Some improvements of the S-adic conjecture (extended abstract)

Julien Leroy

Laboratoire Amiénois de Mathématiques Fondamentales et Appliquées
CNRS-UMR 6140, Université de Picardie Jules Verne, 33 rue Saint Leu,
80039 Amiens Cedex 01, France
(julien.leroy@u-picardie.fr)

Abstract. In [14], 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.

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 [17]) 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 [17]). 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 [16] and Chapter 6 of [15] for surveys on these sequences. In particular, it is proved in [4] that all these sequences can be generated with only four morphisms.

Many other known sequences have a low complexity. By "low complexity" we usually mean "complexity bounded by a linear function". Fixed points of primitive substitutions, automatic sequences, linearly recurrent sequences (see [11]) and Arnoux-Rauzy sequences are examples of sequences with an at most linear 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 a letter-to-letter morphisms of fixed points of uniform substitutions (see [1]), F. Durand proved it in [9] and [10] for linearly recurrent sequences and P. Arnoux and G. Rauzy proved it in [2] for the so-called Arnoux-Rauzy sequences. Following [15], a sequence $\mathbf{w}$ with previous property is said to be S-adic (where S refers to the set of morphisms).

As mentioned in [15], 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 a 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 [18]) and moreover, J. Cassaigne recently showed that there exists a set of morphisms S 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 [6]).

In [14], 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 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 2. Then, adapting his methods, we improve this result by proving Theorem 1 and give some properties on the S-adic representation that could help stating the condition C.

Theorem 1. *Let $\mathbf{w}$ be a sequence over an alphabet A. If $\mathbf{w}$ has an at most linear complexity then $\mathbf{w}$ is a S-adic sequence satisfying Properties 1 to 4 (see Section 5.2) 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 (this property will be called the ω-growth Property).*

Theorem 2 (Ferenczi [14]). *Let $\mathbf{w}$ be a uniformly recurrent sequence over an alphabet A with an at most linear complexity. There exist a finite number of morphisms σ_i, $1 \leq i \leq c$, over an alphabet $D = \{0, \ldots, d-1\}$, an application α from D to A and an infinite sequence $(i_n)_{n \in \mathbb{N}} \in \{1, \ldots, c\}^{\mathbb{N}}$ 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. In Section 2, we recall the definition of S-adicity and present some results and examples about the conjecture and about the complexity of some particular S-adic sequences. In particular, using a similar technique as in [12] we give an upper bound for the complexity of some S-adic sequences. Section 3 talks about Rauzy graphs. We recall their definition and explain how they evolve. Section 4 presents Ferenczi's methods in a general case and we present a sketch or the proof of Theorem 1 in Section 5. Section 6 concludes the paper with some remarks. Observe that due to lack of space, most of the proves are not included to the paper.

2 S-adicity and factor complexity

For basic notions about combinatorics on words, we refer to books like [1], [5], [15], [16] or [19]. In this paper, sequences (or infinite words) are denoted by bold letters and finite words by normal letters.

The notion of S-*adic* sequence generalizes the notion of fixed point of morphism. Let $\mathbf{w}$ be a sequence over a finite alphabet 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 $A_0 = A$ 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 1 (Cassaigne). *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 $\mathbf{w} = \lim_{n\to+\infty} \varphi\sigma_{\mathbf{w}_1}\sigma_{\mathbf{w}_2}\cdots\sigma_{\mathbf{w}_n}(l^\omega)$. □

A word u has an *occurrence* at position i in a (finite or infinite) word w (or *occurs* in w) if $w_iw_{i+1}\cdots w_{i+|u|-1} = u$. The factor $w_iw_{i+1}\cdots w_j$ of a word w, $i \geq 1$, $j \leq |w|$, is denoted by $w[i,j]$. 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 $\mathbf{w}$, i.e., $p_{\mathbf{w}}(n) = \#L_n(\mathbf{w})$. See Chapter 4 of [5] for a survey on this function.

2.1 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. In [7], Cassaigne proved Theorem 3 which is a key point in the proof of Theorem 1.

Theorem 3 (Cassaigne). *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 [8]). 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 a word u in $L(\mathbf{w})$, if δ^+u (resp. δ^-u) denotes the number of letters a in A such that ua (resp. au) is in $L(\mathbf{w})$, we have

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

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

G. A. Hedlund and M. Morse proved in ([17]) that either $p(n)$ is ultimately constant (and corresponds to ultimately periodic sequences) or grows at least like $n+1$. We could easily show that ultimately periodic sequences are S-adic with $\#S = 2$. *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 $\{a,b\}$ defined by $\tau_a(a) = a$, $\tau_a(b) = ab$, $\tau_b(a) = ba$, $\tau_b(b) = b$, $\tau'_a(a) = a$, $\tau'_a(b) = ba$, $\tau'_b(a) = ab$ and $\tau'_b(b) = b$. It is proved in [4] that all Sturmian sequences are $\{\tau_a, \tau'_a, \tau_b, \tau'_b\}$-adic 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 coefficients of the continued fraction of α and by the *Ostrowski expansion* of ρ (see [3] for more details about the Ostrowski expansions).

