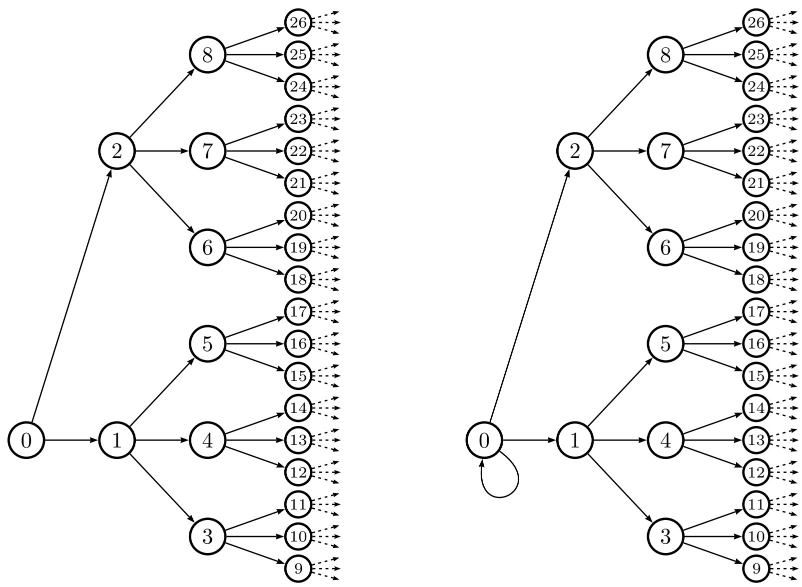
Convention. *The set of nodes of an (i-)tree is always the set* $\mathbb{N}$ *of the non-negative integers.*

With this convention, the root is 0 and n is the $(n+1)$-th node visited by the traversal. For n, m in $\mathbb{N}$, we write

$$n \xrightarrow[\mathcal{T}]{} m$$

whenever m is a child of n in $\mathcal{T}$. We denote by $\mathsf{deg}(n)$ the degree of the node n.

[2] This convention is sometimes taken when implementing tree-like structures (for instance in the unix/linux file system).



(a) A tree that is almost ternary (b) The ternary i-tree

Figure 2: The tree and i-tree associated with the base 3 numeration system

2.2. Signatures of trees

We call *signature* any infinite sequence $\boldsymbol{s} = s_0\, s_1\, s_2 \cdots$ of non-negative integers. Whenever the signature $\boldsymbol{s}$ is obvious from the context, we simply denote by S_j, for every integer j, the partial sum of the j first letters of $\boldsymbol{s}$:

$$\forall j \in \mathbb{N} \qquad S_j \;=\; \sum_{i=0}^{j-1} s_i \;\;,$$

that is, $S_0 = 0$, $S_1 = s_0$ and more generally $S_j = S_{j-1} + s_{j-1}$ for every $j > 0$.

Definition 1. *A signature* $\boldsymbol{s} = s_0 s_1 s_2 \cdots$ *is* valid *if the following holds:*

$$\forall j \in \mathbb{N} \qquad S_{j+1} \;>\; j+1 \;\;. \tag{1}$$

In particular, the validity of $\boldsymbol{s}$ implies that s_0 is greater than, or equal to, 2.

Definition 2.

(i) *The* breadth-first signature *or, for short, the* signature*, of an i-tree* $\mathcal{I}$ *is the sequence* $\boldsymbol{s} = s_0 s_1 s_2 \cdots$ *of the degrees of the nodes of the* i-tree $\mathcal{I}$*:*

$$\forall i \in \mathbb{N} \qquad s_i \;=\; \mathsf{deg}(i) \;\;.$$

(ii) *The breadth-first signature of a tree* $\mathcal{T}$ *is the signature of* the corresponding i-tree.

In other words, we take the convention that the signature is always the one of an i-tree. Figure 2 shows both the tree and the i-tree the signature of which is 3^ω. Valid signatures are in bijection with infinite i-trees of finite degree, as expressed by the next proposition.

Proposition 3.
(i) *Let* $\boldsymbol{s} = s_0\, s_1\, s_1 \cdots$ *be a valid signature. There exists a unique i-tree* $\mathcal{I}_{\boldsymbol{s}}$ *whose signature is* $\boldsymbol{s}$*: the i-tree such that every node* n *has* s_n *children, the* s_n *nodes of the integer interval* $\{S_n,\, S_n+1,\, \ldots,\, S_{n+1}-1\}$.
(ii) *The signature of any infinite (i-)tree of finite degree is valid.*

Proof. The proof of (i) takes essentially the form of a procedure that generates an i-tree from a valid signature $\boldsymbol{s} = s_0 s_1 s_2 \cdots$. It maintains two integers: the node n to be processed and the number m of nodes created so far, both initially set to 0. At step $(n+1)$ of the procedure, s_n nodes are created, namely the nodes $m, m+1, \ldots, (m + s_n - 1)$, and s_n edges are created, all with starting point n, and one for each of these new nodes as end point. Then n is incremented by 1, and m by s_n.

This procedure indeed describes an i-tree. The first node created is 0 and the first arc created is the loop $0 \longrightarrow 0$ on the root. It is verified by induction that at every step, m is equal to S_n. The initial conditions ($n = m = 0$) indeed satisfies this equality since S_0 is an empty sum.

The validity of $\boldsymbol{s}$ ensures that at the end of every step of the procedure $n < m$ holds (but not at the beginning of the first step where $n = m = 0$). It follows that every node is strictly larger than its father, but for the root, whose father is itself.

(ii) Let $\mathcal{I}$ be an infinite i-tree and $\boldsymbol{s} = s_0 s_1 s_2 \cdots$ its signature; S_n is the number of children of the first n nodes of $\mathcal{I}$. If $\boldsymbol{s}$ is not valid, the smallest integer j for which Equation (1) does not hold is such that $S_j = j$, in which case the set of the children of the j first nodes is of cardinal j, hence $\mathcal{I}$ has j nodes and is finite. □

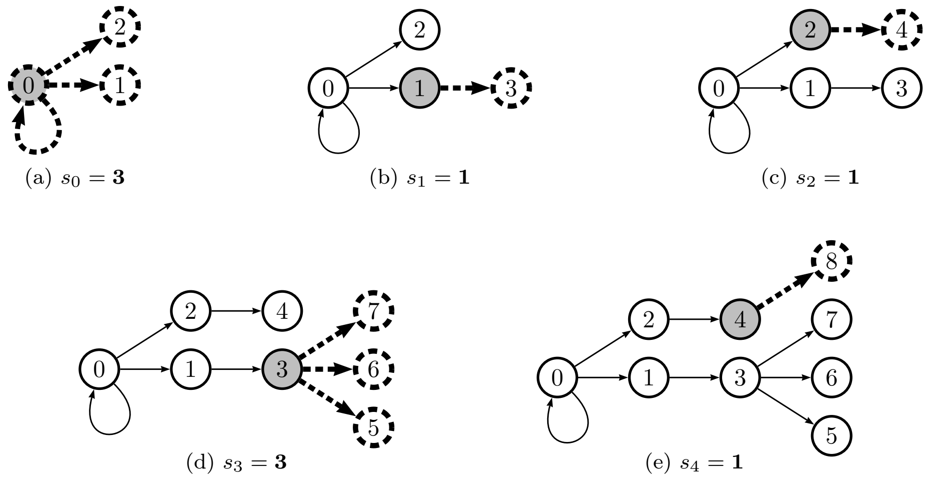
Figure 3 shows the first five steps of the generation process applied to the signature $\boldsymbol{s} = (311)^\omega$. A slightly larger initial part of the resulting infinite i-tree $\mathcal{I}_{(311)^\omega}$ together with a labelling is shown in Figure 4.

2.3. Labelled signatures of labelled trees

In our framework, alphabets are totally ordered. In the case of alphabets of digits, the natural order is of course implicitly used. As usual, the length of a finite word w is denoted by $|w|$.

We say that a word $w = a_0\, a_1 \cdots a_{k-1}$ is *increasing* if its letters are in increasing order, that is if $a_0 < a_1 < \cdots < a_{k-1}$.

A labelled tree $\mathcal{T}$, or i-tree $\mathcal{I}$, is an (i-)tree every arc of which holds a label taken in an alphabet A. Since both A and $\mathcal{T}$ (or $\mathcal{I}$) are ordered, the labels on the arcs have to be *consistent* with these two orders: two arcs originating from the same node n must be labelled by two letters whose order is the same as the



(a) $s_0 = \mathbf{3}$ (b) $s_1 = \mathbf{1}$ (c) $s_2 = \mathbf{1}$ (d) $s_3 = \mathbf{3}$ (e) $s_4 = \mathbf{1}$

Figure 3: The first five steps of the generation of $\mathcal{I}_{(311)^\omega}$

endpoints of the arcs or, more intuitively, an arc to a greater child is labelled by a greater letter. For n, m in $\mathbb{N}$, and a in A, we write

$$n \xrightarrow[\mathcal{I}]{a} m \tag{2}$$

whenever m is a child of n in $\mathcal{I}$ and the arc from n to m holds the label a.

The labelling $\boldsymbol{\lambda} = \lambda_0\,\lambda_1\,\lambda_2 \cdots$ of a labelled i-tree $\mathcal{I}$ (labelled in A) is an infinite word of A^ω. It is the sequence of the labels of the arcs of $\mathcal{I}$ visited in a breadth-first traversal, that is:

for every node m in $\mathbb{N}$, λ_m is *the label of the unique arc incoming* to m.

In particular, λ_0 is the label of the loop on the root of $\mathcal{I}$.

As it is an infinite sequence of non-negative integers, a signature $\boldsymbol{s}$ naturally *determines a factorisation* of any other infinite word $\boldsymbol{\lambda}$: $\boldsymbol{\lambda} = w_0\,w_1\,w_2 \cdots$ by the condition that $|w_n| = s_n$ for every n in $\mathbb{N}$ (and thus $w_n = \varepsilon$ if $s_n = 0$).

Definition 4. *Let $\boldsymbol{s}$ be a signature. An infinite word $\boldsymbol{\lambda}$ in A^ω is* consistent *with $\boldsymbol{s}$ if the factorisation $\boldsymbol{\lambda} = w_0\,w_1\,w_2 \cdots$ determined by $\boldsymbol{s}$ has the property that every w_n is an* increasing word.

A pair $(\boldsymbol{s}, \boldsymbol{\lambda})$ of infinite words is a valid labelled signature *if $\boldsymbol{s}$ is a valid signature and if $\boldsymbol{\lambda}$ is consistent with $\boldsymbol{s}$.*

A simple and formal verification yields the following.

