Some improvements of the S-adic conjecture

Julien Leroy

Université de Picardie Jules Verne,
Laboratoire Amiénois de Mathématiques Fondamentales et Appliquées,
CNRS-UMR 6140, 33 rue Saint Leu, 80039 Amiens Cedex, France.

Abstract

In [Ergodic Theory Dynam. System, 16 (1996) 663–682], S. Ferenczi proved that the language of any uniformly recurrent sequence with an at most linear complexity is S-adic. In this paper we adapt his proof in a more structured way and improve this result stating that any such sequence is itself S-adic. We also give some properties on the constructed morphisms.

Keywords: S-adic, Rauzy graph, factor complexity, special factor

1. Introduction

A usual tool in the study of sequences (or infinite words) over a finite alphabet A is the complexity function p that counts the number of factors of each length n occurring in the sequence. This function is clearly bounded by $d^n, n \in \mathbb{N}$, where d is the number of letters in A but not all functions bounded by d^n are complexity functions. As an example, it is well known (see [19]) that either the sequence is ultimately periodic (and then $p(n)$ is ultimately constant), or its complexity function grows at least like $n + 1$. Non-periodic sequences with minimal complexity $p(n) = n+1$ for all n exist and are called *Sturmian sequences* (see [19]). These words are binary sequences (because $p(1) = 2$) and admit several equivalent definitions: aperiodic balanced sequences, codings of rotations, mechanical words of irrational slope,... See Chapter 2 of [18] and Chapter 6 of [17] for surveys on these sequences. In particular, it is well-known that all these sequences can be generated with only three morphisms.

Many other known sequences have a low complexity. By "low complexity" we usually mean "complexity bounded by a linear or affine function". Fixed points of primitive substitutions, automatic sequences, linearly recurrent sequences (see [12]) and Arnoux-Rauzy sequences are examples of sequences with an at most affine complexity. For any such sequence $\mathbf{w}$, there exists a finite set S of morphisms over an alphabet A, a letter a and a sequence $(\sigma_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}$ such that $\mathbf{w} = \lim_{n\to\infty} \sigma_0 \cdots \sigma_n(a^\omega)$. Indeed, automatic sequences can be seen as images under letter-to-letter morphisms of fixed points of uniform

Email address: julien.leroy@u-picardie.fr (Julien Leroy)

substitutions (see [1]), F. Durand proved it in [10] and [11] for linearly recurrent sequences and P. Arnoux and G. Rauzy proved it in [2] for the so-called Arnoux-Rauzy sequences. Following [17], a sequence $\mathbf{w}$ with previous property is said to be S-adic (where S refers to the set of morphisms).

As mentioned in [17], the S-adic conjecture is the existence of a condition C such that "*a sequence has an at most linear complexity if and only if it is an S-adic sequence verifying C*". It is not possible to avoid considering a particular condition since, for instance, there exist fixed points of morphisms with a quadratic complexity (see [20]) and moreover, J. Cassaigne recently showed that there exists a finite set S of morphisms over an alphabet $A \cup \{l\}$ (where l is a special letter that does not belong to the alphabet A) such that any sequence over A is S-adic (see [8]).

In [16], before Cassaigne's constructions, S. Ferenczi used some other techniques to prove a kind of "only if" part of the conjecture for a weaker version of S-adicity. Indeed, he proved that the language of a uniformly recurrent sequence $\mathbf{w}$ with an at most linear complexity is S-adic in the sense that for any factor u of $\mathbf{w}$, there is a non-negative integer n such that u is a factor of $\sigma_0\sigma_1\cdots\sigma_n(a)$ with $\sigma_0\sigma_1\cdots\sigma_n \in S^*$. Theorem 1.2 states precisely this result which was originally expressed in terms of symbolic dynamical systems. In this paper, we avoid the language of dynamical systems and try to highlight all the key points of the proof of Theorem 1.2. Then, adapting Ferenczi's methods, we improve this result by proving Theorem 1.1 and give some properties on the S-adic representation that could help stating the condition C. In particular, we show that the constructions used make sense in a more general case and are particularly efficient for sequences with an at most linear complexity.

Theorem 1.1. *Let $\mathbf{w}$ be an aperiodic and uniformly recurrent sequence over an alphabet A. If $\mathbf{w}$ has an at most affine complexity then $\mathbf{w}$ is an S-adic sequence satisfying Properties 1 to 5 (see Section 6.2) for a finite set S of non-erasing morphisms such that for all letters a in A, the length of $\sigma_0\sigma_1\cdots\sigma_n(a)$ tends to infinity with n with $(\sigma_n)_n \in S^{\mathbb{N}}$ (this property will be called the ω-growth Property).*

Theorem 1.2 (Ferenczi [16])**.** *Let $\mathbf{w}$ be an aperiodic and uniformly recurrent sequence over an alphabet A with an at most affine complexity. There exist a finite number of morphisms σ_i, $1 \leq i \leq c$, over an alphabet $D = \{0, \dots, d-1\}$, an application α from D to A and an infinite sequence $(i_n)_{n\in\mathbb{N}} \in \{1, \dots, c\}^{\mathbb{N}}$ such that $\inf_{0\leq r\leq d-1} |\sigma_{i_0}\sigma_{i_1}\cdots\sigma_{i_n}(r)|$ tends to infinity if n tends to infinity and any factor of $\mathbf{w}$ is a factor of $\alpha\sigma_{i_0}\sigma_{i_1}\cdots\sigma_{i_n}(0)$ for some n.*

This paper is organized as follows. Section 2 recalls the definition of S-adicity. In Section 3, we present some results and examples about the conjecture and about the complexity of some particular S-adic sequences. In particular, using a technique similar to the technique in [13] we give an upper bound for the complexity of some S-adic sequences. Section 4 deals with Rauzy graphs. We recall their definition and explain how they evolve. Section 5 presents Ferenczi's methods in a general case and Section 6 gives the proof of Theorem 1.1. We conclude the paper with some remarks in Section 7.

2. S-adicity

Let us recall some basic definitions.

An *alphabet* is a finite set A whose elements are called *letters* (or *symbols*). A *word* u over A is a finite sequence of elements of A. The length ℓ of a word $u = u_1 \cdots u_\ell$ is the number of letters of u; it is denoted by $|u|$. The unique word of length 0 is called the *empty word* and is denoted by ε. The set of words of length ℓ over A is denoted by A^ℓ and $A^* = \bigcup_{\ell \in \mathbb{N}} A^\ell$ denotes the set of words over A. The set $A^* \setminus \{\varepsilon\}$ of non-empty words over A is denoted by A^+. The *concatenation* of two words u and v is simply the word uv; u^n is the concatenation of n copies of u. With concatenation, A^* is the free monoid generated by A.

A *sequence* (or *right infinite word*) over A is an element of $A^{\mathbb{N}}$. Recall that with the product topology, the set of sequences $A^{\mathbb{N}}$ is a compact metric space. In the sequel, sequences will be denoted by bold letters and for any non-empty word u over A, the sequence $\mathbf{w}$ composed of consecutive copies of u is denoted by $\mathbf{w} = u^\omega$.

Let A and B be two alphabets. A *morphism* (or a *substitution*) σ is a map from A^* to B^* such that $\sigma(uv) = \sigma(u)\sigma(v)$ for all words u and v over A. It is completely determined by the images of letters and it will be denoted by $\sigma : A \to B^*$. When σ is *non-erasing* (i.e., $\sigma(a) \neq \varepsilon$ for all a in A), it can be extended to a map from $A^{\mathbb{N}}$ to $B^{\mathbb{N}}$. If $\sigma : A \to A^*$ is non-erasing and if there is a letter a in A such that $\sigma(a) = au$ with $u \in A^+$, then the sequence $(\sigma^n(a^\omega))_{n\in\mathbb{N}}$ converges in $A^{\mathbb{N}}$ and the limit $\sigma^\omega(a)$ is called the *fixed point* of σ related to a.

The notion of S*-adic* sequence generalizes this notion. Let $\mathbf{w}$ be a sequence over A. An *adic representation* of $\mathbf{w}$ is given by a sequence $(\sigma_n : A_{n+1} \to A_n^*)_{n\in\mathbb{N}}$ of morphisms and a sequence $(a_n)_{n\in\mathbb{N}}$ of letters, $a_i \in A_i$ for all i such that $A_0 = A$, $\lim_{n\to+\infty} |\sigma_0\sigma_1 \cdots \sigma_n(a_{n+1})| = +\infty$ and

$$\mathbf{w} = \lim_{n\to+\infty} \sigma_0\sigma_1 \cdots \sigma_n(a_{n+1}^\omega).$$

The sequence $(\sigma_n)_{n\in\mathbb{N}}$ is the *directive word* of the representation. Let S be a finite set of morphisms. We say that $\mathbf{w}$ is S*-adic* (or that $\mathbf{w}$ is *directed by* $(\sigma_n)_{n\in\mathbb{N}}$) if $(\sigma_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}$. In the sequel, we will say that a sequence $\mathbf{w}$ is S-adic if there is a finite set S of morphisms such that $\mathbf{w}$ is directed by $(\sigma_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}$.

Proposition 2.1 (Cassaigne [8]). *Every sequence admits an adic representation.*

Proof. Let $\mathbf{w} = \mathbf{w}_0\mathbf{w}_1 \cdots$ be a sequence over a finite alphabet A and let l be a letter that does not belong to A. For each letter a in A we define the morphism σ_a from $(A \cup \{l\})^*$ to itself that maps l to la and maps every other letter b to itself. We also define the morphism φ from $A \cup \{l\}$ to A by $\varphi(l) = \mathbf{w}_0$ and $\varphi(b) = b$ for all b in A. Then we have

$$\mathbf{w} = \lim_{n\to+\infty} \varphi\sigma_{\mathbf{w}_1}\sigma_{\mathbf{w}_2} \cdots \sigma_{\mathbf{w}_n}(l^\omega).$$

□

3. *S*-adicity and factor complexity

A word u has an *occurrence* at position i in a word (or a sequence) w (or *occurs* in w) if $w_i w_{i+1} \cdots w_{i+|u|-1} = u$. It is a *factor* of w if it occurs in w or equivalently, if $w = xuy$ for some words x, y in $A^* \cup A^{\mathbb{N}}$. When $x = \varepsilon$ (resp. $y = \varepsilon$), u is a *prefix* (resp. *suffix*) of w. For a word (or a sequence) w, the factor $w_i w_{i+1} \cdots w_j$, $i \geq 1$, $j \leq |w|$, is denoted by $w[i,j]$. Recall that a sequence $\mathbf{w}$ is *recurrent* if every factor occurs infinitely often in $\mathbf{w}$. It is *uniformly recurrent* if it is recurrent and every factor occurs with bounded gaps and it is *linearly recurrent* if it is uniformly recurrent and if there is a constant K such that for all factors u of $\mathbf{w}$, the gap between two successive occurrences of u is bounded by $K|u|$.

The *language* of a sequence $\mathbf{w}$ is the set of factors of $\mathbf{w}$; it is denoted by $L(\mathbf{w})$. For each $n \in \mathbb{N}$, we note $L_n(\mathbf{w})$ the set of factors of length n in $\mathbf{w}$, i.e., $L_n(\mathbf{w}) = L(\mathbf{w}) \cap A^n$.

The complexity function of a sequence $\mathbf{w}$ is the function $p_{\mathbf{w}}$ (or simply p) that counts the number of factors of a given length in $\mathbf{w}$:

$$p_{\mathbf{w}} : \mathbb{N} \to \mathbb{N} : n \mapsto \# L_n(\mathbf{w}).$$

See Chapter 4 of [5] for a survey on this function.

