AUTOMATIC SEQUENCES: FROM RATIONAL BASES TO TREES

MICHEL RIGO AND MANON STIPULANTI

Abstract. The nth term of an automatic sequence is the output of a deterministic finite automaton fed with the representation of n in a suitable numeration system. In this paper, instead of considering automatic sequences built on a numeration system with a regular numeration language, we consider these built on languages associated with trees having periodic labeled signatures and, in particular, rational base numeration systems. We obtain two main characterizations of these sequences. The first one is concerned with r-block substitutions where r morphisms are applied periodically. In particular, we provide examples of such sequences that are not morphic. The second characterization involves the factors, or subtrees of finite height, of the tree associated with the numeration system and decorated by the terms of the sequence.

2010 *Mathematics Subject Classification*: 68R15, 68Q45, 11A63.
Keywords: Automatic sequences, abstract numeration systems, rational base numeration systems, alternating morphisms, PD0L systems, Cobham's theorem.

1. Introduction

Motivated by a question of Mahler in number theory, the introduction of rational base numeration systems has brought to light a family of formal languages with a rich combinatorial structure [1]. In particular, the generation of infinite trees with a periodic signature has emerged [17, 18, 19, 20]. Marsault and Sakarovitch very quickly linked the enumeration of the vertices of such trees (called breadth-first serialization) to the concept of abstract numeration system built on the corresponding prefix-closed language: the traversal of the tree is exactly the radix enumeration of the words of the language. In this paper, we study automatic sequences associated with that type of numeration systems. In particular, in the rational base $\frac{p}{q}$, a sequence is $\frac{p}{q}$-automatic if its nth term is obtained as the output of a DFAO fed with the base-$\frac{p}{q}$ representation of n. Thanks to a result of Lepistö [13] on factor complexity, we observe that we can get sequences that are not morphic.

We obtain several characterizations of these sequences. The first one boils down to translate Cobham's theorem from 1972 into this setting. In Section 4, we show that any automatic sequence built on a tree language with a purely periodic labeled signature is the image under a coding of an alternate fixed point of uniform morphisms not necessarily of the same length. If all the morphisms had the same

The first author dedicates this paper to the memory of his grandmother Marie Wuidar (1923–2020).

length, as observed in [11], we would only get classical k-automatic sequences. As a consequence, in the rational base $\frac{p}{q}$, if a sequence is $\frac{p}{q}$-automatic, then it is the image under a coding of a fixed point of a q-block substitution whose images all have length p. In the literature, these substitutions are also called PD0L where a periodic control is applied — q different morphisms are applied depending on the index of the considered letter modulo q.

On the other hand, Sturmian trees as studied in [3] also have a rich combinatorial structure where subtrees play a special role analogous to factors occurring in infinite words. In Section 5, we discuss the number of factors, i.e., subtrees of finite height, that may appear in the tree whose paths from the root are labeled by the words of the numeration language and whose vertices are colored according to the sequence of interest. Related to the k-kernel of a sequence, we obtain a new characterization of the classical k-automatic sequences: a sequence $\mathbf{x}$ is k-automatic if and only if the labeled tree of the base-k numeration system decorated by $\mathbf{x}$ is rational, i.e., it has finitely many infinite subtrees. For numeration systems built on a regular language, the function counting the number of decorated subtrees of height n is bounded, and we get a similar result. This is not the case in the more general setting of rational base numeration systems. Nevertheless, we obtain sufficient conditions for a sequence to be $\frac{p}{q}$-automatic in terms of the number of subtrees.

This paper is organized as follows. In Section 2, we recall basic definitions about abstract numeration systems, tree languages, rational base numeration systems, and alternate morphisms. In Section 3, we give some examples of the automatic sequences that we will consider. The parity of the sum-of-digits in base $\frac{3}{2}$ is such an example. In Section 4, Cobham's theorem is adapted to the case of automatic sequences built on tree languages with a periodic labeled signature in Theorem 20 (so, in particular, to the rational base numeration systems in Corollary 21). In Section 5, we decorate the nodes of the tree associated with the language of a rational base numeration system with the elements of a sequence taking finitely many values. Under some mild assumption (always satisfied when distinct states of the deterministic finite automaton with output producing the sequence have distinct output), we obtain a characterization of $\frac{p}{q}$-automatic sequences in terms of the number of trees of some finite height occurring in the decorated tree. In Section 6, we review some usual closure properties of $\frac{p}{q}$-automatic sequences.

2. Preliminaries

We make use of common notions in combinatorics on words, such as alphabet, letter, word, length of a word, language and usual definitions from automata theory. In particular, we let ε denote the empty word. For a finite word w, we let $|w|$ denote its length. For each $i \in \{0, \ldots, |w|-1\}$, we let w_i denote the ith letter of w (and we thus start indexing letters at 0.)

2.1. Abstract numeration systems.

When dealing with abstract numeration systems, it is usually assumed that the language of the numeration system is regular. However the main feature is that words are enumerated by radix order (also called genealogical order: words are first ordered by increasing length and words of the same length are ordered by lexicographical order). The generalization of abstract numeration systems to context-free languages was, for instance, considered in [5].

Rational base numeration systems discussed below in Section 2.3 are also abstract numeration systems built on non-regular languages.

Definition 1. An *abstract numeration system* (or *ANS* for short) is a triple $\mathcal{S} = (L, A, <)$ where L is an infinite language over a totally ordered (finite) alphabet $(A, <)$. We say that L is the *numeration language*. The map $\mathrm{rep}_{\mathcal{S}} : \mathbb{N} \to L$ is the one-to-one correspondence mapping $n \in \mathbb{N}$ onto the $(n+1)$st word in the radix ordered language L, which is then called the *$\mathcal{S}$-representation* of n. The $\mathcal{S}$-representation of 0 is the first word in L. The inverse map is denoted by $\mathrm{val}_{\mathcal{S}} : L \to \mathbb{N}$. For any word w in L, $\mathrm{val}_{\mathcal{S}}(w)$ is its *$\mathcal{S}$-numerical value*.

Positional numeration systems, such as integer base numeration systems, the Fibonacci numeration system, and Pisot numeration systems, are based on the greediness of the representations. They all share the following property: $m < n$ if and only if $\mathrm{rep}(m)$ is less than $\mathrm{rep}(n)$ for the radix order. These numeration systems are thus ANS. As a non-standard example of ANS, consider the language a^*b^* over $\{a, b\}$ and assume that $a < b$. Let $\mathcal{S} = (a^*b^*, \{a, b\}, <)$. The first few words in the numeration language are $\varepsilon, a, b, aa, ab, bb, \ldots$. For instance, $\mathrm{rep}_{\mathcal{S}}(3) = aa$ and $\mathrm{rep}_{\mathcal{S}}(5) = bb$. One can show that $\mathrm{val}_{\mathcal{S}}(a^p b^q) = \frac{(p+q)(p+q+1)}{2} + q$. For details, we refer the reader to [12] or [23].

In the next definition, we assume that most significant digits are read first. This is not real restriction (see Section 6).

Definition 2. Let $\mathcal{S} = (L, A, <)$ be an abstract numeration system and let B be a finite alphabet. An infinite word $\mathbf{x} = x_0x_1x_2 \cdots \in B^{\mathbb{N}}$ is *$\mathcal{S}$-automatic* if there exists a deterministic finite automaton with output (DFAO for short) $\mathcal{A} = (Q, q_0, A, \delta, \mu : Q \to B)$ such that $x_n = \mu(\delta(q_0, \mathrm{rep}_{\mathcal{S}}(n)))$ for all $n \geq 0$.

Let $k \geq 2$ be an integer. We let A_k denote the alphabet $\{0, 1, \ldots, k-1\}$. For the usual base-k numeration system built on the language

$$L_k := \{\varepsilon\} \cup \{1, \ldots, k-1\}\{0, \ldots, k-1\}^*, \tag{2.1}$$

an $\mathcal{S}$-automatic sequence is said to be k-automatic [2]. We also write rep_k and val_k in this context.

2.2. **Tree languages.** Prefix-closed languages define labeled trees (also called *trie* or *prefix-tree* in computer science) and vice-versa. Let $(A, <)$ be a totally ordered (finite) alphabet and let L be a prefix-closed language over $(A, <)$. The set of nodes of the tree is L. If w and wd are words in L with $d \in A$, then there is an edge from w to wd with label d. The children of a node are ordered by the labels of the letters in the ordered alphabet A. In Figure 1, we have depicted the first levels of the tree associated with the prefix-closed language a^*b^*. Nodes are enumerated by breadth-first traversal (or, serialization).

We recall some notion from [18] or [20]. Let T be an ordered tree of finite degree. The *(breath-first) signature* of T is a sequence of integers, the sequence of the degrees of the nodes visited by the (canonical) breadth-first traversal of the tree. The *(breath-first) labeling* of T is the infinite sequence of the labels of the edges visited by the breadth-first traversal of this tree. As an example, with the tree in Figure 1, its signature is $2, 2, 1, 2, 1, 1, 2, 1, 1, 1, 2, \ldots$ and its labeling is $a, b, a, b, b, a, b, b, b, a, b, \ldots$.

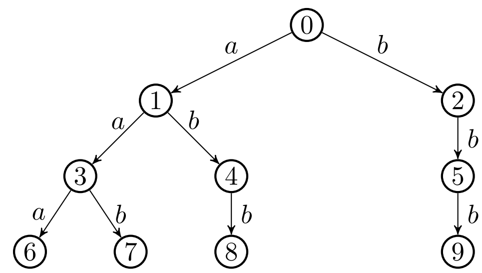

FIGURE 1. The first few levels of the tree associated with a^*b^*.

Remark 3. As observed by Marsault and Sakarovitch [18], it is usually convenient to consider *i-trees*: the root is assumed to be a child of itself. It is especially the case for positional numeration systems when one has to deal with leading zeroes as the words u and $0u$ may represent the same integer.

We now present a useful way to describe or generate infinite labeled i-trees. Let A be a finite alphabet containing 0. A *labeled signature* is an infinite sequence $(w_n)_{n\geq 0}$ of finite words over A providing a signature $(|w_n|)_{n\geq 0}$ and a consistent labeling of a tree (made of the sequence of letters of $(w_n)_{n\geq 0}$). It will be assumed that the letters of each word are in strictly increasing order and that $w_0 = 0x$ with $x \in A^+$. To that aim we let $\mathsf{inc}(A^*)$ denote the set of words over A with increasingly ordered letters. For instance, 025 belongs to $\mathsf{inc}(A_6^*)$ but 0241 does not. Examples of labeled signatures will be given in the next section.

Remark 4. Since a labeled signature s generates an i-tree, by abuse, we say that such a signature defines a prefix-closed language denoted by $L(\mathsf{s})$. Moreover, since we assumed the words of s all belong to $\mathsf{inc}(A^*)$ for some finite alphabet A, the canonical breadth-first traversal of this tree produces an abstract numeration system. Indeed the enumeration of the nodes $v_0, v_1, v_2, \ldots$ of the tree is such that v_n is the nth word in the radix ordered language $L(\mathsf{s})$. The language $L(\mathsf{s})$, the set of nodes of the tree and $\mathbb{N}$ are thus in one-to-one correspondence.