Proposition 5. *A labelled i-tree $\mathcal{I}$ uniquely determines a valid labelled signature and conversely any valid labelled signature $(\boldsymbol{s}, \boldsymbol{\lambda})$ uniquely determines a labelled i-tree $\mathcal{I}_{(\boldsymbol{s},\boldsymbol{\lambda})}$ whose labelled signature is precisely $(\boldsymbol{s}, \boldsymbol{\lambda})$.*

Figure 4 shows the labelling of the i-tree whose signature is $\boldsymbol{s} = (3\,1\,1)^\omega$ by

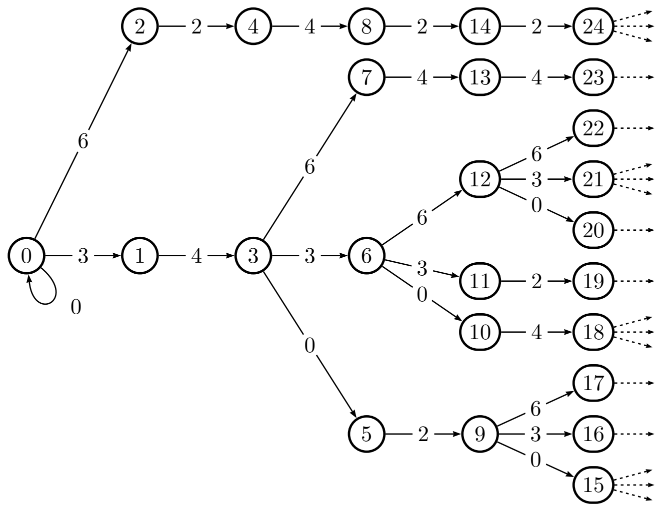


Figure 4: The padded language $0^*L_{(311)}$

the infinite periodic[3] word $\boldsymbol{\lambda} = (0\,3\,6\,.\,4\,.\,2)^\omega$. (This is of course a very special labelling: labellings consistent with $\boldsymbol{s}$ do not need to be periodic, but periodic words are the easiest cases of finitely defined infinite words.)

2.4. Labelled signatures of languages

The *branch language* of a labelled tree is the set of words that label all paths from the root to every node of the tree. It is a *prefix-closed* language. Conversely, every prefix-closed language over a totally ordered alphabet uniquely defines a labelled ordered tree.

The branch language of a labelled *i-tree* is a language of a special form that we call *padded*. The most common example of a padded language is given by the writings of the integers in (an integer) base p. The representation of an integer is a word over the alphabet $A_p = \{0, 1, \ldots, p-1\}$ that does not begin with a 0 (and the set of representations is not a padded language). But there are cases where one wants to have the possibility to write a number differently. For the addition of two numbers for instance, it is convenient to have representations of the same length, and the shorter one is prefixed with the adequate numbers of 0's such that the length of both words match. It is currently said that the shorter representation is *padded* with 0's.

The branch language K of an i-tree has clearly the property that any word of K can be prefixed by an arbitrary number of the label of the loop (on the root) and still be in K. The label of the loop of an i-tree is called *padding*

[3] The dots in the period are written to make obvious the factorisation of the labelling $\boldsymbol{\lambda}$ determined by the signature $\boldsymbol{s}$.

letter. The notion of padded language can be given a purely language-theoretic definition as follows.

Definition 6. *Let A be a (totally ordered) alphabet and let a be a letter in A. A language K over A is said to be a-*padded *if the following conditions hold:*
(i) $u \in K \Leftrightarrow a\,u \in K$;
(ii) *If $b\,u$ is in K, with b in A, then b is not smaller than a.*
A language is padded *if it is a-padded for some letter a.*

If a language is padded, it is a-padded for a unique a: the second condition of Definition 6 implies that if K is both a-padded and a'-padded, then $a = a'$.

Notation. *A padded language is written either as a^*L, or as $\widehat{L}$ if the padding letter does not need to be specified; in both cases L is then implicitly defined as the set of the words of the padded language which do not start with the padding letter.*

It is easy to verify that if $\mathcal{I}$ is a labelled i-tree and $\mathcal{T}$ the corresponding tree, then the branch language of $\mathcal{I}$ is a padded language $\widehat{L}$ where L is the branch language of $\mathcal{T}$. Our notation transfers at the level of branch languages the correspondence between trees and i-trees.

To some extent, there is no difference, between a labelled (i-)tree and the prefix-closed language of its branches. We may thus speak of the labelled signature, and of the signature, of a prefix-closed language and take the corresponding notation: the branch language of a tree $\mathcal{T}_x$ (resp. an i-tree $\mathcal{I}_x$), for some index x, is denoted by L_x (resp. $\widehat{L_x}$). Proposition 5 may then be rephrased in the following way.

Proposition 7. *A prefix-closed padded language $\widehat{L}$ uniquely determines a labelled i-tree and hence a valid labelled signature,* the *labelled signature of $\widehat{L}$ and conversely any valid labelled signature $(\boldsymbol{s}, \boldsymbol{\lambda})$ uniquely determines a labelled i-tree $\mathcal{I}_{(\boldsymbol{s},\boldsymbol{\lambda})}$ and hence a prefix-closed padded language $\widehat{L}_{(\boldsymbol{s},\boldsymbol{\lambda})}$, whose signature is precisely $(\boldsymbol{s}, \boldsymbol{\lambda})$.*

Remark 8. *Any language L over a totally ordered alphabet A* can be made padded *by adding a new letter $\#$ to A and by setting $\#$ smaller than all letters in A. We then consider $K = \#^*L$ instead of L and L is rational if and only if so is K.*

This construction may look awkward but it naturally occurs for instance in the study of ANS (cf. following Section 2.5), where no digit 0 exists in general.

Remark 9. *A very 'simple' tree may produce an artificially 'complex' language when paired with a 'complex' labelling. For instance, the infinite unary tree may be labelled by an infinite word whose prefixes form a non-recursive language. Therefore, any result relative to* languages *defined by signatures will always require some hypothesis to restrict the* labelling*. This is illustrated later on in this article (and in particular in the example shown in Figure 6) by pairing periodic signatures with periodic labellings.*

2.5. Trees, languages and abstract numeration systems

The identification between a prefix-closed language L over a totally ordered alphabet and the ordered labelled tree $\mathcal{T}_L$ whose branch language is L (and whose set of nodes is $\mathbb{N}$) is very close to the notion of *Abstract Numeration Systems* (ANS) introduced by Lecomte and Rigo (*cf.* [9, 10]). In this setting, the language L over the totally ordered alphabet A is ordered by the trace of the *radix order* over A^* and — since it is meant to define a *numeration system* — the representation of an integer n in this system, also called the *L-representation* of n and denoted by $\langle n \rangle_L$, is the $(n+1)$-th word of L in the radix order.

This notion generalises the situation in classical numeration systems. Let us take for instance the numeration in base 3. The usual way for defining the representation of integers in that system is to define an *evaluation function* $\pi_3 \colon {A_3}^* \rightarrow \mathbb{N}$ by the following: if $w = d_k d_{k-1} \cdots d_1 d_0$ is a word of length $k+1$, then

$$\pi_3\,(w) \;=\; \pi_3\,(d_k d_{k-1} \cdots d_1 d_0) \;=\; \sum_{i=0}^{k} d_i\, 3^i \;\; . \tag{3}$$

Note that in this case, it is convenient to have the letters of w indexed *from right to left*.

As said above, every integer n is uniquely represented by a word $\langle n \rangle_3$ of ${A_3}^* = \{0,1,2\}^*$ which does not begin with a 0, that is, the set L_3 of integer representations in base 3 is defined by

$$L_3 \;=\; \{\langle n \rangle_3 \mid n \in \mathbb{N}\} \;=\; \{1,2\}\,\{0,1,2\}^* \cup \{\varepsilon\}$$

(with the convention that the integer 0 is represented by ε rather than by the digit 0, which suits us better). It then turns out that $\langle n \rangle_3$ is the $(n+1)$*-th word* of L_3 in the radix order, that is, the representation of n in base 3 coincides with the representation of n in the ANS defined by L_3 over the ordered alphabet $\{0,1,2\}$:

$$\forall n \in \mathbb{N} \qquad \langle n \rangle_3 \;=\; \langle n \rangle_{L_3} \;\; .$$

On the other hand, since $\mathcal{T}_L$ is visited by a *breadth-first search*, the $(n+1)$-th node of $\mathcal{T}_L$ — labelled with n — is reached from the root by the $(n+1)$-th word — in the radix order — of the branch language of $\mathcal{T}_L$, that is, L itself (under the hypothesis that L is prefix-closed, which is necessary for the identification between L and $\mathcal{T}_L$).

These two descriptions show that considering a prefix-closed language over an ordered alphabet as an ANS or as the branch language of a labelled ordered tree are two ways of expressing the same concept, namely the radix order over the language. The similarity between the two notions is further shown in the following equation

$$\forall n \in \mathbb{N} \qquad 0 \xrightarrow[\mathcal{T}_L]{\langle n \rangle_L} n \;\; , \tag{4}$$

which implies

$$\forall n, m \in \mathbb{N}\,,\; \forall a \in A \qquad \langle n \rangle_L\, a \;=\; \langle m \rangle_L \quad \Longleftrightarrow \quad n \xrightarrow[\mathcal{T}_L]{a} m \;\; . \tag{5}$$

3. Rythmic Trees and Languages

In this paper, we are interested in signatures and labellings that are periodic. We call any period of a periodic signature *a rhythm* and any period of a labelling still *a labelling*, by abuse of language.

3.1. Rhythms and their Geometric Representations

Given two integers n and m such that $m > 0$, we denote by $\frac{n}{m}$ their division in $\mathbb{Q}$; by $n \% m$ the remainder of the Euclidean division of n by m, that is satisfying $n = \lfloor \frac{n}{m} \rfloor m + (n \% m)$, hence $0 \leqslant (n \% m) < m$. We also denote the integer interval $\{n, (n+1), \ldots, m\}$ by $[\![n, m]\!]$.

Definition 10. *Let p and q be two integers with $p > q \geqslant 1$.*

(i) *We call* rhythm *of directing parameter (q, p), a q-*tuple $\mathbf{r}$ *of non-negative integers* whose sum is p*:*

$$\mathbf{r} \;=\; (r_0, r_1, \ldots, r_{q-1}) \qquad \textit{and} \qquad \sum_{i=0}^{q-1} r_i \;=\; p \;\; .$$

(ii) *We say that a rhythm* $\mathbf{r}$ *is* valid *if it satisfies the following equation:*

$$\forall j \in [\![0, q-1]\!] \qquad \sum_{i=0}^{j} r_i \;>\; j+1 \;\; . \tag{6}$$

(iii) *We call* growth ratio *of* $\mathbf{r}$ *the rational number $z = \frac{p}{q}$, also written $z = \frac{p'}{q'}$ where p' and q' are coprime; it is always greater than* 1.