Proposition 2 is a generalization of a result due to Durand (Proposition 2.1 in [10] states it for $d(n) = D \in \mathbb{N}$ for all n).

Proposition 2. *Let* $\mathbf{w}$ *be a* S*-adic sequence over an alphabet* A *such that all morphisms in* S *are non-erasing. Suppose that* $\inf_{c \in A_{n+1}} |\sigma_0\sigma_1 \cdots \sigma_n(c)|$ *tends to infinity and there exists a function* $d : \mathbb{N} \to \mathbb{R}^+$ *such that*

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

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

The proof is similar to the proof of Proposition 2.1 in [10].

If all morphisms σ in S are uniform (that is $|\sigma(a)| = k$ for all letters a), then Proposition 2 holds for $d(n) = D = \max_{\sigma \in S} |\sigma(a)|$ and it is a corollary of it that automatic sequences (that are images under letter-to-letter morphisms of fixed points of uniform morphisms) have an at most linear complexity.

Recall that a 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 2, Durand proved that any primitive S-adic sequence has an at most linear complexity. He also gave a S-adic characterization of linearly recurrent sequences stating that *a sequence is linearly recurrent if and only if it is a primitive and proper* S*-adic sequence.*. Recall that a *linearly recurrent* sequence is a uniformly recurrent sequences for which there is a constant K such that the gaps between two successive occurrences of a factor u is bounded by $K|u|$ and a S-adic sequence is *proper* if for all morphisms σ_n in $(\sigma_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}$, there are two letters a and b in A_n such that $\sigma(c) \in aA_n^*b$ for all letters c in A_{n+1}.

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

As we have seen in Section 2.1, 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 every letter a occurs in every image $\sigma(b)$). 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 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 fixed points with quadratic complexity (see [18])).

Next example presents a set S of morphisms such that the S-adic sequences have or not an at most linear complexity depending on their directive words.

Example 1. Consider $S = \{\alpha, \mu\}$ with α defined by $\alpha(a) = aab$ and $\alpha(b) = b$ and τ 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 1. *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.*

2.3 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 a S-adic sequence satisfying condition C.

In the case of fixed points of morphisms $\sigma^\omega(a)$, it is proved in [18] that the complexity function can only have five asymptotic behaviors: $O(1)$, $O(n)$, $O(n\log n)$ $O(n\log\log n)$ and $O(n^2)$ and that the class of highest complexity $O(n^2)$ 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 1).

In Theorem 1 (and it was already the case in Ferenczi's paper [14]), we show that a sequence with an at most linear complexity is a 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 1).

Although this property is not necessary to have a low complexity (for example, the fixed point $\gamma^\omega(0)$ with γ defined by $\gamma(0) = 0010$ and $\gamma(1) = 1$ has an at most linear complexity (see [13])), 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 $a_n \in A_n$. Note that it is clear from Lemma 1 that the ω-growth Property is not sufficient and that we have to take care not only of the set S but also of the directive words of S-adic sequences.

2.4 Beyond linearity

A consequence of Proposition 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 morphisms (see [18]). However we can still hope to have 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 morphisms σ in S and all letters a in A_n (see Proposition 3 below). Techniques are similar to those used in [12] for D0L systems.

Proposition 3. *Let* $\mathbf{w}$ *be a* S*-adic sequence over an alphabet* A*. Suppose that* $|\sigma(a)| \geq 2$ *for all* σ *in* S *and all letters* a*. Then there is a constant* C *such that* $p(n) \leq Cn \log n$ *for all integers* n*.*

Example 2 shows that this bound is the best one we can obtain.

Example 2. Let β be the morphism on $A = \{a, b, c\}$ 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 condition in Proposition 3 and we know from [18] that its complexity function satisfies $C_1 n \log(n) \leq p_{\mathbf{w}}(n) \leq C_2 n \log(n)$, with $C_1, C_2 > 0$.

3 Rauzy graphs

Proof of Theorem 1 is based on the evolution of Rauzy graphs. In this section, we recall this notion and explain how they evolve. First let us introduce some notation.

Let p be a path in a directed graph $G = (V, E)$ (V is the set of vertices and E the set of edges). We denote the starting vertex by $o(p)$ and the ending vertex by $i(p)$. These vertices are called the *extremities* of p and the other vertices are called *interior vertices*. The length of p is the number of edges composing it.