2.3. **Rational bases.** The framework of rational base numeration systems [1] is an interesting setting giving rise to a non-regular numeration language. Nevertheless the corresponding tree has a rich combinatorial structure: it has a purely periodic labeled signature.

Let p and q be two relatively prime integers with $p > q > 1$. Given a positive integer n, we define the sequence $(n_i)_{i\geq 0}$ as follows: we set $n_0 = n$ and, for all $i \geq 0$, $qn_i = pn_{i+1} + a_i$ where a_i is the remainder of the Euclidean division of qn_i by p. Note that $a_i \in A_p$ for all $i \geq 0$. Since $p > q$, the sequence $(n_i)_{i\geq 0}$ is decreasing and eventually vanishes at some index $\ell + 1$. We obtain

$$n = \sum_{i=0}^{\ell} \frac{a_i}{q} \left(\frac{p}{q}\right)^i.$$

Conversely, for a word $w = w_\ell w_{\ell-1} \cdots w_0 \in A_p^*$, the value of w in base $\frac{p}{q}$ is the rational number

$$\mathrm{val}_{\frac{p}{q}}(w) = \sum_{i=0}^{\ell} \frac{w_i}{q} \left(\frac{p}{q}\right)^i.$$

Note that $\mathrm{val}_{\frac{p}{q}}(w)$ is a not always an integer and $\mathrm{val}_{\frac{p}{q}}(uv) = \mathrm{val}_{\frac{p}{q}}(u)(\frac{p}{q})^{|v|} + \mathrm{val}_{\frac{p}{q}}(v)$ for all $u, v \in A_p^*$. We let $N_{\frac{p}{q}}$ denote the *value set*, i.e., the set of numbers representable in base $\frac{p}{q}$:

$$N_{\frac{p}{q}} = \mathrm{val}_{\frac{p}{q}}(A_p^*) = \left\{ x \in \mathbb{Q} \mid \exists w \in A_p^* : \mathrm{val}_{\frac{p}{q}}(w) = x \right\}.$$

A word $w \in A_p^*$ is a *representation* of an integer $n \geq 0$ in base $\frac{p}{q}$ if $\mathrm{val}_{\frac{p}{q}}(w) = n$. As for integer bases, representations in rational bases are unique up to leading zeroes [1, Theorem 1]. Therefore we let $\mathrm{rep}_{\frac{p}{q}}(n)$ denote the representation of n in base $\frac{p}{q}$ that does not start with 0. By convention, the representation of 0 in base $\frac{p}{q}$ is the empty word ε. In base $\frac{p}{q}$, the numeration language is the set

$$L_{\frac{p}{q}} = \left\{ \mathrm{rep}_{\frac{p}{q}}(n) \mid n \geq 0 \right\}.$$

Hence, rational base numeration systems are special cases of ANS built on $L_{\frac{p}{q}}$: $m < n$ if and only if $\mathrm{rep}_{\frac{p}{a}}(m) < \mathrm{rep}_{\frac{p}{a}}(n)$ for the radix order. It is clear that $L_{\frac{p}{q}} \subseteq A_p^*$ is a prefix-closed language. As a consequence of the previous section, it can be seen as a tree.

Example 5. The alphabet for the base $\frac{3}{2}$ is $A_3 = \{0, 1, 2\}$. The first few words in $L_{\frac{3}{2}}$ are $\varepsilon, 2, 21, 210, 212, 2101, 2120, 2122$, and the associated tree is depicted in Figure 2. If we add an edge of label 0 on the root of this tree (see Remark 3), its signature is $2, 1, 2, 1, \ldots$ and its labeling is $0, 2, 1, 0, 2, 1, 0, 2, 1, \ldots$. Otherwise stated, the purely periodic labeled signature $(02, 1)^\omega$ gives the i-tree of the language $L_{\frac{3}{2}}$; see Figure 2. For all $n \geq 0$, the nth node in the breadth-first traversal is the word $\mathrm{rep}_{\frac{3}{2}}(n)$. Observe that there is an edge labeled by $a \in A_3$ from the node n to the node m if and only if $m = \frac{3}{2} \cdot n + \frac{a}{2}$. This remark is valid for all rational bases.

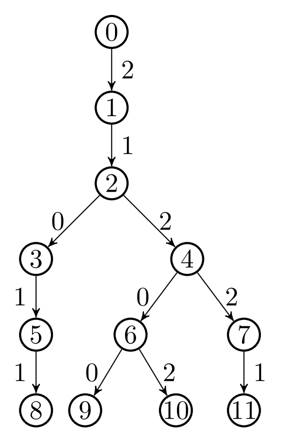

FIGURE 2. The first levels of the tree associated with $L_{\frac{3}{2}}$.

Remark 6. The language $L_{\frac{p}{q}}$ is highly non-regular: it has the bounded left-iteration property; for details, see [17]. In $L_{\frac{p}{q}}$ seen as a tree, no two infinite subtrees are isomorphic, i.e., for any two words $u, v \in L_{\frac{p}{q}}$ with $u \neq v$, the quotients $u^{-1}L_{\frac{p}{q}}$ and $v^{-1}L_{\frac{p}{q}}$ are distinct. As we will see with Lemma 29, this does not prevent the languages $u^{-1}L_{\frac{p}{q}}$ and $v^{-1}L_{\frac{p}{q}}$ from coinciding on words of length bounded by a constant depending on $\mathrm{val}_{\frac{p}{q}}(u)$ and $\mathrm{val}_{\frac{p}{q}}(v)$ modulo a power of q. Nevertheless the associated tree has a purely periodic labeled signature. For example, with $\frac{p}{q}$ respectively equal to $\frac{3}{2}$, $\frac{5}{2}$, $\frac{7}{3}$ and $\frac{11}{4}$, we respectively have the signatures $(02,1)^\omega$, $(024,13)^\omega$, $(036,25,14)^\omega$, $(048,159,26(10),37)^\omega$. Generalizations of these languages (called rhythmic generations of trees) are studied in [20].

Definition 7. We say that a sequence is $\frac{p}{q}$-*automatic* if it is $\mathcal{S}$-automatic for the ANS built on the language $L_{\frac{p}{q}}$, i.e., $\mathcal{S} = (L_{\frac{p}{q}}, A_p, <)$.

2.4. **Alternating morphisms.** The Kolakoski–Oldenburger word [24, `A000002`] is the unique word $\mathbf{k}$ over $\{1,2\}$ starting with 2 and satisfying $\Delta(\mathbf{k}) = \mathbf{k}$ where Δ is the run-length encoding map

$$\mathbf{k} = 2211212212211\cdots.$$

It is a well-known (and challenging) object of study in combinatorics on words. It can be obtained by periodically iterating two morphisms, namely

$$h_0 : \begin{cases} 1 \mapsto 2 \\ 2 \mapsto 22 \end{cases} \quad \text{and} \quad h_1 : \begin{cases} 1 \mapsto 1 \\ 2 \mapsto 11. \end{cases}$$

More precisely, in [7], $\mathbf{k} = k_0k_1k_2\cdots$ is expressed as the fixed point of the iterated morphisms (h_0, h_1), i.e.,

$$\mathbf{k} = h_0(k_0)h_1(k_1)\cdots h_0(k_{2n})h_1(k_{2n+1})\cdots.$$

In the literature, one also finds the terminology PD0L for *D0L system with periodic control* [11, 13].

Definition 8. Let $r \geq 1$ be an integer, let A be a finite alphabet, and let $f_0, \ldots, f_{r-1}$ be r morphisms over A^*. An infinite word $\mathbf{w} = w_0w_1w_2\cdots$ over A is an *alternate fixed point* of $(f_0, \ldots, f_{r-1})$ if

$$\mathbf{w} = f_0(w_0)f_1(w_1)\cdots f_{r-1}(w_{r-1})f_0(w_r)\cdots f_{i \bmod r}(w_i)\cdots.$$

As observed by Dekking [8] for the Kolakoski word, an alternate fixed point can also be obtained by an r-block substitution.

Definition 9. Let $r \geq 1$ be an integer and let A be a finite alphabet. An *r-block substitution* $g : A^r \to A^*$ maps a word $w_0\cdots w_{rn-1} \in A^*$ to

$$g(w_0\cdots w_{r-1})g(w_r\cdots w_{2r-1})\cdots g(w_{r(n-1)}\cdots w_{rn-1}).$$

If the length of the word is not a multiple of r, then the suffix of the word is ignored under the action of g. An infinite word $\mathbf{w} = w_0w_1w_2\cdots$ over A is a *fixed point of the r-block substitution* $g : A^r \to A^*$ if

$$\mathbf{w} = g(w_0\cdots w_{r-1})g(w_r\cdots w_{2r-1})\cdots.$$

Proposition 10. *Let $r \geq 1$ be an integer, let A be a finite alphabet, and let $f_0, \ldots, f_{r-1}$ be r morphisms over A^*. If an infinite word over A is an alternate fixed point of $(f_0, \ldots, f_{r-1})$, then it is a fixed point of an r-block substitution.*

Proof. For every of length-r word $a_0\cdots a_{r-1} \in A^*$, define the r-block substitution $g : A^r \to A^*$ by $g(a_0\cdots a_{r-1}) = f_0(a_0)\cdots f_{r-1}(a_{r-1})$. $\square$

Thanks to the previous result, the Kolakoski–Oldenburger word $\mathbf{k}$ is also a fixed point of the 2-block substitution

$$g : \begin{cases} 11 \mapsto h_0(1)h_1(1) = 21 \\ 12 \mapsto h_0(1)h_1(2) = 211 \\ 21 \mapsto h_0(2)h_1(1) = 221 \\ 22 \mapsto h_0(2)h_1(2) = 2211. \end{cases}$$

Observe that the lengths of images under g are not all equal.

3. Concrete examples of automatic sequences

Let us present how the above concepts are linked with the help of some examples. The first one is our toy example.

Example 11. Let $(s(n))_{n\geq 0}$ be the sum-of-digits in base $\frac{3}{2}$. This sequence was, in particular, studied in [10]. We have

$$(s(n))_{n\geq 0} = 0, 2, 3, 3, 5, 4, 5, 7, 5, 5, 7, 8, 5, 7, 6, 7, 9, \ldots.$$

We let $\mathbf{t}$ denote the sequence $(s(n) \bmod 2)_{n\geq 0}$,

$$\mathbf{t} = 00111011111011011 \cdots.$$

The sequence $\mathbf{t}$ is $\frac{3}{2}$-automatic as the DFAO in Figure 3 generates $\mathbf{t}$ when reading base-$\frac{3}{2}$ representations.

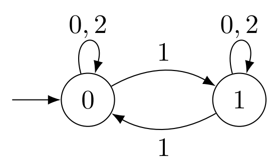

Figure 3. A DFAO generating the sum-of-digits in base $\frac{3}{2}$ modulo 2.

