On sets of indefinitely desubstitutable words

Gwenaël Richomme

On sets of indefinitely desubstitutable words

Gwenaël Richomme

LIRMM,Université Paul-Valéry Montpellier 3,
Université de Montpellier, CNRS, Montpellier, France

November 14, 2020

Abstract

The stable set associated to a given set $\mathcal{S}$ of nonerasing endomorphisms or substitutions is the set of all right infinite words that can be indefinitely desubstituted over $\mathcal{S}$. This notion generalizes the notion of sets of fixed points of morphisms. It is linked to S-adicity and to property preserving morphisms. Two main questions are considered. Which known sets of infinite words are stable sets? Which ones are stable sets of a finite set of substitutions? While bringing answers to the previous questions, some new characterizations of several well-known sets of words such as the set of binary balanced words or the set of episturmian words are presented. A characterization of the set of nonerasing endomorphisms that preserve episturmian words is also provided.

Keywords: S-adicity, fixed points, Sturmian words, episturmian words, property preserving morphisms

1 Introduction

As explained with more details in, for instance, [2, 19], the terminology S-adic was introduced by S. Ferenczi in [6], where he proved that dynamical symbolic systems with subaffine factor complexity are S-adic uniformly minimal symbolic systems. This result was motivated by the so-called S-adic conjecture: there exists a stronger notion of S-adicity which is equivalent to linear factor complexity. Some advances on this conjecture were obtained by J. Leroy [11, 12], especially when the first difference of the factor complexity is bounded by 2 [13]. Also an adaptation for infinite words was obtained [14]. Many classical words are known to be S-adic words such as, *e.g.*, fixed points of morphisms, morphic words (images of fixed points), Sturmian words, 3-interval exchange transformations, Arnoux-Rauzy words and strict episturmian words.

In [2], Berthé and Delecroix write: "*Expansions of S-adic nature have now proved their efficiency for yielding convenient descriptions for highly structured symbolic dynamical systems [...] If one wants to understand such a system [...], it might prove to be convenient to decompose it via a desubstitution process: an S-adic system is a system that can be indefinitely desubstituted.*" The aim of the paper is to study more specifically this desubstitutive process under the approach of limit points as defined by P. Arnoux, M. Mizutani and T. Sella [1] (and also used, for instance, by Justin et al. [5, 10]). More specifically we study stable sets as defined by E. Godelle [9], that is, sets of limit points of a given set of substitutions (or free monoid nonerasing endomorphisms). In other terms, a stable set associated to a set $\mathcal{S}$ of substitutions is the set of all right infinite words that can be indefinitely desubstituted using elements of $\mathcal{S}$.

In [21], answering a question of G. Fici, the author characterizes in terms of limit S-adicity the family of so-called LSP infinite words, that is, the words having all their left special factors as prefixes. For this, he determines a suitable set S_{bLSP} of morphisms and an automaton recognizing the allowed infinite sequences of desubstitutions. As the obtained characterization is quite involved, a second part of [21] considers the question of finding a simpler limit S-adic characterization. Unfortunately, there exists no set $\mathcal{S}$ of endomorphisms such that the set of LSP infinite words is (exactly) the stable set associated to $\mathcal{S}$ except in the binary case. The main motivation of our study is the question: which are the known families of infinite words defined by a combinatorial property P that correspond to

stable sets? In [21], it was observed that, when such a situation arises, morphisms of $\mathcal{S}$ necessarily preserve the property P of infinite words. While bringing answers to the previous question, we present some new characterizations of several families of words such as the set of balanced words or the set of episturmian words. This leads us also to characterize the set of nonerasing endomorphisms that preserve episturmian words.

In Section 2, we present the process of desubstitution as a generalization of the notion of a fixed point of a morphism. We also recall needed definitions such as those of indefinitely desubstituted words, limit points and stable sets. In Section 3, we consider an example of a stable set, proving that the set of binary balanced words is a stable set of a particular set of four morphisms. As far as we know, this result was not stated formally earlier, even if many aspects of the proof are known. The considered process of desubstitution is infinite and an infinite sequence of substitutions of a given set $\mathcal{S}$ is associated to the infinite desubstitution of an infinite word: such a sequence of substitutions is called a directive sequence. We show that, in the general case, any infinite sequence of elements of $\mathcal{S}$ is the directive sequence of at least one infinite word.

In Section 4, we study the possible forms of the elements of a stable set in relation with S-adicity. We show that the characterization of the set of balanced words can be transformed to another characterization in terms of S-adicity even if a general transformation cannot exist for arbitrary stable sets. In Section 5, we provide two more examples of sets of binary infinite words that are stable sets: the set of Sturmian words and the set of Lyndon Sturmian words. For both sets, there exist only infinite sets of substitutions for which the set is a stable set.

In Section 6, we consider episturmian words on arbitrary alphabets. As far as we know, the fact that the set of standard episturmian word is a stable set is the unique result of this form that was previously stated (without adapted terminology) [10]. After recalling this result, we prove that the set of A-strict standard episturmian words, the set of episturmian words and the set of A-strict episturmian words are stable sets, but only of infinite sets of substitutions. For this, we prove and use a characterization of endomorphisms preserving episturmian words. Observe that in [10] there exists a characterization of episturmian words using a desubstitution process associated to a finite set S of substitutions but not all elements of the stable set of S are episturmian (only the recurrent ones are).

2 From fixed points to stable sets

We assume that readers are familiar with combinatorics on words; for omitted definitions see, *e.g.*, [4, 16, 17]. All the infinite words considered in this paper are right infinite words. Let us recall some basic notions on fixed points of morphisms.

Let A be a (finite) alphabet. Let $\#A$ be its cardinality. The set of words over A, usually denoted A^*, equipped with the concatenation operation has a free monoid structure with neutral element the empty word ε. Given two alphabets A and B, a (free monoid) *morphism* f is a map from A^* to B^* that preserves the monoid structure: for all words u and v, $f(uv) = f(u)f(v)$ (and, consequently, $f(\varepsilon) = \varepsilon$). Morphisms are completely defined by images of letters. In what follows we will essentially use *endomorphisms*, that is, morphisms from a free monoid A^* to itself.

Given an alphabet A, the set of right infinite words, that is, infinite sequences of elements of A, is usually denoted A^ω. Images by a morphism of infinite words are defined naturally but to preserve infinity, one may consider only *nonerasing* morphisms (images of nonempty words are never the empty word). As often done, such a nonerasing morphism will be called a *substitution*.

A *fixed point* of an endomorphism is any finite or infinite word w such that $w = f(w)$. For instance, the following morphism has both finite and infinite fixed points. Actually, since it admits a finite fixed point, it has infinitely many finite and infinitely many infinite fixed points.

$$f_1 : \begin{cases} a \mapsto a \\ b \mapsto bac \\ c \mapsto baca \end{cases}$$

The definition of fixed points is not constructive. To determine more precisely prefixes of an infinite fixed point, at least two different approaches are usually considered.

Limit approach

If $\mathbf{w}$ is an infinite fixed point of a morphism f and if p is one of its prefixes, then necessarily, for $n \geq 0$, the word $f^n(p)$ is also a prefix of $\mathbf{w}$ (classically $f^n = f^{n-1} \circ f$, f^0 being the identity morphism). If $\lim_{n \to \infty} |f^n(p)| = \infty$ (where $|u|$ denotes the length, that is, the number of letters, of the word u), then the words $f^n(p)$ provide progressively all the letters of $\mathbf{w}$ that, thus, can be viewed as a limit of finite words. This is usually denoted $\mathbf{w} = \lim_{n \to \infty} f^n(p)$ or $f^\omega(p)$ (in this last case, p is very often a letter).

Observe that the morphism f_1 has exactly one fixed point, the word a^ω, that cannot be obtained as a limit of images of powers of f_1 applied to a word. The other fixed points are the words $\lim_{n \to \infty} f_1^n(a^k b)$ with $k \geq 0$ an integer.

Desubstitutive approach

The second approach consists of firstly considering the fixed point as the image of another word by the morphism. There exists a word $\mathbf{w}'$ such that $\mathbf{w} = f(\mathbf{w}')$: $\mathbf{w}'$ is called a *desubstituted* word. In general, $\mathbf{w}'$ may not be unique. For instance, the unique fixed point of f_1 starting with the letter b can be seen as the image of a word over $\{a, c\}$ as $f_1(c) = f_1(ba)$. Nevertheless in our context, we know that one possibility for $\mathbf{w}'$ is $\mathbf{w}$ itself. And so we can iterate the desubstitution. Hence we can find an infinite sequence $(\mathbf{w}_i)_{i \geq 0}$ of infinite words such that $\mathbf{w} = \mathbf{w}_0$ and, for all $i \geq 0$, $\mathbf{w}_i = f(\mathbf{w}_{i+1})$: the word $\mathbf{w}$ can be *indefinitely desubstituted*.

For obtaining large prefixes of a fixed point, the desubstitutive approach is less constructive than the limit approach but it can be used as follows. Consider the first letter a_0 of $\mathbf{w}$. Then consider all letters whose images by f start with a_0: each of these letters a_1 is a possible letter for the first letter of $\mathbf{w}_1$ whenever $f(a_1)$ is a prefix of $\mathbf{w}$. Then we can iterate looking for potential first letters a_2 of $\mathbf{w}_2$, a_3 of $\mathbf{w}_3$ and so on. The words $f^n(a_n)$, with a_n the first letter of $\mathbf{w}_n$, provide longer and longer prefixes of $\mathbf{w}$ except if this word starts with a finite fixed point. In this case we have to consider the letter occurring after this fixed point and iterate, from this letter, the search for other letters.

It follows from the definition that any fixed point can be indefinitely desubstituted (remember that some fixed points cannot be defined by the limit approach). But, for some morphisms, there may exist some infinite words that can be indefinitely desubstituted without being fixed points. This does not occur with the morphism f_1 but it does occur with the morphism f_2 defined by $f_2(a) = ba$ and $f_2(b) = ab$ (Observe that f_2 has no fixed point). The words that can be indefinitely desubstituted using f_2 are the fixed points of f_2^2 and not of f_2 (these words are also the *Thue-Morse words*, that is, the fixed points of the morphism μ defined by $\mu(a) = ab$ and $\mu(b) = ba$). The previous phenomenon is much more general.

Lemma 2.1 ([9, Prop. 7.1]). *Let $\mathbf{w}$ be a right infinite word and let f be a nonerasing morphism. The word $\mathbf{w}$ can be indefinitely desubstituted by f if and only if $\mathbf{w}$ is the fixed point of f^n for some integer $n \geq 1$.*

Proposition 7.1 in [9] is richer than the previous lemma. Its proof depends on a larger context. For the sake of completeness, we provide a short proof in our context. Let $\text{first}(u)$ denote the first letter of a finite nonempty word or infinite word u.

Proof. First observe that if $\mathbf{w}$ is the fixed point of f^n, then $\mathbf{w}$ can be indefinitely desubstituted. More precisely, as $\mathbf{w} = f^n(\mathbf{w})$, the desubstituted words are $f^{n-1}(\mathbf{w})$, ..., $f(w)$, $\mathbf{w}$, $f^{n-1}(\mathbf{w})$, ..., $f(w)$, $\mathbf{w}$ and so on.

Assume now that $\mathbf{w}$ can be indefinitely desubstituted by f. Let $(\mathbf{w}_n)_{n \geq 0}$ be a sequence of desubstituted words: $\mathbf{w}_0 = \mathbf{w}$ and, for all $n \geq 0$, $w_n = f(\mathbf{w}_{n+1})$. Let $a_n = \text{first}(\mathbf{w}_n)$ for all $n \geq 0$. As the alphabet A is finite, there exist arbitrarily large integers m, n such that $0 \leq m < n \leq m + \#A$ and $a_n = a_m$. Observe that, if $0 < m < n$ and $a_n = a_m$, then $a_{n-1} = a_{m-1}$. Indeed, as f is not erasing, $a_{n-1} = \text{first}(\mathbf{w}_{n-1}) = \text{first}(f(\mathbf{w}_n)) = \text{first}(f(a_n)) = \text{first}(f(a_m)) = a_{m-1}$. Consequently, the sequence $(a_n)_{n \geq 0}$ is periodic with a period smaller than or equals to $\#A$. Let π be this period.

Let $a = a_0$ (a is the first letter of $\mathbf{w}$). For all $n \geq 0$, $a_{\pi n} = a$. Since a is the first letter of $\mathbf{w} = f^\pi(\mathbf{w}_\pi)$ and since it is the first letter of $\mathbf{w}_\pi$, a is the first letter of $f^\pi(a)$. If $|f^\pi(a)| > 1$, then $\lim_{n \to \infty} |f^{n\pi}(a)| = \infty$. For all $n \geq 0$, since a is the first letter of $\mathbf{w}_{n\pi}$ and as $\mathbf{w} = f^{n\pi}(\mathbf{w}_{n\pi})$, the

word $f^{n\pi}(a)$ is a prefix of $\mathbf{w}$. Hence $\mathbf{w} = \lim_{n\to\infty} f^{n\pi}(a)$. But, similarly, $\mathbf{w}_\pi = \lim_{n\to\infty} f^{n\pi}(a)$. Hence $\mathbf{w} = f^\pi(\mathbf{w})$.

Now assume that $|f^\pi(a)| = 1$. Then $\mathbf{w} = a\mathbf{w}'$ with a a letter and $\mathbf{w}'$ an infinite word such that: $a = f^\pi(a)$; $\mathbf{w}'$ can be indefinitely desubstituted using f (the desubstituted words are the words $\mathbf{w}'_n$ obtained from the words $\mathbf{w}_n$ removing their first letters).

Thus, by induction, one can state that one of the two following cases holds.

1. $\mathbf{w} = a_0 \cdots a_N \mathbf{w}'$ where $N \geq -1$ is an integer (when $N = -1$, $\mathbf{w} = \mathbf{w}'$), $\mathbf{w}'$ is an infinite word and $a_0, \ldots, a_N$ are letters such that:

   - there exists an integer p such that $1 \leq p \leq \#A$ and $\mathbf{w}' = f^p(\mathbf{w}')$,
   - for all i, $0 \leq i \leq N$, there exists an integer π_i such that $1 \leq \pi_i \leq \#A$ and $a_i = f^{\pi_i}(a_i)$.

2. $\mathbf{w} = \prod_{i \geq 0} a_i$ where for all $i \geq 0$, $a_i \in A$ and there exists an integer π_i such that $1 \leq \pi_i \leq \#A$ and $a_i = f^{\pi_i}(a_i)$.

In the first case, let $n = gcd(p, \pi_0, \ldots, \pi_N)$ and, in the second case, let $n = gcd((\pi_i)_{i \geq 0})$. In both cases, $\mathbf{w} = f^n(\mathbf{w})$. $\qquad\square$

Before going further studying words that are indefinitely desubstitutable over a set of morphisms, let us recall and introduce some useful terminology. For any set X of finite words, let X^* denote the set of all finite words that can be obtained by concatenation of elements of X (including the empty word), let X^+ denote the set of all finite words that can be obtained by concatenation of at least one elements of X and let X^ω denote the set of all infinite words that can be obtained by concatenation of elements of X. Observe that these notations will be used also with sets of substitutions to denote sets of sequences of substitutions (the set of substitutions is then considered as an alphabet). Each finite sequence $\sigma_1 \sigma_2 \cdots \sigma_k$ of substitutions refers both to the sequence and to the substitution obtained by composition of the σ_i ($\sigma_1 \circ \sigma_2 \circ \ldots \circ \sigma_k$). We will use both interpretations alternatively: the context should be clear when we consider sequences and when we consider substitutions.