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) is the directed graph $(V(n), E(n))$, where
- the set $V(n)$ of vertices is the set $L_n(\mathbf{w})$ 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 four following notation exist:

$$u \xrightarrow{b} v \quad u \xrightarrow[a]{} v \quad u \xrightarrow[a]{b} v \quad u \xrightarrow{av} v.$$

For an edge $(u, (a, b), v) = u \xrightarrow[a]{b} v$, let us call a its *left label*, b its *right label* and $ub = av$ its *full label*. Same definitions hold for labels of paths (left and right labels being words of same length as the considered path). In this paper we will mostly consider left labels.

Remark 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).

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.

By definition of Rauzy graphs, for all words u in $L(\mathbf{w})$ and all integers $n > |u|$, there is a path p in $G_n(\mathbf{w})$ whose full label is u. The contrary is not true, i.e., the full label of a path in $G_n(\mathbf{w})$ is not always a factor of $\mathbf{w}$. Hence, a path in a Rauzy graph is said to be *allowed* if its full label is a word in $L(\mathbf{w})$. Observe that any path $p = (v_0, (a_1, b_1), v_1) \cdots (v_{\ell-1}, (a_\ell, b_\ell), v_\ell)$ that does not contain any subpath $(v_i, (a_{i+1}, b_{i+1}), v_{i+1}) \cdots (v_{j-1}, (a_j, b_j), v_j)$, $i \geq 1$, $j \leq \ell - 1$ with v_i left special and v_j right special is trivially allowed. Moreover, the following trivially holds.

Proposition 4. *Let G_n be a Rauzy graph of order n and let v be a vertex of G_n. For all paths p and p' of length $\ell \leq n$ such that $o(p) = i(p') = v$, the left label of p and the right label of p' are respectively prefix and suffix of v.*

Proof of Theorem 1 is based on the evolution of Rauzy graphs (i.e., going from G_n to G_{n+1}). As edges of G_n are exactly the words of 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 replace $\underset{a}{\rightarrow} v \overset{b}{\rightarrow}$ by $av \overset{b}{\underset{a}{\rightarrow}} vb$ whenever $avb \in L(\mathbf{w})$.

4 Segments, circuits and morphisms

Let $\mathbf{w}$ be an aperiodic and uniformly recurrent sequence over an alphabet A. In this section, we explain Ferenczi's methods to construct an adic representation of any factor of $\mathbf{w}$. In his paper [14], 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 notion of *left n-segment* that is a little bit different. However constructions are mostly the same as described in [14].

For each $n \in \mathbb{N}$, a *left* (reps. *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 $i(p)$. When it is not explicitly stated, *n-segment* means *left n-segment*. Observe that any 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 number of left special vertices in G_n, there exists only a finite number of n-segments.

For sequences with a "reasonably low" complexity, the number of left special factors increases much slower than the complexity. Consequently, we expect that the maximal length of n-segments will grow to infinity. Then, as the left labels fo n-segments are factors of the sequence, 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 (due to the uniform recurrence). 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 of $G_n(\mathbf{w})$ in the following way. For each path p with left label u, $\psi_n(p)$ is the set of paths p' in $G_{n+1}(\mathbf{w})$ whose left label is u and such that $o(p')$ and $i(p')$ admit respectively $o(p)$ and $i(p)$ as a prefix. Roughly speaking, for a path p in $G_n(\mathbf{w})$, $\psi_n(p)$ is the set of paths in $G_{n+1}(\mathbf{w})$ corresponding to p.

Lemma 2 here below — and also Lemmas 3, 4, 5 and 6 in next sections — were already proved in [14]. However, all these lemmas were parts of the proof of Theorem 2. In this paper, we decided to structure the proof to smarten it up.

Lemma 2 (Ferenczi). *Let* $\mathbf{w}$ *be a sequence over an alphabet* A*. Any* $(n+1)$*-segment of* $\mathbf{w}$ *is in the image under* ψ_n *of a concatenation of* n*-segments of* $\mathbf{w}$.

Lemma 2 defines some morphisms σ_n on the alphabets of n-segments. Indeed consider the set of n-segments as an alphabet A_n. We can construct some morphisms $\sigma_n : A_{n+1} \to A_n^*$ that code the $(n+1)$-segments as concatenations of n-segments.