Examples of rhythms of growth ratio $\frac{5}{3}$ are $(2,2,1)$, $(3,1,1)$, $(1,2,2)$, $(3,0,2)$, $(2,1,3,0,0,4)$; all but the third one are valid; the directing parameter is (3,5) for the first four, and (6,10) for the last one.

A simple verification yields that the notions of validity of rhythms and of signatures are consistent: a rhythm $\mathbf{r}$ is valid if and only if the signature $\mathbf{r}^\omega$ is valid.

In the following, whenever the reference to a rhythm $\mathbf{r} = (r_0, r_1, \ldots, r_{q-1})$ is clear, we denote simply by R_j the partial sum of the first j components of $\mathbf{r}^\omega$:

$$\forall j \in \mathbb{N} \qquad R_j \;=\; \sum_{i=0}^{j-1} r_{i \% q} \;\; ,$$

that is, $R_0 = 0$, $R_1 = r_0$ and more generally $R_j = R_{j-1} + r_{(j-1)\% q}$ for every positive integer j. Using previous notation, R_j is equal to S_j for all integers j, if we set $\boldsymbol{s} = \mathbf{r}^\omega$.

In the following, and when it does not produce ambiguity, we rather write rhythms as we write words, that is, as plain sequences of letters (integers): $\mathbf{r} = r_0\, r_1 \cdots r_{q-1}$.

3.2. Generating Trees by Rhythm

As we have already shown in the previous Section 2 (*cf.* proof of Proposition 3), an (i-)tree can be *generated* by its signature $\boldsymbol{s}$, hence in the present case from its rhythm $\mathbf{r}$. In fact, the example (Figure 3) we used to illustrate this generation process had indeed a periodic signature.

Whenever a tree is generated by a rhythm, the procedure becomes periodic and q consecutive steps always create p (consecutive) new nodes and the p arcs reaching them. It follows that, in the resulting tree, q consecutive nodes (in the breadth-first traversal) have p consecutive children, hence the name *growth ratio* given to the *number* $\frac{p}{q}$. More precisely, the following holds.

Lemma 11. *Let $\mathcal{I}_{\mathbf{r}}$ be the i-tree generated by a rhythm $\mathbf{r}$ of directing parameter (q,p). Then, the following equation holds:*

$$\forall n,m \in \mathbb{N} \qquad n \xrightarrow[\mathcal{I}_{\mathbf{r}}]{} m \qquad \Longleftrightarrow \qquad (n+q) \xrightarrow[\mathcal{I}_{\mathbf{r}}]{} (m+p) \quad .$$

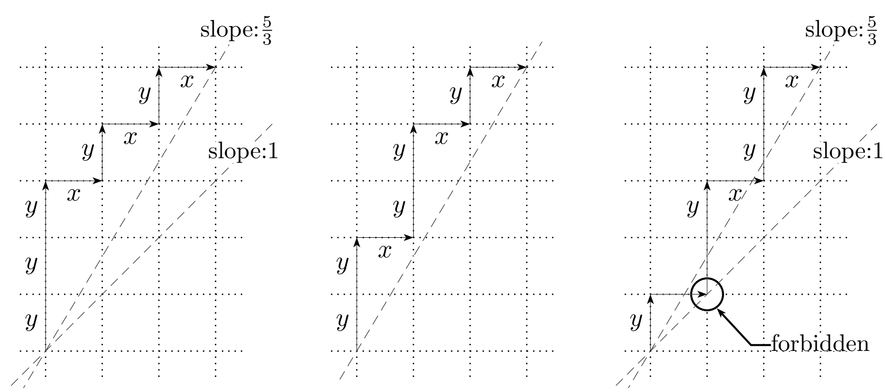
Rhythms are given a very useful geometric representation as *paths* in the $(\mathbb{Z}\times\mathbb{Z})$-lattice and such paths are coded by *words* of $\{x,y\}^*$ where x denotes an horizontal unit segment and y a vertical unit segment. Hence the name *path* given to a *word* associated with a rhythm.

Definition 12. *Let $\mathbf{r} = r_0\, r_1 \cdots r_{q-1}$ be a rhythm of directing parameter (q,p). With $\mathbf{r}$, we associate the word $\mathtt{path}(\mathbf{r})$ of $\{x,y\}^*$:*

$$\mathtt{path}(\mathbf{r}) = y^{r_0} x\, y^{r_1} x\, y^{r_2} \cdots x\, y^{r_{q-1}} x$$

which corresponds to a path from $(0,0)$ to (q,p) in the $(\mathbb{Z}\times\mathbb{Z})$-lattice.

Figure 5 shows the paths associated with three rhythms of directing parameter $(3,5)$. It then appears clearly that Definition 10 (ii) can be restated as *a rhythm is valid if and only if the associated path is strictly above the line of slope 1 passing through the origin.*



(a) rhythm 3 1 1 (b) rhythm 2 2 1 (c) rhythm 1 2 2

Figure 5: Words and paths associated with rhythms of directing parameter $(3,5)$

3.3. Generating (Padded) Languages by Rhythm and Labelling

We consider here periodic signatures $\boldsymbol{s} = \mathbf{r}^\omega$ where $\mathbf{r}$ is a rhythm of directing parameter (q, p). We then will consider pairs $(\boldsymbol{s}, \boldsymbol{\lambda})$ with $\boldsymbol{\lambda} = \boldsymbol{\gamma}^\omega$ where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_{p-1})$ is a sequence of letters (digits) *of length* p. As for rhythms, we rather write the labellings as *words* than with parantheses and commas.

It follows from Lemma 11 that the labelling is consistent on the whole tree if and only if it is consistent on the first q nodes, hence on the first p arcs. By abuse of language, we call in the following $\boldsymbol{\gamma}$ the *labelling* while it is indeed more precisely the *period of the labelling*.

Definition 13. *Let* $\mathbf{r}$ *be a rhythm of directing parameter* (q, p) *and* $\boldsymbol{\gamma}$ *a labelling of length* p.

(i) *We call* factorisation of $\boldsymbol{\gamma}$ induced by $\mathbf{r}$, *the factorisation* $\boldsymbol{\gamma} = u_0\, u_1 \cdots u_{q-1}$ *satisfying* $|u_i| = r_i$ *for every integer* i, $0 \leqslant i < q$. *(Note that* $u_i = \varepsilon$ *if* $r_i = 0$.*)*

(ii) *The labelling* $\boldsymbol{\gamma}$ *is then said* consistent with $\mathbf{r}$ *if each factor* u_i *is increasing.*

(iii) *The pair* $(\mathbf{r}, \boldsymbol{\gamma})$ *is then said* valid *if* $\mathbf{r}$ *is valid and* $\boldsymbol{\gamma}$ *is consistent with* $\mathbf{r}$.

For instance, the labelling $\boldsymbol{\gamma} = 0\,3\,6\,4\,2 = \gamma_0\,\gamma_1 \cdots \gamma_{p-1}$ is consistent with the rhythm $\mathbf{r} = 3\,1\,1$ since $u_0 = 0\,3\,6$, $u_1 = 4$ and $u_2 = 2$ are all increasing and $u_0\, u_1\, u_2$ is the factorisation of $\boldsymbol{\gamma}$ induced by $\mathbf{r}$.

We denote by $\mathcal{I}_{(\mathbf{r},\boldsymbol{\gamma})}$ (resp. $\mathcal{T}_{(\mathbf{r},\boldsymbol{\gamma})}$) the labelled i-tree (resp. tree) *generated by* a rhythm $\mathbf{r}$ of directing parameter (q, p) and a labelling $\boldsymbol{\gamma} = \gamma_0\,\gamma_1 \cdots \gamma_{p-1}$ consistent with $\mathbf{r}$. The following equation gives its complete expression:

$$\forall n, m \in \mathbb{N} \qquad n \xrightarrow[\mathcal{I}_{(\mathbf{r},\boldsymbol{\gamma})}]{a} m \quad \Longleftrightarrow \quad \left\{ \begin{array}{lcl} R_n & \leqslant\ m & <\ R_{n+1} \\ a & = & \gamma_{(m\,\%\,p)} \end{array} \right. \quad . \tag{7}$$

For instance, the labelled tree $\mathcal{T}_{(\mathbf{r},\boldsymbol{\gamma})}$ by $\mathbf{r} = 2\,1$ and $\boldsymbol{\gamma} = 0\,2\,1$ is shown in Figure 1 and the labelled $\mathcal{I}_{(\mathbf{r},\boldsymbol{\gamma})}$ generated by $\mathbf{r} = 3\,1\,1$ and $\boldsymbol{\gamma} = 0\,3\,6\,4\,2$ in Figure 4.

The next statement completes Lemma 11 and follows from previous Equation (7).

Lemma 14. *Let* $\mathcal{I}_{\mathbf{r},\boldsymbol{\gamma}}$ *be the i-tree generated by the rhythm* $\mathbf{r}$ *of directing parameter* (q, p) *and a labelling* $\boldsymbol{\gamma}$ *consistent with* $\mathbf{r}$. *We denote by* A *the alphabet of all the letters appearing in* $\boldsymbol{\gamma}$. *Then, the following equation holds:*

$$\forall n, m \in \mathbb{N}\,,\ \forall a \in A \qquad n \xrightarrow[\mathcal{I}_{\mathbf{r},\boldsymbol{\gamma}}]{a} m \quad \Longleftrightarrow \quad (n+q) \xrightarrow[\mathcal{I}_{\mathbf{r},\boldsymbol{\gamma}}]{a} (m+p) \ \ .$$

As previously mentioned, we denote by $L_{(\mathbf{r},\boldsymbol{\gamma})}$ the branch language of the *tree* $\mathcal{T}_{(\mathbf{r},\boldsymbol{\gamma})}$ (rather than the one of *i-tree* $\mathcal{I}_{(\mathbf{r},\boldsymbol{\gamma})}$), and we call it *the language generated by* $(\mathbf{r}, \boldsymbol{\gamma})$. The branch language of $\mathcal{I}_{(\mathbf{r},\boldsymbol{\gamma})}$ is thus $\widehat{L}_{(\mathbf{r},\boldsymbol{\gamma})} = z^* L_{(\mathbf{r},\boldsymbol{\gamma})}$ where $z = \gamma_0$ is the label of the loop $0 \longrightarrow 0$ in $\mathcal{I}_{(\mathbf{r},\boldsymbol{\gamma})}$ and we call it the *padded* language generated by $(\mathbf{r}, \boldsymbol{\gamma})$.