Let $\mathrm{Subst}(A)$ be the set of substitutions on the alphabet A, that is, the set of nonerasing endomorphisms of the free monoid A^*. Let $\mathcal{S} \subseteq \mathrm{Subst}(A)$. Let $(\sigma_n)_{n \geq 1} \in \mathcal{S}^\omega$. Following P. Arnoux, M. Mizutani and T. Sella [1], a finite or an infinite word $\mathbf{w}$ over A is a *limit point of the sequence* $(\sigma_n)_{n \geq 1}$ if there exists a sequence $(\mathbf{w}_n)_{n \geq 0}$ of infinite words over A such that $\mathbf{w} = \mathbf{w}_0$ and $\mathbf{w}_n = \sigma_{n+1}(\mathbf{w}_{n+1})$ for all $n \geq 0$. In other words $\mathbf{w}$ can be *indefinitely desubstituted* using successively the substitutions in the sequence $(\sigma_n)_{n \geq 1}$. The sequence $(\sigma_n)_{n \geq 1}$ will be called a *directive sequence* of $\mathbf{w}$.

Following Godelle [9], for $\mathcal{S} \subseteq \mathrm{Subst}(A)$, the *stable set* of $\mathcal{S}$, denoted $\mathrm{Stab}(\mathcal{S})$, is the set of all limit points of sequences in $\mathcal{S}^\omega$. Also a set X of infinite words is *stabilized* by $\mathcal{S}$ if $X = \bigcup_{f \in \mathcal{S}} f(X)$.

Lemma 2.2 ([9]).

- $\mathrm{Stab}(\mathcal{S})$ *is stabilized by* $\mathcal{S}$.

- *Any set stabilized by* $\mathcal{S}$ *is included in* $\mathrm{Stab}(\mathcal{S})$.

Thus $\mathrm{Stab}(\mathcal{S})$ is the largest set (w.r.t. inclusion) of infinite words stabilized by $\mathcal{S}$. Hence the stable set of a set of substitutions appears as a natural generalization of the set of fixed points of powers of a morphism. This is coherent with Lemma 2.1 which states that the stable set of a singleton $\{f\}$ is the set of fixed points of the morphisms f^n with $n \geq 1$.

Let P be a property of infinite words. Let X be the set of all infinite words having this property. A morphism is said to *preserve the property* P or *to preserve the set* X if for all elements $\mathbf{w}$ in X, $f(\mathbf{w})$ also belongs to X. By definition of a stable set, for all $\mathcal{S} \subseteq \mathrm{Subst}(A)$, for all $\sigma \in \mathcal{S}$ and $\mathbf{w} \in \mathrm{Stab}(\mathcal{S})$, $\sigma(\mathbf{w}) \in \mathrm{Stab}(\mathcal{S})$. This can be reformulated as the next remark, already made in [21], that will be useful when proving that some sets of infinite words cannot be stable sets of finite sets of substitutions.

Remark 2.3. *If* $X = \mathrm{Stab}(\mathcal{S})$ *for some set* $\mathcal{S}$ *of substitutions, then all elements of* $\mathcal{S}$ *preserve the set* X.

3 Balanced words: an example of a stable set

The desubstitutive approach naturally arises in the study of infinite binary balanced words, especially in the study of aperiodic balanced words, that is, the Sturmian words. Here we consider the whole set of infinite binary balanced words and we show Proposition 3.1 below. Elements of the proof are rather classical. They allow to illustrate some techniques of proof that will be reused later.

A word on an alphabet A is *balanced* if for all letters a in A and for all factors u and v with $|u| = |v|$, the following inequation holds: $||u|_a - |v|_a| \leq 1$ (here $|w|_\alpha$ denotes the number of occurrences of the letter α in the word w). The aperiodic, that is, not ultimately periodic, infinite binary balanced words are called *Sturmian words*. The most famous one is the *Fibonacci word* which is the fixed point of the morphism φ defined by $\varphi(a) = ab$, $\varphi(b) = a$. Since Morse and Hedlund's work [18], periodic infinite binary balanced words are known to be infinite repetitions of conjugates of finite standard words (*e.g.*, $ab(abaab)^\omega = (ababa)^\omega$) (two words x and y are *conjugates* if there exists a word u such that $xu = uy$). But there are also some ultimately periodic balanced words that are not purely periodic as for instance $a^n b a^\omega$, $(ab)^n a(ab)^\omega$, $(abaab)^n aba(abaab)^\omega$, ...

It is well-known that the following four morphisms are intrinsically related to binary balanced words as they naturally occur in the study of Sturmian words.

$$L_a : \begin{cases} a \mapsto a \\ b \mapsto ab \end{cases} \qquad L_b : \begin{cases} a \mapsto ba \\ b \mapsto b \end{cases} \qquad R_a : \begin{cases} a \mapsto a \\ b \mapsto ba \end{cases} \qquad R_b : \begin{cases} a \mapsto ab \\ b \mapsto b \end{cases}$$

Let $\mathcal{S}_{bal} = \{L_a, R_a, L_b, R_b\}$.

Proposition 3.1. *The set of binary balanced infinite words is the stable set of $\mathcal{S}_{\mathrm{bal}}$.*

The previous proposition means that an infinite binary word $\mathbf{w}$ is balanced if and only if there exists some words $(\mathbf{w}_i)_{i \geq 0}$ and morphisms $(\sigma_i)_{i \geq 1}$ in $\{L_a, L_b, R_a, R_b\}$ such that $\mathbf{w}_0 = \mathbf{w}$ and for all $i \geq 0$, $\mathbf{w}_i = \sigma_{i+1}(\mathbf{w}_{i+1})$. One can also see that all words in $(\mathbf{w}_i)_{i \geq 0}$ are balanced words. The proof of the *only if* part illustrates the mechanism of desubstitution.

Proof of the only if part. Given an infinite binary balanced word $\mathbf{w}$ over $\{a, b\}$, one of the two words aa and bb does not occur in $\mathbf{w}$. In each case and whatever is the first letter of $\mathbf{w}$, $\mathbf{w} = \sigma_1(\mathbf{w}_1)$ for a morphism σ_1 in $\{L_a, L_b, R_a, R_b\}$ and an infinite word $\mathbf{w}_1$ over $\{a, b\}$. The following table summarises the four possible cases:

$\mathbf{w}$ starts with	$\mathbf{w}$ does not contain	$\mathbf{w}$ can be decomposed
a	bb	over $\{a, ab\}$: $\mathbf{w} = L_a(\mathbf{w}_1)$
a	aa	over $\{ab, b\}$: $\mathbf{w} = R_b(\mathbf{w}_1)$
b	bb	over $\{ba, a\}$: $\mathbf{w} = R_a(\mathbf{w}_1)$
b	aa	over $\{ba, b\}$: $\mathbf{w} = L_b(\mathbf{w}_1)$

Let us now state that the word $\mathbf{w}_1$ is balanced. Assume by contradiction that $\mathbf{w}_1$ is not balanced. Considering a pair of words of same minimal length such that the first one contains at least two more occurrences of the letter a than the second, we find a word u such that both words aua and bub occur in $\mathbf{w}_1$. In the case where $\mathbf{w}$ starts with b and does not contain bb, we verify that the word $\mathbf{w} = R_a(\mathbf{w}_1)$ contains $aaR_a(u)a$ and $baR_a(u)b$ as factors (the factor aua is not a prefix of $\mathbf{w}_1$, and so, images of its occurrences in $R_a(\mathbf{w}_1)$ are necessarily preceded by a). A similar contradiction with the balance property of $\mathbf{w}$ can be obtained in the three other cases.

As $\mathbf{w}_1$ is balanced, iterating what precedes, one can find a balanced infinite binary word $\mathbf{w}_2$ and a morphism σ_2 in $\{L_a, L_b, R_a, R_b\}$ such that $\mathbf{w}_1 = \sigma_2(\mathbf{w}_2)$. The proof of the only if part ends by induction. $\qquad \square$

The proof of the *if* part needs the following well-known property:

Lemma 3.2 (see, *e.g.*, [17, chap. 2]). *Any morphism in $\{L_a, L_b, R_a, R_b\}^*$ preserves the balanced property.*

Proof of the if part of Proposition 3.1. Assume that $\mathbf{s} = (\sigma_n)_{n \geq 1}$ is a sequence of morphisms in $\{L_a, L_b, R_a, R_b\}$ occurring in an infinite desubstitution of a word $\mathbf{w}$, and let $(\mathbf{w}_n)_{n \geq 0}$ be the corresponding sequence of words: $\mathbf{w}_0 = \mathbf{w}$, $\mathbf{w}_i = \sigma_{i+1}(\mathbf{w}_{i+1})$ for all $i \geq 0$.

Assume first that $\mathbf{s} \in \{L_a, R_a\}^\omega$. Assume that there exists an integer n such that $ba^n b$ is a factor of $\mathbf{w}$. Thus, by induction, one can see that $ba^{n-i}b$ is a factor of $\mathbf{w}_i$ for all i such that $0 \leq i \leq n$. Especially bb is a factor of $\mathbf{w}_n$. This is impossible as it is an image of a word by L_a or R_a. Thus $\mathbf{w}$ contains at most one b, that is, it belongs to the set $\{a^\omega, a^n b a^\omega \mid n \geq 1\}$: it is clearly balanced.

Assume now that more generally, $\mathbf{s}$ contains finitely many elements of $\{L_b, R_b\}$. This means that $\mathbf{w} = f(\mathbf{w}')$ with $f \in \{L_a, L_b, R_a, R_b\}^*$ and $\mathbf{w}'$ as in the previous case. As $\mathbf{w}'$ is balanced and as f preserves balanced words by Lemma 3.2, the word $\mathbf{w}$ is also balanced.

Similarly $\mathbf{w}$ is balanced if $\mathbf{s}$ contains finitely many elements of $\{L_a, R_a\}$.

It remains to study the case where $\mathbf{s}$ contains both infinitely many occurrences of elements of $\{L_a, R_a\}$ and infinitely many occurrences of elements of $\{L_b, R_b\}$. These words are the Sturmian words [3] and so they are balanced. The proof can be given shortly. The sequence $\mathbf{s}$ can be contracted to a sequence $\mathbf{s}'$ of morphisms in $\{L_a, R_a\}^+ \{L_b, R_b\} \cup \{L_b, R_b\}^+ \{L_a, R_a\}$. All the morphisms σ occurring in $\mathbf{s}'$ satisfy $|\sigma(a)| \geq 2$ and $|\sigma(b)| \geq 2$. Thus one can obtain the word $\mathbf{w}$ as the limit of prefixes $\sigma'_1 \cdots \sigma'_n(a_n)$ with $\sigma'_1 \cdots \sigma'_n$ prefixes of $\mathbf{s}'$. Letters a_n are the first letters of the words associated to the infinite desubstitution of $\mathbf{w}$ by morphisms in σ'. As letters are balanced words and as all morphisms σ_i ($i \geq 1$) preserve the property of being balanced by Lemma 3.2, we obtain the balance property of $\mathbf{w}$. □

Let us observe, from the previous proof, that any infinite sequence of morphisms over $\mathcal{S}_{bal}$ is the directive sequence of at least one aperiodic or purely periodic balanced word. This property is general.

Proposition 3.3. *Let $\mathcal{S} \subseteq \mathrm{Subst}(A)$. Any sequence in $\mathcal{S}^\omega$ is the directive sequence of at least one element of $\mathrm{Stab}(\mathcal{S})$.*

This proposition (and its proof) is a generalization of a result by P. Arnoux, M. Mizutani and T. Sella [1, Prop2.1]: *Any primitive sequence of substitutions has a finite and nonzero number of limit points.* Let us recall that a sequence of substitutions $(\sigma_n)_{n \geq 0}$ over an alphabet A is *primitive* if for all $n \geq 0$, there exists an integer k such that, for each letter a in A, all letters of A occur in $\sigma_n \cdots \sigma_{n+k}(a)$. For an arbitrary set of substitutions, there may not exist aperiodic limit points. Indeed when $\mathcal{S} = \{R_b\}$, we have $\mathrm{Stab}(\mathcal{S}) = \{b^\omega\} \cup \{b^n a b^\omega \mid n \geq 0\}$. There may also exist infinitely many aperiodic limit points. Indeed let g be the morphism defined by $g(a) = abab$, $g(b) = b$. The fixed point $g^\omega(a)$ of g starting with a is aperiodic as it contains all words $ab^n a$ (with $n \geq 1$) as factors. The infinite fixed points of g are b^ω and the words $b^n g^\omega(a)$ for all $n \geq 0$.

Proof of Proposition 3.3. Let $(\sigma_n)_{n \geq 1}$ be a sequence in $\mathcal{S}^\omega$. For any $n \geq 1$, let $f_n : A \mapsto A$ be the map that sends any letter $a \in A$ to the first letter of $\sigma_n(a)$. The sequence $f_1(f_2 \cdots (f_n(A)) \cdots)$ of subsets of A is a decreasing sequence of nonempty subsets of A, so their intersection is nonempty. Let X be the set of all words $a_0 \cdots a_n$, with $n \geq 0$, such that $a_i = f_{i+1}(a_{i+1})$ for all $i, 0 \leq i < n$. By Konig's lemma, there exists an infinite sequence of letters $(a_n)_{n \geq 0}$ such that $a_n = f_{n+1}(a_{n+1})$ for all $n \geq 0$.

Consider the sequence of words $(u_n)_{n \geq 0}$ defined by $u_n = \sigma_1 \sigma_2 \cdots \sigma_n(a_n)$. Since a_n is the first letter of $\sigma_{n+1}(a_{n+1})$, u_n is, by construction, a prefix of u_{n+1}. If the sequence $(u_n)_{n \geq 0}$ is not ultimately periodic, the limit $\lim_{n \to \infty} u_n$ defines an infinite word which is a limit point of the sequence $(\sigma_n)_{n \geq 1}$. If the sequence u_n is ultimately periodic, then its ultimate value is a finite word u. Thus the infinite word u^ω is a limit point of $(\sigma_n)_{n \geq 1}$. □

4 About the structure of stable sets

In this section, we study links between stable sets and substitutive-adicity. We also provide a description of elements of a stable set $\mathrm{Stab}(\mathcal{S})$ that are not $\mathcal{S}$-adic. For a survey on S-adicity see, *e.g.*, Berthé and Delecroix [2]. Let us recall that an infinite word $\mathbf{w}$ is S-adic if there exist a sequence $(\sigma_n)_{n \geq 1}$, $\sigma_n : A^*_{n+1} \to A^*_n$, of substitutions and a sequence of letters $(a_n)_{n \geq 1}$ such that $\mathbf{w} = \lim_{n \to \infty} \sigma_1 \cdots \sigma_n(a_n)$. The sequence $(\sigma_n)_{n \geq 1}$ is called a *directive sequence* of $\mathbf{w}$.

Denoting $\mathcal{S} = \{\sigma_n \mid n \geq 1\}$, $\mathbf{w}$ is $\mathcal{S}$-adic (where $\mathcal{S}$ refers to the set of substitutions and not to the term "substitutive"). In what follows we will only consider sets of substitutions that are endomorphisms. As already said, Berthé and Delecroix [2] mentioned that "*an S-adic system is a system that can be indefinitely desubstituted*". The next result formalizes this in our context.

Proposition 4.1. *Let $\mathcal{S} \subseteq \mathrm{Subst}(A)$. Any $\mathcal{S}$-adic word belongs to $\mathrm{Stab}(\mathcal{S})$.*

Moreover, if $(\sigma_n)_{n \geq 1}$ is the directive word of an $\mathcal{S}$-adic word $\mathbf{w}$, then $(\sigma_n)_{n \geq 1}$ is also a directive sequence of $\mathbf{w}$ as an element of $\mathrm{Stab}(\mathcal{S})$.