Remark 2. If a Rauzy graph $G_n(\mathbf{w})$ does not contain any bispecial vertex, it determines exactly the graph $G_{n+1}(\mathbf{w})$. In this case, for all n-segments p in $G_n(\mathbf{w})$, $\#\psi_n(p) = 1$ and the $(n+1)$-segments are exactly the elements of $\bigcup\{\psi_n(p) \mid p = n\text{-segment}\}$. Consequently the morphism σ_n as defined previously is simply a bijective and letter-to-letter morphism. Also note that if the alphabet of $\mathbf{w}$ is $A = \{a_1, \ldots, a_k\}$, the Rauzy graph G_0 is as in Figure 1 so the left labels of the 0-segments are letters of A. In other words, we have $A_0 = A$.

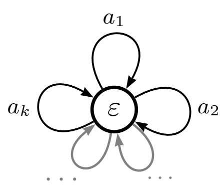

Fig. 1. Rauzy graph G_0 of any sequence over $\{a_1, \ldots a_k\}$

Remark 3. This type of constructions 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. The morphisms coding the $(n+1)$-segments as concatenations of n-segments are therefore uniform of length 1. However we can prove that for sequences with an at most polynomial complexity, the maximal length of n-segments tends to infinity so these constructions make sense.

Lemma 3 below is a key point of Theorem 1 but holds for any uniformly recurrent sequence (not only for those with an at most linear complexity). It is a consequence of Proposition 4 but first we need to recall the notions of n-circuit and of *short* and *long* segments or circuits introduced in [14].

By Lemma 2, 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 $n \geq N$, there is a n-segment in $\psi_n \psi_{n-1} \cdots \psi_N(s)$. 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 then n while a long n-segment will only appear as a proper subpath of m-segments for m larger enough. Note that if p is a short n-segment and if for all integers k, we code a corresponding $(n+k)$-segment by ξ_k in A_{n+k}, we have $\sigma_{n+k}(\xi_{k+1}) = \xi_k$ for all k.

A *n-circuit* is a non-empty path p in $G_n(\mathbf{w})$ such that $o(p) = i(p)$ is a left special vertex and any interior vertex of p is not $o(p)$. It is easy to be convinced that Lemma 2 can be adapted to n-circuits. Hence we can define *short* and *long* n-circuits similarly to *short* and *long* n-segments.

Lemma 3 (Ferenczi). *Let $\mathbf{w}$ be a uniformly recurrent sequence over an alphabet A. For any non-negative integers n, there is no short n-circuit in $G_n(\mathbf{w})$.*

Remark 4. A consequence of Lemma 3 is that for all integers ℓ, there is an integer n_ℓ such that any n_ℓ-circuit has length greater than ℓ.

5 Proof of Theorem 1

Let us recall Theorem 1 (properties will be stated in Section 5.2).

Theorem 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 a S-adic sequence satisfying the ω-growth Property and Properties 1 to 4 for a set S of non-erasing morphisms.*

5.1 S-adicity and ω-growth Property

The next two Lemmas allow us to bound the cardinality of the set of morphisms. First consider that the n-segments are indexed such that we can write $A_n = \{0, 1, \ldots, s(n) - 1\}$.

Lemma 4 (Ferenczi). *Let $\mathbf{w}$ be a sequence over an alphabet A. If $\mathbf{w}$ has an at most linear complexity, then there is constant C such that $s(n) \leq C$ for all n.*

Remark 5. A consequence of Lemma 4 is that the maximal length of n-segments tends to infinity as n increases. In other words, there is at least a long n-segment for each length n. Moreover, as two different n-segments p and q give rise to disjoint sets $\psi_n(p)$ and $\psi_n(q)$, there are at most $K\#A - 1$ shorts segments (all order n included) so we can bound the length of short segments by some constant ℓ.

Now let us improve Lemma 2 stating that the number of n-segments occurring in a $(n+1)$-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 5 (Ferenczi). *Let* $\mathbf{w}$ *be an aperiodic sequence over an alphabet* A*. If* $\mathbf{w}$ *is uniformly recurrent and has an at most linear complexity, then any* $(n+1)$*-segment of* $\mathbf{w}$ *is the image under* ψ_n *of a concatenation of a bounded number of* n*-segments of* $\mathbf{w}$*.*

We need one more lemma to prove the S-adicity Property in Theorem 1. This last one is a consequence of Lemma 3 and it allow us to prove that the S-adic representation satisfies the ω-growth Property.

Lemma 6 (Ferenczi). *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.*

Now we can present a sketch of the proof the S-adicity property in Theorem 1.

Sketch of Proof. We consider the sequence $\mathbf{w}' = \sharp\mathbf{w}$ over $A \cup \{\sharp\}$ with $\sharp \notin A$. For all non-negative integers n, let 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.