4. From Rational Base Numeration Systems to Rhythms

4.1. Integer and Rational Base Numeration Systems

Let p be an integer, $p \geqslant 2$, and $A_p = [\![0, p-1]\!]$ the digit-alphabet of the first p non-negative integers. Every word $w = a_n\, a_{n-1} \cdots a_0$ of A_p^* is given a value n in $\mathbb{N}$ by the *evaluation function* π_{p}:

$$\pi_{\mathrm{p}}(a_n\, a_{n-1} \cdots a_0) \;=\; \sum_{i=0}^{n} a_i\, p^i \;\;,$$

and w is a p-development of n. Every non-negative integer n has a unique *p-development* in A_p^* without leading 0's : it is called the *p-representation* of n and is denoted by $\langle n \rangle_p$. The *p-representation* of n can be computed from left-to-right by a greedy algorithm, but also from *right-to-left* by iterating the Euclidean division of n by p, the digits a_i being the successive remainders. The language of the p-representations of the integers is the regular language

$$L_p \;=\; \{\langle n \rangle_p \mid n \in \mathbb{N}\} \;=\; (A_p \setminus 0)\; A_p^* \;\;.$$

Let p and q be two co-prime integers such that $p > q > 1$. Given a positive integer N, let us write $N_0 = N$ and, for every non negative integer i, let us define the integer N_i by the equation

$$q\, N_i \;=\; p\, N_{i+1} + a_i \;\;, \tag{8}$$

where a_i is the remainder of the Euclidean division of $q\, N_i$ by p, hence belongs to $A_p = [\![0, p-1]\!]$. Since $p > q$, the sequence $(N_i)_{i\in\mathbb{N}}$ is strictly decreasing and eventually stops at $N_{k+1} = 0$. Moreover, it holds that

$$N \;=\; \sum_{i=0}^{k} \frac{a_i}{q} \left(\frac{p}{q}\right)^i \;\;.$$

The *evaluation function* $\pi_{\frac{p}{q}}$ is derived from this formula. Given a word $u = a_n\, a_{n-1} \cdots a_0$ over the alphabet A_p, and indeed over any alphabet of digits, its *value* (in base $\frac{p}{q}$) is defined by

$$\pi_{\frac{p}{q}}(u) \;=\; \pi_{\frac{p}{q}}(a_n a_{n-1} \cdots a_0) \;=\; \sum_{i=0}^{n} \frac{a_i}{q} \left(\frac{p}{q}\right)^i \;\;. \tag{9}$$

Conversely, a word u in A_p^* is called a *$\frac{p}{q}$-development* of a real number x if $\pi_{\frac{p}{q}}(u) = x$. Since the development is unique up to leading 0's (see [1, Theorem 1]) the $\frac{p}{q}$-development of x which does not starts with a 0 is called *the $\frac{p}{q}$-representation of* x, denoted by $\langle x \rangle_{\frac{p}{q}}$, and can be computed with the modified Euclidean division algorithm above. By convention, the representation of 0 is the empty word ε. The set of $\frac{p}{q}$-representations of integers is denoted by $L_{\frac{p}{q}}$:

$$L_{\frac{p}{q}} \;=\; \{\langle n \rangle_{\frac{p}{q}} \mid n \in \mathbb{N}\} \;\;.$$

It is immediate that $L_{\frac{p}{q}}$ is prefix-closed (since, in the modified Euclidean division algorithm $\langle N\rangle_{\frac{p}{q}} = \langle N_1\rangle_{\frac{p}{q}}\, a_0$) and right-extendable (since, for every representation $\langle n\rangle_{\frac{p}{q}}$, there exists (at least) an a in A_p such that q divides $(n\,p+a)$ and then $\langle\frac{n\,p+a}{q}\rangle_{\frac{p}{q}} = \langle n\rangle_{\frac{p}{q}}\, a$). The language $L_{\frac{p}{q}}$ is then considered as a labelled tree whenever it is convenient, in particular when it is drawn; for instance the language $L_{\frac{3}{2}}$ is shown in Figure 1 (introduction, p. 2) and the padded language $0^*L_{\frac{5}{3}} = \widehat{L}_{\frac{5}{3}}$ in Figure 6 later on.

It is shown in [1] that the numeration system with rational base $\frac{p}{q}$ coincides with the *abstract numeration system (ANS)* $L_{\frac{p}{q}}$: the representation of a given integer n is the $(n+1)$-th word of $L_{\frac{p}{q}}$ in the radix order.

It is also shown in the same paper [1] that $L_{\frac{p}{q}}$ is not a rational language and not even context-free language. Indeed, $L_{\frac{p}{q}}$ meets a strong 'non-iteration' condition that we call *Finite Left Iteration Property.*

Definition 15 ([12, 11])**.** *A language L of A^* has the* Finite Left Iteration Property*, or is a* FLIP language *for short, if for all u, v in A^*, $|v| \geqslant 1$, $u\,v^i$ is prefix of a word of L for only* ***finitely*** *many exponents i in $\mathbb{N}$.*[4]

A prefix-closed language is FLIP if and only if it contains no infinite regular subsets, a property that have been called IRS by Greibach [8]. More precisely, the following holds.

Proposition 16 ([11])**.** *For a language L, the following statements are equivalent:*

(i) *L is a FLIP language.*
(ii) *The prefix closure of L is an IRS language.*
(iii) *The topological closure of L does not contain any ultimately periodic words.*

The finite left iteration property is quite a robust property as stated in the next lemma.

Lemma 17 ([11])**.** *The class of FLIP languages is stable by finite union, sublanguage, concatenation and inverse morphic image.*

The very definition of the finite left iteration property immediately implies that a FLIP language, hence indeed $L_{\frac{p}{q}}$, does not meet any kind of iteration lemma and in particular is not context-free.

In many respects, the case of integer base can be seen as a special case of rational base numeration system. Indeed, when $q = 1$, the definitions of $\pi_{\frac{p}{q}}$, $\langle n\rangle_{\frac{p}{q}}$ and $L_{\frac{p}{q}}$ coincide with those of π_{p}, $\langle n\rangle_p$ and L_p respectively. In the sequel, we consider the base $\frac{p}{q}$ where p and q are two coprime integers satisfying $p > q \geqslant 1$,

[4]This property was originally introduced in [12] under the improper name of *Bounded Left Iteration Property*, or *BLIP* for short.

that is, indifferently one numeration system or the other. In particular, the following holds in both integer and rational cases:

$$\forall n \in \mathbb{N}\,,\ \forall m \in \mathbb{N}^+\,,\ \forall a \in A_p \qquad \langle m \rangle_{\frac{p}{q}} \;=\; \langle n \rangle_{\frac{p}{q}}\, a \quad\Longleftrightarrow\quad a \;=\; q\,m - p\,n\ . \tag{10}$$

Note that the previous equivalence would be false if m were equal to 0. Indeed, in the case where $m = n = a = 0$, the left-hand side equality is satisfied while the other equality is not. This is not due to a different behaviour in the case where $m = 0$, but simply to the fact that padding 0's are implicitly excluded in the representation of integers. We are then in a situation where considering the associated padded language (or labelled i-tree) yields a more elegant statement:

$$\forall n, m \in \mathbb{N}\,,\ \forall a \in A_p \qquad n \xrightarrow[\widehat{L}_{\frac{p}{q}}]{a} m \quad\Longleftrightarrow\quad a \;=\; q\,m - p\,n\ . \tag{11}$$

4.2. Rhythm and Labelling of $L_{\frac{p}{q}}$

With every directing parameter (q, p), we associate a particular *rhythm* $\mathbf{r}_{\frac{p}{q}}$ (of directing parameter (q, p), of course) and a canonical *labelling* $\boldsymbol{\gamma}_{\frac{p}{q}}$. The latter results from the generation of $\mathbb{Z}/p\mathbb{Z}$ by q while the former relates to the classical notion of *Christoffel words* if we use the correspondence between rhythms and paths in the $(\mathbb{Z} \times \mathbb{Z})$-lattice that was described in Section 3.2. The remarkable fact is then that the representation language in the $\frac{p}{q}$-numeration system is generated by $(\mathbf{r}_{\frac{p}{q}}, \boldsymbol{\gamma}_{\frac{p}{q}})$.

Christoffel words code the 'best (upper) approximation' of segments in the $\mathbb{Z}\times\mathbb{Z}$-lattice and have been studied in the field of combinatorics of words (*cf.* [2]).

Definition 18 ([2])**.** *The (upper)* Christoffel word *of slope* $\frac{p}{q}$*, denoted by* $\mathbf{w}_{\frac{p}{q}}$*, is the label of the path from* $(0,0)$ *to* (q,p) *on the* $(\mathbb{Z} \times \mathbb{Z})$*-lattice, such that*

- *the path is above the line of slope* $\frac{p}{q}$ *passing through the origin;*
- *the region enclosed by the path and the line contains no point of* $\mathbb{Z} \times \mathbb{Z}$*.*

We translate then Christoffel words into rhythms.

Definition 19. *The* Christoffel rhythm *associated with* $\frac{p}{q}$*, and denoted by* $\mathbf{r}_{\frac{p}{q}}$*, is the rhythm whose path is* $\mathbf{w}_{\frac{p}{q}}$*:* $\mathtt{path}(\mathbf{r}_{\frac{p}{q}}) = \mathbf{w}_{\frac{p}{q}}$ *; its directing parameter is* (q, p)*.*

Figure 5b (p. 12) shows the path of $\mathbf{w}_{\frac{5}{3}} = \overline{y\,y}\,x\,\overline{y\,y}\,x\,\overline{y}\,x$, the Christoffel word associated with $\frac{5}{3}$; the Christoffel rhythm associated with $\frac{5}{3}$ is then $\mathbf{r}_{\frac{5}{3}} = 2\,2\,1$. Other instances of Christoffel rhythms are $\mathbf{r}_{\frac{3}{2}} = 2\,1$, $\mathbf{r}_{\frac{4}{3}} = 2\,1\,1$ and $\mathbf{r}_{\frac{12}{5}} = 3\,2\,3\,2\,2$. The definition of Christoffel words yields the following proposition on rhythms.

Proposition 20. *Let* $\mathbf{r}_{\frac{p}{q}} = r_0\, r_1 \cdots r_{q-1}$ *be the Christoffel rhythm associated with a base* $\frac{p}{q}$*. Then, for every integer* k*,* $0 < k \leqslant q$*, the partial sum of the first* k *components of* $\mathbf{r}_{\frac{p}{q}}$ *is equal to the smallest integer greater than* $k\frac{p}{q}$*:* $R_k = \left\lceil k\frac{p}{q} \right\rceil$*.*

The next lemma rewords Proposition 20 in terms of Christoffel words and follows from their geometrical properties.