Let us observe that Proposition 4.1 is not immediate. If $(\sigma_n)_{n \geq 1}$ is the directive sequence of an $\mathcal{S}$-adic word $\mathbf{w}$, there exists a sequence of letters $(a_n)_{n \geq 1}$ such that $\mathbf{w} = \lim_{n \to \infty} \sigma_1 \sigma_2 \cdots \sigma_n(a_n)$. This does not imply that for all $k \geq 0$, the limit $\lim_{n \to \infty} \sigma_k \sigma_{k+1} \cdots \sigma_n(a_n)$ exists. Hence the fact that $\mathbf{w}$ belongs to $\mathrm{Stab}(\mathcal{S})$ is not immediate. This phenomenon can be illustrated by the following example. Let $f : a \mapsto a, b \mapsto a$ and $g : a \mapsto bb, b \mapsto aa$. For all $n \geq 0$, we have $g^{2n}(a) = a^{2^{2n}}$ $g^{2n+1}(a) = b^{2^{2n+1}}$. Hence taking $a_n = a$ for all $n \geq 1$, we have $\lim_{n \to \infty} fg^n(a) = a^\omega$ but $\lim g^n(a)$ does not exist. Nevertheless taking $\mathbf{w}_{2n} = a^\omega$ and $\mathbf{w}_{2n+1} = b^\omega$ provides a sequence of infinite desubstituted words of a^ω using f and g.

Another difficulty to cope with is that, for some sets $\mathcal{S}$ of substitutions, there may exist a word $\mathbf{w}$ that is $\mathcal{S}$-adic but for which, given any sequence of desubstituted words $(\mathbf{w}_n)_{n \geq 0}$ of $\mathbf{w}$, we do not have $\mathbf{w} = \lim_{n \to \infty} \sigma_1 \cdots \sigma_n(\mathrm{first}(\mathbf{w}_n))$. For instance, let $\mathcal{S} = \{L_a\}$. There is only one word in $\mathrm{Stab}(\mathcal{S})$: the word a^ω. The unique sequence of desubstituted words is $(a^\omega)_{n \geq 0}$ and $a^\omega \neq \lim_{n \to \infty} L_a^n(\mathrm{first}(a^\omega))$. Nevertheless a^ω is $\{L_a\}$-adic since $a^\omega = \lim_{n \to \infty} a^n b = \lim_{n \to \infty} L_a^n(b)$.

The situation in the previous example can be greatly explained by the fact that the finite word a can be indefinitely desubstituted. In Section 4.1, we study the set of finite words that can be indefinitely desubstitutable. In particular, Proposition 4.3 states that, given a directive sequence $\mathbf{s}$, the set of finite words that can be indefinitely desubstituted with $\mathbf{s}$ is finitely generated. The next example shows that this does not mean that the set of finite words that can be indefinitely desubstituted using a given set of substitutions is finitely generated. Set $\mathcal{S} = \{L_a, L_b\}$. The words a and b can be indefinitely desubstituted using, respectively, L_a^ω and L_b^ω as directive sequences. But not all words over $\{a, b\}$ belong to $\mathrm{Stab}(\mathcal{S})$.

After proving Proposition 4.3 in Section 4.1, we use this proposition to prove Proposition 4.1 (in Section 4.2). In Section 4.3, we provide more information on the structure of stable sets. In Section 4.4, we consider a question relative to a converse to Proposition 4.1.

4.1   On desubstitutable finite words

For $\mathbf{s} \in \mathcal{S}^\omega$, let $\mathrm{StabFin}(\mathbf{s})$ and $\mathrm{StabLet}(\mathbf{s})$ denote the sets of, respectively, finite nonempty words and letters that can be indefinitely desubstituted using $\mathbf{s}$ as directive sequence. Set $\mathrm{GenStabFin}(\mathbf{s}) = \mathrm{StabFin}(\mathbf{s}) \setminus (\mathrm{StabFin}(\mathbf{s})\,\mathrm{StabFin}(\mathbf{s})^+)$: this set is the set of all elements of $\mathrm{StabFin}(\mathbf{s})$ that cannot be decomposed as a concatenation of two or more elements of $\mathrm{StabFin}(\mathbf{s})$. Finally, let $\mathrm{StabUltLet}(\mathbf{s})$ be the subset of $\mathrm{StabFin}(\mathbf{s})$ containing words that can be desusbtituted in such a way that desubstituted words are ultimately letters:

$$\mathrm{StabUltLet}(\mathbf{s}) = \{f(\alpha) \mid \mathbf{s} = f\mathbf{s}', f \in \mathcal{S}^*, \mathbf{s}' \in \mathcal{S}^\omega \text{ and } \alpha \in \mathrm{StabLet}(\mathbf{s}')\}$$

Proposition 4.2. *Let $\mathcal{S} \subseteq \mathrm{Subst}(A)$ and let $\mathbf{s} \in \mathcal{S}^\omega$.*

1. $\mathrm{StabFin}(\mathbf{s}) = \mathrm{GenStabFin}(\mathbf{s})^+$

2. $\mathrm{GenStabFin}(\mathbf{s}) \subseteq \mathrm{StabUltLet}(\mathbf{s})$

Proof. Relation 1 follows directly the definition of $\mathrm{StabFin}(\mathbf{s})$ and $\mathrm{GenStabFin}(\mathbf{s})$.

Relation 2 is also a consequence of the definition of the sets $\mathrm{GenStabFin}(\mathbf{s})$ and $\mathrm{StabUltLet}(\mathbf{s})$. Let $w \in \mathrm{GenStabFin}(\mathbf{s})$. Let $\mathbf{s} = (\sigma_n)_{n \geq 1}$ and let $(w_n)_{n \geq 0}$ be the corresponding sequence of finite desubstituted words: $w_0 = w$ and $w_{n-1} = \sigma_n(w_n)$ for $n \geq 1$. Observe that $|w_{n-1}| \geq |w_n| \geq 1$. Hence the sequence $(|w_n|)_{n \geq 0}$ is ultimately constant. If this constant is 2 or more, then we find a contradiction with indecomposability of elements of $\mathrm{GenStabFin}(\mathbf{s})$. Hence $(|w_n|)_{n \geq 0}$ is ultimately 1 and Relation 2 holds. □

Observe that the inverse inclusion of the second assertion of Proposition 4.2 does not hold in general. For instance if $\mathcal{S} = \{f, \mathrm{Id}\}$ with Id the identity morphism and f defined by $f(a) = bc$, $f(b) = b$ and $f(c) = c$, then $bc \notin \mathrm{GenStabFin}(f.\mathrm{Id}^\omega)$ (with $f.\mathrm{Id}^\omega$ the sequence of substitutions beginning with f and followed by Id^ω), even if $bc = f(a)$, as b and c belong to $\mathrm{GenStabFin}(f.\mathrm{Id}^\omega)$. The same phenomenon exists if one consider, instead of Id, any morphism g such that $g(\{a, b, c\}) = \{a, b, c\}$.

Proposition 4.3. *Let $\mathcal{S} \subseteq \mathrm{Subst}(A)$ and $\mathbf{s} \in \mathcal{S}^\omega$. The sets $\mathrm{GenStabFin}(\mathbf{s})$ and $\mathrm{StabUltLet}(\mathbf{s})$ are finite. Their cardinalities are bounded by the cardinality of A.*

More precisely there exist $p \in \mathcal{S}^*$ and $\mathbf{s}' \in \mathcal{S}^\omega$ such that $\mathbf{s} = p\mathbf{s}'$, and, for all elements u in $\mathrm{StabUltLet}(\mathbf{s})$, $u = p(\alpha)$ for some $\alpha \in \mathrm{StabLet}(\mathbf{s}')$.

Proof. By Relation 2 of Proposition 4.2, it is sufficient to prove the result for $\mathrm{StabUltLet}(\mathbf{s})$. Let $x_1, \ldots, x_k$ be distinct elements of $\mathrm{StabUltLet}(\mathbf{s})$. There exist prefixes $p_1, \ldots, p_k$ of $\mathbf{s}$ (each p_i is a composition of morphisms), suffixes $\mathbf{s}_1, \ldots, \mathbf{s}_k$ of $\mathbf{s}$ and letters $\alpha_1, \ldots, \alpha_k$ such that, for all i, $1 \leq i \leq k$, $x_i = p_i(\alpha_i)$, $\alpha_i \in \mathrm{StabLet}(\mathbf{s}_i)$ and $\mathbf{s} = p_i\mathbf{s}_i$. Let p be the longest word among the words $p_1, \ldots, p_k$. Let $\mathbf{s}'$ be the corresponding suffix of $\mathbf{s}$: $\mathbf{s} = p\mathbf{s}'$. For each i, let f_i such that $p = p_i f_i$. As $\alpha_i \in \mathrm{StabLet}(\mathbf{s}_i)$, there exists a letter $\alpha_i' \in \mathrm{StabLet}(\mathbf{s}')$ such that $\alpha_i = f_i(\alpha_i')$. Hence for all i, $x_i = p(\alpha_i')$. As $\mathrm{StabLet}(\mathbf{s}')$ is a subset of A, we conclude that $k \leq \#A$. So $\mathrm{StabUltLet}(\mathbf{s})$ is finite. $\square$

4.2 Proof of Proposition 4.1

Let $\mathbf{w}$ be an $\mathcal{S}$-adic word. There exist a sequence $(\sigma_n)_{n \geq 1}$ of elements of $\mathcal{S}^\omega$ and a sequence of letters $(a_n)_{n \geq 1}$ such that $\mathbf{w} = \lim_{n \to \infty} \sigma_1 \sigma_2 \cdots \sigma_n(a_n)$. We prove that $\mathbf{w} \in \mathrm{Stab}(\mathcal{S})$ and that $(\sigma_n)_{n \geq 1}$ is a directive sequence of $\mathbf{w}$. We have to construct a sequence of desubstituted words.

Note that, for all $m \geq 1$, $\lim_{n \to \infty} |\sigma_m \cdots \sigma_n(a_n)| = \infty$. Let $k \geq 1$ be an integer. We first consider links between prefixes of length k of words $\sigma_m \cdots \sigma_n(a_n)$ (when $k \leq |\sigma_m \cdots \sigma_n(a_n)|$). Informally, the idea of the proof is that if we can find a k and a sequence $(p_m)_{m \geq 0}$ of such prefixes such that, for all m, p_m is a prefix of $\sigma_m(p_{m+1})$ and $\lim_{m \to \infty} |\sigma_1 \sigma_2 \cdots \sigma_m(p_m)| = \infty$, then the words $\lim_{n \to \infty} \sigma_m \cdots \sigma_n(p_n)$ would form a sequence of desubstituted words. If we cannot find any k and any sequence of such prefixes, then $\mathbf{w}$ has infinitely many prefixes that can be indefinitely desubstituted using the directive sequence $(\sigma_n)_{n \geq 1}$ and the construction of desubstituted words would differ.

Let $\mathcal{V}_k$ be the set of all pairs (p, m) with $m \geq 1$ an integer and p a word of length k such that there exists an integer $n \geq m$ for which p is a prefix of $\sigma_m \cdots \sigma_n(a_n)$. Let $\mathcal{G}_k$ be the directed acyclic graph whose vertices are the elements of $\mathcal{V}_k$ and the edges are the pairs $((p, m), (p', m + 1))$ of elements of $\mathcal{V}_k$ such that p is the prefix of length k of $\sigma_m(p')$.

Let $(p', m+1)$ be an element of $\mathcal{V}_k$ with $m \geq 1$. Let $n \geq m+1$ be an integer such that p' is a prefix of $\sigma_{m+1} \cdots \sigma_n(a_n)$. Let p be the prefix of length k of $\sigma_m(p')$: p is also a prefix of $\sigma_m \sigma_{m+1} \cdots \sigma_n(a_n)$. Hence $(p, m) \in \mathcal{V}_k$ and $((p, m), (p', m + 1))$ is an edge of $\mathcal{G}_k$. More precisely, since $\sigma_m(p')$ has a unique prefix of length k, $((p, m), (p', m + 1))$ is the unique edge of $\mathcal{G}_k$ ending on $(p', m + 1)$. It follows that $\mathcal{G}_k$ is an infinite forest.

Let π_k be the prefix of length k of $\mathbf{w}$. As $\mathbf{w} = \lim_{n \to \infty} \sigma_1 \sigma_2 \cdots \sigma_n(a_n)$, there exist infinitely many elements (p, m) in $\mathcal{V}_k$ for which there exists a path from $(\pi_k, 1)$ to (p, m). Let $\mathcal{T}_k$ be the subgraph of $\mathcal{G}_k$ induced by elements of $\mathcal{V}_k$ that are accessible from $(\pi_k, 1)$. The graph $\mathcal{T}_k$ is an infinite tree. By Konig's lemma, there exists an infinite path $(p_i, i)_{i \geq 1}$ starting from $(\pi_k, 1)$.

Let m, n with $n \geq m \geq 1$. By construction, p_m is the prefix of length k of $\sigma_m \cdots \sigma_{n-1}(p_n)$ and, by definition of $\mathcal{V}_k$, p_n is a prefix of $\sigma_n \cdots \sigma_\ell(a_\ell)$ for some $\ell \geq 1$. Hence p_m is the prefix of length k of $\sigma_m \cdots \sigma_\ell(a_\ell)$ for arbitrary large ℓ. It follows that $\sigma_1 \cdots \sigma_{m-1}(p_m)$ is a prefix of $\sigma_1 \cdots \sigma_\ell(a_\ell)$ for arbitrary large ℓ. As $\mathbf{w} = \lim_{n \to \infty} \sigma_1 \sigma_2 \cdots \sigma_n(a_n)$, $\sigma_1 \cdots \sigma_{m-1}(p_m)$ is a prefix of $\mathbf{w}$.

Assume that $\lim_{n \to \infty} |\sigma_1 \cdots \sigma_{n-1}(p_n)| = \infty$. We have $\mathbf{w} = \lim_{n \to \infty} \sigma_1 \cdots \sigma_{n-1}(p_n)$. Let $m \geq 1$. We also have $\lim_{n \to \infty} |\sigma_m \cdots \sigma_{n-1}(p_n)| = \infty$. Moreover, by construction, for all $n \geq m$, $\sigma_m \cdots \sigma_{n-1}(p_n)$ is a prefix of $\sigma_m \cdots \sigma_n(p_{n+1})$. Thus $\lim_{n \to \infty} \sigma_m \cdots \sigma_{n-1}(p_n)$ defines an infinite word. We denote it $\mathbf{w}_m$. It follows from the construction that, for all $m \geq 1$, $\mathbf{w}_m = \sigma_m(\mathbf{w}_{m+1})$ and $\mathbf{w}_1 = \mathbf{w}$. Hence $\mathbf{w} \in \mathrm{Stab}(\mathcal{S})$ and $\mathbf{w}$ has $(\sigma_n)_{n \geq 1}$ as directive sequences.

To end the proof of Proposition 4.1, we consider the case where for all $k \geq 1$, and for all infinite paths $(p_i, i)_{i \geq 1}$ in $\mathcal{T}_k$, $\lim_{n \to \infty} |\sigma_1 \cdots \sigma_{n-1}(p_n)|$ is finite.

Let $k \geq 1$ be an integer and let $(p_i, i)_{i \geq 1}$ be an infinite path in $\mathcal{T}_k$. As, for all $i \geq 1$, p_i is a prefix of $\sigma_i(p_{i+1})$, $\sigma_1 \cdots \sigma_{i-1}(p_i)$ is a prefix of $\sigma_1 \cdots \sigma_i(p_{i+1})$. The fact that $\lim_{n \to \infty} |\sigma_1 \cdots \sigma_{n-1}(p_n)|$ is finite implies that there exists an integer N such that, for all $n \geq N$, $|\sigma_1 \cdots \sigma_{n-1}(p_n)| = |\sigma_1 \cdots \sigma_n(p_{n+1})|$. It follows that, for all $n \geq N$, $|p_n| = |\sigma_n(p_{n+1})|$ and so $p_n = \sigma_n(p_{n+1})$. As for all $n \geq 0$, $\sigma_1 \cdots \sigma_{n-1}(p_n)$ is a prefix of $\sigma_1 \cdots \sigma_n(p_{n+1})$ and a prefix of $\mathbf{w}$, the limit $\lim_{n \to \infty} \sigma_1 \cdots \sigma_{n-1}(p_n)$ converges to a prefix π of $\mathbf{w}$.