As a consequence of Proposition 16, it will turn out that $\mathbf{t}$ is an alternate fixed point of (f_0, f_1) with

$$f_0\colon \begin{cases} 0 \mapsto 00 \\ 1 \mapsto 11 \end{cases} \quad \text{and } f_1\colon \begin{cases} 0 \mapsto 1 \\ 1 \mapsto 0. \end{cases} \tag{3.1}$$

With Proposition 10, $\mathbf{t}$ is also a fixed point of the 2-block substitution

$$g : \begin{cases} 00 \mapsto f_0(0)f_1(0) = 001 \\ 01 \mapsto f_0(0)f_1(1) = 000 \\ 10 \mapsto f_0(1)f_1(0) = 111 \\ 11 \mapsto f_0(1)f_1(1) = 110. \end{cases}$$

Observe that we have a 2-block substitution with images of length 3. This is not a coincidence, as we will see with Corollary 21.

Automatic sequences in integer bases are morphic words, i.e., images, under a coding, of a fixed point of a prolongable morphism [2]. As shown by the next example, there are $\frac{3}{2}$-automatic sequences that are not morphic. For a word $u \in \{0, 1\}^*$, we let $\overline{u}$ denote the word obtained by applying the involution $i \mapsto 1 - i$, $i \in \{0, 1\}$, to the letters of u.

Example 12. Lepistö considered in [13] the following 2-block substitution

$$h_2 : \begin{cases} 00 \mapsto g_0(0)\overline{0} = 011 \\ 01 \mapsto g_0(0)\overline{1} = 010 \\ 10 \mapsto g_0(1)\overline{0} = 001 \\ 11 \mapsto g_0(1)\overline{1} = 000 \end{cases} \quad \text{with } g_0 : 0 \mapsto 01, 1 \mapsto 00,$$

producing the word $\mathbf{F}_2 = 01001100001\cdots$. He showed that the factor complexity $\mathsf{p}_{\mathbf{F}_2}$ of this word satisfies $\mathsf{p}_{\mathbf{F}_2}(n) > \delta n^t$ for some $\delta > 0$ and $t > 2$. Hence, this word cannot be purely morphic nor morphic (because these kinds of words have a factor complexity in $O(n^2)$ [21]). With Proposition 17, we can show that $\mathbf{F}_2$ is a $\frac{3}{2}$-automatic sequence generated by the DFAO depicted in Figure 4.

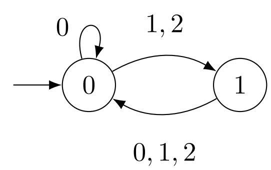

FIGURE 4. A DFAO generating $\mathbf{F}_2$.

Remark 13. Similarly, the non-morphic word $\mathbf{F}_p$ introduced in [13] is $\frac{p+1}{p}$-automatic. It is generated by the p-block substitution defined by $h_p(au) = g_0(a)\overline{u}$ for $a \in \{0, 1\}$ and $u \in \{0, 1\}^{p-1}$, where g_0 is defined in Example 12.

We conclude this section with an example of an automatic sequence associated with a language coming from a periodic signature.

Example 14. Consider the periodic labeled signature $\mathsf{s} = (023, 14, 5)^\omega$ producing the i-tree in Figure 5. The first few words in $L(\mathsf{s})$ are

$$\varepsilon, 2, 3, 21, 24, 35, 210, 212, 213, 241, 244, 355, \ldots$$

which give the representations of the first 12 integers in the abstract numeration system $\mathcal{S} = (L(\mathsf{s}), A_6, <)$. For instance, $\mathrm{rep}_{\mathcal{S}}(15) = 2121$ as the path of label 2121 leads to the node 15 in Figure 5. The sum-of-digits in $\mathcal{S}$ modulo 2, starting with

$$001100110101\cdots,$$

is $\mathcal{S}$-automatic since it is generated by the DFAO in Figure 6. As a consequence of Proposition 16 and Theorem 20, we will see that this sequence is also the coding of an alternate fixed point of three morphisms.

4. COBHAM'S THEOREM

Cobham's theorem from 1972 states that a sequence is k-automatic if and only if it is the image under a coding of the fixed point of a k-uniform morphism [6] (or see [2, Theorem 6.3.2]). This result has been generalized to various contexts: numeration systems associated with a substitution, Pisot numeration systems, Bertrand numeration systems, ANS with regular languages, and so on [4, 9, 14, 22]. Also see [12] or [23] for a comprehensive presentation. In this section, we adapt it to the case of $\mathcal{S}$-automatic sequences built on tree languages with a periodic labeled

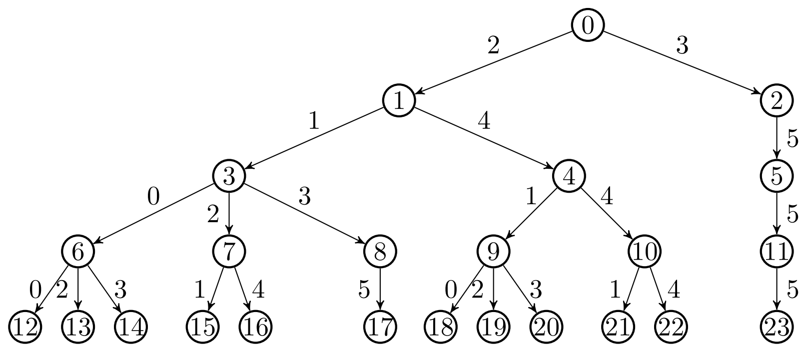

FIGURE 5. The tree associated with the signature $(023, 14, 5)^\omega$.

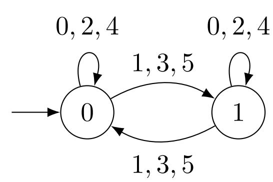

FIGURE 6. A DFAO generating the sum-of-digits modulo 2 in the ANS $\mathcal{S} = (L(\mathsf{s}), A_6, <)$ where $\mathsf{s} = (023, 14, 5)^\omega$.

signature (so, in particular, to the rational base case). We start off with a technical lemma.

Lemma 15. *Let $r \geq 1$ be an integer, let A be a finite alphabet, and let $f_0, \ldots, f_{r-1}$ be morphisms over A^*. Let $\mathbf{x} = x_0x_1x_2\cdots$ be an alternate fixed point of $(f_0, \ldots, f_{r-1})$. For all $m \geq 0$, we have*

$$f_{m \bmod r}(x_m) = x_i \cdots x_{i+|f_{m \bmod r}(x_m)|-1}$$

where $i = \sum_{j=0}^{m-1} |f_{j \bmod r}(x_j)|$.

Proof. Let $m \geq 0$. From the definition of an alternate fixed point, we have the factorization $\mathbf{x} = uf_{m \bmod r}(x_m)f_{(m+1) \bmod r}(x_{m+1})\cdots$ where

$$u = f_0(x_0)f_1(x_1)\cdots f_{r-1}(x_{r-1})f_0(x_r)\cdots f_{(m-1) \bmod r}(x_{m-1}).$$

Now $|u| = \sum_{j=0}^{m-1} |f_{j \bmod r}(x_j)|$, which concludes the proof. □

Given an $\mathcal{S}$-automatic sequence associated with the language of a tree with a purely periodic labeled signature, we can turn it into an alternate fixed point of uniform morphisms.

Proposition 16. *Let $r \geq 1$ be an integer and let A be a finite alphabet of digits. Let $w_0, \ldots, w_{r-1}$ be r non-empty words in $\mathsf{inc}(A^*)$. Consider the language $L(\mathsf{s})$ of the i-tree generated by the purely period signature $\mathsf{s} = (w_0, w_1, \ldots, w_{r-1})^\omega$. Let $\mathcal{A} = (Q, q_0, A, \delta)$ be a DFA. For $i \in \{0, \ldots, r-1\}$, we define the r morphisms from Q^* to itself by*

$$f_i : Q \to Q^{|w_i|}, q \mapsto \delta(q, w_{i,0}) \cdots \delta(q, w_{i,|w_i|-1}),$$

where $w_{i,j}$ denotes the jth letter of w_i. The alternate fixed point $\mathbf{x} = x_0x_1\cdots$ of $(f_0,\ldots,f_{r-1})$ starting with q_0 is the sequence of states reached in $\mathcal{A}$ when reading the words of $L(\mathsf{s})$ in increasing radix order, i.e., for all $n \ge 0$, $x_n = \delta(q_0, \mathrm{rep}_{\mathcal{S}}(n))$ with $\mathcal{S} = (L(\mathsf{s}), A, <)$.

Proof. Up to renaming the letters of w_0, without loss of generality we may assume that $w_0 = 0x$ with $x \in A^+$.

We proceed by induction on $n \ge 0$. It is clear that $x_0 = \delta(q_0, \varepsilon) = q_0$. Let $n \ge 1$. Assume that the property holds for all values less than n and we prove it for n.

Write $\mathrm{rep}_{\mathcal{S}}(n) = a_\ell \cdots a_1 a_0$. This means that in the i-tree generated by s, we have a path of label $a_\ell \cdots a_0$ from the root. We identify words in $L(\mathsf{s})$ with vertices of the i-tree.

Since $L(\mathsf{s})$ is prefix-closed, there exists an integer $m < n$ such that $\mathrm{rep}_{\mathcal{S}}(m) = a_\ell \cdots a_1$. Let $i = m \bmod r$. By definition of the periodic labeled signature s, in the i-tree generated by s, reading $a_\ell \cdots a_1$ from the root leads to a node having $|w_i|$ children that are reached with edges labeled by the letters of w_i. Since $w_i \in \mathsf{inc}(A^*)$, the letter a_0 occurs exactly once in w_i, so assume that $w_{i,j} = a_0$ for some $j \in \{0,\ldots,|w_i|-1\}$. By construction of the i-tree given by a periodic labeled signature (see Figure 7 for a pictorial description), we have that

$$n = \sum_{\substack{v\in L(\mathsf{s})\\ v<\mathrm{rep}_{\mathcal{S}}(m)}} \deg(v) + j = \sum_{k=0}^{m-1} |w_{k \bmod r}| + j. \tag{4.1}$$

By the induction hypothesis, we obtain

$$\delta(q_0, \mathrm{rep}_{\mathcal{S}}(n)) = \delta(\delta(q_0, \mathrm{rep}_{\mathcal{S}}(m)), a_0) = \delta(x_m, a_0)$$

and by definition of f_i, we get $\delta(x_m, a_0) = [f_i(x_m)]_j = [f_{m \bmod r}(x_m)]_j$. From Lemma 15 and Equation (4.1), this is exactly x_n, as desired. □

Given an alternate fixed point of uniform morphisms, we can turn it into an $\mathcal{S}$-automatic sequence for convenient choices of a language of a tree with a purely periodic labeled signature and a DFAO.