Lemma 21. *Let* $\mathbf{w}_{\frac{p}{q}}$ *be the Christoffel word of slope* $\frac{p}{q}$*. If* $u\,x$ *is prefix of* $\mathbf{w}_{\frac{p}{q}}$ *then it corresponds to a path from* $(0,0)$ *to* $(k, \lceil k\,\frac{p}{q} \rceil)$ *in the* $\mathbb{Z}\times\mathbb{Z}$ *lattice, where* k *is the number of* x*'s in* $u\,x$*.*

Proof. From Definition 18 of Christoffel word, there is no integer point between the path and the line of slope $\frac{p}{q}$ and passing through the origin. Since the point $(k, k\,\frac{p}{q})$ is part of this line, the Christoffel path must pass through the point $(k, \lceil k\,\frac{p}{q} \rceil)$. Besides, the prefix of the Christoffel word reaching this point must end with an x; indeed, were it ending with an y, the Christoffel path would pass through the point $(k, \lceil k\,\frac{p}{q} \rceil - 1)$ which is *below* the line of slope $\frac{p}{q}$, a contradiction. □

Since p and q are coprime integers, q is a generator of the additive group $\mathbb{Z}/p\mathbb{Z}$. We denote by $\boldsymbol{\gamma}_{\frac{p}{q}}$ the sequence induced by this generation process:

$$\boldsymbol{\gamma}_{\frac{p}{q}} \;=\; (\,0,\, (q\,\%\,p),\, (2\,q\,\%\,p), \ldots,\, ((p-1)\,q\,\%\,p)\,)\ .$$

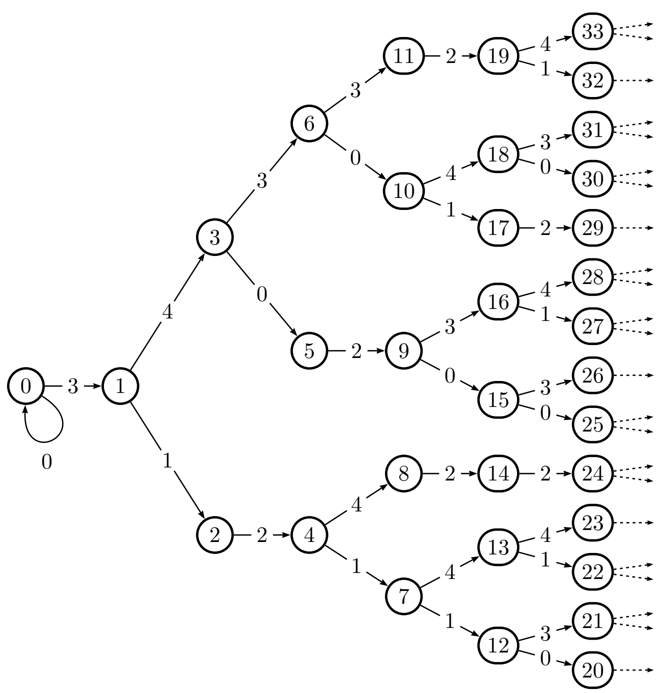


Figure 6: The padded language $0^*L_{\frac{5}{3}}$ of the representation of integers in base $\frac{5}{3}$

Theorem 22. *Let p and q be two coprime integers, $p > q \geqslant 1$. The language $L_{\frac{p}{q}}$ of the $\frac{p}{q}$-representations of the integers is generated by the rhythm $\mathbf{r}_{\frac{p}{q}}$ and the labelling $\boldsymbol{\gamma}_{\frac{p}{q}}$.*

For instance, $L_{\frac{3}{2}}$, shown in Figure 1, is generated by the rhythm $\mathbf{r}_{\frac{3}{2}} = 2\,1$ and the labelling $\boldsymbol{\gamma}_{\frac{3}{2}} = 0\,2\,1$ while the padded language $0^*L_{\frac{5}{3}}$, shown in Figure 6, is generated by the rhythm $\mathbf{r}_{\frac{5}{3}} = 2\,2\,1$ and the labelling $\boldsymbol{\gamma}_{\frac{5}{3}} = 0\,3\,1\,4\,2$.

The next proposition follows directly from Equation (10) and implies the part of Theorem 22 stated by Corollary 24.

Proposition 23. *For every positive integer m, the rightmost digit of $\langle m \rangle_{\frac{p}{q}}$ is equal to $(q\,m)\,\%\,p$.*

Corollary 24. *The labelling of $L_{\frac{p}{q}}$ is $\boldsymbol{\gamma}_{\frac{p}{q}}$.*

The proof of the other part of Theorem 22 requires additional definitions and statements.

Let us consider, for every integer $k < q$, the difference between the approximation $R_k = (r_0 + r_1 + \cdots + r_{k-1})$ and the point of the associated line of the respective abscissa, that is $(k\,\frac{p}{q})$. This difference is a rational number smaller than 1 and whose denominator divides q; we denote by e_k the integer resulting of the multiplication of this difference by q:

$$\forall k \in [\![0, q-1]\!] \qquad e_k \quad = \quad q\,R_k - k\,p \ . \tag{12}$$

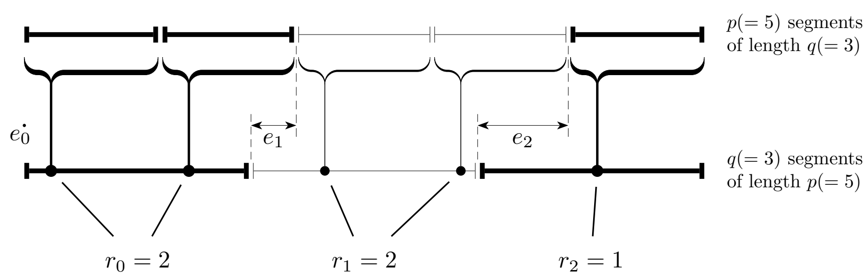


Figure 7: Diagrammatic interpretation of the rhythm $2\,2\,1$ of base $\frac{5}{3}$

Figure 7 describes, on the example of the base $\frac{5}{3}$ (also used in Figure 5b and Figure 6), a more diagrammatic way of characterising Christoffel rhythms. We consider a length of $p \times q$ units divided in two sets of segments: on the top it is divided in p segments of length q and on the bottom by q segments of length p. Every top segment is then associated with the bottom segment in which its leftmost unit lies: for instance the first two top segments, $[0, 3)$ and $[3, 6)$, are associated with the first bottom segment $[0, 5)$ since their respective first unit sub-segments, $[0, 1)$ and $[3, 4)$, lie in it. The r_k's and e_k's are then interpreted as follows.

- The integer r_k is the number of top segments associated with the $(k+1)$-th left-most bottom segment.
- The integer e_k is the difference of length between the k left-most bottom segments (of length $p \times k$) and the total number of top segments associated with them (of length $q \times R_k$).

Figure 8 then shows why this is an equivalent definition of the Christoffel rhythm. In this figure, the unit segments x and y are divided into q subsegments and the e_k's correspond to a number of subsegments.

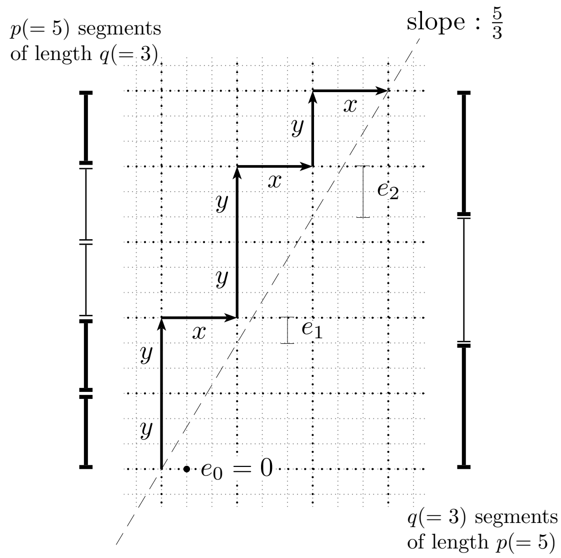


Figure 8: Link with Christoffel rhythm

Basic properties of the r_j's and e_j's that follow directly from Proposition 20 and Equation (12) are compiled in the following statement.

Property 25. *Let* $\mathbf{r}_{\frac{p}{q}} = (r_0, r_1, \ldots, r_{q-1})$ *be the Christoffel rhythm of slope* $\frac{p}{q}$. *For every integer* k *in* $[\![0, q-1]\!]$, *it holds:*

(i) e_k *belongs to* $[\![0, q-1]\!]$;
(ii) r_k *is the smallest integer such that* $q\,r_k + e_k \geqslant p$;
(iii) $e_{k+1} = e_k + q\,r_k - p$.

Lemma 26. *For every integer* n, *the smallest digit labelling an outgoing arc of* n *in the i-tree* $\widehat{L}_{\frac{p}{q}}$ *is* $e_{(n\,\%\,q)}$.

Proof. Let n be an integer and k its congruence class modulo q. It follows from (11) that the digits labelling arcs going out from n are congruent modulo q. Since e_k is in $[\![0, q-1]\!]$ (Property 25 (i)), hence strictly smaller than q, it is enough to show that e_k is an outgoing label of n. From (11), it is the case if and only if $(n\,p + e_k)$ is a multiple of q or, equivalently if $(k\,p + e_k)$ is a multiple of q, which follows from the definition of e_k (Equation (12)). $\square$

Proposition 27. *For every integer n, there are exacly $r_{(n\,\%q)}$ arcs going out from n in the i-tree $\widehat{L}_{\frac{p}{q}}$.*

Proof. Let n be an integer and k its congruence class modulo q. From Property 25 (ii), r_k is the smallest integer such that $q\,r_k + e_k > p$. In other words, the two following equations hold:

$$\forall j \in [\![0, r_k - 1]\!] \qquad (e_k + q\,j) < p \;\; ; \qquad e_k + q\,r_k > p \;\; .$$

The set $S = \{e_k, (e_k + q), \ldots, (e_k + q\,(r_k - 1))\}$ consists of all digits in A_p that are congruent to e_k modulo q. Since e_k labels an arc going out from n (Lemma 26), it follows from Equation (11) that

$$S \;=\; \{a \in A_p \mid a \text{ labels an arc going out from } n\} \;\; .$$

The set S is of cardinal r_k, concluding the proof. □

Proposition 27 states that the rhythm of $L_{\frac{p}{q}}$ is indeed $\mathbf{r}_{\frac{p}{q}}$ and Corollary 24 that its labelling is $\boldsymbol{\gamma}_{\frac{p}{q}}$, hence establishing Theorem 22.

The next statement gives a different way to compute $\boldsymbol{\gamma}_{\frac{p}{q}}$; its generalisation in the next section (Definition 29) to arbitrary rhythms will be instrumental in the proof of Theorem 32.