For any $n \geq 1$, $|p_n| = k$. Let us decompose p_n over letters. For i with $1 \leq i \leq k$, let $p_{n,i} \in A$ such that $p_n = p_{n,1} \cdots p_{n,k}$. We have $\pi = [\sigma_1 \cdots \sigma_{N-1}(p_{N,1})] \cdots [\sigma_1 \cdots \sigma_{N-1}(p_{N,k})]$ and for all $n \geq N$ and for all i, $1 \leq i \leq k$, $p_{n,i} = \sigma_n(p_{n+1,i})$. In other words, the prefix π of $\mathbf{w}$ can be decomposed into k factors that are indefinitely desubstitutable with $(\sigma_n)_{n \geq 1}$ as directive sequence: $\pi \in \mathrm{StabUltLet}(\mathbf{s})$ with $\mathbf{s} = (\sigma_n)_{n \geq 1}$.

What is described before holds for arbitrary $k \geq 1$ since we have assumed: for all $k \geq 1$, and for all infinite paths $(p_i, i)_{i \geq 1}$ in $\mathcal{T}_k$, $\lim_{n \to \infty} |\sigma_1 \cdots \sigma_{n-1}(p_n)|$ is finite. This means that $\mathbf{w}$ has infinitely many prefixes in $(\mathrm{StabUltLet}(\mathbf{s}))^*$. By Proposition 4.3, $\mathrm{StabUltLet}(\mathbf{s})$ is finite. Thus $\mathbf{w}$ belongs to $(\mathrm{StabUltLet}(\mathbf{s}))^\omega$.

Let us decompose $\mathbf{w}$ over $\mathrm{StabUltLet}(\mathbf{s})$. Let $(u_i)_{i \geq 1}$ in $\mathrm{StabUltLet}(\mathbf{s})$ such that $\mathbf{w} = \prod_{i \geq 1} u_i$. For each $i \geq 1$, let $(u_{i,j})_{j \geq 0}$ be the sequence of desubstituted words of u_i: $u_i = u_{i,0}$ and $u_{i,j} = \sigma_{j+1}(u_{i,j+1})$ for $j \geq 0$. Let $\mathbf{w}_j = \prod_{i \geq 1} u_{i,j}$. By construction, $\mathbf{w} = \mathbf{w}_0$ and, for $j \geq 0$, $\mathbf{w}_j = \sigma_{j+1}(\mathbf{w}_{j+1})$. So $\mathbf{w}$ belongs to $\mathrm{Stab}(\mathcal{S})$ and $(\sigma_n)_{n \geq 0}$ is a directive sequence of $\mathbf{w}$.

4.3 More on the structure on stable sets

For $\mathbf{s} \in \mathcal{S}^\omega$, let $\mathrm{Stab}(\mathbf{s})$ denote the set of infinite words that can be indefinitely desubstituted using $\mathbf{s}$ as directive sequence. Let $\mathrm{adic}(\mathbf{s})$ be the set of infinite words that are S-adic with $\mathbf{s}$ as directive sequence.

Proposition 4.4. *Let $\mathcal{S} \subseteq \mathrm{Subst}(A)$ and let $\mathbf{s} \in \mathcal{S}^\omega$. We have:*

$$\mathrm{Stab}(\mathbf{s}) = \mathrm{GenStabFin}(\mathbf{s})^\omega \cup (\mathrm{GenStabFin}(\mathbf{s}))^* \mathrm{adic}(\mathbf{s})$$

Proof. The inclusion $\mathrm{GenStabFin}(\mathbf{s})^\omega \cup (\mathrm{GenStabFin}(\mathbf{s}))^* \mathrm{adic}(\mathbf{s}) \subseteq \mathrm{Stab}(\mathcal{S})$ follows directly the definitions. For the converse, assume that $\mathbf{w} \in \mathrm{Stab}(\mathbf{s})$. Let $\mathbf{s} = (\sigma_n)_{n \geq 1}$ and let $(\mathbf{w}_n)_{n \geq 0}$ be the corresponding sequence of desubstituted words. Let also $(a_n)_{n \geq 0}$ be the sequence $(\mathrm{first}(\mathbf{w}_n))_{n \geq 0}$. If there exist arbitrarily large integers n such that $|\sigma_n(a_n)| \geq 2$, then $\mathbf{w}(= \mathbf{w}_0)$ belongs to $\mathrm{adic}(\mathbf{s})$. Otherwise we ultimately have $a_{n-1} = \sigma_n(a_n)$. Set $u = \lim_{n \to \infty} \sigma_1 \cdots \sigma_n(a_n)$. This word u belongs to $\mathrm{StabFin}(\mathbf{s})$. Thus, by Relation 1 of Proposition 4.2, $\mathbf{w} = u\mathbf{w}'$ with $u \in \mathrm{GenStabFin}(\mathbf{s})^+$ and $\mathbf{w}' \in \mathrm{Stab}(\mathbf{s})$. Hence the proof can end by induction on the length of prefixes of $\mathbf{w}$. $\square$

Let us observe that $\mathrm{GenStabFin}(\mathbf{s})^\omega \cup (\mathrm{GenStabFin}(\mathbf{s}))^* \mathrm{adic}(\mathbf{s})$ may not be empty. This is the case for instance when $\mathbf{s} = L_a^\omega$ since $a^\omega = \lim_{n \to \infty} L_a^n(b)$ and $\mathrm{GenStabFin}(\mathbf{s}) = \{a\}$.

The next result is a direct consequence of Propositions 4.4 and 4.3.

Corollary 4.5. *Let $\mathcal{S} \subseteq \mathrm{Subst}(A)$. An infinite word $\mathbf{w}$ belongs to $\mathrm{Stab}(\mathcal{S})$ if and only if one of the two following cases holds:*

- $\mathbf{w} = f(u\mathbf{w}')$ *with $f \in \mathcal{S}^*$ and there exists a sequence $\mathbf{s} \in \mathcal{S}^\omega$ such that $u \in \mathrm{StabLet}(\mathbf{s})^*$ and $\mathbf{w}' \in \mathrm{adic}(\mathbf{s})$.*

- $\mathbf{w} = f(\mathbf{w}')$ *with $f \in \mathcal{S}^*$ and $\mathbf{w}' \in (\mathrm{StabLet}(\mathbf{s}))^\omega$ for a sequence $\mathbf{s} \in \mathcal{S}^\omega$.*

The next result is a direct consequence of Corollary 4.5.

Corollary 4.6. *Let $\mathcal{S} \subseteq \mathrm{Subst}(A)$. If $\bigcup_{s \in \mathcal{S}^\omega} \mathrm{StabLet}(\mathbf{s})$ is empty, then the set $\mathrm{Stab}(\mathcal{S})$ is exactly the set of $\mathcal{S}$-adic words.*

Let us observe that the converse of this lemma does not hold, as shown by Proposition 4.8.

Let us observe also that the condition in Corollary 4.6 is decidable when $\mathcal{S}$ is finite. Indeed the set $\mathrm{StabLet}(\mathcal{S})$ can be algorithmically determined as follows. First construct the graph whose vertices

are letters and whose labeled oriented edges are the triples (α, f, β) with α, β letters and $f \in \mathcal{S}$ such that $\alpha = f(\beta)$. Elements of $\mathrm{StabLet}(\mathcal{S})$ are the vertices from which go at least one infinite path in the graph. Labels of such paths are directive sequences of the desubstitutions. Finally, $\bigcup_{s \in \mathcal{S}^\omega} \mathrm{StabLet}(\mathbf{s})$ is empty if and only if there is no circuit in the graph. An open question is to decide, given a finite set $\mathcal{S}$ of substitutions, whether the set $\mathrm{Stab}(\mathcal{S})$ is exactly the set of $\mathcal{S}$-adic words.

4.4 Substitutive-adicity of balanced words

After Corollary 4.6, a natural question is: given a set $\mathcal{S} \subseteq \mathrm{Subst}(A)$, does there exist a set $\mathcal{S}' \subseteq \mathrm{Subst}(A)$ such that $\mathrm{Stab}(\mathcal{S})$ is the set of all $\mathcal{S}'$-adic words. Although Proposition 4.8 below provides an example of positive answer, this does not hold in general as shown by the set $\{\mathrm{Id}\}$ with Id the identity morphism over A^*: $\mathrm{Stab}(\{\mathrm{Id}\}) = A^\omega$ but this set does not correspond to a set of $\mathcal{S}'$-adic words as shown by the next lemma. We say that a morphism is a *permutation* if $\{f(a) \mid a \in A\} = A$.

Lemma 4.7. *Let A be an alphabet containing at least two letters. There exists no set $\mathcal{S} \subseteq \mathrm{Subst}(A)$ such that A^ω is exactly the set of all $\mathcal{S}$-adic words. Moreover any set $\mathcal{S}$ such that $A^\omega = \mathrm{Stab}(\mathcal{S})$ must contain a permutation.*

Proof. Let A be an alphabet with $k \geq 2$ distinct letters $a_1, \ldots, a_k$. Let $\mathcal{S} \subseteq \mathrm{Subst}(A)$ be such that A^ω is the set of all $\mathcal{S}$-adic words. By Proposition 4.1, $A^\omega = \mathrm{Stab}(\mathcal{S})$. Let $\mathbf{w}$ be an infinite word containing all finite words over A as factors. There exists a substitution f in $\mathcal{S}$ and an infinite word $\mathbf{w}'$ such that $\mathbf{w} = f(\mathbf{w}')$. For each i, $1 \leq i \leq k$, $\mathbf{w}$ has arbitrary large factors that are powers of a_i. This implies that there exists a letter $\alpha_i \in A$ and an integer ℓ_i with $f(\alpha_i) = a_i^{\ell_i}$. Note that $A = \{\alpha_i \mid 1 \leq i \leq k\}$. Hence $\mathbf{w} = f(\mathbf{w}')$ belongs to $\{a_1^{\ell_1}, \ldots, a_k^{\ell_k}\}^\omega$. But there also exist arbitrarily large factors of $\mathbf{w}$ in $(a_1 a_2 \cdots a_k)^*$. As $k \geq 2$, for all i, $\ell_i = 1$. The morphism f is a permutation.

Observe that $\mathbf{w}'$ contains all finite words as factors. Thus, by induction, we can see that any directive sequence of $\mathbf{w}$ as an $\mathcal{S}$-adic word is a sequence of permutations. This contradicts the fact that, for any directive sequence $(\sigma_n)_{n \geq 0}$ of $\mathbf{w}$ as an $\mathcal{S}$-adic word, it should exist a sequence of letters $(a_n)_{n \geq 0}$ such that $\mathbf{w} = \lim_{n \to \infty} \sigma_1 \cdots \sigma_n(a_n)$. $\square$

Let us recall that J. Cassaigne provides an example showing that all words of A^ω are $\mathcal{S}$-adic with $\mathcal{S}$ a finite set of substitutions but the substitutions used in this example are defined on a larger alphabet than A and the stable set $\mathrm{Stab}(\mathcal{S})$ contains elements that are not in A^ω (see, *e.g.*, [2, Remark 3]). Let us recall this example. Let $\mathbf{w} = (a_i)_{i \geq 1}$ in A^ω ($a_i \in A$). We have $\mathbf{w} = \lim_{n \to \infty} f \sigma_2 \cdots \sigma_n(\ell)$ with ℓ a letter that does not belong to A and the morphisms f and $(\sigma_i)_{i \geq 2}$ defined as follows.

$$f : \begin{cases} \ell \mapsto a_1 \\ \alpha \mapsto \alpha \end{cases} (\alpha \in A) \qquad \sigma_i : \begin{cases} \ell \mapsto \ell a_i \\ \alpha \mapsto \alpha \end{cases} (\alpha \in A)$$

Note also that the previous lemma is not true for a one-letter alphabet as a^ω is the fixed point of the morphism d defined by $d(a) = aa$.

The next result characterizes balanced words in terms of $\mathcal{S}$-adicity.

Proposition 4.8. *An infinite word over $\{a, b\}$ is balanced if and only if it is $\mathcal{S}_{bal}$-adic.*

Proof. First assume that a word $\mathbf{w}$ is $\mathcal{S}_{bal}$-adic. By Proposition 4.1, $\mathbf{w}$ belongs to $\mathrm{Stab}(\mathcal{S}_{bal})$. Then by Proposition 3.1, it is balanced.

From now on, assume that $\mathbf{w}$ is balanced. By Proposition 3.1, $\mathbf{w} \in \mathrm{Stab}(\mathcal{S}_{bal})$. Let $\mathbf{s} = (\sigma_n)_{n \geq 1} \in \mathcal{S}^\omega$ be a directive sequence of $\mathbf{w}$. As already mentioned in the proof of Proposition 3.1, if $\mathbf{s}$ contains infinitely many elements in $\{L_a, R_a\}$ and infinitely many elements in $\{L_b, R_b\}$, then $\mathbf{w}$ is $\mathcal{S}_{bal}$-adic. In the remaining cases, there exist an integer $N \geq 1$ and a letter $\alpha \in \{a, b\}$ such that, for all $n \geq N$, $\sigma_n \in \{L_\alpha, R_\alpha\}$. Let $f = \sigma_1 \cdots \sigma_{N-1}$ and let $\mathbf{w}'$ be the element of $\mathrm{Stab}(\mathcal{S})$ directed by $(\sigma_n)_{n \geq N}$ such that $\mathbf{w} = f(\mathbf{w}')$. Let β be the letter such that $\{\alpha, \beta\} = \{a, b\}$. Observe that $\mathbf{w}'$ contains at most one occurrence of β. Indeed if it contains a factor $\beta \alpha^n \beta$ for some integer n, then for any i, $1 \leq i \leq n$, the ith desubstituted word of $\mathbf{w}'$ (remember that $\sigma_i \in \{L_\alpha, R_\alpha\}$) would contain the factor $\beta \alpha^{n-i} \beta$. But the nth desubstituted word cannot contain the factor $\beta\beta$ since it can itself be desubstituted using L_α or R_α. Hence $\mathbf{w}' = \alpha^\omega = \lim_{n \to \infty} L_\alpha^n(\beta)$ or $\mathbf{w}' = \alpha^k \beta \alpha^\omega = L_\alpha^k(\lim_{n \to \infty} R_\alpha^n(\beta))$ for some $k \geq 0$. Hence $\mathbf{w}'$ and $\mathbf{w} = f(\mathbf{w}')$ are $\mathcal{S}_{bal}$-adic. $\square$

5 Two other examples of stable sets on binary alphabets

After Remark 2.3, while searching for examples of stable sets defined by a combinatorial property, it is natural to consider properties for which we know the morphisms that preserve them. But this does not provide systematically a stable set as shown by the next example.

An overlap-free word is a word that contains no factor of the form $\alpha u \alpha u \alpha$ with α a letter and u a word. Since Thue [22], it is known that the morphisms that preserve overlap-free words are the morphisms obtained by compositions of the morphism μ ($\mu(a) = ab$, $\mu(b) = ba$) and the exchange morphisms E ($E(a) = b$, $E(b) = a$). As $\mu E = E\mu$, one can check that $\mathrm{Stab}(\{\mu\}) = \mathrm{Stab}(\{\mu, E\}^* \setminus \{E, \mathrm{Id}\}) = \{\mathbf{M}, E(\mathbf{M})\}$ where $\mathbf{M}$, the Thue-Morse word, is the fixed point of μ starting with letter a. It follows that no stable set corresponds to the set of binary overlap-free words. Also $\mathbf{M}$ and $E(\mathbf{M})$ are the unique overlap-free words that can be defined using S-adicity for a set S of morphisms preserving overlap-freeness (as by Proposition 4.1, S-adic words belong to $\mathrm{Stab}(\mathcal{S})$).

5.1 Sturmian words

As mentioned previously the Sturmian words are the binary balanced aperiodic words. We show below that the set of Sturmian words can be defined as a stable set but only with infinite sets of substitutions.