3.1. Some examples of sub-linear S-adic sequences

A sequence is said to have an *at most linear complexity* (or *sub-linear complexity*) if there is a constant C such that $p(n) \leq Cn$ for all $n \geq 1$. One could equivalently say that the sequence has an at most affine (or sub-affine) complexity, i.e., $p(n) \leq Cn + D$ for all $n \in \mathbb{N}$. In [6], Cassaigne proved Theorem 3.1 which is a key point in the proof of Theorem 1.1.

Theorem 3.1 (Cassaigne [6])**.** *A sequence* $\mathbf{w}$ *has an at most linear complexity if and only if the first difference of its complexity* $(p_{\mathbf{w}}(n+1) - p_{\mathbf{w}}(n))$ *is bounded.*

Note that the first difference of complexity is closely related to special factors (see [7]). A factor u of $\mathbf{w}$ is *right special* (resp. *left special*) if there are two letters a and b in A such that ua and ub (resp. au and bu) belong to $L(\mathbf{w})$. It is *bispecial* if it is right and left special. For u in $L(\mathbf{w})$, denoting by $\delta^+ u$ (resp. by $\delta^- u$) the number of letters a in A such that ua (resp. au) is in $L(\mathbf{w})$ we have

$$p_{\mathbf{w}}(n+1) - p_{\mathbf{w}}(n) \quad = \sum_{u \in L_n(\mathbf{w}),\ u \text{ right special}} \underbrace{(\delta^+ u - 1)}_{\geq 1} \tag{1}$$

and when $\mathbf{w}$ is recurrent, we also have

$$= \sum_{u \in L_n(\mathbf{w}),\ u \text{ left special}} \underbrace{(\delta^- u - 1)}_{\geq 1} \tag{2}$$

A sequence $\mathbf{w}$ is *periodic* (resp. *ultimately periodic*) if it can be written as $\mathbf{w} = u^\omega$ (resp. $\mathbf{w} = vu^\omega$) with u a non-empty word over A. Such sequences are S-adic with

$\#S = 2$. Indeed, let $\mathbf{w} = uv^\omega$, $u, v \in A^+$. Let l be a letter that does not belong to A and let φ and ψ be the morphisms defined by $\psi(l) = lv$, $\psi(a) = vv$ for all letters a in A and $\varphi(l) = u$ and $\varphi(a) = a$ for all letters a in A. Then $\mathbf{w} = \varphi(\psi^\omega(l^\omega))$.

G. A. Hedlund and M. Morse proved in ([19]) that either $p(n)$ is ultimately constant (and corresponds to ultimately periodic sequences) or grows at least like $n + 1$. *Sturmian sequences* are binary infinite aperiodic sequences with minimal complexity $p(n) = n + 1$ for all n. Let τ_a, τ'_a, τ_b and τ'_b be morphisms over the alphabet $\{a, b\}$ as defined below:

$$\tau_a : \begin{cases} a \mapsto a \\ b \mapsto ab \end{cases} \qquad \tau_b : \begin{cases} a \mapsto ba \\ b \mapsto b \end{cases} .$$

$$\tau'_a : \begin{cases} a \mapsto a \\ b \mapsto ba \end{cases} \qquad \tau'_b : \begin{cases} a \mapsto ab \\ b \mapsto b \end{cases} .$$

It is well-known that Sturmian sequences are $\{\tau_a, \tau'_a, \tau_b, \tau'_b\}$-adic sequences such that if $\mathbf{w}$ is a Sturmian sequence coding the line $y = \alpha x + \rho$, then its directive word is completely determined by the coefficient of the continued fraction of α and by the *Ostrowski expansion* of ρ (see [4] and see [3] for more details about Ostrowski expansions). Observe that one could also define another set S of 3 morphisms such that all Sturmian sequences are S-adic. However, since the four morphisms defined above can be recovered with the techniques given in Section 5, we prefer working with them.

Next results are due to Durand (see [10] and [11]).

Proposition 3.2 (Durand [11])**.** *Let $\mathbf{w}$ be an S-adic sequence over an alphabet A such that all morphisms in S are non-erasing. Suppose that the minimal length $\inf_{c \in A_{n+1}} |\sigma_0\sigma_1 \cdots \sigma_n(c)|$ tends to infinity and there exists a constant D such that*

$$|\sigma_0\sigma_1 \cdots \sigma_{n+1}(b)| \leq D|\sigma_0\sigma_1 \cdots \sigma_n(c)|$$

for all $b \in A_{n+2}$, $c \in A_{n+1}$ and $n \in \mathbb{N}$. Then $p_y(n) \leq (\#A)^2 Dn$.

If all morphisms σ in S are uniform (that is $|\sigma(a)| = \ell \in \mathbb{N}_0$ for all letters a), then the result holds for $D = \max_{\sigma \in S} |\sigma(a)|$. We recover, as a corollary, that automatic sequences (that are images under letter-to-letter morphisms of fixed points of uniform substitutions) have an at most linear complexity (see [9] for the original proof).

Recall that an S-adic sequence is *primitive* if there is an integer s_0 such that for all integers r and all letters b in A_r and c in A_{r+s_0+1}, the letter b occurs in $\sigma_r\sigma_{r+1} \cdots \sigma_{r+s_0}(c)$. Using Proposition 3.2, Durand also proved Proposition 3.3.

Proposition 3.3 (Durand [10])**.** *Any primitive S-adic sequence has an at most linear complexity.*

He also proved Proposition 3.4 that gives an S-adic characterization of linearly recurrent sequences. In particular, it is proved in [12] that if $\mathbf{w}$ is a linearly recurrent sequence (with constant K), then $p_{\mathbf{w}}(n) \leq Kn$. A morphism $\sigma : A \to B^*$ is said to be *proper* if there exist two letters a and b in B such that $\sigma(c) \in aB^*b$ for all letter c in A; an S-adic sequence is *proper* if all morphisms in S are proper.

Proposition 3.4 (Durand [10])**.** *A sequence is linearly recurrent if and only if it is a primitive and proper S-adic sequence.*

Observe that although the condition *primitive and proper S-adic* is too restrictive for our goal (as there are some sequences with an at most linear complexity that are not linearly recurrent), it is this kind of characterization that we are looking for.

3.2. On the importance of directive words for some S-adic sequences

As we have seen in Section 3.1, for some set S of morphisms, any S-adic sequence has an at most linear complexity. Examples of such sets are those containing only uniform morphisms or only strongly primitive morphisms (that is all letters a occur in all images $\sigma(b)$). We could also prove that any $\{\varphi, \mu\}$-adic sequence is linearly recurrent, with φ and μ being respectively the *Fibonacci morphism* and the *Thue-Morse morphism* defined by $\varphi(a) = ab$, $\varphi(b) = a$, $\mu(a) = ab$ and $\mu(b) = ba$. However this is not true for any set S. There are some sets for which the directive words are important (think to Proposition 2.1) and even some sets for which any S-adic sequence does not have an at most linear complexity (for example the sets $S = \{\sigma\}$ such that σ has only fixed points with quadratic complexity (see [20])).

Example 3.1. Consider $S = \{\alpha, \mu\}$ with α defined by $\alpha(a) = aab$ and $\alpha(b) = b$ and μ which is the Thue-Morse morphism. Let $(k_n)_{n\in\mathbb{N}}$ be a sequence of non-negative integers and let $\mathbf{w}$ be the S-adic sequence

$$\mathbf{w} = \lim_{n\to+\infty} \alpha^{k_0}\mu\alpha^{k_1}\mu\cdots\alpha^{k_n}(a^\omega). \tag{3}$$

Lemma 3.5. *The S-adic sequence $\mathbf{w}$ defined in (3) has an at most linear complexity if and only if the sequence $(k_n)_{n\in\mathbb{N}}$ is bounded. However, even for unbounded sequences $(k_n)_{n\in\mathbb{N}}$, there is an increasing sequence $(m_n)_{n\in\mathbb{N}}$ of non-negative integers such that $p(m_n) \leq 4m_n$ for all n.*

Proof. First suppose $k_n \leq K$ for all integers n. For all integers i such that $0 \leq i \leq K$, we define the morphism $\mu^{(i)} = \alpha^i\mu$. Each such morphism $\mu^{(i)}$ is strongly primitive and we can write the sequence $\mathbf{w}$ as a primitive S'-adic sequence with $S' = \{\mu^{(i)} \mid 0 \leq i \leq K\}$. From Proposition 3.3, its complexity is at most linear.

Now consider that the sequence $(k_n)_{n\in\mathbb{N}}$ is unbounded and let us show that the complexity is not at most linear. We know from [20] that the fixed point $\alpha^\omega(a)$ has a quadratic complexity. From Theorem 3.1 and (1) we deduce that the number of right special factors of $\alpha^\omega(a)$ of a given length is unbounded. Moreover we can show that all the right special factors of length n of $\alpha^\omega(a)$ occurs in $\alpha^{n+1}(a)$. Now let us show that if u is a right special factor of length n in $\alpha^{k_n}(a)$, then $\alpha^{k_0}\mu\alpha^{k_1}\mu\cdots\alpha^{k_{n-1}}\mu(u)$ is a right special factor of $\mathbf{w}$ of length $n2^q$ with $q = \sum_{i=0}^{n-1}(k_i + 1)$. Indeed, as $\mu(a)$ and $\alpha(a)$ start with a and $\mu(b)$ and $\alpha(b)$ start with b, the image of u is still a right special factor. Moreover, $\mu(u)$ contains exactly n letters a and n letters b, and both α and μ map a word with the same number of a and b to a word of double length with the same number of a and b. Hence $|\alpha^{k_0}\mu\alpha^{k_1}\mu\cdots\alpha^{k_{n-1}}\mu(u)| = |u|2^q$ with q defined as previously. Now, if u and v are two distinct right special factors of length n of $\alpha^\omega(a)$, then $\alpha^{k_0}\mu\alpha^{k_1}\mu\cdots\alpha^{k_{n-1}}\mu(u)$ and $\alpha^{k_0}\mu\alpha^{k_1}\mu\cdots\alpha^{k_{n-1}}\mu(v)$ are two distinct special factors of length $n2^{q(n)}$ of $\mathbf{w}$. As the number of right special factors of a given length of $\alpha^\omega(a)$ is unbounded, the number of right special factors of a given length of $\mathbf{w}$ is also unbounded. Using Theorem 3.1, we see that the complexity is not at most linear.

The last step is to show that, for infinitely many integers m_n, the complexity is at most linear. For all non-negative integers n, we already know that $|\alpha^{k_0}\mu\alpha^{k_1}\mu\cdots\alpha^{k_{n-1}}\mu(a)| = |\alpha^{k_0}\mu\alpha^{k_1}\mu\cdots\alpha^{k_{n-1}}\mu(b)| = m_n = 2^q$ with q as defined previoulsy by $\sum_{i=0}^{n-1}(k_i + 1)$. Consequently, all factors u of length m_n are factors of $|\alpha^{k_0}\mu\alpha^{k_1}\mu\cdots\alpha^{k_{n-1}}\mu(v)|$ for some words v of length 2. As there are only 4 possible binary words of length 2 and as there are less that m_n distinct factors of length m_n is a word of length $2m_n$, this concludes the proof. □

3.3. An interesting condition

As mentioned in the introduction, the S-adic conjecture is the existence of a condition C such that $\mathbf{w}$ has an at most linear complexity if and only if $\mathbf{w}$ is an S-adic sequence satisfying condition C.