Lemma 2 can be adapted to paths in A'_{n+1}. Hence we can define some morphisms $\tau_n : A'_{n+1} \to A'^*_n$ that code the paths in A'_{n+1} with the paths in A'_n. Let S be the set of morphisms $\{\tau_n \mid n \in \mathbb{N}\}$. Using Lemmas 4, 5 and 6, we prove that $\#S < +\infty$

As $\mathbf{w}$ is recurrent, all prefixes $\mathbf{w}[0,n]$ are left special vertices in $G_n(\mathbf{w}')$. Let B_n denotes the set of paths p in A'_n with $o(p) = \mathbf{w}[0, n-1]$. For all n there is a paths p in B_n whose left label is a prefix of $\mathbf{w}$ and we have $\tau_n(B_{n+1}) \in B_n^+ A'^*_n$. We conclude by stating the ω-growth Property. □

5.2 Properties of the morphisms

In this section we consider notation introduced in the proof of Theorem 1: $\mathbf{w}$ is an aperiodic and uniformly recurrent sequence over an alphabet A with an at most linear complexity, $\mathbf{w}'$ is $\sharp\mathbf{w}$ and τ_n denotes the morphism from A'_{n+1} to A'^*_n. We also write $A' = \bigcup_{n\in\mathbb{N}} A'_n = \{1, \ldots, D-1\}$ with $D < +\infty$ and for all integers n and all words u in A'^+_n, $p^{(u)}$ denotes the path of $G_n(\mathbf{w}')$ coded by u. Finally, N is the smallest integer such that all the short segments already exist in G_N. More precisely, N is such that if p is a short m-segment for $m > N$ then there is a short N-segment q such that $p \in \psi_{m-1}\psi_{m-2}\cdots\psi_N(q)$.

Lemma 7. *Let* $\left((\tau_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \in A'^{\mathbb{N}}\right)$ *be the adic representation of* $\mathbf{w}$ *given by Theorem 1. Let* a *be a letter in* A'_n *that codes a path* $p_s p_l$ *in* G_n *such* *that the path p_s composed of consecutive short n-segments is non-empty. If there is a letter b in A'_{n+1} and a word u in A'^{+}_{n} such that $\tau_n(b) = ua$, then the subpath of $p^{(b)}$ that belongs to $\psi_n(p^{(u)})$ is a concatenation of short $(n+1)$-segment.*

Property 1 is a consequence of Lemma 7 and of the definition of N.

Property 1. Let $\big((\tau_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \in A'^{\mathbb{N}}\big)$ be the adic representation of $\mathbf{w}$ given by Theorem 1. For all integers $n \geq N$, if a is a letter in A'_n that codes a path $p_s p_l$ in G_n such that the path p_s composed of consecutive short n-segments is non-empty, then for all letters b in A'_{n+1}, $\tau_n(b) \notin A'^{+}_{n} a$.

Properties 2 and 2 are consequence of Property 1 and of the uniform recurrence.

Property 2. Let $\big((\tau_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \in A'^{\mathbb{N}}\big)$ be the adic representation of $\mathbf{w}$ given by Theorem 1. For all integers $n \geq N$ and all letters $a \in A'_{n+1}$ and b in A'_n, $\tau_n(a) \notin A'^{*}_{n} b A'^{*}_{n} b A'^{*}_{n}$.

Property 3. Let $\big((\tau_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \in A'^{\mathbb{N}}\big)$ be the adic representation of $\mathbf{w}$ given by Theorem 1. For all integers $n \geq N$, all letters b and c in A'_n and all letters a and d in A'_{n+1}, $(\tau_n(a), \tau_n(d)) \notin A'^{*}_{n} b A'^{*}_{n} c A'^{*}_{n} \times A'^{*}_{n} c A'^{*}_{n} b A'^{*}_{n}$.

Property 4 is a consequence of the uniform recurrence and of the ω-growth Property.

Property 4. Let $\big((\tau_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \in A'^{\mathbb{N}}\big)$ be the adic representation of $\mathbf{w}$ given by Theorem 1. For any non-negative integer r there is an integer $s > r$ such that all letters b in A'_r occurs in $\tau_r \tau_{r+1} \cdots \tau_s(c)$ for all letters c in A'_{s+1}.

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 [9]), where the letters b_n are defined at the end of the proof of Theorem 1.

6 Conclusions

First it is easy to see that Theorem 2 is a consequence of Theorem 1.