In the proof of Proposition 3.1 we have already recalled that Sturmian words are (exactly) the elements of $\mathrm{Stab}(\mathcal{S}_{bal})$ such that any directive sequence contains infinitely many elements of $\{L_a, R_a\}$ and infinitely many elements of $\{L_b, R_b\}$ [3]. Using the contraction method already used in the proof of Proposition 3.1, this can be reformulated using the set $\mathcal{S}_{Sturm} = \{L_a, R_a\}^+\{L_b, R_b\} \cup \{L_b, R_b\}^+\{L_a, R_a\}$. Thus we have proved the following proposition.

Proposition 5.1. *A word is Sturmian if and only if it belongs to* $\mathrm{Stab}(\mathcal{S}_{Sturm})$.

Proposition 5.2. *There exists no finite set $\mathcal{S}$ of substitutions such that the set of Sturmian words is equal to* $\mathrm{Stab}(\mathcal{S})$.

For the proof of this proposition we need the next lemma.

Lemma 5.3 (see, *e.g.*, [17, chap. 2]). *Morphisms that preserve Sturmian words are the elements of the set* $(\mathcal{S}_{bal} \cup \{E\})^*$.

Proof of Proposition 5.2. Assume that the set of Sturmian words is the set $\mathrm{Stab}(\mathcal{S})$ for some finite set $\mathcal{S}$ of substitutions. First observe that, by Remark 2.3, for all f in $\mathcal{S}$, f preserves the family of Sturmian words. Hence by Lemma 5.3, $\mathcal{S} \subseteq \{L_a, L_b, R_a, R_b, E\}^*$. As $EL_a = L_bE$, $EL_b = L_aE$, $ER_a = R_bE$, $ER_b = R_aE$ and $EE = \mathrm{Id}$, we have $\mathcal{S} \subseteq \{L_a, L_b, R_a, R_b\}^*\{\mathrm{Id}, E\}$. Hence there exist two subsets $\mathcal{F}$ and $\mathcal{G}$ of $\mathcal{S}_{bal}^*$ such that $\mathcal{S} = \mathcal{F} \cup \mathcal{G}E$. As $\mathcal{S}$ is finite, let $\mathcal{F} = \{f_i \mid 1 \leq i \leq \ell\}$ and let $\mathcal{G} = \{g_i \mid 1 \leq i \leq k\}$.

The following fact is important.

Fact. For $i \geq 0$ an integer, for $f \in \mathcal{S}_{bal}^i$ and for an infinite word $\mathbf{w}$, if the word $f(\mathbf{w})$ contains a factor $ba^{j+2}b$ with $j \geq i$, then $f \in \{L_a, R_a\}^i$.

Let us prove this fact by induction on i. It is basically true for $i = 0$. Assume $i \geq 1$. Observe that for any infinite word $\mathbf{u}$, both words $L_b(\mathbf{u})$ and $R_b(\mathbf{u})$ does not contain aa. As aa is a factor of $f(\mathbf{w})$ and $f \in \mathcal{S}_{bal}^i$, it follows that $f = L_a g$ or $f = R_a g$ with $g \in \mathcal{S}_{bal}^{i-1}$. Whatever is the decomposition of f, $g(\mathbf{w})$ contains a factor $ba^{j+1}b = ba^{(j-1)+2}b$. As $i - 1 \leq j - 1$, by induction, $g \in \{L_a, R_a\}^{i-1}$. hence $f \in \{L_a, R_a\}^i$.

Now let $M = 2\max(\{|f_i| \mid 1 \leq i \leq \ell\} \cup \{|g_i| \mid 1 \leq i \leq k\})$ (here, for $f \in \mathcal{S}_{bal}^*$, $|f|$ is the length of f considered as a word over the alphabet $\mathcal{S}_{bal}$). Let $\mathbf{w}$ be a Sturmian word that contains the factor $ba^{M+2}b$. Let $(\sigma_n)_{n \geq 1} \in \mathcal{S}^\omega$ be a directive word of $\mathbf{w}$ (as an element of $\mathrm{Stab}(\mathcal{S})$).

For any morphism f in $(\mathcal{S}_{bal} \cup E)^*$, let $\bar{f}$ be the morphism obtained from the decomposition of f over $\{L_a, L_b, R_a, R_b\}$ replacing each occurrence of L_a with L_b, each occurrence of R_a with R_b, each occurrence of L_b with L_a and each occurrence of R_b with R_a. Observe that $fE = E\bar{f}$.

Let σ_1', σ_2' be the morphisms defined as follows.

- if $\sigma_1 \in \mathcal{F}$ and $\sigma_2 \in \mathcal{F}$, set $\sigma_1' = \sigma_1$, $\sigma_2' = \sigma_2$ (note that $\sigma_1'\sigma_2' = \sigma_1\sigma_2$);

- if $\sigma_1 \in \mathcal{F}$ and $\sigma_2 \in \mathcal{G}E$, set $\sigma_1' = \sigma_1$, $\sigma_2' = \sigma_2 E$ (note that $\sigma_1' \sigma_2' = \sigma_1 \sigma_2 E$);

- if $\sigma_1 \in \mathcal{G}E$ and $\sigma_2 \in \mathcal{F}$, set $\sigma_1' = \sigma_1 E$, $\sigma_2' = \overline{\sigma_2}$ (note that $\sigma_1' \sigma_2' = \sigma_1 \sigma_2 E$);

- if $\sigma_1 \in \mathcal{G}E$ and $\sigma_2 \in \mathcal{G}E$, set $\sigma_1' = \sigma_1 E$, $\sigma_2' = \overline{\sigma_2}E$ (note that $\sigma_1' \sigma_2' = \sigma_1 \sigma_2$).

In all cases, σ_1' and σ_2' belong to $\mathcal{S}_{bal}^*$ and $\mathbf{w} = \sigma_1' \sigma_2'(\mathbf{w}')$ for some infinite word $\mathbf{w}'$. As $|\sigma_1' \sigma_2'| \leq M$ and as $\mathbf{w}$ contains the factor $ba^{M+2}b$, by the initial fact, we must have $\sigma_1' \sigma_2' \in \{L_a, R_a\}^*$, and so σ_1', $\sigma_2' \in \{L_a, R_a\}^*$

If $\sigma_1 \in \mathcal{F}$, we get $\sigma_1 \in \{L_a, R_a\}^*$. If $\sigma_1 \notin \mathcal{F}$ and $\sigma_2 \in \mathcal{F}$, from $\sigma_2' \in \{L_a, R_a\}^*$, we get $\sigma_2 \in \{L_b, R_b\}^*$. If $\sigma_1 \notin \mathcal{F}$ and $\sigma_2 \notin \mathcal{F}$, we have $\sigma_1 \sigma_2 \in \{L_a, R_a\}^*$. But then, one of the words a^ω or b^ω belongs to $\mathrm{Stab}(\mathcal{S})$ with σ_1^ω, σ_2^ω or $(\sigma_1 \sigma_2)^\omega$ as a directive sequence. As these words are not Sturmian, we have a contradiction with the hypothesis that $\mathrm{Stab}(\mathcal{S})$ is the set of Sturmian words. $\square$

The previous result shows that the set of aperiodic words of a stable set of a finite set of substitutions may not form also a stable set of a finite set of substitutions by themselves.

5.2 Lyndon Sturmian words

Let us recall that an *infinite Lyndon word* is a word smaller, with respect to the lexicographic order, than its suffixes. Here, without loss of generality, we restrict our attention on the ordered alphabet $\{a < b\}$, that is on the alphabet $\{a, b\}$ with $a < b$. Results for the order $\{b < a\}$ can be obtained exchanging the roles of a and b. Let us show that the set of binary balanced infinite Lyndon words form a stable set. Set $\mathcal{S}_{Lynd} = \{L_a^n R_b, R_b^n L_a \mid n \geq 1\}$.

Proposition 5.4.

- *A word is a Lyndon Sturmian word over $\{a < b\}$ if and only if it belongs to $\mathrm{Stab}(\mathcal{S}_{Lynd})$ if and only if it is $\mathcal{S}_{Lynd}$-adic.*

- *There is no finite set $\mathcal{S}$ such that the set of Lyndon Sturmian words is $\mathrm{Stab}(\mathcal{S})$.*

Proof. The first part is a direct consequence of Theorems 5.6 and 6.5 in [15] from which we have: *A Sturmian word $\mathbf{w}$ is a Lyndon word over $\{a < b\}$ or over $\{b < a\}$ if and only if there exists a sequence $(d_n)_{n \geq 0}$ of integers such that $d_1 \geq 0$, $d_k \geq 1$ for all $k \geq 2$ and*

$$\mathbf{w} = \lim_{n \to \infty} L_a^{d_1} R_b^{d_2} L_a^{d_3} R_b^{d_4} \cdots L_a^{d_{2n-1}} R_b^{d_{2n}}(a)$$

or

$$\mathbf{w} = \lim_{n \to \infty} L_b^{d_1} R_a^{d_2} L_b^{d_3} R_a^{d_4} \cdots L_b^{d_{2n-1}} R_a^{d_{2n}}(a)$$

Consequently a Sturmian word $\mathbf{w}$ is a Lyndon word over $\{a < b\}$ if and only if it is $\mathcal{S}_{Lynd}$-adic. Observe that, for any f in $\mathcal{S}_{Lynd}$, $|f(a)| \geq 2$ and $|f(b)| \geq 2$. Hence the set $\bigcup_{\mathbf{s} \in \mathcal{S}_{Lynd}^\omega} \mathrm{StabLet}(\mathbf{s})$ is empty. By Corollary 4.6, the set of $\mathcal{S}_{Lynd}$-adic words is equal to $\mathrm{Stab}(\mathcal{S}_{Lynd})$.

The second part is a consequence of the fact that morphisms that preserve Lyndon Sturmian words are the elements of $\{L_a, R_b\}^*$ [20]. Assume that the set of Lyndon Sturmian words is $\mathrm{Stab}(\mathcal{S})$ for some $\mathcal{S} \subseteq \mathrm{Subst}(\{a, b\})$. By Remark 2.3, $\mathcal{S} \subseteq \{L_a, R_b\}^*$. But no element of $\mathcal{S}$ can belong to L_a^* or R_b^* as a^ω and b^ω (that are indefinitely desubstitutable over L_a and R_b respectively) are not Lyndon words. Hence any element $f \in \mathcal{S}$ should be a concatenation of elements of $\{L_a, R_b\}$ with at least one occurrence of L_a and one occurrence of R_b. But then, as in the proof of Proposition 5.2, $\mathcal{S}$ cannot be finite as otherwise the number of occurrences of the letter a between two occurrences of the letter b would be bounded. $\square$

6 Episturmian words

In [5, 10] an *episturmian word* is defined as an infinite word whose set of factors is closed under reversal and that has at most one left special factor of each length (Let us recall that u is a *left special factor* of a word $\mathbf{w}$ if au and bu are factors of $\mathbf{w}$ for a and b two different letters). The subset of the *strict episturmian words* (or *A-strict episturmian words* with A the alphabet) corresponds to the *Arnoux-Rauzy words*. It is the set of infinite words having exactly one left special factor of each length, whose set of factors is closed under reversal and such that if u is a left special factor then all words αu, for $\alpha \in A$, are also factors. Also an episturmian word is *standard* if its left special factors are its prefixes.

The following morphisms are important when studying episturmian words. They extend to arbitrary alphabet the morphisms of $\mathcal{S}_{bal}$, and the exchange morphism E. For $a, b \in A$, let $E_{a,b}$ be the substitution defined by $E_{a,b}(a) = b$, $E_{b,a}(b) = a$ and $E_{b,a}(c) = c$ for $c \in A \setminus \{a, b\}$. For $\alpha \in A$, let L_α and R_α be the following substitutions:

$$L_\alpha : \begin{cases} \alpha \mapsto \alpha \\ \beta \mapsto \alpha\beta \text{ for } \beta \neq \alpha,\ \beta \in A \end{cases} \qquad R_\alpha : \begin{cases} \alpha \mapsto \alpha \\ \beta \mapsto \beta\alpha \text{ for } \beta \neq \alpha,\ \beta \in A \end{cases}$$

Let $\mathcal{L} = \{L_\alpha \mid \alpha \in A\}$ and $\mathcal{R} = \{R_\alpha \mid \alpha \in A\}$. Let Exch be the set $\{E_{\alpha,\beta} \mid \alpha, \beta \in A\}$. Observe that elements of Exch* are the permutations. The set Exch* is finite and its cardinality is $\#A!$. Here follows some useful relations. For pairwise distinct letters α, β, γ, we have

$$L_\alpha E_{\beta,\gamma} = E_{\beta,\gamma} L_\alpha \tag{1}$$
$$R_\alpha E_{\beta,\gamma} = E_{\beta,\gamma} R_\alpha \tag{2}$$
$$L_\alpha E_{\alpha,\beta} = E_{\alpha,\beta} L_\beta \tag{3}$$
$$R_\alpha E_{\alpha,\beta} = E_{\alpha,\beta} R_\beta \tag{4}$$

6.1 Standard episturmian words and stable sets

As far as we know, the next statement is the unique one that characterized explicitly a combinatorial family of words in terms of a stable set before the current paper.

Proposition 6.1 ([10, Cor. 2.7])**.** *The set of standard episturmian words is* $\mathrm{Stab}(\mathcal{L})$.

Strict standard episturmian words correspond to the words whose directive sequence contains infinitely many L_α for each letter α (see [10]). In the binary case they correspond to the standard Sturmian words, also called characteristic words, that is, to the infinite binary words such that their prefixes are exactly their special factors. It is known that the standard Sturmian words are the $\mathcal{L}$-adic words (see, *e.g.*, [17, Prop. 2.2.24]). This does not extend to strict standard episturmian words for alphabets with three letters or more. Indeed over an alphabet containing three letters or more, being $\mathcal{L}$-adic does not imply that there exists a directive sequence such that, for each letter α, L_α occurs infinitely often in the directive sequence. For instance, the episturmian word directed by $L_c(L_a L_b)^\omega$ is $\mathcal{L}$-adic but it not a strict standard episturmian word.

Let $\mathcal{L}_{StrictStand}$ be the set of all elements of $\mathcal{L}^*$ whose decompositions on $\mathcal{L}$ contain at least one occurrence of each element of $\mathcal{L}$ and the last substitution in the decomposition has no previous occurrence in the decomposition: $f \in \mathcal{L}_{StrictStand}$ if $f = L_{a_1} L_{a_2} \cdots L_{a_k}$ for some letters $a_1, \ldots, a_k$, if for each $\alpha \in A$, there exists i, $1 \leq i \leq k$ such that $a_i = \alpha$, and if $a_k \notin \{a_1, \ldots, a_{k-1}\}$. (Observe that the decomposition $L_{a_1} L_{a_2} \cdots L_{a_k}$ is unique: This can be verified directly using the fact that, for an infinite word $\mathbf{w}$ containing a, $L_a(\mathbf{w})$ contains aa) For instance, when $A = \{a, b\}$, $\mathcal{L}_{StrictStand}$ $= \{L_a^n L_b, L_b^n L_a \mid n \geq 1\}$. From the definition of strict standard episturmian words and from the definition of the set $\mathcal{L}_{StrictStand}$, the first part of the next proposition may be directly verified.

Proposition 6.2. *Let A be an alphabet containing at least two letters.*

- *The set of A-strict standard episturmian words is the set* $\mathrm{Stab}(\mathcal{L}_{StrictStand})$

- *There exists no finite set* $\mathcal{S} \subseteq \mathrm{Subst}(A)$ *such that* $\mathrm{Stab}(\mathcal{L}_{StrictStand}) = \mathrm{Stab}(\mathcal{S})$.

- *The set of standard episturmian words is the set of $\mathcal{L}$-adic words.*

Proof. The proof of the second part is similar to the proof of Proposition 5.2 but we have to face a combinatorial problem due to the fact that the alphabet may have more than two letters.