Proposition 17. *Let $r \ge 1$ be an integer and let A be a finite alphabet. Let $f_0,\ldots,f_{r-1} : A^* \to A^*$ be r uniform morphisms of respective length $\ell_0,\ldots,\ell_{r-1}$ such that f_0 is prolongable on some letter $a \in A$, i.e., $f_0(a) = ax$ with $x \in A^+$. Let $\mathbf{x} = x_0x_1\cdots$ be the alternate fixed point of $(f_0,\ldots,f_{r-1})$ starting with a. Consider the language $L(\mathsf{s})$ of the i-tree generated by the purely periodic labeled signature*

$$\mathsf{s} = \left(0\cdots(\ell_0-1), \ell_0(\ell_0+1)\cdots(\ell_0+\ell_1-1), \ldots, \left(\sum_{j<r-1}\ell_j\right)\cdots\left(\sum_{j<r}\ell_j - 1\right)\right)^\omega,$$

which is made of consecutive non-negative integers. Define a DFA $\mathcal{A}$ having

- *A as set of states,*
- *a as initial state,*
- *$B = \{0,\ldots,\sum_{j<r}\ell_j - 1\}$ as alphabet,*
- *its transition function $\delta : A \times B \to A$ defined as follows: For all $i \in B$, there exist a unique $j_i \ge 0$ and a unique $t_i \ge 0$ such that $i = \sum_{k\le j_i-1}\ell_k + t_i$ with $t_i < \ell_{j_i}$, and we set*

$$\delta(b, i) = [f_{j_i}(b)]_{t_i}, \quad \forall b \in A.$$

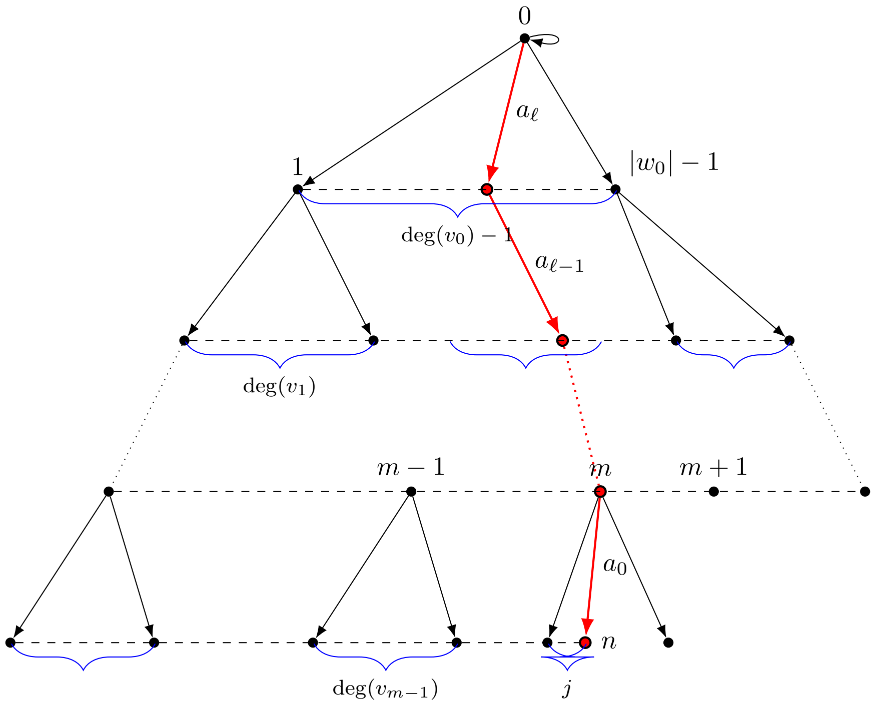

FIGURE 7. Illustration of Equation (4.1).

Then the word $\mathbf{x}$ *is the sequence of the states reached in* $\mathcal{A}$ *when reading the words of* $L(\mathsf{s})$ *by increasing radix order, i.e.,* $x_n = \delta(a, \mathrm{rep}_{\mathcal{S}}(n))$ *with* $\mathcal{S} = (L(\mathsf{s}), B, <)$.

Proof. We again proceed by induction on $n \geq 0$. It is clear that $x_0 = a = \delta(a, \varepsilon)$. Let $n \geq 1$. Assume the property holds for all values less than n and we prove it for n.

Write $\mathrm{rep}_{\mathcal{S}}(n) = a_\ell \cdots a_1 a_0$. This means that in the i-tree with a periodic labeled signature s, we have a path of label $a_\ell \cdots a_0$ from the root. We identify words in $L(\mathsf{s}) \subseteq B^*$ with vertices of the i-tree.

Since $L(\mathsf{s})$ is prefix-closed, there exists $m < n$ such that $\mathrm{rep}_{\mathcal{S}}(m) = a_\ell \cdots a_1$. Let $j = m \bmod r$. In the i-tree generated by s, reading $a_\ell \cdots a_1$ from the root leads to a node having ℓ_j children that are reached with edges labeled by

$$\sum_{k \leq j-1} \ell_k, \ \sum_{k \leq j-1} \ell_k + 1, \ \ldots, \ \sum_{k \leq j} \ell_k - 1.$$

Observe that the words in s belong to $\mathsf{inc}(B^*)$. Therefore the letter a_0 occurs exactly once in B and in particular amongst those labels, assume that $a_0 = \sum_{k \leq j-1} \ell_k + t$ for some $t \in \{0, \ldots, \ell_j - 1\}$. By construction of the i-tree, we have that

$$n = \sum_{\substack{v \in L(\mathsf{s}) \\ v < \mathrm{rep}_{\mathcal{S}}(m)}} \deg(v) + t = \sum_{i=0}^{m-1} \ell_{i \bmod r} + t. \tag{4.2}$$

By the induction hypothesis, we obtain

$$\delta(a, \mathrm{rep}_{\mathcal{S}}(n)) = \delta(\delta(a, \mathrm{rep}_{\mathcal{S}}(m)), a_0) = \delta(x_m, a_0)$$

and by definition of the transition function, $\delta(x_m, a_0) = [f_j(x_m)]_t = [f_{m \bmod r}(x_m)]_t$. From Lemma 15 and Equation (4.2), this is exactly x_n. □

Remark 18. What matters in the above statement is that two distinct words of the signature s do not share any common letter. It mainly ensures that the choice of the morphism to apply when defining δ is uniquely determined by the letter to be read.

Example 19. If we consider the morphisms in (3.1), Proposition 17 provides us with the signature $\mathsf{s} = (01, 2)^\omega$ instead of the signature $(02, 1)^\omega$ of $L_{\frac{3}{2}}$. We will produce the sequence $\mathbf{t}$ using the language $h(L_{\frac{3}{2}})$ where the coding h is defined by $h(0) = 0$, $h(1) = 2$ and $h(2) = 1$ and in the DFAO in Figure 3, the same coding is applied to the labels of the transitions. What matters is the form of the tree (i.e., the sequence of degrees of the vertices) rather than the labels themselves.

Theorem 20. *Let A, B be two finite alphabets. An infinite word over B is the image under a coding $g : A \to B$ of an alternate fixed point of uniform morphisms (not necessarily of the same length) over A if and only if it is $\mathcal{S}$-automatic for an abstract numeration system $\mathcal{S}$ built on a tree language with a purely periodic labeled signature.*

Proof. The forward direction follows from Proposition 17: define a DFAO where the output function τ is obtained from the coding $g : A \to B$ defined by $\tau(b) = g(b)$ for all b in A. The reverse direction directly follows from Proposition 16. □

We are able to say more in the special case of rational bases. The tree language associated with the rational base $\frac{p}{q}$ has a periodic signature of the form $(w_0, \dots, w_{q-1})^\omega$ with $\sum_{i=0}^{q-1} |w_i| = p$ and $w_i \in A_p^*$ for all i. See Remark 6 for examples.

Corollary 21. *If a sequence is $\frac{p}{q}$-automatic, then it is the image under a coding of a fixed point of a q-block substitution whose images all have length p.*

Proof. Let $(w_0, \dots, w_{q-1})^\omega$ denote the periodic signature in base $\frac{p}{q}$. Proposition 16 provides q morphisms f_i that are respectively $|w_i|$-uniform. By Proposition 10, the alternate fixed point of $(f_0, \dots, f_{q-1})$ is a fixed point of a q-block substitution g such that, for any length-q word $a_0 \cdots a_{q-1}$,

$$|g(a_0 \cdots a_{q-1})| = |f_0(a_0) f_1(a_1) \cdots f_{q-1}(a_{q-1})| = \sum_{i=0}^{q-1} |w_i| = p.$$ □

5. Decorating trees and subtrees

As already observed in Section 2.2, a prefix-closed language L over an ordered (finite) alphabet $(A, <)$ gives an ordered labeled tree $T(L)$ in which edges are labeled by letters in A. Labels of paths from the root to nodes provide a one-to-one correspondence between nodes in $T(L)$ and words in L. We now add an extra information, such as a color, on every node. This information is provided by a sequence taking finitely many values.

Definition 22. Let $T = (V, E)$ be a rooted ordered infinite tree, i.e., each node has a finite (ordered) sequence of children. As observed in Remark 4, the canonical breadth-first traversal of T gives an abstract numeration system — an enumeration of the nodes: $v_0, v_1, v_2, \ldots$. Let $\mathbf{x} = x_0x_1\cdots$ be an infinite word over a finite alphabet B. A *decoration* of T by $\mathbf{x}$ is a map from V to B associating with the node v_n the decoration (or color) x_n, for all $n \geq 0$.

To be consistent and to avoid confusion, we refer respectively to *label* and *decoration* the labeling of the edges and nodes of a tree.

Example 23. In Figure 8 are depicted a prefix of $T(L_{\frac{3}{2}})$ decorated with the sequence $\mathbf{t}$ of Example 11 and a prefix of the tree $T(L_2)$ associated with the binary numeration system (see (2.1)) and decorated with the Thue–Morse sequence $0110100110010110\cdots$. In these trees, the symbol 0 (respectively 1) is denoted by a black (respectively red) decorated node.

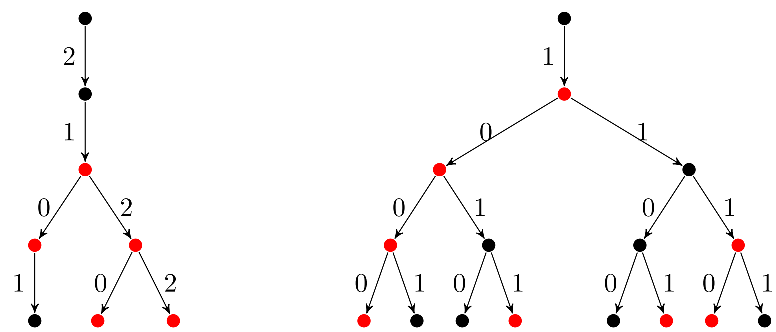

FIGURE 8. Prefixes of height 4 of two decorated trees.

We use the terminology of [3] where Sturmian trees are studied; it is relevant to consider (labeled and decorated) factors occurring in trees.

Definition 24. The *domain* $\mathrm{dom}(T)$ of a labeled tree T is the set of labels of paths from the root to its nodes. In particular, $\mathrm{dom}(T(L)) = L$ for any prefix-closed language L over an ordered (finite) alphabet. The *truncation* of a tree at height h is the restriction of the tree to the domain $\mathrm{dom}(T) \cap A^{\leq h}$.

Let L be a prefix-closed language over $(A, <)$ and $\mathbf{x} = x_0x_1\cdots$ be an infinite word over some finite alphabet B. (We could use an *ad hoc* notation like $T_{\mathbf{x}}(L)$ but in any case we only work with decorated trees and it would make the presentation cumbersome.) From now on, we consider the labeled tree $T(L)$ decorated by $\mathbf{x}$. For all $n \geq 0$, the nth word w_n in L corresponds to the nth node of $T(L)$ decorated by x_n. Otherwise stated, for the ANS $\mathcal{S} = (L, A, <)$ built on L, if $w \in L$, the node corresponding to w in $T(L)$ has decoration $x_{\mathrm{val}_{\mathcal{S}}(w)}$.