Now let us explain how this last result can easily be extended to bi-infinite sequences (i.e., to elements of $A^{\mathbb{Z}}$). For any uniformly recurrent sequence $\mathbf{w}$ in $A^{\mathbb{Z}}$ with an at most linear complexity, we consider the sequences $\mathbf{w_l} = \mathbf{w}[-\infty, -1]\sharp \in (A \cup \{\sharp\})^{-\mathbb{N}}$ and $\mathbf{w_r} = \sharp\mathbf{w}[0, +\infty] \in (A \cup \{\sharp\})^{\mathbb{N}}$. Then we can construct some adic representations $\big((\tau_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \in B'^{\mathbb{N}}\big)$ and $\big((\sigma_n)_{n\in\mathbb{N}}, (c_n)_{n\in\mathbb{N}} \in C'^{\mathbb{N}}\big)$ respectively for $\mathbf{w_r}$ and $\mathbf{w_l}$. Indeed, the adic representation of $\mathbf{w_r}$ is given by Theorem 1 and for the adic representation of $\mathbf{w_l}$, it suffices to replace the left n-segments by the right n-segments (with right labels). Then for all non-negative integers n, we extend the morphisms τ_n and σ_n to $B' \cup C'$ by fixing the new letters and we get the following adic representation of $\mathbf{w}$: $\big((\sigma_n \tau_n)_{n\in\mathbb{N}}, (c_n.b_n)_{n\in\mathbb{N}} \in (C' \times B')^{\mathbb{N}}\big)$.

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 and J. Shallit. *Automatic sequences*. Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.
2. P. Arnoux and G. Rauzy. Représentation géométrique de suites de complexité $2n+1$. *Bull. Soc. Math. France*, 119(2):199–215, 1991.
3. V. Berthé. Autour du système de numération d'Ostrowski. *Bull. Belg. Math. Soc. Simon Stevin*, 8(2):209–239, 2001. Journées Montoises d'Informatique Théorique (Marne-la-Vallée, 2000).
4. V. Berthé, C. Holton, and L. Q. Zamboni. Initial powers of Sturmian sequences. *Acta Arith.*, 122(4):315–347, 2006.
5. V. Berthé and M. Rigo (Eds). *Combinatorics, Automata and Number Theory*, volume 135 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, to appear.
6. J. Cassaigne. S-adiques. [Online]. Available: http://www.lirmm.fr/arith/wiki/PytheasFogg/S-adiques.
7. J. Cassaigne. Special factors of sequences with linear subword complexity. In *Developments in language theory, II (Magdeburg, 1995)*, pages 25–34. World Sci. Publ., River Edge, NJ, 1996.
8. J. Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin*, 4(1):67–88, 1997. Journées Montoises (Mons, 1994).
9. F. Durand. Linearly recurrent subshifts have a finite number of non-periodic subshift factors. *Ergodic Theory Dynam. Systems*, 20(4):1061–1078, 2000.
10. F. Durand. Corrigendum and addendum to: "Linearly recurrent subshifts have a finite number of non-periodic subshift factors". *Ergodic Theory Dynam. Systems*, 23(2):663–669, 2003.
11. F. Durand, B. Host, and C. Skau. Substitutional dynamical systems, Bratteli diagrams and dimension groups. *Ergodic Theory Dynam. Systems*, 19(4):953–993, 1999.
12. A. Ehrenfeucht, K. P. Lee, and G. Rozenberg. Subword complexities of various classes of deterministic developmental languages without interactions. *Theor. Comput. Sci.*, 1(1):59–75, 1975.
13. S. Ferenczi. Les transformations de Chacon: combinatoire, structure géométrique, lien avec les systèmes de complexité $2n+1$. *Bull. Soc. Math. France*, 123(2):271–292, 1995.
14. S. Ferenczi. Rank and symbolic complexity. *Ergodic Theory Dynam. Systems*, 16(4):663–682, 1996.
15. N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.
16. M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.
17. M. Morse and G. A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.*, 62:1–42, 1940.
18. J.-J. Pansiot. Complexité des facteurs des mots infinis engendrés par morphismes itérés. In *Automata, languages and programming (Antwerp, 1984)*, volume 172 of *Lecture Notes in Comput. Sci.*, pages 380–389. Springer, Berlin, 1984.
19. M. Queffélec. *Substitution dynamical systems—spectral analysis*, volume 1294 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1987.

Annexe pour les rapporteurs

Proof (of Lemma 1). First suppose $k_n \leq K$ for all integers n. For all integers $0 \leq i \leq K$, we define the morphism $\mu^{(i)} = \alpha^i \mu$. Each such morphism $\mu^{(i)}$ is strongly primitive. Moreover, we can rewrite the sequence $\mathbf{w}$ as a primitive S'-adic sequence with $S' = \{\mu^{(i)} \mid 0 \leq i \leq K\}$ hence 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 [18] that the fixed point $\alpha^\omega(a)$ has a quadratic complexity. From Theorem 3 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 we can conclude: if u and v are two different 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 different 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, we conclude 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, for all factors u of length m_n is a factor of $|\alpha^{k_0}\mu\alpha^{k_1}\mu\cdots\alpha^{k_{n-1}}\mu(v)|$ where v is a word of length 2. As there are only 4 possible binary words of length 2 and there are less that m_n different factors of length m_n is a word of length $2m_n$, we can conclude. $\square$