Proposition 28. *Let $\mathbf{r}_{\frac{p}{q}}$ and $\boldsymbol{\gamma}_{\frac{p}{q}} = \gamma_0\,\gamma_1 \cdots \gamma_{p-1}$ be the Christoffel rhythm and labelling of directing parameter (q, p). We denote by $\boldsymbol{\gamma}_{\frac{p}{q}} = u_0\,u_1 \cdots u_{q-1}$ the factorisation of $\boldsymbol{\gamma}_{\frac{p}{q}}$ induced by $\mathbf{r}_{\frac{p}{q}}$. For every index $i < p$, it holds $\gamma_i = i\,q - j\,p$, where j is the index of the factor u_j containing γ_i.*

Proof. Let us denote by $c_0\,c_1 \cdots c_{p-1}$ the sequence of integers computed by the formula:

$$c_i \;=\; q\,i - p\,j \qquad \text{if } \gamma_i \text{ is a letter of the factor } u_j \;\; .$$

This definition implies that $c_i \equiv i\,q\;[p]$, hence $c_i \equiv \gamma_i\;[p]$ for every i, $0 \leqslant i < p$. The proof will be complete when we have shown that $0 \leqslant c_i < p$ for every integer $i < p$.

Let us take $i, j > 0$ such that $\gamma_0\,\gamma_1 \cdots \gamma_{i-1} = u_0\,u_1 \cdots u_{j-1}$, a word of length $i = R_j$. It follows from Proposition 20 that $i = \lceil j\frac{p}{q} \rceil$, or, in other word, that

$$j\,p - q \;\leqslant\; (i-1)\,q \;<\; j\,p \;\; . \tag{13}$$

Since from the choice of i and j, γ_{i-1} is the last letter of u_{j-1}, it holds

$$c_{i-1} \;=\; (i-1)\,q - (j-1)\,p \quad \text{hence, from (13),} \quad c_{i-1} \;<\; j\,p - (j-1)\,p \;=\; p \;\; .$$

Since $|u_j| = r_j > 1$, γ_i is the first letter of u_j which implies that

$$c_i \;=\; (i\,q - j\,p) \quad \text{hence, from (13),} \quad c_i \;\geqslant\; (j\,p - q) + q - j\,p \;=\; 0 \;\; .$$

We have thus shown that:

- the first letter of every factor u_j is non-negative (note that we have previously shown the case where $j > 0$ only, the remaining case is trivial: the first letter of the first factor is always equal to 0);
- the last letter of every factor u_j is strictly smaller than p.

Since every factor is an increasing word (each letter being equal to the previous letter plus q), every letter a of every factor satisfies $0 \leqslant a < p$.

As previously mentioned, for each integer $i < p$, since $c_i \equiv \gamma_i\ [p]$, the previous paragraph implies that $c_i = \gamma_i$. □

5. From Rhythms Back to Rational Bases

We now establish a kind of converse of Theorem 22. With an arbitrary rhythm is associated a rational base (its growth ratio) and a *special* labelling. We consider the language generated by this labelled rhythm as an abstract numeration system and show that it is simply a rational base numeration system on a non-canonical alphabet (Theorem 32)

In this section, p and q are two integers, $p > q \geqslant 1$, *not necessarily coprime anymore*, and $\mathbf{r}$ is a rhythm of directing parameter (q, p). As in Definition 10, we denote by p' and q' the respective quotients of p and q by their gcd, denoted by d.

5.1. Special Labelling

The next definition is a generalisation of the labelling of a Christoffel rhythm for arbitrary rhythms; it is based on the characterisation given by Proposition 28 but is more complicated in order to take into account the possible components equal to 0 appearing in the rhythm.

Definition 29. *The* special labelling $\boldsymbol{\gamma}_{\mathbf{r}} = \gamma_0\,\gamma_1 \cdots \gamma_{p-1}$ *associated with* $\mathbf{r}$*, is defined as follows. Let* $\boldsymbol{\gamma}_{\mathbf{r}} = u_0\,u_1 \cdots u_{q-1}$ *be the factorisation of* $\boldsymbol{\gamma}_{\mathbf{r}}$ *induced by* $\mathbf{r}$ *(for all i, $0 \leqslant i < p$, $|u_i| = r_i$, cf. Definition 13 (i)).*

For every integer i, $0 \leqslant i < p$, let j be the index such that the digit γ_i belongs to u_j; then γ_i is defined by $\gamma_i \;=\; i\,q' - j\,p'$.

The previous Definition 29 may be rewritten by means of a recurrence formula as follows. First $\gamma_0 = 0$. Then, for every i, $0 \leqslant i < p-1$, if k and j are the indices such that γ_i belongs to u_j and γ_{i+1} belongs to u_{j+k}, then $\gamma_{i+1} = \gamma_i + q' - k\,p'$. This alternate definition of the special labelling is easier to showcase, it is therefore used in examples.

Example 30. *Let* $\mathbf{r} = 3\,1\,1$*; its directing parameter is* $(3, 5)$*, hence* $p = p' = 5$*,* $q = q' = 3$ *and the computation of* $\boldsymbol{\gamma}_{\mathbf{r}}$ *is given below. Within a factor* u_i*, the difference between two consecutive digits is* 3 $(= q')$*, otherwise it is* -2 $(= q' - p')$*.*

$$\begin{array}{rccc} \mathbf{r} = & 3 & 1 & 1 \\ \boldsymbol{\gamma}_{\mathbf{r}} = & \overbrace{0\,3\,6}^{u_0} & \overbrace{4}^{u_1} & \overbrace{2}^{u_2} \end{array}$$

Example 31. *Let now* $\mathbf{r} = 4\,0\,0\,2$*; its directing parameter is* $(4,6)$*,* $p' = 3$*,* $q' = 2$ *and the computation of* $\boldsymbol{\gamma}_{\mathbf{r}}$ *is given below. Within a factor* u_i*, the difference between two consecutive digits is* $+2$ $(= +q')$*; the fourth digit belongs to* u_0 *and the fifth to* u_3*: the difference between the two is* -7 $(= +q' - 3\,p')$*.*

$$\begin{array}{rcccc} \mathbf{r} = & 4 & 0 & 0 & 2 \\ \boldsymbol{\gamma}_{\mathbf{r}} = & \overbrace{0\,2\,4\,6}^{u_0} & \overbrace{\varepsilon}^{u_1} & \overbrace{\varepsilon}^{u_2} & \overbrace{-1\;\;1}^{u_3}\;) \end{array}$$

Definition 29 implies that two consecutive digits in the same factor are in increasing order ($\gamma_{i+1} = (\gamma_i + q)$ since j is unchanged), hence that $\boldsymbol{\gamma}_{\mathbf{r}}$ is consistent with $\mathbf{r}$.

Notation. *We denote by* $L_{\mathbf{r}}$ *the language generated by a rhythm* $\mathbf{r}$ *and the associated special labelling* $\boldsymbol{\gamma}_{\mathbf{r}}$*, that is,* $L_{\mathbf{r}} = L_{(\mathbf{r},\boldsymbol{\gamma}_{\mathbf{r}})}$*.*

5.2. Non-Canonical Representation of Integers

If $\mathbf{r}$ happens to be a Christoffel rhythm, then, by Theorem 22, $L_{\mathbf{r}}$ is equal to $L_{\frac{p'}{q'}}$ (which, in this case, is also $L_{\frac{p}{q}}$). The key result of this work is that $L_{\mathbf{r}}$ and $L_{\frac{p'}{q'}}$ are indeed of the same kind.

Theorem 32. *Let* $\mathbf{r}$ *be a rhythm of directing parameter* (q,p) *and* $\frac{p'}{q'}$ *the reduced fraction of* $\frac{p}{q}$*. Then, the language* $L_{\mathbf{r}}$ *is a set of representations of the integers in the rational base* $\frac{p'}{q'}$ *using a non-canonical set of digits.*

Let us now call $\mathbf{r}$-representation of an integer n, and denote it by $\langle n \rangle_{\mathbf{r}}$, the representation of n in the abstract numeration system $L_{\mathbf{r}}$. We know from Equation (4) that $\langle n \rangle_{\mathbf{r}}$ labels the path from the root 0 to the node n in the labelled tree defined by $L_{\mathbf{r}}$. First we show that the existence of arcs in $L_{\mathbf{r}}$ has a necessary condition similar to those of $L_{\frac{p'}{q'}}$ (*cf.* Equation (10)).

Lemma 33. *For every integers* n *and* m *and every letter* a*, it holds:*

$$\langle n \rangle_{\mathbf{r}}\, a \;=\; \langle m \rangle_{\mathbf{r}} \quad\Longrightarrow\quad a \;=\; q'\,m - p'\,n \;\;.$$

Proof. Let n and m be two integers and a be a letter such that $\langle n \rangle_{\mathbf{r}}\, a = \langle m \rangle_{\mathbf{r}}$. From Equation (5) follows that

$$n \xrightarrow[L_{\mathbf{r}}]{a} m \;\;.$$

We then apply the converse direction of Lemma 14 iteratively as many times as possible and write k the number of times it was used. It yields two integers n', m' and an integer k such that

$$n' \xrightarrow[L_{\mathbf{r}}]{a} m' \;\;, \tag{14a}$$

$$n' \;=\; n - k\,q \;\;, \qquad\qquad m' \;=\; m - k\,p \;\;, \tag{14b}$$

$$m' < p \qquad\qquad \text{and} \qquad\qquad n' < q \;\;. \tag{14c}$$

Note that Lemma 14 directly implies only one of the two conditions of Equation (14c), but since Equation (14a) holds, they are equivalent.

It then follows from Equation (7) that

$$R_{n'} \leqslant m' < R_{n'+1} \qquad \text{and} \qquad a = \gamma_{m' \% q} \ ,$$

hence from Definition 29 that

$$a = q'\, m' - p'\, n' \ .$$

Combining the previous equation with (14b) yields (d is the gcd of p and q):

$$\begin{aligned} a &= q'\,(m - k\,p) - p'\,(n - k\,q) &&= q'\,m - k\,(q'p - p'q) - p'\,n \\ &= q'\,m - k\,(q'p'd - p'q'd) - p'\,n &&= q'\,m - p'\,n \end{aligned}$$

which concludes the proof. □

The converse of Lemma 33 does not hold in general; it holds only for rhythms (of directing parameter (q, p)) such that p and q are coprime, and for powers of such rhythms. Otherwise, the alphabet of the letters appearing in $\boldsymbol{\gamma}_{\mathbf{r}}$ contains at least two different digits congruent modulo p'; the incoming arc of a given node then depends on its congruence class modulo p (and not only modulo p').

The following proposition rewords Theorem 32 in a more precise way.