Definition 25. Let $w \in L$. We let $T[w]$ denote the subtree of T having w as root. Its domain is $w^{-1}L = \{u \mid wu \in L\}$. We say that $T[w]$ is a *suffix* of T.

For any $h \geq 0$, we let $T[w, h]$ denote the *factor of height h rooted at w*, which is the truncation of $T[w]$ at height h. The *prefix of height h* of T is the factor $T[\varepsilon, h]$. Two factors $T[w, h]$ and $T[w', h]$ of the same height are *equal* if they have the same domain and the same decorations, i.e., $x_{\mathrm{val}_{\mathcal{S}}(wu)} = x_{\mathrm{val}_{\mathcal{S}}(w'u)}$ for all $u \in \mathrm{dom}(T[w,h]) = \mathrm{dom}(T[w',h])$. We let

$$F_h = \{T[w,h] \mid w \in L\}$$

denote the set of factors of height h occurring in T. The tree T is *rational* if it has finitely many suffixes.

Note that, due to Remark 6, with any decoration, even constant, the tree $T(L_{\frac{p}{q}})$ is not rational.

In Figure 9, we have depicted the factors of height 2 occurring in $T(L_{\frac{3}{2}})$ decorated by $\mathbf{t}$. In Figure 10, we have depicted the factors of height 2 occurring in $T(L_2)$ decorated by the Thue–Morse sequence. In this second example, except for the prefix of height 2, observe that a factor of height 2 is completely determined by the decoration of its root.

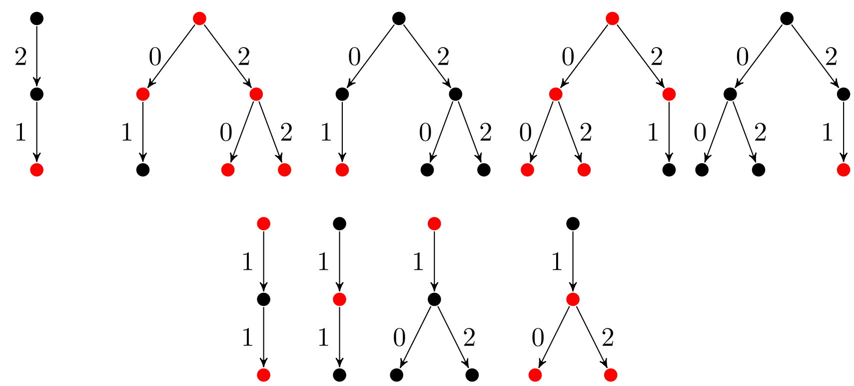

FIGURE 9. The 9 factors of height 2 in $T(L_{\frac{3}{2}})$ decorated by $\mathbf{t}$. The first one is the prefix occurring only once.

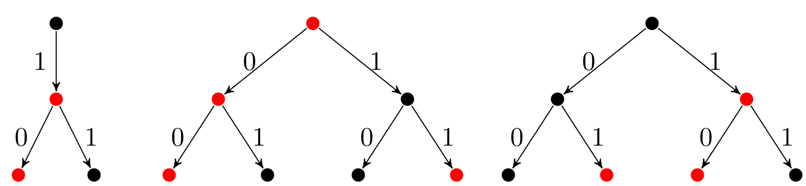

FIGURE 10. The 3 factors of height 2 in $T(L_2)$ decorated by the Thue–Morse sequence. The first one is the prefix occurring only once.

Since every factor of height h is the prefix of a factor of height $h+1$, we trivially have $\#F_{h+1} \geq \#F_h$. This is quite similar to factors occurring in an infinite word: any factor has at least one extension. In particular, ultimately periodic words are characterized by a bounded factor complexity.

Lemma 26. [3, Proposition 1] *Let L be a prefix-closed language over $(A,<)$ and let $\mathbf{x} = x_0x_1\cdots$ be an infinite word over some finite alphabet B. Consider the labeled tree $T(L)$ decorated by $\mathbf{x}$. The tree $T(L)$ is rational if and only if $\#F_h = \#F_{h+1}$ for some $h \geq 0$. In particular, $\#F_h = \#F_{h+n}$ for all $n \geq 0$.*

We can characterize $\mathcal{S}$-automatic sequences built on a prefix-closed regular language L in terms of the decorated tree $T(L)$. For the sake of presentation, we mainly focus on the case of k-automatic sequences. The reader can relate our construction to the k-kernel of a sequence. Roughly, each element of the k-kernel corresponds to reading one fixed suffix u from each node w of the tree $T(L_k)$. We have $\mathrm{val}_k(wu) = k^{|u|}\,\mathrm{val}_k(w) + \mathrm{val}_k(u)$ and an element from the k-kernel is a sequence of the form $(x_{k^{|u|}n+\mathrm{val}_k(u)})_{n\geq 0}$.

Theorem 27. *Let $k \geq 2$ be an integer. A sequence $\mathbf{x}$ is k-automatic if and only if the labeled tree $T(L_k)$ decorated by $\mathbf{x}$ is rational.*

Proof. Let us prove the forward direction. If $\mathbf{x}$ is k-automatic, there exists a DFAO $\mathcal{A} = (Q, q_0, A_k, \delta, \tau)$ producing it when fed with base-k representations of integers. Let $w \in L_k$ be a non-empty base-k representation and let $h \geq 1$ be an integer. The factor $T[w, h]$ is completely determined by the state $\delta(q_0, w)$. Indeed, it is a full k-ary tree of height h and the decorations are given by $\tau(\delta(q_0, wu))$ for u running through $A_k^{\leq h}$ in radix order. For the empty word, however, the prefix $T[\varepsilon, h]$ is decorated by $\tau(\delta(q_0, u))$ for u running through $\{\varepsilon\} \cup \{1, \dots, k-1\}A_k^{<h}$. Hence $\#F_h$ is bounded by $\#Q + 1$, for all $h \geq 0$. Since $h \mapsto \#F_h$ is non-decreasing, there exists $H \geq 0$ such that $\#F_H = \#F_{H+1}$. We conclude by using Lemma 26.

Let us prove the other direction. Assume that the tree $T(L_k)$ is rational. In particular, there exists an integer $h \geq 1$ such that $\#F_h = \#F_{h+1}$. This means that any factor of height h can be extended in a unique way to a factor of height $h+1$, i.e., if $T[w, h] = T[w', h]$ for two words $w, w' \in L_k$, then $T[w, h+1] = T[w', h+1]$. This factor of height $h+1$ is made of a root and k subtrees of height h attached to it. So, for each copy of $T[w, h]$ in the tree $T(L_k)$, to its root are attached the same k trees $T[w0, h], \dots, T[w(k-1), h]$. The same observation holds for the prefix of the tree except that to the root are attached the $k-1$ trees $T[1, h], \dots, T[k-1, h]$. We thus define a DFAO $\mathcal{F}$ whose set of states is F_h and whose transition function is given by

$$\forall i \in A_k : \delta(T[w, h], i) = T[wi, h].$$

The initial state is given by the prefix $T[\varepsilon, h]$ and we set $\delta(T[\varepsilon, h], 0) = T[\varepsilon, h]$. Finally the output function maps a factor $T[w, h]$ to the decoration of its root w, that is, $x_{\mathrm{val}_k(w)}$. For each $n \geq 0$, x_n is the decoration of the nth node in $T(L_k)$ by definition. To conclude the proof of the backward direction, we have to show that x_n is the output of $\mathcal{F}$ when fed with $\mathrm{rep}_k(n)$. This follows from the definition of $\mathcal{F}$: starting from the initial state $T[\varepsilon, h]$, we reach the state $T[\mathrm{rep}_k(n), h]$ and the output is $x_{\mathrm{val}_k(\mathrm{rep}_k(n))} = x_n$. □

We improve the previous result to ANS with a regular numeration language.

Theorem 28. *Let $\mathcal{S} = (L, A, <)$ be an ANS built on a prefix-closed regular language L. A sequence $\mathbf{x}$ is $\mathcal{S}$-automatic if and only if the labeled tree $T(L)$ decorated by $\mathbf{x}$ is rational.*

Proof. The proof follows exactly the same lines as for integer base numeration systems. The only refinement is the following one. A factor $T[w, h]$ of $T(L)$ is determined by $w^{-1}L \cap A^{\leq h}$ and $\delta(q_0, w)$. Since L is regular, the set $\{w^{-1}L \cap A^{\leq h} \mid w \in A^*\}$ is finite. Thus $\#F_h$ is bounded by $\#Q$ times the number of states of the minimal automaton of L. □

5.1. **Rational bases.** We now turn to rational base numeration systems. A factor of height h in $T(L_{\frac{3}{2}})$ only depends on the value of its root modulo 2^h. This result holds for any rational base numeration system.

Lemma 29. [16, Lemme 4.14] *Let $w, w' \in L_{\frac{p}{q}}$ be non-empty words and let $u \in A_p^*$ be a word of length h.*

- *If $\mathrm{val}_{\frac{p}{q}}(w) \equiv \mathrm{val}_{\frac{p}{q}}(w') \bmod q^h$, then $u \in w^{-1}L_{\frac{p}{q}}$ if and only if $u \in (w')^{-1}L_{\frac{p}{q}}$.*
- *If $u \in (w^{-1}L_{\frac{p}{q}} \cap (w')^{-1}L_{\frac{p}{q}})$, then $\mathrm{val}_{\frac{p}{q}}(w) \equiv \mathrm{val}_{\frac{p}{q}}(w') \bmod q^h$.*

In the previous lemma, the empty word behaves differently. For a non-empty word $w \in L_{\frac{p}{q}}$ with $\mathrm{val}_{\frac{p}{q}}(w) \equiv 0 \bmod q^h$, a word $u \in A_p^h$ not starting with 0 verifies $u \in \varepsilon^{-1}L_{\frac{p}{q}}$ if and only if $u \in w^{-1}L_{\frac{p}{q}}$. Therefore the prefix of the tree $T(L_{\frac{p}{q}})$ has to be treated separately.

Lemma 30. [16, Corollaire 4.17] *Every word $u \in A_p^*$ is suffix of a word in $L_{\frac{p}{q}}$.*

As a consequence of these lemmas $\{w^{-1}L_{\frac{p}{q}} \cap A_p^h \mid w \in A_p^+\}$ is a partition of A_p^h into q^h non-empty languages. Otherwise stated, in the tree $T(L_{\frac{p}{q}})$ with no decoration or, equivalently with a constant decoration for all nodes, there are q^h+1 factors of height $h \geq 1$ (we add 1 to count the height-h prefix, which has a different shape). For instance, if the decorations in Figure 9 are not taken into account, there are $5 = 2^2+1$ height-2 factors occurring in $T(L_{\frac{3}{2}})$.