In the case of fixed points of morphisms $\sigma^\omega(a)$, Pansiot proved in [20] that the complexity function can only have five asymptotic behaviors. More precisely, p satisfies one of the 5 following inequalities for all $n \geq 1$:

$$\begin{aligned}
C_1 &\leq p(n) \leq C_2 \\
C_1 n &\leq p(n) \leq C_2 n \\
C_1 n \log n &\leq p(n) \leq C_2 n \log n \\
C_1 n \log\log n &\leq p(n) \leq C_2 n \log\log n \\
C_1 n^2 &\leq p(n) \leq C_2 n^2
\end{aligned}$$

for some positive constants C_1 and C_2. In particular, he proved that the class of highest complexity n^2 can be reached only by morphisms σ admitting bounded letters, i.e., letters c such that the sequence $(|\sigma^n(c)|)_{n\in\mathbb{N}}$ is bounded (as for the morphism α in Example 3.1).

In Theorem 1.1 (and it was already the case in Ferenczi's paper [16]), we show that a sequence with an at most linear complexity is an S-adic sequence such that the length of $\sigma_0\sigma_1\cdots\sigma_n(a_{n+1})$ tends to infinity as n increases for all sequences $(a_n)_{n\in\mathbb{N}}$ of letters $a_n \in A_n$. Moreover, observe that almost all examples treated in previous sections satisfy this property: the only constructions that do not satisfy it are Cassaigne's constructions (Proposition 2.1). Although this property is not necessary to have a low complexity (for example, the fixed point $\sigma^\omega(0)$ of the substitution σ defined by $\sigma(0) = 0010$ and $\sigma(1) = 1$ has an at most linear complexity (see [15])), it is interesting to remark that up to now, Cassaigne's constructions are the only S-adic constructions that allow us to construct sequences with an arbitrary high complexity.

As a conclusion, the growth of letters seems to be an important condition to have a *reasonably low* complexity. Let us call the *ω-growth Property* the fact that the length of $\sigma_0\sigma_1\cdots\sigma_n(a_{n+1})$ tends to infinity with n for all sequences $(a_n)_{n\in\mathbb{N}}$ of letters $a_n \in A_n$. Observe that it is clear from Lemma 3.5 that the ω-growth Property is not sufficient to have an at most linear complexity. Hence we have to take care not only of the set S of morphisms but also of the directive words of S-adic sequences.

3.4. Beyond linearity

Another consequence of Proposition 2.1 is that we cannot have an upper bound on the complexity of S-adic sequences like the one we have for fixed points of substitutions (see [20]). However we can still hope having such a bound for S-adic sequences satisfying the ω-growth Property. This question seems to be a new non-trivial problem. Although its study is not the purpose of this paper, we give an upper bound for S-adic sequences such that $|\sigma(a)| \geq 2$ for all σ in S and all letters a in $A(\sigma)$ (see Proposition 3.6 below). Techniques are similar as those used in [13] for D0L systems. Recall that a *D0L system* (which means *deterministic L-system without interaction*) is essentially equivalent to a substitution. Roughly speaking, the main difference is that for D0L systems, we are only interested in the language of the fixed point. In the same way that S-adic sequences are a generalization of fixed points of substitutions, *DT0L systems* (which means *deterministic table system without interaction*) are a generalization of D0L systems. However there is a more important difference between DT0L and S-adic sequences than between D0L and substitutive sequences. Indeed, for DT0L systems, the language one is usually interested in is the set of words occurring in $\sigma_0\sigma_1\cdots\sigma_k(a)$ for any finite sequence in S^* (where S denotes also the set of rules of the system). In other words, we consider the language of all S-adic sequences (i.e., we consider all directive words). It is proved in [14] that *everywhere growing* DT0L systems (which mean $|\sigma(a)| \geq 2$ for all σ and a) have an at most polynomial complexity. For S-adic sequences built upon the same hypothesis, we have a better upper bound as shown in Proposition 3.6.

First let us recall the definition of the *radix order* $\preceq^*$. Let $\preceq$ be an order on the alphabet A and let u and v be in A^*, $u \neq v$. We have $u \prec^* v$ if either $|u| < |v|$ or $|u| = |v|$ and there is a smallest integer $i \in [1, |u|]$ such that $u_i \prec v_i$.

Proposition 3.6. *Let* $\mathbf{w}$ *be an S-adic sequence over an alphabet A such that* $|\sigma(a)| \geq 2$ *for all σ in S and all letters a in $A(\sigma)$. Then there is a constant C such that* $p_{\mathbf{w}}(n) \leq Cn\log n$ *for all integers* $n \geq 1$.

Proof. Let ℓ denotes the maximal length of $\sigma(a)$ for σ in S and a in $A' = \bigcup_{n\in\mathbb{N}} A_n$. Consider an integer n greater than 2ℓ. For all words u in $L_n(\mathbf{w})$, we construct a sequence $(u_k)_{k\in\mathbb{N}}$ of words in the following way:

- $u_0 = u$;
- for all integers k, u_{k+1} is the smallest word in $L(\mathbf{w}_{k+1})$ (with respect to the radix

order) such that $u_k \in L(\sigma_k(u_{k+1}))$, with $\mathbf{w}_{k+1} = \lim_{n\to+\infty} \sigma_{k+1}\sigma_{k+2}\cdots\sigma_n(a_{n+1}^\omega)$.
We can easily see that the sequence $(|u_k|)_{k\in\mathbb{N}}$ is decreasing until a smallest integer r such that $|u_r| \leq 2$. We have $2 < r < 1+C\log n$, the first inequality being trivial from the choice of n. For the second one, observe that $|u_{r-1}|$ is at least 3. Then, writing $u_{r-1} = av_{r-1}b$ with $a, b \in A$, we see that $\sigma_0\sigma_1\cdots\sigma_{r-2}(v_{r-1})$ is a proper factor of length at least 2^{r-1} of u. Therefore we have $n > 2^{r-1}$ and then $r < C\log n + 1$.

Now for all words u in A'^* of length smaller than or equal to 2, we define $W_n(u)$ as the set of words of length n in $L(\mathbf{w})$ such that the construction previously described gives $u_r = u$. Obviously, $\bigcup_{u\in L_{\leq 2}(\mathbf{w})} W_n(u) = L_n(\mathbf{w})$. To conclude the proof, it suffices to check that there are no more than n words in $\sigma_0\sigma_1\cdots\sigma_{r-1}(u_r)$ that admit $\sigma_0\sigma_1\cdots\sigma_{r-2}(v_{r-1})$ as a factor. $\square$

Example 3.2 shows that this bound is the best one we can obtain.

Example 3.2. Let β be the substitution over $A = \{a, b, c\}$ defined by $\beta(a) = abca$, $\beta(b) = bb$ and $\beta(c) = ccc$ and consider its fixed point $\mathbf{w} = \beta^\omega(a)$. It can be seen as a $\{\beta\}$-adic sequence satisfying the ω-growth Property and we know from [20] that its complexity function satisfies $C_1 n\log(n) \leq p_{\mathbf{w}}(n) \leq C_2 n\log(n)$, with $C_1, C_2 > 0$.

4. Rauzy graphs

The proof of Theorem 1.1 is based on the evolution of Rauzy graphs. In this section, we recall this notion. First let us recall some definitions.

A *directed graph* G is a couple (V, E) where V is the set of vertices and $E \subset V \times V$ is the set of edges. Edges may be labeled by elements of a set A and then $E \subset V \times A \times V$. If $e = (u, a, v)$ is an edge of G, we let $o(e) = u$ denote its starting vertex (o for *outgoing*) and $i(e) = v$ its ending vertex (i for *incoming*). A *path* p in G is a sequence $(v_0, a_1, v_1)(v_1, a_2, v_2)\ldots(v_{\ell-1}, a_\ell, v_\ell)$ of consecutive edges. The label of p is the ℓ-tuple $(a_1, a_2, \ldots, a_\ell)$. However in the sequel we will simply denote it by concatenating the labels of each edge. We also let $o(p)$ denote the starting vertex v_0 of p and by $i(p)$ its ending vertex v_ℓ; they are called the *extremities* of p and $v_1, \ldots, v_{\ell-1}$ are called *interior vertices*. The *length* of a path is the number of edges composing it. A *subpath* of $p = (v_0, a_1, v_1)(v_1, a_2, v_2)\ldots(v_{\ell-1}, a_\ell, v_\ell)$ is a path $q = (u_0, b_1, u_1)(u_1, b_2, u_2)\ldots(u_{k-1}, b_k, u_k)$ such that $k \leq \ell$ and there exists an integer $i \in [0, \ell-k]$ such that $(v_{i+j}, a_{i+j+1}, v_{i+j+1}) = (u_j, b_{j+1}, u_{j+1})$ for all integers $j \in [0, k-1]$. It is a *proper subpath* if $k < \ell$.

4.1. Rauzy graphs and allowed paths

Let $\mathbf{w}$ be a sequence over an alphabet A. For each non-negative integer n, we define the *Rauzy graph of order* n of $\mathbf{w}$ (also called *graph of words of length* n), denoted by $G_n(\mathbf{w})$ (or simply G_n) as the directed graph $(V(n), E(n))$, where

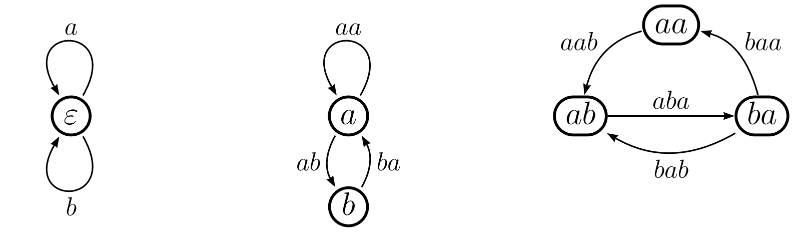
- the set $V(n)$ of vertices is the set L_n of factors of length n of $\mathbf{w}$ and
- there is an edge from u to v if there are two letters a and b in A such that $ub = av \in L_{n+1}$.

In the literature, there are different ways of labeling the edges. Indeed, the edges are sometimes labeled by the letter a, by the letter b, by the couple (a, b) or by the word av, i.e., the following four notations exist:

$$u \xrightarrow{b} v \qquad u \underset{a}{\rightarrow} v \qquad u \underset{a}{\xrightarrow{b}} v \qquad u \xrightarrow{av} v.$$

For an edge $e = (u, (a, b), v) = u \underset{a}{\xrightarrow{b}} v$, let us call $\lambda_L(e) = a$ its *left label*, $\lambda_R(e) = b$ its *right label* and $\lambda(e) = ub = av$ its *full label*. Same definitions hold for labels of paths (left and right labels being words of same length as the considered path) where we naturaly extend the map λ to the set of paths by $\lambda\left((u_0, (a_1, b_1), u_1)(u_1, (a_2, b_2), u_2) \cdots (u_{\ell-1}, (a_\ell, b_\ell), u_\ell)\right) = u_0 b_1 b_2 \cdots b_\ell = a_1 a_2 \cdots a_\ell u_\ell$. In this paper we will mostly consider left labels.

Example 4.1. Let f be the *Fibonacci sequence* that is the fixed point $\varphi^\omega(a)$ of the substitution φ previously defined. Figure 1 represents the first three Rauzy graphs of f (with full labels on the edges).



(a) $G_0(f)$ (b) $G_1(f)$ (c) $G_2(f)$

Figure 1: First Rauzy graphs of the Fibonacci sequence

Remark 4.1. A sequence is recurrent if and only if all its Rauzy graphs are *strongly connected* (that is for all vertices u and v of G_n there is a path p from u to v, i.e., $o(p) = u$ and $i(p) = v$.).