Proposition 34. *Let* $\mathbf{r}$ *be a rhythm of directing parameter* (q, p), $\frac{p'}{q'}$ *the reduced fraction of* $\frac{p}{q}$ *and* $\pi_{\frac{p'}{q'}}$ *the evaluation function in the* $\frac{p'}{q'}$*-numeration system. Then,* $\pi_{\frac{p'}{q'}}$ *is the evaluation function of the ANS* $L_{\mathbf{r}}$*, that is:*

$$\forall n \in \mathbb{N} \qquad \pi_{\frac{p'}{q'}}(\langle n \rangle_{\mathbf{r}}) = n \ .$$

Proof. By induction on the length of $\langle n \rangle_{\mathbf{r}}$. The equality is obviously verified for $\langle 0 \rangle_{\mathbf{r}} = \varepsilon$. Let m be a positive integer and $\langle m \rangle_{\mathbf{r}} = a_{k+1}\, a_k\, a_{k-1} \cdots a_1\, a_0$ its $\mathbf{r}$-representation, that is, a word of $L_{\mathbf{r}}$. The word $a_{k+1}\, a_k\, a_{k-1} \cdots a_1$ is also in $L_{\mathbf{r}}$; it is the $\mathbf{r}$-representation of an integer n strictly smaller than m, satisfying $\langle n \rangle_{\mathbf{r}}\, a_0 = \langle m \rangle_{\mathbf{r}}$, hence $n \xrightarrow[L_{\mathbf{r}}]{a_0} m$. On the right hand, by induction hypothesis, $n = \pi_{\frac{p'}{q'}}(\langle n \rangle_{\mathbf{r}})$ and on the other hand, it follows from the previous Lemma 33 that $a_0 = q'm - p'n$, or, equivalently, that $m = \frac{np' + a_0}{q'}$, hence

$$m = \frac{p'}{q'}\, \pi_{\frac{p'}{q'}}(\langle n \rangle_{\mathbf{r}}) + \frac{a_0}{q} = \pi_{\frac{p}{q}}(\langle n \rangle_{\mathbf{r}}\, a_0) = \pi_{\frac{p'}{q'}}(\langle m \rangle_{\mathbf{r}}) \ .$$

□

It is shown in [1] that in spite of the 'complexity' of $L_{\frac{p}{q}}$, the *conversion* from any digit-alphabet B into the canonical alphabet A_p is realised by a *finite transducer* exactly as in the case of an integer numeration system (*cf.* also [7]). More precisely:

Theorem 35 ([1])**.** *For every digit alphabet B, the function $\chi\colon B^* \to A_{p'}^*$ which maps every word w of B^* onto the word of $A_{p'}^*$ which has the same value in the $\frac{p'}{q'}$-numeration system — that is, $\pi_{\frac{p'}{q'}}(w) = \pi_{\frac{p'}{q'}}(\chi(w))$ — is a (right sequential) rational function.*

If we write B for the set of digits appearing in $\boldsymbol{\gamma}_{\mathbf{r}}$, Theorem 35 implies in particular that $\chi(L_{\mathbf{r}}) = L_{\frac{p'}{q'}}$. Since χ is a rational function, $L_{\mathbf{r}}$ cannot be 'simpler' than $L_{\frac{p'}{q'}}$. More precisely, the following holds.

Corollary 36. *Let $\mathbf{r}$ be a rhythm of directing parameter (q,p). If $\frac{p}{q}$ is not an integer, then $L_{\mathbf{r}}$ is a FLIP language.*

Proof. We denote by B the alphabet of $L_{\mathbf{r}}$, that is, the set of the digits appearing in $\boldsymbol{\gamma}_{\mathbf{r}}$. Since B is finite, it follows from Theorem 35 that there exists a letter-to-letter, sequential and right transducer $\mathcal{T}$ such that

$$\forall w \in L_{\mathbf{r}} \qquad \pi_{\frac{p'}{q'}}(w) \;=\; \pi_{\frac{p'}{q'}}(\mathcal{T}(w)) \quad . \tag{15}$$

Moreover, since the $\frac{p'}{q'}$-developments of a given integer are unique up to leading 0's, (see [1, Theorem 1]), it follows from Proposition 34, that

$$\forall w \in L_{\mathbf{r}} \qquad \mathcal{T}(w) \in 0^* L_{\frac{p'}{q'}} \quad . \tag{16}$$

For sake of contradiction, let us assume that $L_{\mathbf{r}}$ is not a FLIP language. Since it is a prefix-closed and infinite language, it then contains an infinite rational language that we denote by K. In addition, the words of K have pairwise distinct values. It follows from (16) that the rational language $\mathcal{T}(K)$ is contained in $0^* L_{\frac{p'}{q'}}$ and from (15) that the words of $\mathcal{T}(K)$ have pairwise distinct values. Then, $(0^*)^{-1}\mathcal{T}(K) \cap (A_p^* \setminus 0\,A_p^*)$ is an infinite rational language contained in $L_{\frac{p'}{q'}}$, a contradiction with the property that $L_{\frac{p'}{q'}}$ is a FLIP language. $\square$

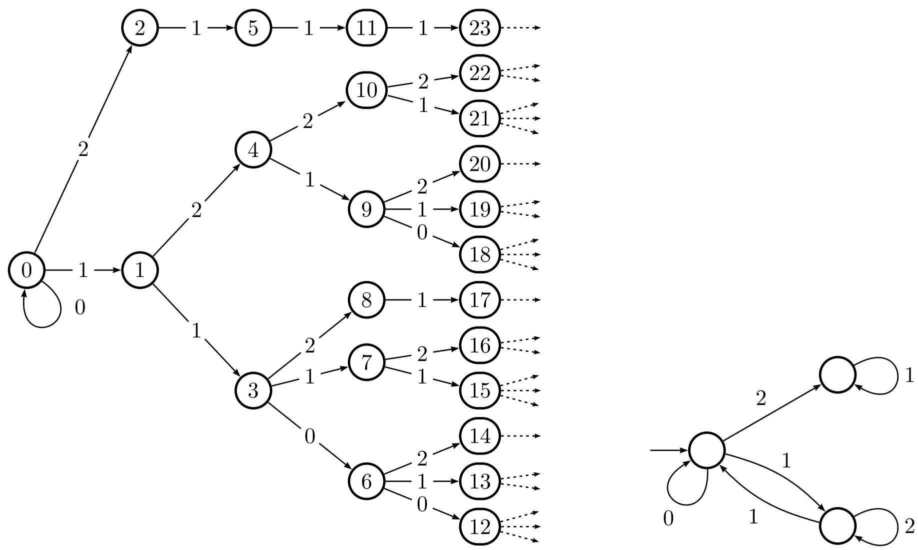
Conversely, an easy proof that may be found in [13, 11] establishes the next statement.

Theorem 37 ([13])**.** *Let $\mathbf{r}$ be a rhythm of directing parameter (q,p). If $\frac{p}{q}$ is an integer, then $L_{\mathbf{r}}$ is a regular language.*

For instance, Figure 9 shows $\widehat{L}_{321}$ and the automaton accepting it.

5.3. The FK Variant and its Associated Rhythm

Let p and q be two coprime integers such that $p > q \geqslant 1$. Let $\mathbf{r}$ be the *extreme* rhythm of directing parameter (q,p) where all components are 0 but one which is p. For $\mathbf{r}$ to be valid, the positive digit needs to be the first one: $\mathbf{r} = p\,0\cdots 0$ and the associated special labelling is then $\boldsymbol{\gamma}_{\mathbf{r}} = 0\,q\,(2\,q)\,\cdots\,((p-1)\,q)$. Since every letter of $\boldsymbol{\gamma}_{\mathbf{r}}$ is a multiple of q, we perform a component-wise division of $\boldsymbol{\gamma}_{\mathbf{r}}$ by q and obtain $\boldsymbol{\gamma} = 0\,1\,2\,\cdots\,(p-1)$.



(a) The padded language $\widehat{L}_{321}$ (b) Automaton accepting it

Figure 9: The padded language $\widehat{L}_{321}$ is rational

The language $L_{(\mathbf{r},\boldsymbol{\gamma})}$ generated by $(\mathbf{r},\boldsymbol{\gamma})$ is then the language of the representations of the integers in a variant (that we call FK after its authors) of $\frac{p}{q}$-numeration systems considered in [6]. In the FK variant, the value of a word u, denoted by $\theta_{\frac{p}{q}}(u)$, is q times its standard evaluation: $\theta_{\frac{p}{q}}(u) = q \times \pi_{\frac{p}{q}}(u)$. This is exactly the behaviour described by Proposition 34, since all digits have been divided by q. This example highlights the soundness of the relationship between rational base numeration system and periodic signature.

Figure 10 shows the language of the representations of integers in the variant FK of base $\frac{3}{2}$, that is the language $L_{(30,\,012)}$ generated by the rhythm $(3\,0)$ and labelling $(0\,1\,2)$.

6. The Case of Ultimately Periodic Signature

We now generalise the results Section 5 to ultimately periodic signatures. Let us first recall a few facts related to (non-necessarily periodic) signatures and previously accounted in Section 2. First, we recall that whenever a signature $\boldsymbol{s} = s_0\,s_1\,s_2\cdots$ is clear, S_i denotes the partial sum $s_0 + s_1 \cdots + s_{i-1}$ of the first i terms of $\boldsymbol{s}$. Moreover, the explicit definition of the edges of the tree I_s generated by a (valid) signature $\boldsymbol{s}$ (*cf.* Proposition 3 (i), page 6):

$$\forall n, m \in \mathbb{N} \qquad n \xrightarrow[\mathcal{I}_s]{} m \quad \Longleftrightarrow \quad S_n \leqslant m < S_{n+1} \ . \tag{17}$$

Finally, if moreover $\boldsymbol{\lambda} = \lambda_0\,\lambda_1\,\lambda_2\cdots$ denotes a labelling, then its factorisation $\boldsymbol{\lambda} = u_0\,u_1\,u_2\cdots$ determined by $\boldsymbol{s}$ (that is, satisfying $|u_i| = s_i$, for all i)

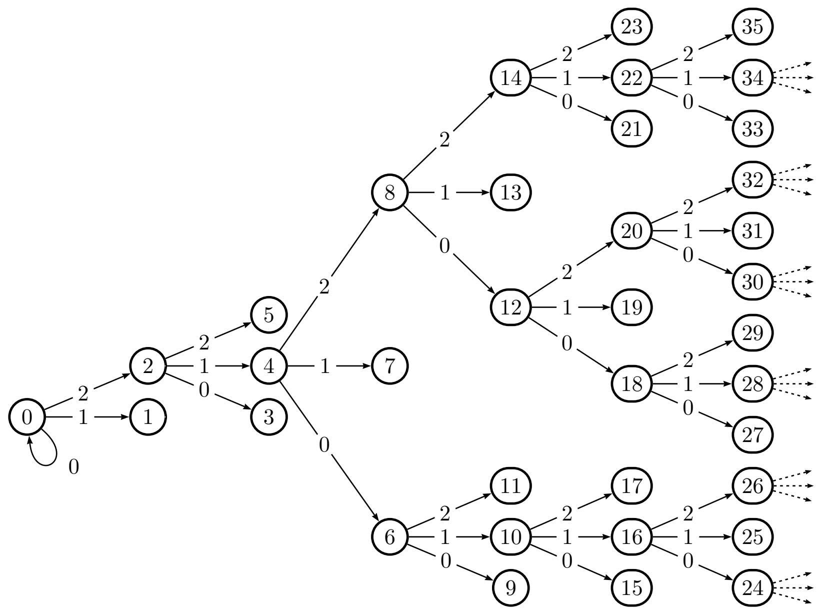


Figure 10: Padded language of the representations of the integers in the FK variant of base $\frac{3}{2}$

obviously meets :

$$\forall n, m \in \mathbb{N} \qquad S_n \leqslant m < S_{n+1} \qquad \Longleftrightarrow \qquad \lambda_m \text{ belongs to the factor } u_n \quad . \tag{18}$$

We first consider arbitrary signature and show that with the appropriate definition of special labelling (below), the extended versions of Lemma 33 and Proposition 34 hold. However, as detailed by Remark 39, this definition goes beyond the framework of language theory and the results presented here are only meaningful when applied on ultimately periodic signatures.