Proof (of Lemma 2). Let p be a $(n+1)$-segment in $G_{n+1}(\mathbf{w})$. We note $u = o(p)$, $u' = i(p)$ and v its left label. Then v is a return word to $LS_{\mathbf{w}}(n+1)$ such that u and u' are the two left special factors of length $n+1$ in $\mathbf{w}$ occurring in vu'. It is clear that the respective prefixes $u[1,n]$ and $u'[1,n]$ of length n of u and u' are left special factors of length n of $\mathbf{w}$. As there might be some left special factors of length n that have an occurrence at position $2 \leq i \leq |v|-1$ in $vu'[1,n]$, v is a concatenation of return words to $LS_{\mathbf{w}}(n)$. Moreover, v is the left label of a path p' in $G_n(\mathbf{w})$ such that $o(p') = u[1,n]$ and $i(p') = u'[1,n]$. Consequently p' is a concatenation of n-segments and we conclude the proof by noting that $p \in \psi_n(p')$. $\square$

Proof (of Proposition 3). 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 $0 < i \leq |u|$ such that $u_i \prec v_i$.

Let ℓ denotes the maximal length of $\sigma(a)$ for σ in S and a in A. Consider an integer n strictly 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 $L(\mathbf{w})$ (with respect to the radix order) such that $u_k \in L(\sigma_k(u_{k+1}))$.

We can easily see that the sequence $(|u_k|)_{k\in\mathbb{N}}$ is ultimately decreasing. Let r be the smallest integer such that $|u_r| \leq 2$. We have $2 \leq r \leq 1 + C\log n$. Indeed, the first inequality is trivial from the choice of n and for the second one, we can see notice that $|u_{r-1}|$ is at least 3. 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 factor of u and $|\sigma_0\sigma_1\cdots\sigma_{r-2}(v_{r-1})|$ is at least 2^{r-1}. Therefore we have $n > 2^{r-1}$ and then $r < C\log n + 1$.

Now for all words u in A^* of length smaller or equal to 2, we define $W_n(u)$ as the set of words of length n in $L(\mathbf{w})$ such that $u_r = u$. Obviously, we have $\bigcup_{u\in L_{\leq 2}(w)} W_n(u) = L_n(\mathbf{w})$. To conclude with 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. □

Proof (of Lemma 3). 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 $\mathbf{w}$ 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 a $(n+k)$-circuit in G_{n+k} with left label u. Now let v_k denotes the extremity of the $(n+k)$-circuit in $\psi_{n+k-1}\psi_{n+k-2}\cdots\psi_n(p)$ and $e_k = \left\lfloor \frac{n+k}{|u|} \right\rfloor$. The word u^{e_k} is the left label of a path of length smaller or equal to $n+k$ in G_{n+k} starting in v_k. From Proposition 4 it is a prefix of v_k and consequently a factor of $\mathbf{w}$. As e_k tends to infinity as k increases, there are arbitrary large powers of u in $L(\mathbf{w})$ and this contradicts the uniform recurrence. □

Proof (of Lemma 4). By Theorem 3, there exists a constant K such that $p(n+1) - p(n) \leq K$ for all n. From (2) 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 $K\#A$. □

Proof (of Lemma 5). Consider a $(n+1)$-segment p in $G_{n+1}(\mathbf{w})$. The number of n-segments composing it is equal to 1 plus the number of vertices va in p, $a \in A$, such that v is a left special factor of length n of $\mathbf{w}$ and va not. Moreover, the path p cannot pass through one of these vertices more than once. Indeed, as none of these vertices is left special, it would create a loop in $G_{n+1}(\mathbf{w})$ that would be inaccessible from vertices that are not in this loop and the graph would not be strongly connected (we know that there are some vertices that are not in the loop since the sequence is aperiodic). Finally, as there exist at most K left special vertices v in $G_n(\mathbf{w})$ (where K is given by $p(n+1) - p(n) \leq K$), there exist at most $K\#A$ vertices va as considered just above. Consequently, the number of n-segments in p is bounded by $1 + K\#A$. □

Proof (of Lemma 6). Let K be such that $p(n+1) - p(n) \leq K$. 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 in 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 length bounded by ℓ (see Remark 5). 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 $K\ell$. Remark 4 states that this can not happen for n large enough. □

Proof (of Theorem 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 $A \cup \{\sharp\}$. For all non-negative integers n, let 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. Suppose that these paths are indexed such that we can write $A'_n = \{0, \ldots, P(n)-1\}$. Remark 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

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

and 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.