We say that a vertex v is *right special* (resp. *left special*, *bispecial*) if it corresponds to a right special (resp. left special, bispecial) factor.

Remark 4.2. By definition of Rauzy graphs, $(u, (a, b), v)$ is an edge in $G_n(\mathbf{w})$ if and only if the word ub is in the language $L(\mathbf{w})$. It is also clear that for any word u in $L(\mathbf{w})$, for all non-negative integers $n < |u|$ there is a non-empty path p in $G_n(\mathbf{w})$ whose full label $\lambda(p)$ is u. The contrary is not true: not every path in $G_n(\mathbf{w})$ has a full label that is a factor of $\mathbf{w}$. Indeed, in the Rauzy graph $G_1(f)$ of the Fibonacci sequence (see Figure 1(b)), the full label of the path $(a, (a, a), a)^n$ is a^{n+1} for each n and this word is not in the language as soon as $n \geq 2$. The reason is that once we have reached the vertex a coming from some edge, we have two possibilities: either we stay in this vertex passing through the loop $(a, (a, a), a)$, or we go in the vertex b with the edge $(a, (a, b), b)$. These possibilities exist because the word a is a right special factor of the Fibonacci sequence, but this particularity only implies that, starting at vertex a, we can read a a or a b. In other words, it does not take care of what happened before (i.e., from which edge we arrived in this vertex) although we have to. Indeed, if we come from the loop, this mean that the previous vertex of the path was the vertex a and the full label of this path is aa. Then the only possibility that you really have is to go into the vertex b (because $aaa \notin L(\mathbf{w})$).

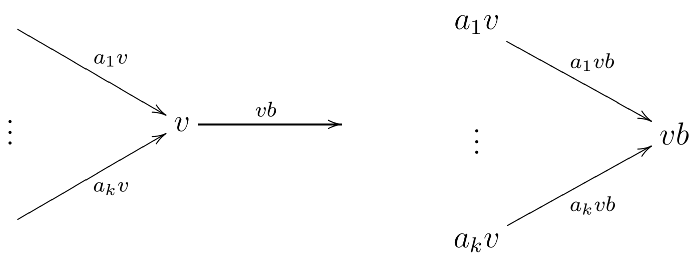
A path in a Rauzy graph is said to be *allowed* if its full label is a word in $L(\mathbf{w})$. Observe that, by definition, any path $p = (v_0, (a_1, b_1), v_1) \dots (v_{\ell-1}, (a_\ell, b_\ell), v_\ell)$ that does not contain any subpath $(v_i, (a_{i+1}, b_{i+1}), v_{i+1}) \dots (v_{j-1}, (a_j, b_j), v_j)$, $i \geq 1$, $j \leq \ell - 1$ with v_i left special and v_j right special is allowed. Moreover, the following trivially holds.

Proposition 4.1. *Let G_n be a Rauzy graph of order n. For all paths p of length $\ell \leq n$ in G_n, the left (resp. right) label of p is a prefix (resp. a suffix) of $o(p)$ (resp. of $i(p)$).*

4.2. Evolution of Rauzy graphs

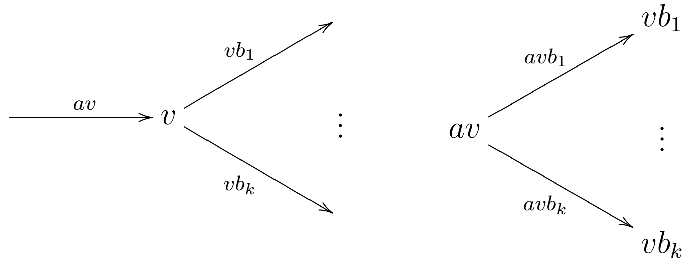
To prove Theorem 1.1 we will need to let the Rauzy graphs evolve, i.e., we will need to go from $G_n(\mathbf{w})$ to $G_{n+1}(\mathbf{w})$. Let us see how it goes. As the set of edges of G_n is in bijection with L_{n+1}, we can write G_n as the directed graph (L_n, L_{n+1}). Then to get the Rauzy graph of order $n+1$, it suffices to replace each edge of G_n by a vertex and to define the edges in the following way:

- for each non special vertex v in G_n, we replace $\xrightarrow{av} v \xrightarrow{vb}$ by $av \xrightarrow{avb} vb$;
- for each left special vertex v in G_n that is not right special we make the following changes



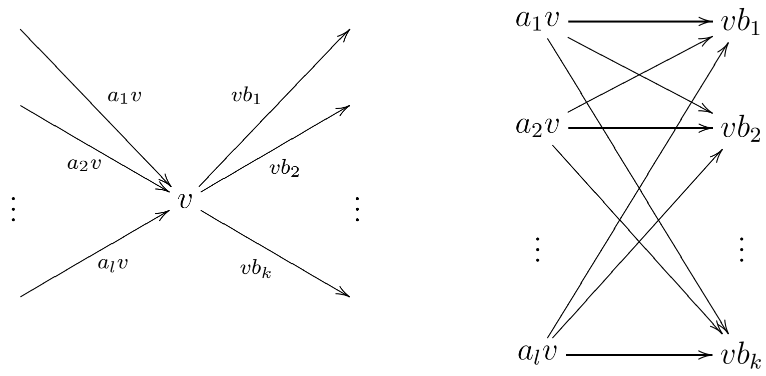
Transitions in G_n Transitions in G_{n+1}

- for each right special vertex v in G_n that is not left special, we make the following changes



Transitions in G_n Transitions in G_{n+1}

- finally, for each bispecial vertex v in G_n, we have among the transitions in G_{n+1} represented here below, only those whose label a_ivb_j is a factor of $\mathbf{w}$.



Transitions in G_n Possible transitions in G_{n+1}

Remark 4.3. It is a direct consequence of what precedes that for each non-negative integer n, if there is no bispecial factor in L_n, then the Rauzy graph of order n determines exactly the Rauzy graph of order $n+1$. Moreover, in this case the length of the smallest path from a left special vertex to a right special vertex decreases by 1 as n increases by 1. Consequently, there exists a smallest non-negative integer k_n such that the Rauzy graph G_{n+k_n} contains a bispecial vertex v and we have to check which labels a_ivb_j belongs to $L(\mathbf{w})$ to construct the Rauzy graph G_{n+k_n+1}.

5. Segments, circuits and morphisms

Let $\mathbf{w}$ be an aperiodic and uniformly recurrent sequence over an alphabet A. The proof of Theorem 1.1 is similar to the proof of Ferenczi's result (Theorem 1.2). Hence, in this section, we explain Ferenczi's methods to construct an adic representation of any factor of $\mathbf{w}$. In his paper [16], Ferenczi defined the notion of *n-segments* (that we will call *right n-segment*; see below for the definition). For our result we need to define the symmetric notion of *left n-segment*. However constructions are mostly the same as those described in [16].

For each $n \in \mathbb{N}$, a *left* (resp. *right*) *n-segment* is a non-empty path p in $G_n(\mathbf{w})$ whose only left (resp. right) special vertices are its extremities $o(p)$ and $i(p)$. If not explicitly stated, *n-segment* means *left n-segment*.

Observe that any (left or right) n-segment is trivially allowed. As the Rauzy graphs of recurrent sequences are strongly connected, the set of n-segments is a covering of the set of edges of G_n in the sense that each edge belongs to at least one n-segment. Moreover, for each n, as there exists only a finite (possibly unbounded) number of left special vertices in G_n, there exists only a finite (possibly unbounded) number of n-segments.

Remark 5.1. The notion of n-segment is related to the notion of *return word*. Recall that if a word $u \in A^*$ has at least two occurrences in a sequence $\mathbf{w}$, a *return word* to u in $\mathbf{w}$

is a word v such that vu is a factor of $\mathbf{w}$, u is a prefix of vu and vu contains only two occurrences of u. We can extend this notion to a set of words in the following way: for a sequence $\mathbf{w}$ and a set F of factors of $\mathbf{w}$, a return word to F is a word v such that there are two words u and u' in F (they might be the same) such that vu' is a factor of $\mathbf{w}$ that admits u as a prefix and u and u' are the only words in F that occur in vu'. Now it is easy to be convinced that the set of left labels of n-segments is exactly the set of return words to $LS_{\mathbf{w}}(n)$ where $LS_{\mathbf{w}}(n)$ denotes the set of left special factors of length n in $\mathbf{w}$.

For sequences with a "reasonably low" complexity, the number of left special factors increases much more slowly than the complexity. Consequently, we expect that the maximal length of n-segments will grow to infinity. Then, due to the uniform recurrence, all factors of $\mathbf{w}$ of length smaller than some ℓ will be factors of the label of the longest n_ℓ-segment for some n_ℓ large enough. So now, let us study the behavior of n-segments as n increases. To this aim we define a map ψ_n on the set of paths $\mathcal{P}_{n+1}$ of $G_{n+1}(\mathbf{w})$ in the following way. For each path p in $\mathcal{P}_{n+1}$ with left label u, $\psi_n(p)$ is the path p' in $\mathcal{P}_n$ whose left label is u and such that $o(p')$ and $i(p')$ are the prefixes of length n of respectively $o(p)$ and $i(p)$. Roughly speaking, $\psi_n(p)$ is the corresponding path in $G_n(\mathbf{w})$ of the path p in $G_{n+1}(\mathbf{w})$. Observe that ψ_n is not one-to-one. Indeed, if for example p is a path in $G_n(\mathbf{w})$ such that its full label does not contain any bispecial factor as a proper factor and such that $i(p)$ is right special, then $\#\psi_n^{-1}(p) = \delta^+ i(p) > 1$. To be coherent with Definition 5.1 below, we consider the concatenation over $\mathcal{P}_n$ allowing that a concatenation of paths might not be a element of $\mathcal{P}_n$, i.e., $pq \in \mathcal{P}_n^*$ even if $pq \notin \mathcal{P}_n$.

Lemma 5.1 here below — and also Lemmas 5.2, 6.1, 6.2 and 6.3 in next sections — was already proved in [16]. However, all these lemmas were parts of the proof of Theorem 1.2. In this paper, we decided to structure the proof in several lemmas.

Lemma 5.1 (Ferenczi [16])**.** *Let $\mathbf{w}$ be a sequence over an alphabet A. For any $(n+1)$-segment p of $\mathbf{w}$, $\psi_n(p)$ is a concatenation of n-segments of $\mathbf{w}$.*

Proof. Let p be a $(n+1)$-segment in $G_{n+1}(\mathbf{w})$ and $p' = \psi_n(p)$. As a prefix of a left special factor is still a left special factor, $o(p')$ and $i(p')$ are left special. Hence p' is a concatenation of n-segments. □

Definition 5.1 (Definition of morphisms)**.** Lemma 5.1 allows to define some morphisms σ_n over the alphabets of n-segments. Indeed, for each non-negative integer n, let $\mathcal{A}_n$ be the set of n-segments, A_n be the set $\{0, 1, \ldots, \#\mathcal{A}_n - 1\}$ and let us consider a bijection $\theta_n : A_n \to \mathcal{A}_n$. We can extend θ_n to an isomorphism $\theta_n : A_n^* \to \mathcal{A}_n^*$ considering the concatenation of paths as explained after Remark 5.1 and putting $\theta_n(ab) = \theta_n(a)\theta_n(b)$, $a, b \in A_n$. Now for all n, we define the morphism $\sigma_n : A_{n+1} \to A_n$ as the only map that satisfies $\theta_n \circ \sigma_n = \psi_n \circ \theta_{n+1}$.