Definition 38. *Let $\boldsymbol{s}$ be a signature and p, q two integers such that $p > q \geqslant 1$. We write as usual $\frac{p'}{q'}$ the reduced fraction of $\frac{p}{q}$.*

The special labelling $\boldsymbol{\lambda}_{\mathbf{s},(q,p)} = \lambda_0\,\lambda_1\,\lambda_2\,\cdots$ *of directing parameter (q,p) and associated with $\boldsymbol{s}$ is defined as follows. We denote by $\boldsymbol{\lambda}_{\mathbf{s},(q,p)} = u_0\,u_1\,u_2\,\cdots$ the factorisation of $\boldsymbol{\lambda}_{\mathbf{s},(q,p)}$ induced by $\boldsymbol{s}$. For every integer i, let j be the index such that the letter λ_i belongs to u_j and then we define $\lambda_i \;=\; i\,q' - j\,p'$.*

Remark 39. *The preceding Definition 38 of special labelling yields an infinite sequence of integers that* ***is not bounded*** *in general. The generated (padded) language will in that case be over an* ***infinite alphabet****, hence beyond the framework set in Section 2. Later on, we will restrict signatures to be ultimately periodic and show that in this case the special labelling is bounded (Lemma 42).*

The next lemma follows directly from the previous Definition 38 together with Equations 17 and 18.

Lemma 40. *Let $\boldsymbol{s}$ be a signature and p, q be two integers such that $p > q \geqslant 1$. We denote by $L = L_{\boldsymbol{s},(q,p)}$ the language generated by the signature $\boldsymbol{s}$ and the labelling $\boldsymbol{\lambda}_{\mathbf{s},(q,p)}$. Then, the following implication holds:*

$$\langle m \rangle_L \;=\; \langle n \rangle_L \, a \quad \Longrightarrow \quad a \;=\; q' \, m - p' \, n \;\; .$$

The next proposition may then be proven using the Lemma 40 much like Proposition 34 has been shown using Lemma 33.

Proposition 41. *Let $\boldsymbol{s}$ be a signature and p, q two integers such that $p > q \geqslant 1$. We write $\frac{p'}{q'}$ the reduced fraction of $\frac{p}{q}$ and denote by $L = L_{\boldsymbol{s},(q,p)}$ the language generated by the signature $\boldsymbol{s}$ and the labelling $\boldsymbol{\lambda}_{\mathbf{s},(q,p)}$. Then, the following equation holds:*

$$\forall n \in \mathbb{N} \qquad \pi_{\frac{p'}{q'}}\big(\langle n \rangle_L\big) \;=\; n \;\; .$$

In the remainder of this section, we consider an ultimately periodic signature $\boldsymbol{s} = \mathbf{t}\,\mathbf{r}^\omega$, where $\mathbf{t}$ and $\mathbf{r}$ are two rhythms; the former will be called *pre-period* and its directing parameter is denoted by (g, h); the later, $\mathbf{r}$, is called *period*, its directing parameter is denoted as usual by (q, p). By convention, the special labelling associated with an ultimately periodic signature always has a directing parameter equal to the one of the period, that is (q, p); it is then denoted simply by $\boldsymbol{\lambda}_{\boldsymbol{s}}$ for short.

Lemma 42. *The special labelling associated with an ultimately periodic signature is written on a finite alphabet of digits.*

Proof. Let $\boldsymbol{s} = \mathbf{t}\,\mathbf{r}^\omega$ be an ultimately periodic signature, (g, h) the directing parameter of $\mathbf{t}$ and (q, p) the directing parameter of $\mathbf{r}$.

Let us first show that the following equation holds:

$$\forall j\,,\; j \geqslant g \qquad S_{j+q} \;=\; S_j + p \;\; . \tag{19}$$

Let j be an integer greater than g. The partial sum S_j then contains all the terms of $\mathbf{t}$, hence $S_j = T_g + R_{j-g}$; the same reasoning applied to $(j + q)$ yields that $S_{j+q} = T_g + R_{j+q-g}$. Since for all integer k, $R_{k+q} = R_k + p$, the proof of Equation (19) is complete.

We write the special labelling associated with $\boldsymbol{s}$, and its factorisation determined by $\boldsymbol{s}$ as follows:

$$\boldsymbol{\lambda}_{\boldsymbol{s}} \;=\; \lambda_0 \, \lambda_1 \, \lambda_2 \cdots \;=\; u_0 \, u_1 \, u_2 \cdots$$

The whole statement is then a consequence of the following equation that we prove thereafter:

$$\forall i\,,\; i \geqslant h \qquad \lambda_i \;=\; \lambda_{i+p} \;\; . \tag{20}$$

Let i be an integer greater than or equal to h. We denote by j the integer such that $S_j \leqslant i < S_{j+1}$. Since i is greater than $h = T_g$, j is necessarily greater than g. Applying Equation (19) hence yields that $S_{j+q} \leqslant i + p < S_{j+1+q}$.

It follows that the letters λ_i and λ_{i+p} belong respectively to the factors u_j and u_{j+q}. From Definition 38 these two letters are then defined as follows:

$$\lambda_i \;=\; i\,q' - j\,p' \qquad \text{and} \qquad \lambda_{i+p} \;=\; (i+p)\,q' - (j+q)\,p' \;\; .$$

Since $p'\,q = p\,q'$, it follows that $\lambda_i = \lambda_{i+p}$. □

The next corollary then follows from Theorem 35.

Corollary 43. *Let* $\boldsymbol{s} = \mathbf{tr}^\omega$ *be an ultimately periodic signature, let* (q, p) *be the directing parameter of* $\mathbf{r}$ *and let* $L_{\boldsymbol{s}}$ *be the language generated by the pair* $(\boldsymbol{s}, \boldsymbol{\lambda}_{\boldsymbol{s}})$*. If* $\frac{p}{q}$ *is not an integer, then* $L_{\boldsymbol{s}}$ *is a FLIP language .*

Once again, an easy proof that may be found in [13, 11] establishes the next statement.

Theorem 44. *Let* $\boldsymbol{s} = \mathbf{tr}^\omega$ *be an ultimately periodic signature, let* (q, p) *be the directing parameter of* $\mathbf{r}$ *and let* $L_{\boldsymbol{s}}$ *be the language generated by the pair* $(\boldsymbol{s}, \boldsymbol{\lambda}_{\boldsymbol{s}})$*. If* $\frac{p}{q}$ *is an integer, then* $L_{\boldsymbol{s}}$ *is a regular language.*

7. Conclusion

We have established in this paper that the infinite trees or languages generated by periodic signatures are completely determined (up to very simple transformations — that is, rational sequential functions) by the growth ratio of the period only and independent of the actual components of the period.

There is certainly still much to be understood on the relationship between the 'high regularity' of periodic signatures and the apparent disorder or complexity of trees that are generated by these periodic signatures. Some questions, such as statistics of labels along infinite branches, are indeed related to identified problems in number theory that are recognised as very difficult.

Using rhythm often sheds light on problems related to rational base. It is the case for the question of representation of the negative integers, tackled in [6], that may be given a new approach in terms of Christoffel words and their properties.

This 'characterisation' of rational base numeration system by the period was somehow unexpected. It makes the scenery simpler but the call for further investigations on the subject even stronger.

References

[1] Shigeki Akiyama, Christiane Frougny, and Jacques Sakarovitch. Powers of rationals modulo 1 and rational base number systems. *Israel J. Math.*, 168:53–91, 2008.

[2] Jean Berstel, Aaron Lauve, Christophe Reutenauer, and Franco Saliola. *Combinatorics on Words: Christoffel Words and Repetition in Words*, volume 27 of *CRM monograph series.* American Math. Soc., 2008.

[3] Valérie Berthé and Michel Rigo, editors. *Combinatorics, Automata and Number Theory*. Number 135 in Encyclopedia Math. Appl. Cambridge University Press, 2010.

[4] Reinhard Diestel. *Graph Theory*. Springer, 1997.

[5] Tom Edgar, Hailey Olafson, and James van Alstine. The combinatorics of rational base representations. to appear.

[6] Christiane Frougny and Karel Klouda. Rational base number systems for p-adic numbers. *RAIRO Theor. Informatics and Appl.*, 46(1):87–106, 2012.

[7] Christiane Frougny and Jacques Sakarovitch. *Number representation and finite automata*, chapter 2, pages 34–107. In Berthé and Rigo [3], 2010.

[8] Sheila A. Greibach. One counter languages and the IRS condition. *J. Comput. Syst. Sci.*, 10(2):237–247, 1975.

[9] Pierre Lecomte and Michel Rigo. Numeration systems on a regular language. *Theory Comput. Syst.*, 34:27–44, 2001.

[10] Pierre Lecomte and Michel Rigo. *Abstract numeration systems*, chapter 3, pages 108–162. In Berthé and Rigo [3], 2010.

[11] Victor Marsault. *Énumération et Numération*. PhD thesis, Télécom–ParisTech, 2016.

[12] Victor Marsault and Jacques Sakarovitch. On sets of numbers rationally represented in a rational base number system. In Traian Muntean, Dimitrios Poulakis, and Robert Rolland, editors, *CAI 2013*, number 8080 in Lect. Notes in Comput. Sci., pages 89–100, 2013.

[13] Victor Marsault and Jacques Sakarovitch. The signature of rational languages. *Theoret. Computer Sci.*, 2016. To appear.

[14] Victor Marsault and Jacques Sakarovitch. Trees and languages with periodic signature. In Evangelos Kranakis, Gonzalo Navarro, and Edgar Chávez, editors, *LATIN 2016*, number 9644 in Lect. Notes in Comput. Sci., pages 605–618, 2016.