Assume, by contradiction, that the set of A-strict standard episturmian words, $(\mathrm{Stab}(\mathcal{L}_{StrictStand})$ by the first part of the proposition) is equal to $\mathrm{Stab}(\mathcal{S})$ for some finite set $\mathcal{S}$ of substitutions. First observe that, by Remark 2.3, for all f in $\mathcal{S}$, f preserves A-strict standard episturmian words. Theorem 10 in [5] states that *a substitution $\sigma \in \mathrm{Subst}(A)$ preserves A-strict standard episturmian words if and only $\sigma \in (\mathcal{L} \cup \mathrm{Exch})^*$.* Hence $\mathcal{S} \subseteq (\mathcal{L} \cup \mathrm{Exch})^*$. By Equations (1) and (3), $\mathcal{S}^* \subseteq \mathcal{L}^* \mathrm{Exch}^*$: any element σ of $\mathcal{S}^*$ can be decomposed into $\sigma = f\pi$ with $f \in \mathcal{L}^*$ and π a permutation.

Let a and b be two different letters. The following fact is important.

Fact. Let i, j be integers such that $i \le j$. Let $\mathbf{w} \in A^\omega$ and $f \in \mathcal{L}^i$. If $f(\mathbf{w})$ contains a factor $ba^{j+2}b$, then $f = L_a^i$.

We prove this fact by induction. It is basically true for $i = 0$. Assume $i \ge 1$. Observe that for $c \ne a$ and for any infinite word $\mathbf{u}$, $L_c(\mathbf{u})$ does not contain aa. As aa is a factor of $f(\mathbf{w})$ and as $f \in \mathcal{L}^i$, it follows that $f = L_a g$ with $g \in \mathcal{L}^{i-1}$. Consequently, $g(\mathbf{w})$ contains a factor $ba^{(j-1)+2}b$ and $i - 1 \le j - 1$. Hence $g = L_a^{i-1}$ by induction. So $f = L_a^i$.

Now let $M_1 = \max(\{|f| \mid f\pi \in \mathcal{S}, f \in \mathcal{L}^*, \pi \in \mathrm{Exch}^*\}$ (here, for $f \in \mathcal{L}^*$, $|f|$ is the length of f considered as a word over the alphabet $\mathcal{L}$). Let $M = (\#A! + 1)M_1$ Let $\mathbf{w}$ be a strict standard episturmian word containing the factor $ba^{M+2}b$. Let $(\sigma_n)_{n \ge 1} \in \mathcal{S}^\omega$ be a directive word of $\mathbf{w}$ (as an element of $\mathrm{Stab}(\mathcal{S})$).

Let i, $1 \le i \le \#A! + 1$. For $1 \le j \le i$, as $\sigma_j \in \mathcal{L}^* \mathrm{Exch}^*$, there exist $g_j \in \mathcal{L}^*$ and $\pi'_j \in \mathrm{Exch}^*$ such that $\sigma_j = g_j \pi'_j$. Using Equations (1) and (3), we can see that there exist $f_i \in \mathcal{L}^*$ and $\pi_i \in \mathrm{Exch}^*$ such that $\sigma_1 \cdots \sigma_i = f_i \pi_i$. Moreover $|f_i| = \sum_{j=1}^i |g_j|$ and so $|f_i| \le iM_1 \le M$. From the previous fact, $f_i \in \{L_a\}^*$. Since the cardinality of Exch^* is $\#A!$, there exists i and j such that $1 \le i < j \le \#A! + 1$, such that $\pi_i = \pi_j$. Set $\pi = \pi_i$.

Let ℓ_i and ℓ_j be the integers such that $\sigma_1 \ldots \sigma_i = L_a^{\ell_i} \pi$ and $\sigma_1 \ldots \sigma_j = L_a^{\ell_j} \pi$. We have $L_a^{\ell_j} \pi = \sigma_1 \ldots \sigma_j = L_a^{\ell_i} \pi \sigma_{i+1} \cdots \sigma_j$. Thus $\pi \sigma_{i+1} \cdots \sigma_j = L_a^{\ell_j - \ell_i} \pi$. By Equations (1) and (3), there exists a letter c such that $L_a^{\ell_j - \ell_i} \pi = \pi L_c^{\ell_j - \ell_i}$. So $\sigma_{i+1} \cdots \sigma_j = L_c^{\ell_j - \ell_i}$.

Now observe that the word c^ω can be indefinitely desubstituted using the morphism $L_c^{\ell_j - \ell_i}$. Thus $c^\omega \in \mathrm{Stab}(\mathcal{S})$. But c^ω is not a strict standard episturmian word as it does not contain all the letters of A. This contradicts the fact that $\mathrm{Stab}(\mathcal{S})$ is the set of strict standard episturmian words.

Let us prove the third part of Proposition 6.2. By Proposition 6.1, standard episturmian words are the elements of $\mathrm{Stab}(\mathcal{L})$. If an element $\mathbf{w}$ of $\mathrm{Stab}(\mathcal{L})$ has a directive sequence $(\sigma_n)_{n \ge 1}$ containing infinitely many occurrences of at least two elements of $\mathcal{L}$ and if $(\mathbf{w}_n)_{n \ge 0}$ is the corresponding sequence of desubstituted words, then one can verify that $\lim_{n \to \infty} |\sigma_1 \cdots \sigma_n(\mathrm{first}(\mathbf{w}_n))| = \infty$, and so, $\mathbf{w}$ is $\mathcal{L}$-adic. If it is ultimately L_α^ω, then $\mathbf{w} = f(\alpha^\omega)$ for some $f \in \mathcal{L}^*$. As there exists a letter β different from α, $\alpha^\omega = \lim_{n \to \infty} L_\alpha^n(\beta)$: $f(\alpha^\omega)$ is $\mathcal{L}$-adic.

Conversely, by Proposition 4.1, the set of $\mathcal{L}$-adic words is a subset of $\mathrm{Stab}(\mathcal{L})$. $\qquad\square$

Let us recall some results on the family of *LSP words*. These words were introduced by G. Fici in 2011 [7] as the words having all their left special factors as prefixes. Standard episturmian words are particular LSP words. The set of factors of these words is not necessarily closed by reversal. In [21] it was proved that the set of LSP words over $\{a, b\}$ is the set $\mathrm{Stab}(\{L_a, L_b\})$. By Proposition 6.1 this means that on a binary alphabet, a word is standard episturmian if and only if it is LSP. This does not hold on larger alphabets. Indeed by [21] there exist no finite set $\mathcal{S} \subseteq \mathrm{Subst}(A)$ such that the set of LSP words over A is $\mathrm{Stab}(\mathcal{S})$.

6.2   Recurrence, stable sets and Episturmian words

Justin and Pirillo [10] obtained also the next result on desubstitutions of episturmian words. Let us recall that a word is *recurrent* if all its factors occur infinitely often, or equivalently, if all its factors occur at least twice.

Theorem 6.3 ([10, Th. 3.10]). *An infinite word $\mathbf{w}$ is episturmian if and only if $\mathbf{w} \in \mathrm{Stab}(\mathcal{L} \cup \mathcal{R})$ and it has a sequence $(\mathbf{w}_n)_{n \geq 0}$ of **recurrent** desubstituted words.*

The previous theorem and Proposition 3.1 show that, in the binary case, the episturmian words are the recurrent balanced words. On arbitrary alphabet, this characterization of episturmian words can be simplified as follows.

Proposition 6.4. *An infinite word $\mathbf{w}$ is episturmian if and only if it is a recurrent element of $\mathrm{Stab}(\mathcal{L} \cup \mathcal{R})$.*

To prove this proposition we need the next two lemmas. The first one is a slight variation of Theorem 5.2 in [8] that concerns only the directive sequence and not the sequence of desubstituted words.

Lemma 6.5. *Let $\mathbf{w}$ be an element of $\mathrm{Stab}(\mathcal{L} \cup \mathcal{R})$. There exist a sequence of infinite words $(\mathbf{w}_n)_{n \geq 0}$ and a sequence $(\sigma_n)_{n \geq 1}$ in $(\mathcal{L} \cup \mathcal{R})^\omega$ such that $\mathbf{w}_0 = \mathbf{w}$, for all $n \geq 0$, $\mathbf{w}_n = \sigma_{n+1}(\mathbf{w}_{n+1})$, and, for all $k \geq 0$, if $\mathbf{w}_k$ begins with the letter α, then $\sigma_{k+1} \neq R_\alpha$.*

Proof. As $\mathbf{w} \in \mathrm{Stab}(\mathcal{L} \cup \mathcal{R})$, there exist sequences $(\mathbf{w}_n)_{n \geq 0}$ and $(\sigma_n)_{n \geq 1}$ such that $\mathbf{w}_0 = \mathbf{w}$ and for all $n \geq 0$, $\mathbf{w}_n = \sigma_{n+1}(\mathbf{w}_{n+1})$.

We construct sequences $(\mathbf{w}'_n)_{n \geq 0}$ and $(\sigma'_n)_{n \geq 1}$ such that $\mathbf{w}'_0 = \mathbf{w}$, $\mathbf{w}'_n = \sigma'_{n+1}(\mathbf{w}'_{n+1})$ for all $n \geq 0$, and, for all $k \geq 0$, if $\mathbf{w}'_k$ begins with the letter α, then $\sigma'_{k+1} \neq R_\alpha$. It may be emphasize that for all $k \geq 0$, there will exist an integer $n_k \geq 0$, such that $\mathbf{w}_k = \mathrm{first}(\mathbf{w}_k)^{n_k} \mathbf{w}'_k$. The construction is done by an infinite induction. Set $\mathbf{w}'_0 = \mathbf{w}_0$.

Let $k \geq 0$ be an integer such that $(\mathbf{w}'_i)_{0 \leq i \leq k}$ and $(\sigma'_i)_{1 \leq i \leq k}$ are already defined.

Assume first that $\mathbf{w}'_k = \mathbf{w}_k$.

Let $\alpha = \mathrm{first}(\mathbf{w}_k)$. If $\sigma_{k+1} \neq R_\alpha$ (that is $\sigma_{k+1} = R_\beta$ with $\beta \neq \alpha$ or $\sigma_{k+1} = L_\beta$ (possibly $\beta = \alpha$)), set $\mathbf{w}'_{k+1} = \mathbf{w}_{k+1}$ and $\sigma'_{k+1} = \sigma_{k+1}$.

If $\sigma_{k+1} = R_\alpha$, from $\mathbf{w}_k = R_\alpha(\mathbf{w}_{k+1})$, we get $\mathrm{first}(\mathbf{w}_{k+1}) = \alpha$. Let $\mathbf{w}'_{k+1}$ be the word such that $\mathbf{w}_{k+1} = \alpha \mathbf{w}'_{k+1}$. Set $\sigma'_{k+1} = L_\alpha$. We have $\mathbf{w}'_k = \mathbf{w}_k = R_\alpha(\mathbf{w}_{k+1}) = R_\alpha(\alpha \mathbf{w}'_{k+1}) = \alpha R_\alpha(\mathbf{w}'_{k+1}) = L_\alpha(\mathbf{w}'_{k+1}) = \sigma'_{k+1}(\mathbf{w}'_{k+1})$.

To continue the induction step, we now have to consider cases where $\mathbf{w}_k = \alpha^n \mathbf{w}'_k$ for some letter α and some integer $n \geq 1$. Observe that, since $\alpha^n \mathbf{w}'_k = \mathbf{w}_k = \sigma_{k+1}(\mathbf{w}_{k+1})$, we cannot have $\sigma_{k+1} = L_\beta$ with $\beta \neq \alpha$. Observe also that if $\sigma_{k+1} = L_\alpha$, then $\mathbf{w}_{k+1}$ begins with $\alpha^{n-1}\beta$ where $\beta = \mathrm{first}(\mathbf{w}'_k)$.

If $\sigma_{k+1} = L_\alpha$ and $\alpha \neq \beta$, let $\mathbf{w}'_{k+1}$ be the word such that $\mathbf{w}_{k+1} = \alpha^{n-1} \mathbf{w}'_{k+1}$. Set $\sigma'_{k+1} = R_\alpha$. From $\alpha^n \mathbf{w}'_k = \mathbf{w}_k = L_\alpha(\mathbf{w}_{k+1}) = L_\alpha(\alpha^{n-1} \mathbf{w}'_{k+1}) = \alpha^{n-1} L_\alpha(\mathbf{w}'_{k+1})$, we get $\alpha \mathbf{w}'_k = L_\alpha(\mathbf{w}'_{k+1})$ and so $\mathbf{w}'_k = R_\alpha(\mathbf{w}'_{k+1}) = \sigma'_{k+1}(\mathbf{w}'_{k+1})$.

If $\sigma_{k+1} = L_\alpha$ and $\alpha = \beta$, let $\mathbf{w}'_{k+1}$ be the word such that $\mathbf{w}_{k+1} = \alpha^n \mathbf{w}'_{k+1}$. Set $\sigma'_{k+1} = L_\alpha$. From $\alpha^n \mathbf{w}'_k = \mathbf{w}_k = L_\alpha(\mathbf{w}_{k+1}) = L_\alpha(\alpha^n \mathbf{w}'_{k+1}) = \alpha^n L_\alpha(\mathbf{w}'_{k+1})$, we get $\mathbf{w}'_k = L_\alpha(\mathbf{w}'_{k+1}) = \sigma'_{k+1}(\mathbf{w}'_{k+1})$.

If $\sigma_{k+1} = R_\beta$ for some letter $\beta \neq \alpha$, from $\alpha^n \mathbf{w}'_k = \mathbf{w}_k = R_\beta(\mathbf{w}_{k+1})$, we deduce successively that $\mathrm{first}(\mathbf{w}_{k+1}) = \alpha$, $\mathbf{w}_k$ begins with $\alpha\beta$ and $n = 1$. Let $\mathbf{w}'_{k+1}$ be the word such that $\mathbf{w}_{k+1} = \alpha \mathbf{w}'_{k+1}$. Set $\sigma'_{k+1} = L_\beta$. From $\alpha \mathbf{w}'_k = R_\beta(\alpha \mathbf{w}'_{k+1}) = \alpha\beta R_\beta(\mathbf{w}'_{k+1})$, we get $\mathbf{w}'_k = \beta R_\beta(\mathbf{w}'_{k+1}) = L_\beta(\mathbf{w}'_{k+1}) = \sigma'_{k+1}(\mathbf{w}'_{k+1})$.

If $\sigma_{k+1} = R_\alpha$, from $\alpha^n \mathbf{w}'_k = \mathbf{w}_k = R_\alpha(\mathbf{w}_{k+1})$, we deduce that $\mathbf{w}_{k+1}$ begins with $\alpha^n\beta$ with β the first letter of $\mathbf{w}'_k$. If $\alpha \neq \beta$, let $\mathbf{w}'_{k+1}$ be the word such that $\mathbf{w}_{k+1} = \alpha^n \mathbf{w}'_{k+1}$. Set $\sigma'_{k+1} = R_\alpha$. From $\alpha^n \mathbf{w}'_k = R_\alpha(\alpha^n \mathbf{w}'_{k+1}) = \alpha^n R_\alpha(\mathbf{w}'_{k+1})$, we get $\mathbf{w}'_k = R_\alpha(\mathbf{w}'_{k+1}) = \sigma'_{k+1}(\mathbf{w}'_{k+1})$. If $\mathbf{w}_{k+1}$ begins with α^{n+1}, let $\mathbf{w}'_{k+1}$ be the word such that $\mathbf{w}_{k+1} = \alpha^{n+1} \mathbf{w}'_{k+1}$. Set $\sigma'_{k+1} = L_\alpha$. From $\alpha^n \mathbf{w}'_k = R_\alpha(\alpha^{n+1} \mathbf{w}'_{k+1}) = \alpha^n \alpha R_\alpha(\mathbf{w}'_{k+1})$, we get $\mathbf{w}'_k = \alpha R_\alpha(\mathbf{w}'_{k+1}) = L_\alpha(\mathbf{w}'_{k+1}) = \sigma'_{k+1}(\mathbf{w}'_{k+1})$.