Remark 5.2. If the alphabet of $\mathbf{w}$ is $A = \{a_1, \ldots, a_k\}$, the Rauzy graph G_0 is as in Figure 2 so that for all 0-segments p, we have $\lambda(p) = \lambda_L(p) = \lambda_R(p) \in A$. Hence we could have chosen λ_R or λ instead of λ_L. However it is more convenient to work with λ_L for the proof of Theorem 1.1. In the sequel we will denote by γ the map $\lambda_L \circ \theta_0$.

Remark 5.3. It is a consequence of the constructions described above that $|\sigma_n(i)| \geq 2$ means that there are at least two n-segments occurring in $\theta_{n+1}(i)$. Suppose that p and q are such n-segments with $i(p) = o(q)$. Then $\sigma_n(i) \in A_n^* \theta_n^{-1}(p)\theta_n^{-1}(q)A_n^*$ and as any interior vertex of a $(n+1)$-segment cannot be left special, the only possibility is that the vertex $i(p) = o(q)$ is a bispecial vertex such that its right extension that is a vertex of $\theta_{n+1}(i)$ is not left special. Hence if a Rauzy graph $G_n(\mathbf{w})$ does not contain any bispecial vertex, then the n-segments are exactly the elements of $\{\psi_n(p) \mid p \text{ is an } (n+1)\text{-segment}\}$ and the morphism σ_n is simply a bijective and letter-to-letter morphism.

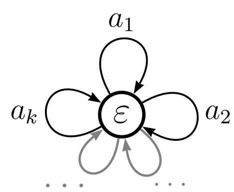


Figure 2: Rauzy graph G_0 of any sequence over $\{a_1, \dots, a_k\}$

Remark 5.4. Morphisms in Definition 5.1 may be uninteresting. Indeed consider the case of sequences with maximal complexity (like the *Champernowne sequence* for example). As $L(\mathbf{w}) = A^*$ for these sequences, all factors are left special and so all edges in G_n are n-segments. For all n, the morphism σ_n is therefore uniform of length 1 so $|\sigma_0\sigma_1 \cdots \sigma_n(a)| = 1$ for all n and all letters a. However construction of Definition 5.1 makes sense as soon as there is a sequence $(a_n)_{n\in\mathbb{N}}$ of letters $a_n \in A_n$ such that $|\sigma_0 \cdots \sigma_n(a_{n+1})|$ tends to infinity as n increases. Indeed in this case $L(\mathbf{w}) = \bigcup_{n\in\mathbb{N}} L(\gamma\sigma_0 \cdots \sigma_n(a_{n+1}))$ (due to the uniform recurrence). We can easily see that for $|\sigma_0 \cdots \sigma_n(a_{n+1})|$ to converge to infinity for at least one sequence of letters $(a_n)_{n\in\mathbb{N}}$, $a_n \in A_n$, it suffices that the sequence $\frac{p(n)}{s(n)}$ is unbounded, where $s(n)$ denotes the number of n-segments. As $s(n) \leq \#A(p(n+1) - p(n))$ (as any n-segment is completely determined by its last edge $(u, (a, b), v)$, with v left special and a extending v to the left), it suffices that $\limsup_{n\to+\infty} \frac{p(n)}{p(n+1)-p(n)} = +\infty$ and so that $\liminf_{n\to+\infty} \frac{p(n+1)}{p(n)} = 1$. Note that sequences with an at most polynomial complexity satisfy this property although for sequences with higher complexity, it is not always the case.

Lemma 5.2 below is a key point in the proof of Theorem 1.1 but it is true for any uniformly recurrent sequence (not only for those with an at most linear complexity). We first need to recall the notions of n-circuit and of *short* and *long* segments or circuits introduced in [16].

By Lemma 5.1, the minimal length of n-segments is non-decreasing. If it is bounded, there is an integer N and a N-segment s such that for all integers $n > N$, there is a n-segment s' such that $s = \psi_N\psi_{N+1} \cdots \psi_{n-1}(s')$. Such a segment is said to be *short*. Non-short n-segments are said to be *long*. Roughly speaking, a short n-segment will be a m-segment for all m greater than n while a long n-segment will only appear as a proper subpath of m-segments for m large enough. Note that if p is a short n-segment then for all positive integers k and all $(n+k)$-segment q_k such that $\psi_n \cdots \psi_{n+k-1}(q_k) = p$ we have $|\sigma_{n+k-1}(\theta_{n+k}(q_k))| = 1$.

A *n-circuit* is a non-empty path p in $G_n(\mathbf{w})$ such that $v = o(p) = i(p)$ is a left special vertex and any interior vertex of p is not $o(p)$. It is easy to be convinced that Lemma 5.1 can be adapted to n-circuits. Hence we can define *short* and *long* n-circuits similarly to *short* and *long* n-segments.

Lemma 5.2 (Ferenczi [16]). *Let $\mathbf{w}$ be a uniformly recurrent sequence over an alphabet A. For any non-negative integer n, there is no short n-circuit in $G_n(\mathbf{w})$.*

Proof. As the sequence $\mathbf{w}$ is uniformly recurrent, if it is ultimately periodic, it is periodic. Hence in this case there is no left special factor of length greater than some N and so no n-circuit for $n > N$. Now suppose that $\mathbf{w}$ is aperiodic and let p be a short n-circuit of left label u in $G_n(\mathbf{w})$. By definition, for all positive integers k, there is an $(n+k)$-circuit q_k such that $p = \psi_n \cdots \psi_{n+k-1}(q_k)$. As the left label of q_k is u by definition, from Proposition 4.1 we deduce that, for all k large enough, $o(q_k)$ is equal to u followed by a prefix of itself and so u^{e_k} is a prefix of $o(q_k)$ with $e_k = \left\lfloor \frac{|o(q_k)|}{|u|} \right\rfloor$. Since $o(q_k)$ is a factor of $\mathbf{w}$ for all k and $|o(q_k)|$ tends to infinity with k, there are arbitrary large powers of u in $L(\mathbf{w})$ and this contradicts aperiodicity and uniform recurrence. ☐

Remark 5.5. A consequence of Lemma 5.2 is that for all integers ℓ, there is an integer n_ℓ such that any n_ℓ-circuit has length greater than ℓ.

6. Proof of Theorem 1.1

First let us recall Theorem 1.1. The S-adicity Property and the ω-growth Property are proved in Section 6.1 and Properties 1 to 5 will be stated in Section 6.2.

Theorem 1.1. *Let $\mathbf{w}$ be an aperiodic and uniformly recurrent sequence over an alphabet A. If $\mathbf{w}$ has an at most linear complexity then it is an S-adic sequence satisfying the ω-growth Property and Properties 1 to 5 for some finite set S of non-erasing morphisms.*

6.1. S-adicity and ω-growth Property

First, next three lemmas allow us to bound the cardinality of the set of morphisms.

Lemma 6.1 (Ferenczi [16]). *Let $\mathbf{w}$ be a sequence over an alphabet A. If $\mathbf{w}$ has an at most linear complexity, then the number of n-segments of $\mathbf{w}$ is bounded.*

Proof. By Theorem 3.1, there exists a constant K such that $p(n+1) - p(n) \leq K$ for all n. From Equality 2 in Section 3.1 we deduce that the number of left special factors of length n is also bounded by K and as a n-segment is completely determined by its last edge, the number of n-segments is bounded by $K\#A$. ☐

A consequence of Lemma 6.1 is that the maximal length of n-segments tends to infinity as n increases (as the number of edges in $G_n(\mathbf{w})$ tends to infinity). In other words, there is at least one long n-segment for each length n. Moreover, as the number of n-segments is bounded and as two different n-segments p and q give rise to disjoint sets $\psi_n^{-1}(p)$ and $\psi_n^{-1}(q)$, there is only a bounded number of shorts segments (all order n included) so we can bound the length of short segments by some constant ℓ.

When the sequence has an at most linear complexity, we can improve Lemma 5.1 stating that the number of n-segments occurring in an $(n+1)$-segment is bounded. In this case we will construct only a finite number of morphisms because this only gives rise to morphisms of bounded length over a bounded alphabet.

Lemma 6.2 (Ferenczi [16])**.** *Let* $\mathbf{w}$ *be an aperiodic sequence over an alphabet* A*. If* $\mathbf{w}$ *has an at most linear complexity, then for any* $(n+1)$*-segment* p *of* $\mathbf{w}$*,* $\psi_n(p)$ *is a bounded concatenation of* n*-segments.*

Proof. Let K be such that $p(n+1) - p(n) \leq K$ for all n (see Theorem 3.1). Consider a $(n+1)$-segment p in $G_{n+1}(\mathbf{w})$ (we know it exists since $\mathbf{w}$ is aperiodic). The number of n-segments in $\psi_n(p)$ is equal to 1 plus the number of vertices va in p, $a \in A$, such that v is a left special factor of $\mathbf{w}$ and va not. Moreover, as these vertices are not left special, the path p cannot pass through one of them more than once. Since there exist at most K left special vertices v in $G_n(\mathbf{w})$, there exist at most $K\#A$ vertices va as considered just above. Consequently, the number of n-segments in p is bounded by $1 + K\#A$. □

We need one more lemma to prove the S-adicity Property in Theorem 1.1. This last one will also allow us to prove that the S-adic representation satisfies the ω-growth Property.

Lemma 6.3 (Ferenczi [16])**.** *Let* $\mathbf{w}$ *be a uniformly recurrent sequence over an alphabet* A*. If* $\mathbf{w}$ *has an at most linear complexity, then in any path in* $G_n(\mathbf{w})$*, the number of consecutive short* n*-segments is bounded.*

Proof. Let K be such that $p(n+1) - p(n) \leq K$ for all n (see Theorem 3.1). As any edge of $G_n(\mathbf{w})$ appears in at least one n-segment, any finite path in $G_n(\mathbf{w})$ can be decomposed into a finite number of n-segments, the first one and the last one being possibly truncated. In this decomposition, some segments may be short and so have bounded length, say by ℓ. Now if a path p composed of consecutive short n-segments has length greater than $K\ell$, the path contains at least $K+1$ occurrences of left special vertices. Consequently some vertices v_i and v_j of p are equal and the graph contains a n-circuit whose length is smaller than $K\ell$. As $\mathbf{w}$ is uniformly recurrent, by Lemma 5.2, this is impossible for n large enough (see Remark 5.5). □

Now we can prove the S-adicity Property in Theorem 1.1.

Proof of the S-adicity property and of the ω-growth Property in Theorem 1.1. Let $\mathbf{w}$ be a uniformly recurrent sequence over an alphabet A. Let $\sharp$ be a symbol that is not in A and consider the sequence $\mathbf{w}' = \sharp\mathbf{w}$ over $A \cup \{\sharp\}$. Observe that $\mathbf{w}'$ is not uniformly recurrent and so its Rauzy graphs are not strongly connected. However, if we have $G_n(\mathbf{w}) = (V(n), E(n))$, the graph $G_n(\mathbf{w}')$ is simply the graph $(V'(n), E'(n))$ with

$$V'(n) = V(n) \cup \{\sharp \mathbf{w}[0, n-2]\} \text{ and}$$
$$E'(n) = E(n) \cup (\sharp \mathbf{w}[0, n-2], (\sharp, \mathbf{w}_{n-1}), \mathbf{w}[0, n-1]) \, .$$

Also remark that the edge $(\sharp \mathbf{w}[0, n-2], (\sharp, \mathbf{w}_{n-1}), \mathbf{w}[0, n-1])$ of $G_n(\mathbf{w}')$ does not appear in any n-segment of $\mathbf{w}'$ for $n \geq 1$. As a consequence, Lemma 6.3 still hold.