Except for the height-h prefix, each factor of height h is extended in exactly q ways to a factor of height $h+1$. To the first (leftmost) leaf of a factor of height h are attached children corresponding to one of the q words of the periodic signature. To the next leaves on the same level are periodically attached as many nodes as the length of the different words of the signature. For instance, in the case $\frac{p}{q} = \frac{3}{2}$, the first (leftmost) leaf of a factor of height h becomes a node of degree either 1 (label 1) or 2 (labels 0 and 2) to get a factor of height $h+1$. The next leaves on the same level periodically become nodes of degree 2 or 1 accordingly. An example is depicted in Figure 11.

Lemma 31. *Let $\mathbf{x}$ be a $\frac{p}{q}$-automatic sequence produced by the DFAO $\mathcal{A} = (Q, q_0, A_p, \delta, \tau)$ and let $T(L_{\frac{p}{q}})$ be decorated by $\mathbf{x}$. For all $h \geq 1$, the number $\#F_h$ of height-h factors of $T(L_{\frac{p}{q}})$ is bounded by $1 + q^h \cdot \#Q$.*

Proof. Let $w \in L_{\frac{p}{q}}$ be a non-empty base-$\frac{p}{q}$ representation and let $h \geq 1$. We claim that the factor $T[w,h]$ is completely determined by the word w. First, from Lemma 29, the labeled tree $T[w,h]$ of height h with root w and in particular, its domain, only depends on $\mathrm{val}_{\frac{p}{q}}(w)$ modulo q^h. Indeed, if $w, w' \in L_{\frac{p}{q}}$ are such that $\mathrm{val}_{\frac{p}{q}}(w) \equiv \mathrm{val}_{\frac{p}{q}}(w') \bmod q^h$, then

$$\mathrm{dom}(T[w,h]) = w^{-1}L_{\frac{p}{q}} \cap A_p^{\leq h} = w'^{-1}L_{\frac{p}{q}} \cap A_p^{\leq h} = \mathrm{dom}(T[w',h]).$$

Second, the decorations of the factor $T[w,h]$ are given by $\tau(\delta(q_0, wu))$ for u running through $\mathrm{dom}(T[w,h]) = w^{-1}L_{\frac{p}{q}} \cap A_p^{\leq h}$ enumerated in radix order. So the decorations only depend on the state $\delta(q_0, w)$ of $\mathcal{A}$. Hence the number of such factors is bounded by $q^h \cdot \#Q$.

Similarly, the height-h prefix $T[\varepsilon, h]$ is decorated by $\tau(\delta(q_0, u))$ for u running through $\mathrm{dom}(T[\varepsilon, h]) = L_{\frac{p}{q}} \cap A_p^{\leq h}$.

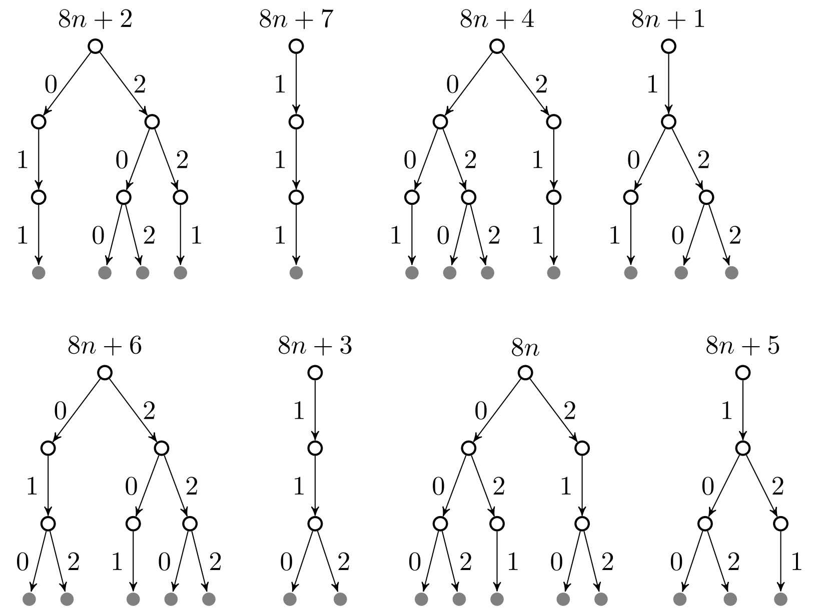

FIGURE 11. For the rational base $\frac{3}{2}$, each factor of height $h = 2$ gives 2 factors of height $h + 1 = 3$.

Hence $\#F_h$ is bounded by $1 + q^h \cdot \#Q$, for all $h \geq 1$. □

Definition 32. A tree of height $h \geq 0$ has nodes on $h + 1$ levels: the *level* of a node is its distance to the root. Hence, the root is the only node of level 0 and the leaves have level h.

For instance, in Figure 11, each tree of height 3 has four levels.

Definition 33. Let T be a labeled decorated tree and let $h \geq 0$. We let $F_h^\infty \subseteq F_h$ denote the set of factors of height h occurring infinitely often in T. For any suitable letter a in the signature of T, we let $F_{h,a}^\infty \subseteq F_h^\infty$ denote the set of factors of height h occurring infinitely often in T such that the label of the edge between the first node on level $h-1$ and its first child is a. Otherwise stated, the first word of length h in the domain of the factor ends with a.

Example 34. In Figure 11, assuming that they occur infinitely often, the first four trees belong to $F_{3,1}^\infty$ and the last four on the second row belong to $F_{3,0}^\infty$.

Even though the language $L_{\frac{p}{q}}$ is highly non-regular, we can still handle a subset of $\frac{p}{q}$-automatic sequences. Roughly, with the next two theorems, we characterize $\frac{p}{q}$-automatic sequences in terms of the number of factors of a fixed height occurring infinitely often. As mentioned below, the first result can be notably applied when distinct states of the DFAO producing the sequence have distinct outputs.

In the remaining of the section, we let $(w_0, \dots, w_{q-1})$ denote the signature of $T(L_{\frac{p}{q}})$. For all $0 \leq j \leq q-1$ and all $0 \leq i \leq |w_j| - 1$, we also let $w_{j,i}$ denote the ith letter of w_j.

Theorem 35. *Let* $\mathbf{x}$ *be a* $\frac{p}{q}$*-automatic sequence over a finite alphabet* B *generated by a DFAO* $\mathcal{A} = (Q, q_0, A_p, \delta, \tau : A_p \to B)$ *with the following property: there exists* *an integer h such that, for all distinct states $q, q' \in Q$ and all words $w \in L_{\frac{p}{q}}$, there exists a word u in $w^{-1}L_{\frac{p}{q}}$ of length at most h such that $\tau(\delta(q,u)) \neq \tau(\delta(q',u))$. Then in the tree $T(L_{\frac{p}{q}})$ decorated by $\mathbf{x}$, we have for all $0 \le j \le q-1$,*

$$\#F^{\infty}_{h+1,w_{j,0}} \le \#F^{\infty}_{h}.$$

Proof. Consider a factor of height h occurring infinitely often, i.e., there is a sequence $(u_i)_{i\ge 1}$ of words in $L_{\frac{p}{q}}$ such that $T[u_1,h] = T[u_2,h] = T[u_3,h] = \cdots$. From Lemma 29, all values $\mathrm{val}_{\frac{p}{q}}(u_i)$ are congruent to r modulo q^h for some $0 \le r < q^h$. Thus the values of $\mathrm{val}_{\frac{p}{q}}(u_i)$ modulo to q^{h+1} that appear infinitely often take at most q values (among $r, r+q^h, \ldots, r+(q-1)q^h$).

The extra assumption on the DFAO means that if two words $v, w \in L_{\frac{p}{q}}$ with $\mathrm{val}_{\frac{p}{q}}(v) \equiv \mathrm{val}_{\frac{p}{q}}(w) \mod q^h$ are such that $\delta(q_0,v) \neq \delta(q_0,w)$, then $T[v,h] \neq T[w,h]$. Indeed, by assumption, there exists $u \in v^{-1}L_{\frac{p}{q}} \cap A_p^{\le h} = w^{-1}L_{\frac{p}{q}} \cap A_p^{\le h}$ such that $\tau(\delta(q_0,vu)) \neq \tau(\delta(q_0,wu))$. Hence, by contraposition, since $T[u_i,h] = T[u_j,h]$, then $\delta(q_0,u_i) = \delta(q_0,u_j)$. Consequently, if $T[u_i,h+1]$ and $T[u_j,h+1]$ have the same domain, then $T[u_i,h+1] = T[u_j,h+1]$ because $\delta(q_0,u_iv) = \delta(q_0,u_jv)$ for all words $v \in \mathrm{dom}(T[u_i,h+1])$.

Consequently, no two distinct factors of height $h+1$ occurring infinitely often and having the same domain can have the same prefix of height h. Therefore, each factor U of height h occurring infinitely often gives rise to at most one factor U' of height $h+1$ in every $\#F^{\infty}_{h+1,w_{j,0}}$ for $0 \le j \le q-1$ (U and the first letter $w_{j,0}$ uniquely determine the domain of U'). □

Remark 36. In the case of a k-automatic sequence, the assumption of the above theorem is always satisfied. We may apply the usual minimization algorithm about undistinguishable states to the DFAO producing the sequence: two states q, q' are *distinguishable* if there exists a word u such that $\tau(\delta(q,u)) \neq \tau(\delta(q',u))$. The pairs $\{q,q'\}$ such that $\tau(q) \neq \tau(q')$ are distinguishable (by the empty word). Then proceed recursively: if a not yet distinguished pair $\{q,q'\}$ is such that $\delta(q,a) = p$ and $\delta(q',a) = p'$ for some letter a and an already distinguished pair $\{p,p'\}$, then $\{q,q'\}$ is distinguished. The process stops when no new pair is distinguished and we can merge states that belong to undistinguished pairs. In the resulting DFAO, any two states are distinguished by a word whose length is bounded by the number of states of the DFAO. We can thus apply the above theorem. Notice that for a k-automatic sequence, there is no restriction on the word distinguishing states since it belongs to A_k^*. The extra requirement that $w \in L_{\frac{p}{q}}$ is therefore important in the case of rational bases and is not present for base-k numeration systems.

Remark 37. For a rational base numeration system, the assumption of the above theorem is always satisfied if the output function τ is the identity; otherwise stated, if the output function maps distinct states to distinct values. This is for instance the case of our toy example $\mathbf{t}$. However the assumption is not readily satisfied on examples such as the following one with the DFAO depicted in Figure 12 reading base-$\frac{3}{2}$ representations.

For instance the words $u = 212001220110220$ and $v = 212022000012021$ are such that $q_0.u = q_1$, $q_0.v = q_0$, $u^{-1}L_{\frac{3}{2}} \cap A_3^4 = v^{-1}L_{\frac{3}{2}} \cap A_3^4 = \{1111\}$ and $u^{-1}L_{\frac{3}{2}} \cap A_3^5 = v^{-1}L_{\frac{3}{2}} \cap A_3^5 = \{11110, 11112\}$. So $T[u,4] = T[v,4]$ because reading 1's from q_0 or q_1

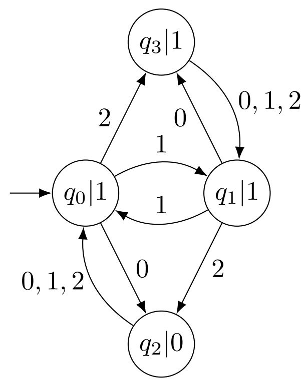

FIGURE 12. A DFAO with two distinct outputs but four states.

leads to one of these two states with the same output. But $T[u,5] \neq T[v,5]$ because $q_0.u1^40 = q_1.0 = q_3$ and $q_0.v1^40 = q_0.0 = q_2$, and the corresponding outputs are different.