Lemma 2 can be adapted to paths in A'_{n+1}. Hence we can define some morphisms $\tau_n : A'_{n+1} \to A'^*_n$ that code the paths in A'_{n+1} with the paths in A'_n. Let S be the set of morphisms $\{\tau_n \mid n \in \mathbb{N}\}$ and let us show that $\mathbf{w}$ is S-adic.

First, we have to proved that $\#S < +\infty$. It is a consequence of Lemmas 4 and 6 that there is a constant D such that $P(n) \leq D$ for all n. Moreover, the number of n-segments composing a path in A'_n is bounded by Lemma 6 hence Lemma 5 implies that the number of paths of A'_n occurring in a path of A'_{n+1} is bounded.

Now let us show that $\mathbf{w}$ admits a S-adic representation. First, 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 B_n denotes the set of paths p in A'_n such that $o(p) = \mathbf{w}[0, n-1]$. 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 the path corresponding to b_n has a left label that is a prefix of $\mathbf{w}$ and $\tau_n(b_{n+1}) \in b_n A'^*_n$ (it is clear from 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 $\mathbf{w} = \lim_{n\to+\infty} \tau_0\tau_1\cdots\tau_n(b_{n+1})$ and this ends the proof of the S-adicity. $\square$

Proof (of Lemma 7). Indeed, as p_s is non-empty, the interior vertex of $p^{(b)}$ that admits $o(p_s)$ as prefix is left special. Hence the subpath of $p^{(b)}$ that belongs to $\psi_n(p^{(u)})$ has to be a concatenation of short $(n+1)$-segments $\square$

Proof (of Property 1). It is a direct consequence of Lemma 7. Indeed, suppose that $\tau_n(b) = ua$ with $u \neq \varepsilon$. By definition, there is at least one subpath p of $p^{(u)}$ that is a long n-segment. However, by Lemma 7, the subpath of $p^{(b)}$ that belongs to $\psi_n(p^{(u)})$ is a concatenation of short $(n+1)$-segment. Hence, there is a path in $\psi_n(p)$ that is a subpath of a short $(n+1)$-segment s such that for all short n-segments r, $s \notin \psi_n(r)$ and this contradicts the definition of N. $\square$

Proof (of Property 2). Suppose $\tau_n(a) = u_1bu_2bu_3$ with $b \in A'_n$ and $u_1, u_2, u_3 \in A'_n$. From Property 1, the path $p^{(b)}$ is a long n-segment. Moreover, from the way the morphisms τ_n are constructed, we can deduce that the interior vertices of $p^{(a)}$ that admit respectively $o(p^{(b)})$ and $i(p^{(b)})$ as prefixes are not left special. Hence the subpath of $p^{(a)}$ that belongs to $\psi_n\left(p^{(bu_2b)}\right)$ is a loop in $G_{n+1}(\mathbf{w}')$ that is inaccessible from vertices that are not in it. As no n-segment of $\mathbf{w}'$ contains the "added edges" $(\sharp\mathbf{w}[0, n-2], (\sharp, \mathbf{w}_{n-1}), \mathbf{w}[0, n-1])$, this loop exists in $G_{n+1}(\mathbf{w})$ and this last graph is not strongly connected. $\square$

Proof (of Property 3). Suppose that $\tau_n(a) = u_1bu_2cu_3$ and $\tau_n(d) = v_1cv_2bv_3$ with $a \in A'_{n+1}$, $b, c \in A'_n$ and $u_1, u_2, u_3, v_1, v_2, v_3 \in A'^*_n$. With the same reasoning as in the proof of Property 2, the respective subpaths p and q of $p^{(a)}$ and $p^{(d)}$ that respectively belong to $\psi_n\left(p^{(bu_2c)}\right)$ and $\psi_n\left(p^{(cv_2b)}\right)$ do not contain any left special vertex. Consequently the path pq in $G_{n+1}(\mathbf{w}')$ is a loop which is inaccessible from vertices that are not in it. With the same reasoning as at the end of the proof of Property 2, the graph $G_{n+1}(\mathbf{w})$ is not strongly connected and this contradicts the uniform recurrence. $\square$

Proof (of Property 4). 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. For all non-negative integers n, the paths coded by letters in A'_n are allowed. Hence $\tau_0\tau_1\cdots\tau_n(a)$ is a factor of $\mathbf{w}$ for all letters a in A'_{n+1}. For all non-negative integers n, let M_n and m_n be respectively the maximal and minimal lengths of a path in G_n coded by a letter of A'_n. Let i be a non-negative integer. As 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 $a_n \in A'_n$, there is a non-negative integer j such that all factors of length at most M_i occurs in all factors of length at least m_j and this conclude the proof. $\square$