For all non-negative integers n, let $\mathcal{A}'_n$ be the set of allowed paths $p = p_s p_l$ in $G_n(\mathbf{w}')$ where p_l is a long n-segment and p_s is composed of consecutive short n-segments. We want to build an adic representation of $\mathbf{w}$ similarly to what is described in Section 5 but with the sets $\mathcal{A}'_n$ instead of $\mathcal{A}_n$. Observe that, as the symbol $\sharp$ does not occur in the label of any path of $\mathcal{A}'_n$ for $n \geq 1$, the adic representation built will really be an adic representation of $\mathbf{w}$ (not of $\mathbf{w}'$).

First, it is easy to be convinced that Lemma 5.1 can be adapted to the sets $\mathcal{A}'_n$. Moreover, as $p_{\mathbf{w}'}(n) = p_{\mathbf{w}}(n) + 1$ for all n, Lemmas 6.1 and 6.2 still hold and Lemma 6.3 implies that any path in $\mathcal{A}'_n$ contains a bounded number of short n-segments. Consequently, $\#\mathcal{A}'_n$ is bounded by some constant C.

Next, for all non-negative integers n, we define the alphabet $A'_n = \{0, 1, \cdots, \#\mathcal{A}'_n - 1\}$ and consider a bijection $\Theta_n : A'_n \to \mathcal{A}'_n$. We can extend it to an isomorphism by concatenation: $\Theta_n(ab) = \Theta_n(a)\Theta_n(b)$. Similarly to what is done in Definition 5.1, for all n we define the morphism $\tau_n : A'_{n+1} \to A'^*_n$ as the unique morphism that satisfies $\Theta_n \circ \tau_n = \psi_n \circ \Theta_{n+1}$. Observe that, as the edge $(\sharp \mathbf{w}[0, n-2], (\sharp, \mathbf{w}_{n-1}), \mathbf{w}[0, n-1])$ of $G_n(\mathbf{w}')$ does not appear in any n-segment of $\mathbf{w}'$ for $n \geq 1$, $\tau_0(a) \notin A'^*_0 x A'^*_0$ for all letters a in A'_1, where $\lambda_L(\Theta_0(x)) = \sharp$. Consequently, as the paths in $\mathcal{A}'_n$ are allowed, any word $\lambda_L \Theta_0 \tau_0 \tau_1 \tau_2 \cdots \tau_n(a)$ is a factor of $\mathbf{w}$, $a \in A'_{n+1}$.

Now let us prove that $\mathbf{w}$ is S-adic with $S = \{\gamma'\} \cup \{\tau_n \mid n \in \mathbb{N}\}$ and $\gamma' = \lambda_L \Theta_0$. Clearly, as $\#\mathcal{A}'_n \leq C$ for all n, Lemma 6.2 implies that $\#S < +\infty$. Let us show that $\mathbf{w}$ admits an S-adic representation.

As $\mathbf{w}$ is recurrent, all prefixes $\mathbf{w}[0, n]$ are left special factors of $\mathbf{w}'$. Consequently for all n there are some n-segments p of $\mathbf{w}'$ such that $o(p) = \mathbf{w}[0, n-1]$ and some of these n-segments have a left label that is a prefix of $\mathbf{w}$. Let $\mathcal{B}_n$ denotes the set of paths p in $\mathcal{A}'_n$ such that $o(p) = \mathbf{w}[0, n-1]$ and let $B_n = \Theta_n^{-1}(\mathcal{B}_n)$. For all non-negative integers n, $\tau_n(B_{n+1}) \in B_n^+ A'^*_n$. Let $(b_n)_{n \in \mathbb{N}}$ be a sequence of letters $b_n \in B_n$ such that $\lambda_L \Theta_n(b_n)$ is a prefix of $\mathbf{w}$ and $\tau_n(b_{n+1}) \in b_n A'^*_n$ (it is a consequence of the constructions that such a sequence exists). As any letter a_n in A'_n corresponds to a path containing a long n-segment, the length of $\tau_0 \tau_1 \cdots \tau_n(a_{n+1})$ tends to infinity with n for all sequences $(a_n)_{n \in \mathbb{N}}$ of letters $a_n \in A'_n$ (that is the ω-growth-Property). In particular, $\inf_{b \in B_{n+1}} |\tau_0 \tau_1 \cdots \tau_n(b)|$ tends to infinity as n increases so we have $\mathbf{w} = \lim_{n \to +\infty} \gamma' \tau_0 \tau_1 \cdots \tau_n(b_{n+1})$ and this ends the proof of S-adicity. $\square$

Remark 6.1. Observe that the adic representation built in the proof of the S-adicity property of Theorem 1.1 is not unique. Indeed, it depends on the bijections Θ_n that we choose and so it is unique up to exchanging the letters. Also, if $((\gamma', \tau_0, \tau_1, \dots), (b_n)_{n\in\mathbb{N}}) \in S^{\mathbb{N}} \times A'^{\mathbb{N}}$ is an adic representation of a sequence $\mathbf{w}$ as given by Theorem 1.1, then for all non-negative integers k, $((\gamma'_k, \tau_k, \tau_{k+1}, \dots), (b_n)_{n\geq k})$ is also an adic representation of $\mathbf{w}$ where $\gamma'_k = \lambda_L \Theta_k$. Actually we have $\gamma'_k = \gamma' \tau_0 \tau_1 \cdots \tau_{k-1}$. Observe that if we had worked with λ instead of λ_L, the map γ'_k would not have been a morphism as soon as $k \geq 1$. Indeed, for a given path p we have $\lambda(p) = \lambda_L(p)i(p)$. Moreover, if p_1 and p_2 are two paths, we also have $\lambda_L(p_1p_2) = \lambda_L(p_1)\lambda_L(p_2)$ and this is why γ'_k is a morphism (since Θ_k is bijective for all k). However if we replace λ_L by λ, the equation $\lambda(p_1p_2) = \lambda(p_1)\lambda(p_2)$ is true only if $i(p_1) = \varepsilon$, i.e., only if $i(p_1)$ has length 0. Consequently γ'_k would have been a morphism only if $k = 0$.

6.2. Properties of the morphisms

In this section we use the notation introduced in the proof of Theorem 1.1: $\mathbf{w}$ is an aperiodic and uniformly recurrent sequence over an alphabet A with an at most linear complexity, $\sharp$ is a letter that does not belong to A, $\mathbf{w}'$ is $\sharp\mathbf{w}$ and τ_n denotes the morphism from A'_{n+1} to A'^*_n. We also write $A' = \bigcup_{n\in\mathbb{N}} A'_n = \{0, \dots, D-1\}$ with $D < +\infty$ and we let N denote the smallest integer such that all the short segments already exist in G_N. More precisely, N is the smallest integer such that if p is a short m-segment for $m > N$ then $\psi_N \cdots \psi_{m-1}(p)$ is a short N-segment.

Lemma 6.4. *Let $((\gamma', \tau_0, \tau_1, \dots), (b_n)_{n\in\mathbb{N}}) \in S^{\mathbb{N}} \times A'^{\mathbb{N}}$ be an adic representation of $\mathbf{w}$ given by Theorem 1.1. Let a be a letter in A'_n such that $\Theta_n(a)$ contains a non-empty concatenation of short n-segments. If there is a letter c in A'_{n+1} and a word u in A'^+_n such that $\tau_n(c) \in uaA'^*_n$, then the subpath of $\Theta_{n+1}(c)$ that belongs to $\psi_n^{-1}(\Theta_n(u))$ is a concatenation of short $(n+1)$-segments.*

Proof. Indeed, let p_s be a short n-segment such that $\Theta_n(a)$ starts with p_s. By definition, $\psi_n^{-1}(p_s)$ contains an $(n+1)$-segment and for any path q in $\psi_n^{-1}(p_s)$, $o(q)$ is a left special vertex. Hence the path $\Theta_{n+1}(c)$ contains a proper subpath in $\psi_n^{-1}(\Theta_n(u))$ that ends with a left special vertex. By definition of the paths in $\mathcal{A}'_{n+1}$, this path has to be a concatenation of short $(n+1)$-segments. □

Next lemma states that for $n \geq N$, if a letter in A'_n occurs at a position $i \geq 2$ in an image by τ_n, then the corresponding path is a long n-segment that is not a proper subpath of another path in $\mathcal{A}'_n$.

Lemma 6.5. *Let $((\gamma', \tau_0, \tau_1, \dots), (b_n)_{n\in\mathbb{N}}) \in S^{\mathbb{N}} \times A'^{\mathbb{N}}$ be an adic representation of $\mathbf{w}$ given by Theorem 1.1. For all integers $n \geq N$, if $\tau_n(c) \in A'^+_n a A'^*_n$ for some letters $a \in A'_n$ and $c \in A'_{n+1}$, then $\Theta_n(a)$ is a long n-segment and there is no path in $\mathcal{A}'_n$ that contains $\Theta_n(a)$ as a proper subpath.*

Proof. First let us prove that if $\tau_n(c) \in A'^+_n a$ for some letters, then $\Theta_n(a)$ is a long n-segment (not a concatenation of short n-segments followed by a long n-segment). This is a direct consequence of Lemma 6.4. Indeed, suppose that $\Theta_n(a)$ contains a non-empty concatenation of short n-segments and that $\tau_n(c) = ua$ with $u \neq \varepsilon$. By definition, there is at least one subpath p of $\Theta_n(u)$ that is a long n-segment. However, by Lemma 6.4, the subpath of $\Theta_{n+1}(c)$ that belongs to $\psi_n^{-1}(\Theta_n(u))$ is a concatenation of short $(n+1)$-segments. Hence, there is a path in $\psi_n^{-1}(p)$ that is a subpath of a short $(n+1)$-segment s such that for all short n-segments r, $\psi_n(s) \neq r$ and this contradicts the definition of N.