We can generalize the above example with the suffix 1^4. Let $h \geq 1$ and consider the word 1^h. From Lemma 30, it occurs as a suffix of words in $L_{\frac{3}{2}}$. One may thus find words similar to u and v in the above computations. Actually, $\mathrm{val}_{\frac{3}{2}}(u) = 591$ and $\mathrm{val}_{\frac{3}{2}}(v) = 623$ are both congruent to $15 = 2^4 - 1$ modulo 2^4 (so, they can be followed by the suffix 1^4), and $\mathrm{val}_{\frac{3}{2}}(u1^4)$ and $\mathrm{val}_{\frac{3}{2}}(v1^4)$ are both even (so, they can be followed by either 0 or 2). To have a situation similar to the one with u and v above, we have to look for numbers n which are congruent to $2^h - 1$ modulo 2^h and such that

$$n\left(\frac{3}{2}\right)^h + \mathrm{val}_{\frac{3}{2}}(1^h) = n\left(\frac{3}{2}\right)^h + \left(\frac{3}{2}\right)^h - 1$$

is an even integer. Numbers of the form $n = (2j+1)2^h - 1$ are convenient. We moreover have to ensure that reading the representation of n ends either in q_0 or q_1.

Theorem 38. *Let* $\mathbf{x}$ *be a sequence over a finite alphabet* B*, and let the tree* $T(L_{\frac{p}{q}})$ *be decorated by* $\mathbf{x}$*. If there exists some* $h \geq 0$ *such that* $\#F^\infty_{h+1,w_{j,0}} \leq \#F^\infty_h$ *for all* $0 \leq j \leq q-1$*, then* $\mathbf{x}$ *is* $\frac{p}{q}$*-automatic.*

Proof. For the sake of readability, write $T = T(L_{\frac{p}{q}})$. The length-h factors of T occurring only a finite number of times appear in a prefix of the tree. Let $t \geq 0$ be the least integer such that all nodes at any level $\ell \geq t$ are roots of a factor in F^∞_h.

We first define a NFA $\mathcal{T}$ in the following way. An illustration that we hope to be helpful is given below in Example 39. It is made (nodes and edges) of the prefix $T[\varepsilon, t+h-1]$ of height $t+h-1$ and a copy of every element in F^∞_h. So the set of states is the union of the nodes of the prefix $T[\varepsilon, t+h-1]$ and the nodes in the trees of F^∞_h. Final states are all the nodes of the prefix $T[\varepsilon, t+h-1]$ and the nodes of level exactly h in every element of F^∞_h, i.e., the leaves of every element of F^∞_h. The unique initial state is the root of the prefix $T[\varepsilon, t+h-1]$. We define the following extra transitions between these elements.

- If a node m of level $t-1$ in the prefix $T[\varepsilon, t+h-1]$ has a child n reached through an arc with label d, then in the NFA we add an extra transition with the same label d from m to the root of the element of F_h^∞ equal to $T[n,h]$. This is well defined because n has level t.
- Let r be the root of an element $T[r,h]$ of F_h^∞. Suppose that r has a child s reached through an arc with label d. The assumption in the statement means that the element $T[r,h]$ in F_h^∞ can be extended in at most one way to an element U_c in $F_{h+1,c}^\infty$ for some $c \in \{w_{0,0}, \dots, w_{q-1,0}\}$. The tree U_c with root r has a subtree of height h with root $rd = s$ denoted by $V_{c,d} \in F_h^\infty$ (as depicted in Figure 13). In the NFA, we add extra transitions with label d from r to the root of $V_{c,d}$ (there are at most q such trees).

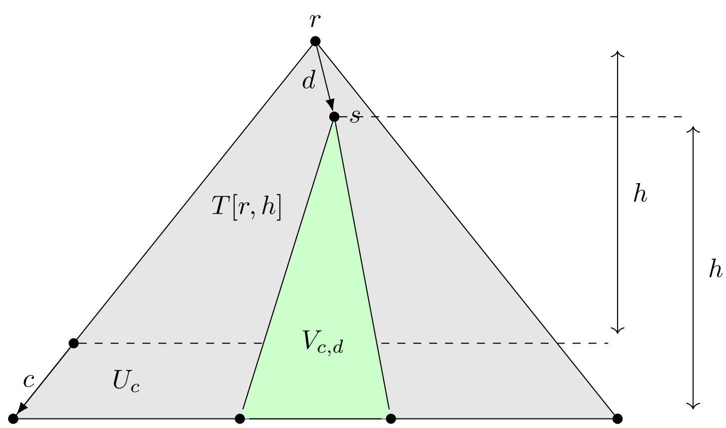

FIGURE 13. Extension of a tree in F_h^∞.

We will make use of the following *unambiguity property* of $\mathcal{T}$. Every word $u \in L_{\frac{p}{q}}$ is accepted by $\mathcal{T}$ and there is exactly one successful run for u in $\mathcal{T}$. If the length of $u \in L_{\frac{p}{q}}$ is less than $t+h$, there is one successful run and it remains in the prefix $T[\varepsilon, t+h-1]$. If a run uses a transition between a node of level $t-1$ in the prefix $T[\varepsilon, t+h-1]$ and the root of an element in F_h^∞, then the word has to be of length at least $t+h$ to reach a final state by construction. Now consider a word $u \in L_{\frac{p}{q}}$ of length $t+h+j$ with $j \geq 0$ and write

$$u = u_0 \cdots u_{t-1} u_t u_{t+1} \cdots u_{t+h-1} \cdots u_{t+h+j-1}.$$

Reading the prefix $u_0 \cdots u_{t-1}$ leads to the root of an element U in F_h^∞. Assume that this element can be extended in (at least) two ways to a tree of height $h+1$. This means that in $\mathcal{T}$, we have two transitions from the root of U with label u_{t-1}: one going to the root of some $V_1 \in F_{h,c_1}^\infty$ and one going to the root of some $V_2 \in F_{h,c_2}^\infty$. Note that V_1 and V_2 have the same prefix of height $h-1$. The difference appears precisely at level h where the labeling is periodically $(w_e, w_{e+1}, \dots, w_q, w_1, \dots, w_{e-1})$ and $(w_f, w_{f+1}, \dots, \dots, w_q, w_1, \dots, w_{f-1})$ respectively where w_e (respectively w_f) starts with c_1 (respectively c_2) and the two q-tuples of words are a cycle shift of the signature $(w_0, \dots, w_{q-1})$ of T. Nevertheless, if x has length $h-1$ and belongs to the domain of V_1 and thus of V_2, then xc_1 belongs to the domain of V_1 if and only if xc_2 belongs to the domain of V_2. So if we non-deterministically make the wrong choice of transition at step t, we will not be able to process the letter at position $t+h$. The choice of a transition determines the words of length h that can be read from that point on. The same reasoning occurs for the decision taken at step $t+j$ and the letter at position $t+h+j$.

We still have to turn $\mathcal{T}$ into a DFAO producing $\mathbf{x} \in B^{\mathbb{N}}$. To do so, we determinize $\mathcal{T}$ with the classical subset construction. Thanks to the unambiguity property of $\mathcal{T}$, if a subset of states obtained during the construction contains final states of $\mathcal{T}$, then they are all decorated by the same letter $b \in B$. The output of this state is thus set to b. If a subset of states obtained during the construction contains no final state, then its output is irrelevant (it can be set to any value). ☐

Example 39. Consider the rational base $\frac{3}{2}$. Our aim is to illustrate the above theorem: we have information about factors of a decorated tree $T(L_{\frac{3}{2}})$ — those occurring infinitely often and those occurring only a finite number of times — and we want to build the corresponding $\frac{3}{2}$-automatic sequence. Assume that $t = h = 1$ and that factors of length 1 can be extended as in Figure 9. We assume that the last eight trees of height 2 occur infinitely often. Hence their four prefixes of height 1 have exactly two extensions. We assume that the prefix given by the first tree in Figure 9 occurs only once.

From this, we build the NFA $\mathcal{T}$ depicted in Figure 14. The prefix tree of height $t+h-1 = 1$ is depicted on the left and its root is the initial state. The single word 2 of length 1 is accepted by a run staying in this tree. Then, are represented the four trees of F_1^∞. Their respective leaves are final states. Finally, we have to inspect Figure 9 to determine the transitions connecting roots of these trees. For instance, let us focus on state 7 in Figure 14. On Figure 9, the corresponding tree can be extended in two ways: the second and the fourth trees on the first row. In the first of these trees, the tree hanging to the child 0 (respectively 2) of the root corresponds to state 5 (respectively 7). Hence, there is a transition of label 0 (respectively 2) from 7 to 5 (respectively 7) in Figure 14. Similarly, the second tree gives the extra transitions of label 0 from 7 to 7 and of label 2 from 7 to 5.

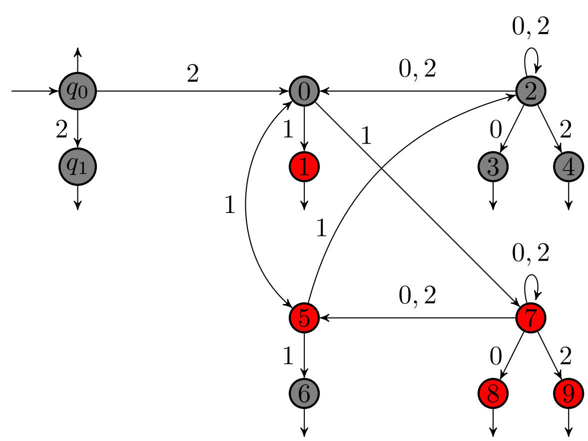

FIGURE 14. A NFA $\mathcal{T}$.

Take the word $210 \in L_{\frac{3}{2}}$. Starting from q_0, the only successful run is $q_0 \xrightarrow{2} 0 \xrightarrow{1} 7 \xrightarrow{0} 8$. If we had reached 0 with $q_0 \xrightarrow{2} 0$ and chose the other transition of label 1, we would have the run $q_0 \xrightarrow{2} 0 \xrightarrow{1} 5$, but from state 5 there is no transition with label 0. The successful runs of the first few words in $L_{\frac{3}{2}}$ are given below:

ε	q_0
2	$q_0 \to q_1$
21	$q_0 \to 0 \to 1$
210	$q_0 \to 0 \to 7 \to 8$
212	$q_0 \to 0 \to 7 \to 9$
2101	$q_0 \to 0 \to 7 \to 5 \to 6$
2120	$q_0 \to 0 \to 7 \to 7 \to 8$
2122	$q_0 \to 0 \to 7 \to 7 \to 9$
21011	$q_0 \to 0 \to 7 \to 5 \to 0 \to 1$
21200	$q_0 \to 0 \to 7 \to 7 \to 7 \to 8$
21202	$q_0 \to 0 \to 7 \to 7 \to 7 \to 9$
21221	$q_0 \to 0 \to 7 \to 7 \to 5 \to 6$

We may now determinize this NFA $\mathcal{T}$. We apply the classical subset construction to get a DFAO. If a subset of states contains a final state of $\mathcal{T}$ from $\{1, 8, 9\}$ (respectively $\{q_0, q_1, 3, 4, 6\}$), the corresponding decoration being 1 (respectively 0), the output for this state is 1 (respectively 0). Indeed, as explained in the proof, a subset of states of $\mathcal{T}$ obtained during the determinization algorithm cannot contain states with two distinct decorations. After determinization, we obtain the (minimal) DFAO depicted in Figure 15. In the latter figure, we have not set any output for state 2 because it corresponds to a subset of states in $\mathcal{T}$ which does not contain any final state. Otherwise stated, that particular output is irrelevant as no valid representation will end up in that state.