Observe that, in all cases, we have by construction $\sigma'_{k+1} \neq R_{\mathrm{first}(w'_k)}$. $\qquad \square$

Lemma 6.6. *Let $\mathbf{w}$ be an infinite word and let α be a letter.*

- *If $L_\alpha(\mathbf{w})$ is recurrent then $\mathbf{w}$ is recurrent.*

- *If $R_\beta(\alpha\mathbf{w})$ is recurrent then $\alpha\mathbf{w}$ is recurrent for any letter $\beta \neq \alpha$.*

Proof. Assume by contradiction that $\mathbf{w}$ is not recurrent. It has a factor u that occurs only once. Let $v = L_\alpha(u)\alpha$. Any occurrence of v in $L_\alpha(\mathbf{w})$ must come from an occurrence of u in $\mathbf{w}$. This contradicts the recurrence property of $L_\alpha(\mathbf{w})$.

Assume by contradiction that $\alpha\mathbf{w}$ is not recurrent. As $\alpha \neq \beta$, we have $R_\beta(\alpha\mathbf{w}) = \alpha\beta R_\beta(\mathbf{w}) = \alpha L_\beta(\mathbf{w})$. As $R_\beta(\alpha\mathbf{w})$ is recurrent, $L_\beta(\mathbf{w})$ is recurrent and, from what precedes, $\mathbf{w}$ is also recurrent. Hence $\alpha\mathbf{w}$ has a prefix v such that αv occurs only once in $\alpha\mathbf{w}$. As $R_\beta(\alpha\mathbf{w})$ is recurrent, the word $R_\beta(\alpha v)$ occurs at a non-prefix position in $R_\beta(\alpha\mathbf{w})$. This implies that αv also occurs at a non-prefix position in $\alpha\mathbf{w}$. A contradiction. $\square$

Proof of Proposition 6.4. Assume that $\mathbf{w}$ is a recurrent element of $\mathrm{Stab}(\mathcal{L} \cup \mathcal{R})$. There exist sequences $(\mathbf{w}_n)_{n\geq 0}$ and $(\sigma_n)_{n\geq 1}$ as in the statement of Lemma 6.5. By induction, using Lemma 6.6 we get the recurrence of all words $\mathbf{w}_n$. By Theorem 6.3, $\mathbf{w}$ is episturmian.

The converse also follows immediately from Theorem 6.3. $\square$

One can observe that while desubstituting balanced words in the proof of Proposition 3.1, we get directly a sequence of desubstituted words and a directive sequence as in Lemma 6.5.

6.3 Characterizing the set of episturmian words as a stable set

Let $\mathcal{E}$ be the set of episturmian words. Proposition 6.8 provides characterizations of $\mathcal{E}$ as a stable set and as a set of adic words. We first need a technical result.

Lemma 6.7. *Any episturmian word, considered as an element of* $\mathrm{Stab}(\mathcal{L} \cup \mathcal{R})$*, has a directive sequence containing infinitely many occurrences of elements of* $\mathcal{L}$*. In other words, any episturmian word belongs to* $\mathrm{Stab}(\mathcal{R}^*\mathcal{L})$*.*

Proof. Let $\mathbf{w}$ in $\mathrm{Stab}(\mathcal{L} \cup \mathcal{R})$ and let $(\mathbf{w}_n)_{n\geq 0}$ and $(\sigma_n)_{n\geq 1}$ as in Lemma 6.5. Observe that, for any infinite word $\mathbf{u}$ and any letters α, β, the word $R_\beta(\mathbf{u})$ begins with α if and only if $\mathbf{u}$ begins with α.

Assume that $(\sigma_n)_{n\geq \ell+1} \in \mathcal{R}^\omega$ for some integer ℓ. Let α be the first letter of $\mathbf{w}_\ell$. From the previous observation, it follows that α is the first letter of $\mathbf{w}_k$ for all $k \geq \ell$. The hypothesis "if $\mathbf{w}_k$ begins with the letter α, then $\sigma_{k+1} \neq R_\alpha$", implies that $\sigma_k \neq R_\alpha$ for all $k \geq \ell + 1$. Hence, for any $k \geq \ell$, α occurs only as a prefix of the words $\sigma_\ell \cdots \sigma_k(\alpha)$ whose lengths grow to infinity. As these words $\sigma_\ell \cdots \sigma_k(\alpha)$ are prefixes of $\mathbf{w}_\ell$, α has only one occurrence in $\mathbf{w}_\ell$ and the word $\mathbf{w}_\ell$ is not recurrent. Using Lemma 6.6 and the previously recalled hypothesis, by inverse induction, we can show that $\mathbf{w}_i$ is also not recurrent for every i, $\ell \geq i \geq 0$. Hence, by Proposition 6.4, $\mathbf{w}_0$ is not episturmian. Thus, for all $\ell \geq 1$, there exists $k \geq \ell$ such that $\sigma_k \in \mathcal{L}$, that is, $\mathbf{w} \in \mathrm{Stab}(\mathcal{R}^*\mathcal{L})$. $\square$

Proposition 6.8. *Consider an alphabet containing at least two letters. The set $\mathcal{E}$ of episturmian words is the set* $\mathrm{Stab}(\mathcal{R}^*\mathcal{L})$ *and it is also the set of all* $\mathcal{R}^*\mathcal{L}$*-adic words.*

Proof. From Theorem 6.3 and Lemma 6.7, $\mathcal{E} \subseteq \mathrm{Stab}(\mathcal{R}^*\mathcal{L})$.

Assume now that $\mathbf{w} \in \mathrm{Stab}(\mathcal{R}^*\mathcal{L})$. Let $(\sigma_n)_{n\geq 1}$ be a directive sequence of $\mathbf{w}$ ($\sigma_n \in \mathcal{R}^*\mathcal{L}$ for all $n \geq 1$). If there exists a letter a such that $\sigma_n \in R_a^*L_a$ for arbitrarily large n, then, $\mathbf{w} = f(a^\omega)$ for some f in $(\mathcal{R}^*\mathcal{L})^*$. As there exists a letter $b \neq a$, $a^\omega = \lim_{n\to\infty} L_a^n(b)$: $\mathbf{w}$ is $\mathcal{R}^*\mathcal{L}$-adic. Otherwise, let $(\mathbf{w}_n)_{n\geq 0}$ be a sequence of desubstituted words of $\mathbf{w}$ using $(\sigma_n)_{n\geq 1}$. For any integer $n \geq 0$, there exists m, $m \geq n$, such that $|\sigma_{n+1} \cdots \sigma_m(\mathrm{first}(\mathbf{w}_m))| \geq 2$. Hence $\mathbf{w}$ is $(\mathcal{R}^*\mathcal{L})$-adic.

Assume now that $\mathbf{w}$ is $\mathcal{R}^*\mathcal{L}$-adic. A first case is that $\mathbf{w} = f(a^\omega)$ for some letter a and some f in $(\mathcal{L} \cup \mathcal{R})^*$. In this case $\mathbf{w}$ is recurrent as any periodic word is. Moreover it belongs to $\mathrm{Stab}(\mathcal{L} \cup \mathcal{R})$ with fL_a^ω as a directive sequence. By Proposition 6.4, $\mathbf{w}$ is episturmian.

Assume now that the first case does not hold (but still $\mathbf{w}$ is $\mathcal{R}^*\mathcal{L}$-adic). By Proposition 4.1, $\mathbf{w} \in \mathrm{Stab}(\mathcal{R}^*\mathcal{L})$. Let $\mathbf{s} = (\sigma_n)_{n\geq 1}$ be the directive sequence of $\mathbf{w}$ as an element of this stable set (for all $n \geq 0$, σ_n belongs to $\mathrm{Stab}(\mathcal{R}^*\mathcal{L})$) and let $(\mathbf{w}_n)_{n\geq 0}$ be the sequence of desubstituted words. Observe that there cannot exist a letter α and an integer N such that, for all $n \geq N$, $\sigma_n \in (R_\alpha^*L_\alpha)$. Indeed, for any infinite word $\mathbf{u}$ and any element f of $(R_\alpha^*L_\alpha)^k$, $f(\mathbf{u})$ begins with α^k. Hence the existence of α and N would implies that $\mathbf{w}_N = \alpha^\omega$ and that we are back in the first case. From the previous observation, we deduce that, for all $n \geq 0$, there exist an integer $m \geq 1$ and a letter $b \neq \mathrm{first}(\mathbf{w}_n)$ such that

$|\sigma_{n+1}\cdots\sigma_m(\text{first}(\mathbf{w}_m))|_b \neq 0$. This implies that, for all $n \geq 0$, $\text{first}(\mathbf{w}_n)$ has at least two occurrences in $\mathbf{w}_n$. Observe that any factor of $\mathbf{w}$ is a factor of a prefix of $\mathbf{w}$ and so a factor of $\sigma_1 \cdots \sigma_n(\text{first}(\mathbf{w}_n))$ for some integer n. Hence any factor of $\mathbf{w}$ occurs at least twice in $\mathbf{w}$: $\mathbf{w}$ is recurrent. By Proposition 6.4, $\mathbf{w}$ is episturmian. $\qquad\square$

As $\mathcal{R}^*\mathcal{L}$ is infinite, the next proposition answers a natural question.

Proposition 6.9. *Assume that the alphabet A contains at least two letters. There is no finite set $\mathcal{S} \subseteq \text{Subst}(A)$ such that the set $\mathcal{E}$ of episturmian words is $\text{Stab}(\mathcal{S})$.*

For proving this result, we need a characterization of the set of endomorphisms preserving episturmian words.

Let $\mathcal{P}$ be the set of all endomorphisms f for which there exists a letter a such that, for all letter b in A, $f(b) \in a^+$. For any infinite word $\mathbf{w}$ and $f \in \mathcal{P}$, $f(\mathbf{w})$ is periodic: it is the word a^ω for some letter a. Finally, let $\mathcal{S}_{\text{epi}} = (\mathcal{L} \cup \mathcal{R} \cup \text{Exch} \cup \mathcal{P})^*$.

Proposition 6.10. *The set of endomorphisms preserving $\mathcal{E}$ is $\mathcal{S}_{\text{epi}}$.*

Proof. Let us first verify that any element of $\mathcal{S}_{\text{epi}}$ preserves $\mathcal{E}$. It is sufficient to consider elements of $\mathcal{L} \cup \mathcal{R} \cup \text{Exch} \cup \mathcal{P}$. As for any letter α the word α^ω is episturmian, this holds for elements of $\mathcal{P}$. Changing the alphabet of a word does not change the property of being episturmian. Hence this holds also for elements of Exch. Finally, as $\mathcal{E} = \text{Stab}(\mathcal{R}^*\mathcal{L})$ by Proposition 6.8, elements of $(\mathcal{L} \cup \mathcal{R})$ also preserve $\mathcal{E}$.

To prove that any endomorphisms preserving $\mathcal{E}$ is an element of $\mathcal{S}_{\text{epi}}$, we need the next two lemmas.

Lemma 6.11. *Let f, g in $\text{Subst}(A)$ and $\alpha \in A$ such that $f = L_\alpha g$ or $f = R_\alpha g$. We have: f preserves $\mathcal{E}$ if and only if g preserves $\mathcal{E}$.*

Proof. We have just seen that elements of $\mathcal{L} \cup \mathcal{R}$ preserve $\mathcal{E}$. Hence if g preserves $\mathcal{E}$, f also preserves $\mathcal{E}$.

From now on, assume that f preserves $\mathcal{E}$ but not g. This means that there exists $\mathbf{w}$ in $\mathcal{E}$ such that $g(\mathbf{w}) \notin \mathcal{E}$ and $f(\mathbf{w}) \in \mathcal{E}$. Observe that $g(\mathbf{w})$ is recurrent as $\mathbf{w}$ is recurrent. By definition of episturmian words, $g(\mathbf{w})$ has two left special factors of the same length or its set of factors is not closed by reversal.

First assume the existence of two different left special factors. Considering them of minimal length we may assume that these words are ua_1 and ua_2 for a word u and different letters a_1 and a_2. There exist letters b_1, b_2, c_1, c_2 with $b_1 \neq b_2$, $c_1 \neq c_2$ such that b_1ua_1, b_2ua_1, c_1ua_2, c_2ua_2 are factors of $g(\mathbf{w})$. As $g(\mathbf{w})$ is recurrent, we can consider non-prefix occurrences of the previous words.

If $f = L_\alpha g$, then the words $b_1L_\alpha(u)\alpha a_1$, $b_2L_\alpha(u)\alpha a_1$, $c_1L_\alpha(u)\alpha a_2$, $c_2L_\alpha(u)\alpha a_2$ are factors of $L_\alpha(g(\mathbf{w}))$. To verify this assertion it may be observed that (for instance) $b_1L_\alpha(u)\alpha a_1$ is a factor of $L_\alpha(b_1ua_1)\alpha$ which is a factor of $L_\alpha(g(\mathbf{w}))$ (note that moreover $a_1\alpha = \alpha a_1$ when $a_1 = \alpha$).

If $f = R_\alpha g$, then the words $b_1\alpha R_\alpha(u)a_1$, $b_2\alpha R_\alpha(u)a_1$, $c_1\alpha R_\alpha(u)a_2$, $c_2\alpha R_\alpha(u)a_2$ are factors of $R_\alpha(g(u))$. To verify this assertion it may be observed that (for instance) $b_1\alpha R_\alpha(u)a_1$ is a factor of $\alpha R_\alpha(b_1ua_1)$ which is also a factor of $R_\alpha(g(\mathbf{w}))$ as we consider a non-prefix occurrence of b_1ua_1.

In both cases, $f(\mathbf{w})$ has two left special factors of the same length and consequently it is not episturmian. A contradiction.

From now on, assume that $g(\mathbf{w})$ contains a factor u but not its reversal $\tilde{u}$. The word $L_\alpha(g(\mathbf{w}))$ contains the factor $L_\alpha(u)\alpha$ but not its reversal. Indeed if it contains $\widetilde{L_\alpha(u)}\alpha = \alpha\widetilde{L_\alpha(u)} = \alpha R_\alpha(\tilde{u}) = L_\alpha(\tilde{u})\alpha$, then the word $g(\mathbf{w})$ contains the factor $\tilde{u}$: a contradiction. Similarly the word $R_\alpha(g(\mathbf{w}))$ contains the factor $\alpha R_\alpha(u)$ (remember that $g(\mathbf{w})$ is recurrent and so there exists a non-prefix occurrence of u) but not its reversal $\alpha R_\alpha(\tilde{u})$. As $f(\mathbf{w}) = L_\alpha(g(\mathbf{w}))$ or $f(\mathbf{w}) = R_\alpha(g(\mathbf{w}))$, $f(\mathbf{w})$ is not episturmian. This contradicts the facts that $\mathbf{w}$ is episturmian and that f preserves $\mathcal{E}$. Hence g preserves $\mathcal{E}$. $\qquad\square$

Lemma 6.12. *Let $f \in \text{Subst}(A) \setminus \text{Exch}^*$ be a morphism that preserves $\mathcal{E}$. All images of letters by f begin with the same letter, or, all images of letters end with the same letter.*

Proof. Let last(w) be the last letter of a non-empty finite word w.

Let $f \in \text{Subst}(A) \backslash \text{Exch}^*$ be a morphism that preserves $\mathcal{E}$. We start with the following observation: For any episturmian word $\mathbf{w}$, there exists a letter α such that any factor of $\mathbf{w}$ of length 2 contains α. This follows directly from Theorem 6.3 that states, that for any episturmian word $\mathbf{w}$, there exists a letter α such that $\mathbf{w} = L_\alpha(\mathbf{w}')$ or $\mathbf{w} = R_\alpha(\mathbf{w}')$ for some episturmian word $\mathbf{w}'$.