Now let us prove that $\Theta_n(a)$ is not a proper subpath of any path in $\mathcal{A}'_n$. Let $\tau_n(c) = ua$, $u = u_1 \cdots u_\ell$ and let p and q be the subpaths of $\Theta_{n+1}(c)$ such that $\psi_n(p) = \Theta_n(u_1 \cdots u_\ell)$ and $\psi_n(q) = \Theta_n(a)$. Since $\psi_n(q)$ is a long n-segment (by definition of N), we deduce from Remark 5.3 that $o(q)$ is not left special and that its unique incoming edge e_1 in G_n is the last edge of p. Moreover, by definition of $\mathcal{A}'_n$, $\psi_n(p)$ ends with a long n-segment l. Now suppose that $b \in A'_n$ is such that $\Theta_n(a)$ is a proper subpath of $\Theta_n(b)$. By definition of $\mathcal{A}_n$, there exists $k \geq 1$ and some short n-segments $q_1, \dots, q_k$ such that $\Theta_n(b) = q_1 \cdots q_k \Theta_n(a)$. Since $\Theta_n(b)$ is allowed (by definition), $\psi_n^{-1}(\Theta_n(b))$ is non-empty. Hence $o(q)$ has an incoming edge e_2 in G_n that is the last edge of a path in $\psi_n^{-1}(q_k)$. Since q_k is short and l is long, these segments do not have any common edge. Hence $e_1 \neq e_2$ and $o(q)$ has to be left special although it is not. $\square$

Property 1. *Let $((\gamma', \tau_0, \tau_1, \dots), (b_n)_{n\in\mathbb{N}}) \in S^{\mathbb{N}} \times A'^{\mathbb{N}}$ be an adic representation of $\mathbf{w}$ given by Theorem 1.1. For all integers $n \geq N$ and all letters $a \in A'_n$ and c in A'_{n+1}, $\tau_n(c) \notin A'^*_n a A'^*_n a A'^*_n$.*

Proof. Suppose $\tau_n(c) = uavaw$ with $a \in A'_n$ and $u, v, w \in A'^*_n$. From Lemma 6.5, the paths $\Theta_n(a)$ and $\Theta_n(v_i)$, $1 \leq i \leq |v|$, are long n-segments. Moreover, from Remark 5.3 we can deduce that the interior vertices of $\Theta_{n+1}(c)$ that admit respectively $o(\Theta_n(a)) = i(\Theta_n(v_{|v|}))$, $i(\Theta_n(a)) = o(\Theta_n(v_1))$ and $o(\Theta_n(v_i))$ for $2 \leq i \leq |v| - 1$ as prefixes are not left special. Hence the subpath of $\Theta_{n+1}(c)$ that belongs to $\psi_n^{-1}(\Theta_n(au_2a))$ is a path p in $G_{n+1}(\mathbf{w}')$ such that $i(p) = o(p)$ and that does not go through any left special vertex. Hence it is inaccessible from vertices that are not in it. As no n-segment of $\mathbf{w}'$ contains the "added edge" $(\sharp\mathbf{w}[0, n-2], (\sharp, \mathbf{w}_{n-1}), \mathbf{w}[0, n-1])$, this loop is composed of edges of $G_{n+1}(\mathbf{w})$. Hence this last graph is not strongly connected and this contradicts the uniform recurrence. $\square$

Property 2. *Let $((\gamma', \tau_0, \tau_1, \dots), (b_n)_{n\in\mathbb{N}}) \in S^{\mathbb{N}} \times A'^{\mathbb{N}}$ be an adic representation of $\mathbf{w}$ given by Theorem 1.1. For all integers $n \geq N$, if there is $a \in A'_n$, $u = u_1 u_2 \cdots u_\ell \in A'^+_n$ and $c \in A'_{n+1}$ such that $\tau_n(c) \in uaA'^*_n$, then for all letters $d \in A'_{n+1}$, $\tau_n(d) \in (A'^*_n \setminus A'^*_n a A'^*_n) \cup (A'_n u_2 \cdots u_\ell a A'^*_n)$. Moreover, if $\Theta_n(u_1)$ is a long n-segment such that there is no path p in $\mathcal{A}'_n$ that contains it as a proper subpath, then $\tau_n(d) \in (A'^*_n \setminus A'^*_n a A'^*_n) \cup (u_1 \cdots u_\ell a A'^*_n)$ for all letters $d \in A'_{n+1}$.*

Proof. For $i = 1, \dots, \ell$, let p_i be the subpath of $\Theta_{n+1}(c)$ such that $\psi_n(p_i) = \Theta_n(u_i)$. Let us also write p the subpath of of $\Theta_{n+1}(c)$ such that $\psi_n(p) = \Theta_n(a)$. From Lemma 6.5 we know that $\Theta_n(a)$ and $\Theta_n(u_i)$ are long n-segments for $i = 2, \dots, \ell$. Moreover, from Remark 5.3 we deduce that the vertices $i(p_i)$, $i = 1, \dots \ell$, are not left special. Consequently, $p_2 \cdots p_\ell p$ does not contain any left special vertex so it is the only path in $G_{n+1}(\mathbf{w}')$ from $o(p_2)$ to $i(p_\ell)$ (supposing that we do not consider paths containing twice the vertex $o(p_2)$). Roughly speaking, starting from $i(p)$ and going backwards through the path $p_2 \cdots p_\ell a$, the only right special vertex (actually left special but since we are going backwards we see them as right special) that we go through is the first one $i(p)$ so the letters $u_2, u_3, \dots, u_\ell$ and a are "glued together" in any image by τ_n. Moreover, as $o(p_2)$ is not left special, there is a unique long n-segment l such that $i(l) = o(p_2)$. However, as there might be several letters b in A'_n such that l is a subpath of $\Theta_n(b)$ (those with a different concatenation of short n-segments), the first letter of an image $\tau_n(d)$ containing a is unique only if $\Theta_n(u_1)$ is a long n-segment and there is no path in $\mathcal{A}'_n$ that contains it as a proper subpath. $\square$

Property 3. *Let $((\gamma', \tau_0, \tau_1, \dots), (b_n)_{n\in\mathbb{N}}) \in S^{\mathbb{N}} \times A'^{\mathbb{N}}$ be an adic representation of $\mathbf{w}$ given by Theorem 1.1. For all integers $n \geq N$, all letters $a_1, \dots, a_k$ in A'_n and all letters $c_1, \dots, c_k$ in A'_{n+1}, $(\tau_n(c_1), \dots, \tau_n(c_k)) \notin A'^*_n a_1 A'^*_n a_2 A'^*_n \times A'^*_n a_2 A'^*_n a_3 A'^*_n \times \cdots \times A'^*_n a_{k-1} A'^*_n a_k A'^*_n \times A'^*_n a_k A'^*_n a_1 A'^*_n$.*

Proof. From Lemma 6.5 we know that $\Theta_n(a_i)$ is a long n-segment for $i = 1, \dots, k$ and no other path in $\mathcal{A}'_n$ contains them as subpaths. Then, from Property 2 we deduce that all letters a_i occur twice in all images $\tau_n(c_i)$, $i = 1, \dots, k$. Indeed, consider for example the letter a_1. Property 2 and $\tau_n(c_k) \in A'^*_n a_k A'^*_n a_1 A'^*_n$ imply that in all images containing a_1, a_1 is preceded by a_k. With the same reasoning, a_k is always preceded by a_{k-1}, so by $a_{k-2}, a_{k-3}, \dots, a_2$ and then by a_1. From Property 1, this is forbidden. $\square$

The following two properties are true only under the additional condition that there is no short n-segment for all n. In particular, this implies that the alphabets we work with are simply those defined in Section 5, i.e., $\mathcal{A}'_n = \mathcal{A}_n$.

Property 4. *Let $((\gamma', \tau_0, \tau_1, \dots), (b_n)_{n\in\mathbb{N}}) \in S^{\mathbb{N}} \times A'^{\mathbb{N}}$ be an adic representation of $\mathbf{w}$ given by Theorem 1.1. If there is no short segments, then for any non-negative integer r there is an integer $s > r$ such that all letters a in A'_r occur in $\tau_r \tau_{r+1} \cdots \tau_s(c)$ for all letters c in A'_{s+1}.*

Proof. Indeed, as the sequence $\mathbf{w}$ is uniformly recurrent, for all integers ℓ there is an integer k_ℓ such that any factor of length ℓ occurs in any factor of length k_ℓ. Moreover, for all n-segments p and q, $\lambda(p)$ is not a factor of $\lambda(q)$ (which is not true for the paths in $\mathcal{A}'_n$ when there are some short segments), where $\lambda(p)$ denotes the full label of p. For all non-negative integers n, let M_n and m_n be respectively the maximal and minimal lengths of an n-segment. Let i be a non-negative integer. As the length of $\gamma' \tau_0 \tau_1 \cdots \tau_n(a_{n+1})$ tends to infinity with n for all sequences $(a_n)_{n\in\mathbb{N}}$ of letters $a_n \in A'_n$, there is a non-negative integer j such that all factors of length at most $M_i + i$ occurs in all factors of length at least $m_j + j$. Consequently, since the label of any i-segment is not a factor of the label of another i-segment, all i-segments occur in the image under $\psi_i \cdots \psi_{j-1}$ of all j-segments so $\tau_i \cdots \tau_{j-1}(c)$ contains all letters in $\mathcal{A}'_i$ for all letters c in $\mathcal{A}'_j$. $\square$

As a corollary of Property 4, for all non-negative integers r, the sequence

$$\mathbf{w}_r = \lim_{n\to+\infty} \tau_r \tau_{r+1} \cdots \tau_n(b_n)$$

is uniformly recurrent (see Lemma 7 in [10]), whenever there is no short segments in $\mathbf{w}'$ and $((\gamma', \tau_0, \tau_1, \dots), (b_n)_{n\in\mathbb{N}}) \in S^{\mathbb{N}} \times A'^{\mathbb{N}}$ is an adic representation of $\mathbf{w}$ given by Theorem 1.1.

For the proof of next property, we need to recall some basic notions of graph theory. Let G be a graph. A path p in G is a *cycle* if its extremities are equal. Let v be a vertex of graph G. The *neighbors* of v are the vertices u such that there is an edge between u and v.

A *tree* is an undirected graph in which any two vertices are connected by exactly one simple path, i.e., a path that does not pass twice through a same vertex. In other words, any connected graph with no cycle (except the cycles $(u,v)(v,u)$ where u and v are vertices) is a tree. A tree is said to be *rooted* if one particular vertex v_0 is designated the *root*. In this case, the vertices v can be ordered with respect to the length of the unique simple path between v_0 and v. If the length of the simple path between v_0 and v is i, we say that v is a vertex of *level* i. The *children* of a vertex v of level i are the neighbors of level $i+1$ of v. A vertex u is a *successor* of a vertex v if there is a sequence of vertices $v = v_1, v_2, \dots, v_k = u$ such that v_{i+1} is a child of v_i for all i, $1 \leq i \leq k-1$. The set of successors of v in G is denoted by $succ_G(v)$. In the same idea, the *parent* of v is the neighbor of level $i-1$ of v and the *ancestors* of v are the vertices u such that $v \in succ_G(u)$. A vertex v is a *leaf* if it has no child.

A *forest* is an undirected graph whose connected component are trees. When the trees of a forest F are rooted, the *roots* (resp. the *leaves*) of F are the respective roots (resp. the respective leaves) of its connected components.

Property 5. *Let* $((\gamma', \tau_0, \tau_1, \dots), (b_n)_{n\in\mathbb{N}}) \in S^{\mathbb{N}} \times A'^{\mathbb{N}}$ *be an adic representation of* $\mathbf{w}$ *given by Theorem 1.1. If there is no short segments, then for all* n*,* τ_n *belongs to* T^* *with* $T = \{G\} \cup \{E_{ij} \mid i,j \in A'\} \cup \{M_i \mid i \in A'\}$ *a set of morphisms such that:*

- ▶ $G(0) = 10$ *and* $G(i) = i$ *for all letters* $i \neq 0$*;*
- ▶ E_{ij} *exchange* i *and* j *and fix the other letters;*
- ▶ M_i *maps* i *to* 0 *and fix the other letters.*

Proof. Let n be an integer. The main idea to decompose the morphism τ_n is the following. Let F be the graph whose set of vertices are the couples (a,n) with a in A'_n and the couples $(c, n+1)$ with c in A'_{n+1}; the set of edges is defined as follows:

- for $c \in A'_{n+1}$ and a in A'_n, there is an edge between $(c, n+1)$ and (a,n) if $\tau_n(c) \in A_n^* a$;
- for $a, b \in A'_n$, there is an edge between (a,n) and (b,n) whenever there is a letter c in A'_{n+1} such that ba occurs in $\tau_n(c)$.

As there is no short segments in $\mathbf{w}'$, Properties 2 (last part) and 3 imply that F is a forest such that the number of connected components (that are trees) of F is the number of letters a in A'_n such that $\tau_n(c) \in aA_n'^*$ for some letter c in A'_{n+1}. We suppose that the root of such a tree is the vertex (a,n). Consequently, the leaves of F are the vertices $(c, n+1)$ and we can check that the set of images in $\tau_n(A'_{n+1})$ is the set of words $a_1 \cdots a_k$, $k \geq 0$, $a_1, \dots, a_k \in A'_n$ being the respective first components of the vertices of a simple path in F from a root to the parent of a leaf.