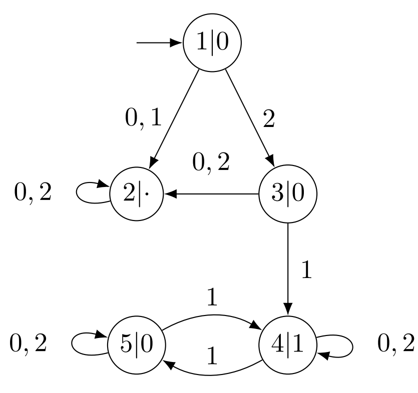

Figure 15. Determinization of $\mathcal{T}$.

6. Recognizable sets and stability properties

In this short section, our aim is to present some direct closure properties of automatic sequences in ANS built on tree languages. These statements should not surprise the reader used to constructions of automata and automatic sequences.

In [15], a subset X of $N_{\frac{p}{q}}$ is said to be $\frac{p}{q}$*-recognizable* if there exists a DFA over A_p accepting a language L such that $\mathrm{val}_{\frac{p}{q}}(L) = X$. Since $L_{\frac{p}{q}}$ is not regular, the set $\mathbb{N}$ is not $\frac{p}{q}$-recognizable.

Proposition 40. *A sequence $\mathbf{x} = x_0x_1 \cdots$ over A is $\frac{p}{q}$-automatic if and only if, for every $a \in A$, there exists a $\frac{p}{q}$-recognizable set R_a such that $\{i \in \mathbb{N} : x_i = a\} = R_a \cap \mathbb{N}$.*

Proof. In the DFAO producing the sequence, consider as final the states having output a. The accepted set is R_a. □

For k-automatic sequences, the above result can also be expressed in terms of fibers (see, for instance, [2, Lemma 5.2.6]). The $\frac{p}{q}$*-fiber* of an infinite sequence $\mathbf{x}$ is the language $I_{\frac{p}{q}}(\mathbf{x}, a) = \{\mathrm{rep}_{\frac{p}{q}}(i) : i \in \mathbb{N} \text{ and } x_i = a\}$. A sequence $\mathbf{x} = x_0x_1 \cdots$ over A is $\frac{p}{q}$-automatic if and only if, for every $a \in A$, there exists a regular language S_a such that $I_{\frac{p}{q}}(\mathbf{x}, a) = S_a \cap L_{\frac{p}{q}}$.

We can verbatim take several robustness or closure properties of automatic sequences. They use classical constructions of automata such as reversal or compositions.

Proposition 41. *Let $\mathcal{S}$ be an abstract numeration system built on a tree language with a purely periodic labeled signature. The set of $\mathcal{S}$-automatic sequences is stable under finite modifications.*

Proof. One has to adapt the DFAO to take into account those finite modifications. Suppose that these modifications occur for representations of length at most ℓ. Then the DFAO can have a tree-like structure for words of length up to ℓ and we enter the original DFAO after passing through this structure encoding the modifications. □

Proposition 42. *Let $\mathcal{S}$ be an abstract numeration system built on a tree language with a purely periodic labeled signature. The set of $\mathcal{S}$-automatic sequences is stable under codings.*

Automatic sequences can be produced by reading least significant digits first. Simply adapt the corresponding result in [22].

Proposition 43. *Let $\mathcal{S} = (L, A, <)$ be an abstract numeration system built on a tree language with a purely periodic labeled signature. A sequence $\mathbf{x}$ is $\mathcal{S}$-automatic if and only if there exists a DFAO $(Q, q_0, A, \delta, \tau)$ such that, for all $n \geq 0$, $x_n = \tau(\delta(q_0, (\mathrm{rep}_{\mathcal{S}}(n))^R))$.*

Adding leading zeroes does not affect automaticity. Simply adapt the proof of [2, Theorem 5.2.1].

Proposition 44. *A sequence $\mathbf{x}$ is $\frac{p}{q}$-automatic if and only if there exists a DFAO $(Q, q_0, A_p, \delta, \tau)$ such that, for all $n \geq 0$ and all $j \geq 0$, $x_n = \tau(\delta(q_0, 0^j\, \mathrm{rep}_{\frac{p}{q}}(n)))$.*

For any finite alphabet $D \subset \mathbb{Z}$ of digits, we let χ_D denote the *digit-conversion* map defined as follows: for all $u \in D^*$ such that $\mathrm{val}_{\frac{p}{q}}(u) \in \mathbb{N}$, $\chi_D(u)$ is the unique word $v \in L_{\frac{p}{q}}$ such that $\mathrm{val}_{\frac{p}{q}}(u) = \mathrm{val}_{\frac{p}{q}}(v)$. In [1], it is shown that χ_D can be realized by a finite letter-to-letter right transducer. As a consequence of this result, multiplication by a constant $a \geq 1$ is realized by a finite letter-to-letter right transducer. Indeed take a word $u = u_0 \cdots u_t \in L_{\frac{p}{q}}$ and consider the alphabet

$D = \{0, a, 2a, \ldots, (p-1)a\}$. Feed the transducer realizing χ_D with $au_t, \ldots, au_0$. The output is the base-$\frac{p}{q}$ representation of $a \cdot \mathrm{val}_{\frac{p}{q}}(u)$. Similarly, translation by a constant $b \geq 0$ is realized by a finite letter-to-letter right transducer. Consider the alphabet $D' = \{0, \ldots, p+b-1\}$. Feed the transducer realizing $\chi_{D'}$ with $(u_t + b)$, $u_{t-1}, \ldots, u_0$. The output is the base-$\frac{p}{q}$ representation of $\mathrm{val}_{\frac{p}{q}}(u) + b$. Combining these results with the DFAO producing a $\frac{p}{q}$-automatic sequence, we get the following result.

Corollary 45. *Let $a \geq 1, b \geq 0$ be integers. If a sequence $\mathbf{x}$ is $\frac{p}{q}$-automatic, then the sequence $(x_{an+b})_{n\geq 0}$ is also $\frac{p}{q}$-automatic.*

Remark 46. Ultimately periodic sequences are k-automatic for any integer $k \geq 2$ [2, Theorem 5.4.2]. They are also $\mathcal{S}$-automatic for any abstract numeration system $\mathcal{S}$ based on a regular language [12]. In general, this is not the case for $\frac{p}{q}$-automaticity: the characteristic sequence of multiples of q is not $\frac{p}{q}$-automatic [15, Proposition 5.39]. Nevertheless when the period length of an ultimately periodic sequence is coprime with q, then the sequence is $\frac{p}{q}$-automatic [15, Théorème 5.34].

Acknowledgment

Manon Stipulanti is supported by the FNRS Research grant 1.B.397.20.

References

[1] S. Akiyama, Ch. Frougny, and J. Sakarovitch, Powers of rationals modulo 1 and rational base number systems, *Israel J. Math.* **168** (2008), 53–91.

[2] J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, (2003).

[3] J. Berstel, L. Boasson, O. Carton, and I. Fagnot, Sturmian Trees, *Theoret. Comput. Sci.* **46** (2010), 443–478.

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

[5] É. Charlier, M. Le Gonidec, and M. Rigo, Representing real numbers in a generalized numeration system, *J. Comput. System Sci.* **77** (2011), no. 4, 743–759.

[6] A. Cobham, Uniform tag sequences, *Math. Systems Theory* **6** (1972), 164–192.

[7] K. Culik, J. Karhumäki, and A. Lepistö, Alternating iteration of morphisms and the Kolakovski sequence, in *Lindenmayer systems*, 93–106, Springer, Berlin, (1992).

[8] F. M.Dekking, Regularity and irregularity of sequences generated by automata, *Sém. Th. Nombres Bordeaux* 79–80 (1980), 901–910.

[9] J.-M. Dumont and A. Thomas, Systèmes de numération et fonctions fractales relatifs aux substitutions, *Theoret. Comput. Sci.* **65** (1989), 153–169.

[10] T. Edgar, H. Olafson, and J. Van Alstine, Some combinatorics of rational base representations, preprint.

[11] J. Endrullis and D. Hendriks, On periodically iterated morphisms, *Proc. CSL-LICS'14* in Vienna (2014), 1–10.

[12] P. Lecomte and M. Rigo, Abstract numeration systems, Ch. 3, in *Combinatorics, Automata and Number Theory*, Encyclopedia Math. Appl. **135**, Cambridge University Press, (2010).

[13] A. Lepistö, On the power of periodic iteration of morphisms, ICALP 1993, 496–506, *Lect. Notes Comp. Sci* **700**, (1993).

[14] J. Peltomäki, A. Massuir, and M. Rigo, Automatic sequences based on Parry or Bertrand numeration systems, *Adv. Appl. Math.* **108** (2019), 11–30.

[15] V. Marsault, On $\frac{p}{q}$-recognisable sets, `arXiv:1801.08707`.

[16] V. Marsault, *Énumération et numération*, Ph.D. thesis, Télecom-Paristech, 2015.

[17] V. Marsault and J. Sakarovitch, On sets of numbers rationally represented in a rational base number system. Algebraic informatics, *Lect. Notes Comp. Sci.* **8080**, 89–100, Springer, Heidelberg, 2013.

[18] V. Marsault and J. Sakarovitch, Breadth-first serialisation of treesand rational languages, Developments in Language Theory - 18th International Conference, 2014, Ekaterinburg, Russia, August 26-29, 2014, *Lect. Notes Comp. Sci.* **8633**, 252–259.

[19] V. Marsault and J. Sakarovitch, Trees and languages with periodic signature, *Indagationes Mathematicae* **28** (2017), 221–246.

[20] V. Marsault and J. Sakarovitch, The signature of rational languages, *Theor. Comput. Sci.* **658** (2017), 216–234.

[21] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés. Automata, languages and programming (Antwerp, 1984), 380–389, *Lect. Notes Comp. Sci.* **172**, Springer, Berlin, (1984).

[22] M. Rigo and A. Maes, More on generalized automatic sequences, *J. Autom. Lang. Comb.* **7** (2002), 351–376.

[23] M. Rigo, *Formal Languages, Automata and Numeration Systems*, ISTE–Wiley, (2014).

[24] N. Sloane et al., The On-Line Encyclopedia of Integer Sequences, `http://oeis.org`.

Department of Mathematics, University of Liège, Allée de la Découverte 12, 4000 Liège, Belgium, {`m.rigo,m.stipulanti`}`@uliege.be`