Now let us explain how we can build τ_n with F. The idea is to start from the leaves, to move towards the roots and to build τ_n reading the letters on the vertices, i.e., the first components of them. The first step (from the leaves to their respective parents) is simply to map each letter c in A'_{n+1} to the last letter of $\tau_n(c)$. This can be realized with the morphisms E_{ij} and M_i. Indeed, for any n-segment p, let $\chi(p) = \{Xx \mid X =$

$i(p)$ and $x \in A$ such that $Xx \in L(\mathbf{w})\}$. As a segment is completely determined by its last edge, there is a bijection between the set $\mathcal{A}'_{n+1}$ of $(n+1)$-segments and the set $\{Xx \in \chi(p) \mid Xx \text{ is left special and } p \in \mathcal{A}'_n\}$. We write

$$\mathcal{A}'_{n+1} \cong \{Xx \in \chi(p) \mid Xx \text{ is left special and } p \in \mathcal{A}'_n\}\,. \tag{4}$$

Let p be a n-segment and let $k(p)$ be the number of vertices Xx in $\chi(p)$ that are left special. If $k(p) = 1$, we deduce from Equation 4 that there is a unique $(n+1)$-segment q_p such that

$$\tau_n \circ \Theta_{n+1}^{-1}(q_p) \in A_n'^{*}\Theta_n^{-1}(p).$$

Consequently, there is a bijection between P^* and $\{q_p \mid p \in P\}^*$ with

$$P = \{p \in \mathcal{A}'_n \mid \exists! Xx \in \chi(p) \text{ that is left special}\}\,.$$

This bijection is realized by a bijective and letter-to-letter morphism $\mathcal{E}$ and it is clear that such a morphism can be decomposed in a finite product of morphisms E_{ij} (see for instance Lemma 2.2 in [21]).

Now, if $k(p) > 1$, Once again we deduce from Equation 4 that there are $k(p)$ $(n+1)$-segments $q_{p,1}, \ldots, q_{p,k(p)}$ such that

$$\tau_n \circ \Theta_{n+1}^{-1}(q_{p,i}) \in A_n'^{*}\Theta_n^{-1}(p)$$

for all i, $1 \le i \le k(p)$. For all i, $1 \le i \le k(p)$, the letter $\Theta_{n+1}^{-1}(q_{p,i})$ must be mapped to $\Theta_n^{-1}(p)$. This is realized by the following product of morphisms:

$$\mathcal{M} = \prod_{\substack{p \in \mathcal{A}'_n \text{ such} \\ \text{that } k(p)>1}} E_{0\Theta_n^{-1}(p)} \left(\prod_{1 \le i \le k(p)} M_{\Theta_{n+1}^{-1}(q_{p,i})} \right) E_{0\Theta_n^{-1}(p)}.$$

Observe that, by construction, the morphisms $\mathcal{E}$ and $\mathcal{M}$ respectively act on disjoints subset of A'_{n+1}. Consequently, we have

$$\mathcal{E} \circ \mathcal{M}(A'_{n+1}) = \mathcal{M} \circ \mathcal{E}(A'_{n+1})$$

and this morphism realizes the step from the leaves of F to their respective parents.

Now let us show that we can keep moving towards the roots of F and build τ_n reading the letters on the vertices. Let us define the morphism $\tau_{\text{temp}} = \mathcal{E} \circ \mathcal{M}$ and the graph $F_{\text{temp}} = F$. Since we have already build the morphism realizing the step from the leaves to their respective parents, we remove them (the leaves) from F_{temp}. Once this is done, their might be some new leaves in F_{temp} that are also roots of F_{temp}. For these vertices (a, n), this means that for any child $(c, n+1)$ of (a, n) in F we have $\tau_{\text{temp}}(c) = \tau_n(c) = a$ (otherwise there would be an edge between (a, n) and another vertex (b, n)). Hence the work is done for these letters so we remove the corresponding vertices from F_{temp}. Consequently, the remaining vertices in F_{temp} correspond to the letters a in A'_n that occur in images $\tau_n(c)$ of length at least 2. Observe that since we have only removed some leaves from F_{temp}, the graph is still a forest and we can repeat the process until F_{temp} is empty. This is formalized by the algorithm below.

Algorithm:
While F_{temp} is non-empty:

1. Consider a leaf (a, n) in F_{temp}. Let (b, n) be the parent of (a, n) in F_{temp}. Remove (a, n) from F_{temp}.
2. Replace τ_{temp} by $E_{0a} \circ E_{1b} \circ G \circ E_{1b} \circ E_{0a} \circ \tau_{\text{temp}}$.
3. If (b, n) is a root of F_{temp}, remove (b, n) from F_{temp}.

This algorithm clearly stops since any vertex of F can be reached (so removed from F_{temp}) in a finite number of steps. Moreover, when it stops, we have $\tau_{\text{temp}} = \tau_n$ (by construction of F). $\square$

7. Conclusions

First it is easy to see that Theorem 1.2 is a consequence of Theorem 1.1. However, the periodic case can be included in Theorem 1.2 since the only necessary condition that we have in this one is the ω-growth Property that can be satisfied for ultimately periodic sequences (see Section 3.1).

Next, one can regret that Properties 4 and 5 do not hold in general. Maybe it would be interesting to study the meaning (in terms of dynamical systems for example) of the existence of short segments. However, we can show that the conditions listed in Properties 1 to 5 are not sufficient to get an S-adic sequence with an at most linear complexity. Indeed, consider the S-adic sequence studied in Example 3.1. This one does not have an at most linear complexity as soon as the sequence of exponents $(k_n)_{n\in\mathbb{N}}$ is unbounded (see Lemma 3.5). Moreover, we can decompose the 2 morphisms α and μ appearing in the adic representation with 3 morphisms that satisfy the conditions expressed in Properties 1 to 5: $\alpha = \beta \circ \alpha'$ and $\mu = \beta \circ \mu'$ with

$$\alpha' : \begin{cases} a \mapsto abc \\ b \mapsto d \end{cases} \qquad \mu' : \begin{cases} a \mapsto ac \\ b \mapsto db \end{cases} \qquad \beta : \begin{cases} a \mapsto a \\ b \mapsto a \\ c \mapsto b \\ d \mapsto b \end{cases}$$

One last thing is that Theorem 1.1 can easily be extended to bi-infinite sequences (i.e., to elements of $A^{\mathbb{Z}}$). Indeed, for any sequence $\mathbf{w}$ in $A^{\mathbb{Z}}$ we consider the sequence $\mathbf{w}' = \mathbf{w}[-\infty, -1].\sharp\mathbf{w}[0, +\infty]$ (recall that the dot determines the position of the letter indexed by 0). Then any prefix of $\mathbf{w}[0, +\infty]$ is left special and any suffix of $\mathbf{w}[-\infty, -1]$ is right special. For any non-negative integer n, we consider two sets $\mathcal{A}'_n$ and $\mathcal{B}'_n$ whose elements are respectively some paths qp and st with q a bounded concatenation of left short n-segments, q a left long n-segment, s a right long n-segment and t a bounded concatenation of right short n-segments. Then we define some morphisms $\tau_n : A'_{n+1} \cup B'_{n+1} \to A'^*_n \cup B'^*_n$ and $\sigma_n : A'_{n+1} \cup B'_{n+1} \to A'^*_n \cup B'^*_n$ as defined in Definition 5.1 such that

▶ the morphisms τ_n are coding the "left paths" of order $n+1$ by the "left paths" of order n and fix the letters in B'_n;

▶ the morphisms σ_n are coding the "right paths" of order $n+1$ by the "right paths" of order n and fix the letters in A'_n.
Then there are two sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ of letters $a_n \in A'_n$ and $b_n \in B'_n$ such that $\mathbf{w} = \lim_{n\to+\infty} \gamma'\tau_0\sigma_0\tau_1\sigma_1 \cdots \tau_n\sigma_n(b_{n+1}^{\omega}.a_{n+1}^{\omega})$, with γ' as defined in the proof of the S-adicity property in Theorem 1.1.

Aknowledgements: I would like to thank Fabien Durand and Gwénaël Richomme for introducing me to this subject and for their numerous advices and remarks. I also want to thank Michel Rigo for his remarks.

References

[1] J.-P. Allouche, J. Shallit, Automatic sequences: Theory, applications, generalizations, Cambridge University Press, Cambridge, 2003.

[2] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité $2n + 1$, Bull. Soc. Math. France 119 (1991) 199–215.

[3] V. Berthé, Autour du système de numération d'Ostrowski, Bull. Belg. Math. Soc. Simon Stevin 8 (2001) 209–239, Journées Montoises d'Informatique Théorique (Marne-la-Vallée, 2000).

[4] V. Berthé, C. Holton, L. Q. Zamboni, Initial powers of Sturmian sequences, Acta Arith. 122 (2006) 315–347.

[5] V. Berthé, M. Rigo (eds.), Combinatorics, Automata and Number Theory, vol. 135 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2010.

[6] J. Cassaigne, Special factors of sequences with linear subword complexity, in: Developments in language theory, II (Magdeburg, 1995), World Sci. Publ., River Edge, NJ, 1996, pp. 25–34.

[7] J. Cassaigne, Complexité et facteurs spéciaux, Bull. Belg. Math. Soc. Simon Stevin 4 (1997) 67–88, Journées Montoises (Mons, 1994).

[8] J. Cassaigne, S-adiques, [Online], private communication (2009).
URL `http://www.lirmm.fr/arith/wiki/PytheasFogg/S-adiques`

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

[10] F. Durand, Linearly recurrent subshifts have a finite number of non-periodic subshift factors, Ergodic Theory Dynam. Systems 20 (2000) 1061–1078.

[11] F. Durand, Corrigendum and addendum to: "Linearly recurrent subshifts have a finite number of non-periodic subshift factors", Ergodic Theory Dynam. Systems 23 (2003) 663–669.

[12] F. Durand, B. Host, C. Skau, Substitutional dynamical systems, Bratteli diagrams and dimension groups, Ergodic Theory Dynam. Systems 19 (1999) 953–993.

[13] A. Ehrenfeucht, K. P. Lee, G. Rozenberg, Subword complexities of various classes of deterministic developmental languages without interactions, Theor. Comput. Sci. 1 (1975) 59–75.

[14] A. Ehrenfeucht, K. P. Lee, G. Rozenberg, On the number of subwords of everywhere growing DT0L languages, Discrete Math. 15 (1976) 223–234.

[15] S. Ferenczi, Les transformations de Chacon : combinatoire, structure géométrique, lien avec les systèmes de complexité $2n + 1$, Bull. Soc. Math. France 123 (1995) 271–292.

[16] S. Ferenczi, Rank and symbolic complexity, Ergodic Theory Dynam. Systems 16 (1996) 663–682.

[17] N. P. Fogg, Substitutions in dynamics, arithmetics and combinatorics, vol. 1794 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2002, edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[18] M. Lothaire, Algebraic combinatorics on words, vol. 90 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2002.

[19] M. Morse, G. A. Hedlund, Symbolic dynamics II. Sturmian trajectories, Amer. J. Math. 62 (1940) 1–42.

[20] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, in: Automata, languages and programming (Antwerp, 1984), vol. 172 of Lecture Notes in Comput. Sci., Springer, Berlin, 1984, pp. 380–389.

[21] G. Richomme, Conjugacy and episturmian morphisms, Theoret. Comput. Sci. 302 (2003) 1–34.