Case 1. Assume first that there exist letters α and a such that $\alpha\alpha$ is a factor of $f(a)$. Then, by the previous observation, for any episturmian word $\mathbf{w}$ containing a, $f(\mathbf{w})$ cannot contain a factor $\beta\gamma$ with β, γ two letters with $\beta \neq \alpha$ and $\gamma \neq \alpha$. This implies that there cannot exist letters b and c such that $\text{first}(f(c)) \neq \alpha$ and $\text{last}(f(b)) \neq \alpha$. Indeed there exist episturmian words containing both a and bc: their images by f would contain both $\alpha\alpha$ and the word $\text{last}(f(b))\,\text{first}(f(c))$ which is impossible. Thus all images of letters begin with α or all images of letters end with α.

Case 2. Assume now that the previous case does not hold but that there exist pairwise distinct letters α, such that $\alpha\beta$ or $\beta\alpha$ is a factor of an element of $f(A)$, and, $\alpha\gamma$ or $\gamma\alpha$ is also a factor of an element of $f(A)$.

Assume first that there exist an episturmian word $\mathbf{w}$ containing all letters, and, two letters δ_1 and δ_2 different from α such that $\delta_1\delta_2$ is a factor of $f(\mathbf{w})$. Then by initial observation, as the previous case does not hold, $\delta_1\delta_2 \in \{\beta\gamma, \gamma\beta\}$. It follows that no other letter than α, β and γ occurs in words of $f(A)$. Indeed otherwise, $f(\mathbf{w})$ would contain two factors written on different letters: a contradiction with the initial observation. Assume now by contradiction that there exist images of letters that do not begin with the same letter and images of letters that do not end with the same letter. Then there exist a letter $x \in \{\alpha, \beta, \gamma\}$ such that $x = \text{first}(f(b))$ and $x = \text{last}(f(a))$ for some letters a and b. Let $\mathbf{w}$ be an episturmian word containing all letters and the factor ab. The episturmian word $f(\mathbf{w})$ contains xx and a factor of length 2 that does not contains x. This contradicts the initial observation.

Hence, for any $\mathbf{w} \in \mathcal{E}$, $f(\mathbf{w})$ cannot contain a factor $\delta_1\delta_2$ with δ_1, δ_2 two letters different from α. We end as in the Case 1.

Case 3. Assume now that the previous cases do not hold but that there exists a letter b such that $|f(b)| \geq 2$.

Note that, as the two previous cases do not hold, there must exist distinct letters α and β such that factors of length 2 of elements of $f(A)$ are necessarily $\alpha\beta$ and $\beta\alpha$. This implies that the first letter of $f(b)$ is α or β. Without loss of generality, assume $f(b)$ begins with α. Hence, for some $k \geq 1$, $f(b) = (\alpha\beta)^k$ of $f(b) = (\alpha\beta)^k\alpha$.

Case $f(b) = (\alpha\beta)^k$. There cannot exist letters a, c, γ, δ such that $\gamma \neq \beta$, $\alpha \neq \delta$, $\gamma = \text{last}(f(a))$, $\delta = \text{first}(f(c))$. Indeed otherwise $f(abc)$ would contain the factor $\gamma(\alpha\beta)^k\delta$. Since β is not a factor of $\gamma\alpha$ and α is not a factor of $\beta\delta$, for any episturmian word $\mathbf{w}$ containing abc, the word $f(\mathbf{w})$ would not be an episturmian word by the initial observation. Hence all images of letters begin with α, or, all images of letters end with β.

Case $f(b) = (\alpha\beta)^k\alpha$. Assume that there exists a letter a such that $\text{last}(f(a)) = \gamma$ with $\gamma \neq \alpha$. As all length 2 factors of images of letters are $\alpha\beta$ and $\beta\alpha$, if $\gamma \neq \beta$, then $f(a) = \gamma$. In this case, for any episturmian word containing aab, $f(\mathbf{w})$ contains both $\gamma\gamma$ and $\alpha\beta$ contradicting the fact that it is episturmian. So $\gamma = \beta$. Moreover $f(a) = (\alpha\beta)^\ell$ or $f(a) = (\beta\alpha)^\ell\beta$ for some $\ell \geq 0$. The case $f(a) = (\alpha\beta)^\ell$ has already been considered. Hence assume $f(a) = (\beta\alpha)^\ell\beta$. Let $\mathbf{w}$ be an episturmian word containing the factor $bbbabb$. The word $f(\mathbf{w})$ contains the factor

$$\alpha(\alpha\beta)^k\alpha(\alpha\beta)^k\alpha(\beta\alpha)^\ell\beta(\alpha\beta)^k\alpha\alpha$$

which is equals to

$$\alpha(\alpha\beta)^k\alpha(\alpha\beta)^{k+\ell+1+k}\alpha\alpha$$

As $f(\mathbf{w}) \in \text{Stab}(\mathcal{L} \cup \mathcal{R})$ by Theorem 6.3, $f(\mathbf{w}) = L_\alpha(\mathbf{w}')$ or $f(\mathbf{w}) = R_\alpha(\mathbf{w}')$ for some episturmian word $\mathbf{w}'$. Thus $\mathbf{w}'$ should contain the factor $\alpha\beta^k\alpha\beta^{k+\ell+1+k}\alpha$. As $k + \ell + 1 + k \geq k + 2$, this implies that both β^{k+1} and $\beta^k\alpha$ are left special factors of $\mathbf{w}'$: this contradicts the fact that $\mathbf{w}'$ is episturmian. Thus for each letter a, $\text{last}(f(a)) = \alpha$.

Case 4. It remains to consider the case where all images of letters are of length 1. As $f \notin \text{Exch}^*$, if all images are not the same, then there exist pairwise different letters a, b and c such that $f(a) = f(b)$ and $f(a) \neq f(c)$. Set $x = f(a)$, $y = f(c)$ (x and y are letters). Let $\mathbf{w}$ be an episturmian word containing

the factor acc. The word $L_aL_bL_c(\mathbf{w})$ contains the factor $cabaabacabac$ and the word $fL_aL_bL_c(\mathbf{w})$ contains the factor yx^6yx^3y. Hence x^3y and x^4 are both left special factors of $fL_aL_bL_c(\mathbf{w})$. This is not possible for an episturmian word: we have a contradiction with the fact that f preserves $\mathcal{E}$. Thus all images of letters are equal. $\qquad\square$

Let us continue the proof of Proposition 6.10.

From now on, let f be an endomorphism preserving episturmian words. Let $||f|| = \sum_{\alpha \in A} |f(\alpha)|$. We prove by induction on $||f||$ that $f \in \mathcal{S}_{\text{epi}}$.

Assume first that $||f|| = \#A$. If $f(A) = A$, then $f \in \text{Exch}^* \subseteq \mathcal{S}_{epi}$. Otherwise by Lemma 6.12, $f \in \mathcal{P} \subseteq \mathcal{S}_{epi}$.

From now on, assume that $||f|| > \#A$ and assume $f \notin \mathcal{P}$. By Lemma 6.12, all images of letters begin with the same letter, or, all images of letters end with the same letter. Assume that the first case holds (the second case is symmetric) and let α be the first letter of images of letters. Let $\mathbf{w} \in \mathcal{E}$. By hypothesis $f(\mathbf{w}) \in \mathcal{E}$. As $\mathcal{E} \subseteq \text{Stab}(\mathcal{L} \cup \mathcal{R})$ by Theorem 6.3, $f(\mathbf{w}) = L_\beta(\mathbf{w}')$ or $f(\mathbf{w}) = R_\beta(\mathbf{w}')$ for some letter β and some word $\mathbf{w}'$.

When $f(\mathbf{w}) = L_\beta(\mathbf{w}')$, we have $\alpha = \beta$ and we can find a morphism g such that $f = L_\alpha g$. By Lemma 6.11, g preserves $\mathcal{E}$. As $f \notin \mathcal{P}$, $||g|| < ||f||$ and by induction $g \in \mathcal{S}_{\text{epi}}$. So $f \in \mathcal{S}_{\text{epi}}$.

Assume now that $f(\mathbf{w}) = R_\beta(\mathbf{w}')$. If all images of letters by f end with β, then $f = R_\beta g$ and, as in the case $f = L_\alpha g$ above, $f \in \mathcal{S}_{\text{epi}}$. If some image of a letter by f does not end with β, as the image of letters is followed by α in $f(\mathbf{w})$ and by β in $R_\beta(\mathbf{w}')$, we get $\alpha = \beta$. Hence $f = L_\alpha h$ for some morphism h, and as previously $f \in \mathcal{S}_{\text{epi}}$. $\qquad\square$

Proof of Proposition 6.9. The proof is similar to the proof of the second part of Proposition 6.2 but we have to take care of elements in $\mathcal{R}$. Assume by contradiction that there exists a finite set $\mathcal{S} \subseteq \text{Subst}(A)$ such that the set of episturmian words is $\text{Stab}(\mathcal{S})$. By Remark 2.3, elements of $\mathcal{S}$ preserve $\mathcal{E}$ and so, by Proposition 6.10, $\mathcal{S} \subseteq \mathcal{S}_{\text{epi}} = (\mathcal{L} \cup \mathcal{R} \cup \text{Exch} \cup \mathcal{P})^*$.

Observe that $\mathcal{S} \cap \text{Exch}^* = \emptyset$. Indeed if $f \in \mathcal{S} \cap \text{Exch}^*$, f acts on the alphabet as a permutation and there exists an integer n such that f^n is the identity. Then any infinite word belongs to $\text{Stab}(\mathcal{S})$: a contradiction.

From now on, we consider only aperiodic episturmian words. These words can be desubstituted only on elements of $\mathcal{S} \cap (\mathcal{L} \cup \mathcal{R} \cup \text{Exch})^*$. From Relations (1) to (4), any element of $(\mathcal{L} \cup \mathcal{R} \cup \text{Exch})^*$ can be decomposed into σe with $\sigma \in (\mathcal{L} \cup \mathcal{R})^*$ and $e \in \text{Exch}^*$, that is, it can be viewed as an element of $(\mathcal{L} \cup \mathcal{R})^* \text{Exch}^*$.

Let a and b be two different letters. The following fact is important.

Fact. Let i, j be integers such that $i \leq j$. Let $\mathbf{w} \in A^\omega$ and $f \in (\mathcal{L} \cup \mathcal{R})^i$. If $f(\mathbf{w})$ begins with $ab^{j+2}a$, then $f = R_b^i$.

We prove this fact by induction on i. It is basically true for $i = 0$. Assume $i \geq 1$. As $f(\mathbf{w})$ begins with abb and $f \in (\mathcal{L} \cup \mathcal{R})^i$, necessarily $f = R_b g$ for some $g \in (\mathcal{L} \cup \mathcal{R})^{i-1}$. Moreover $g(\mathbf{w})$ has a prefix $ab^{(j-1)+2}a$ and $i - 1 \leq j - 1$. Hence $g = R_b^{i-1}$ by induction. So $f = R_b^i$.

Now let $M_1 = \max(\{|f| \mid f\pi \in \mathcal{S}, f \in (\mathcal{L} \cup \mathcal{R})^*, \pi \in \text{Exch}^*\}$. (here, for $f \in (\mathcal{L} \cup \mathcal{R})^*$, $|f|$ is the length of f considered as a word over the alphabet $\mathcal{L} \cup \mathcal{R}$). Let $M = (\#A! + 1)M_1$. Let $\mathbf{w}$ be an episturmian word having $ab^{M+2}a$ as a prefix. Let $(\sigma_n)_{n \geq 1} \in \mathcal{S}^\omega$ be a directive word of $\mathbf{w}$ (as an element of $\text{Stab}(\mathcal{S})$).

Let i, $1 \leq i \leq \#A! + 1$. For $1 \leq j \leq i$, as $\sigma_j \in (\mathcal{L} \cup \mathcal{R})^* \text{Exch}^*$, there exist $g_j \in (\mathcal{L} \cup \mathcal{R})^*$ and $\pi'_j \in \text{Exch}^*$ such that $\sigma_j = g_j \pi'_j$. Using Equations (2) and (4), we can see that there exist $f_i \in (\mathcal{L} \cup \mathcal{R})^*$ and $\pi_i \in \text{Exch}^*$ such that $\sigma_1 \cdots \sigma_i = f_i \pi_i$. Moreover $|f_i| = \sum_{j=1}^i |g_j|$ and so $|f_i| \leq iM_1 \leq M$. From the previous fact, $f_i \in \{R_b\}^*$. Since the cardinality of Exch^* is $\#A!$, there exists i and j, $1 \leq i < j \leq \#A! + 1$, such that $\pi_i = \pi_j$. Set $\pi = \pi_i$.

Let ℓ_i and ℓ_j be the integers such that $\sigma_1 \ldots \sigma_i = R_b^{\ell_i} \pi$ and $\sigma_1 \ldots \sigma_j = R_b^{\ell_j} \pi$. We have $R_b^{\ell_j} \pi = \sigma_1 \ldots \sigma_j = R_b^{\ell_i} \pi \sigma_{i+1} \cdots \sigma_j$. Thus $\pi \sigma_{i+1} \cdots \sigma_j = R_b^{\ell_j - \ell_i} \pi$. By Equations (3) and (4), there exists a letter c such that $R_b^{\ell_j - \ell_i} \pi = \pi R_c^{\ell_j - \ell_i}$. So $\sigma_{i+1} \cdots \sigma_j = R_c^{\ell_j - \ell_i}$.

Observe that, for d a letter different from c (the alphabet A contains at least two letters), R_c^ω is a directive sequence of the word dc^ω. Thus $(R_c^{\ell_j - \ell_i})^\omega$ is a directive sequence of the word dc^ω. Hence this word dc^ω belongs to $\mathrm{Stab}(\mathcal{S})$ although it is not episturmian (by Theorem 6.3 since it is not recurrent). We have obtained a contradiction with $\mathrm{Stab}(\mathcal{S}) = \mathcal{E}$. $\qquad\square$

6.4 Strict episturmian words

Theorem 6.3 recalls only a part of Theorem 3.10 in [10]. This latter theorem also implies that an infinite word $\mathbf{w}$ is A-strict episturmian if and only if there exist an infinite sequence of *recurrent* infinite words $(\mathbf{w}_n)_{n \geq 0}$ and a sequence $(\sigma_n)_{n \geq 1}$ in $(\mathcal{L} \cup \mathcal{R})^\omega$ such that, for each letter α in A, L_α or R_α occurs infinitely often in $(\sigma_n)_{n \geq 1}$. Let $\mathcal{S}_{strictepi}$ be the set $(\mathcal{L} \cup \mathcal{R})^* \mathcal{L} (\mathcal{L} \cup \mathcal{R})^* \cap \cap_{\alpha \in A} (\mathcal{L} \cup \mathcal{R})^* \{L_\alpha, R_\alpha\} (\mathcal{L} \cup \mathcal{R})^*$ of all elements of $(\mathcal{L} \cup \mathcal{R})^*$ having a decomposition over $\mathcal{L} \cup \mathcal{R}$ with at least one element of $\mathcal{L}$ and at least one element of $\{L_\alpha, R_\alpha\}$ for each letter α.

From what precedes, using Lemma 6.5, we can see that the set of A-strict episturmian words is included in $\mathrm{Stab}(\mathcal{S}_{strictepi})$. Conversely, acting as in the proof of Proposition 6.8, we can deduce that any element of $\mathrm{Stab}(\mathcal{S}_{strictepi})$ is recurrent. The next result follows.

Proposition 6.13. *The set of A-strict episturmian words is $\mathrm{Stab}(\mathcal{S}_{strictepi})$.*

In [10, Th. 3.13], it is proved that $(\mathcal{L} \cup \mathcal{R} \cup Exch)^*$ is the set of endomorphisms of A^* that preserve A-strict episturmian words. Using this result, as done for the proof of Proposition 6.9, we can prove:

Proposition 6.14. *There is no finite set $\mathcal{S} \subseteq \mathrm{Subst}(A)$ such that $\mathrm{Stab}(\mathcal{S})$ is the set of A-strict episturmian words.*

7 Conclusion

Stable sets formalize the concept of infinite desubstitutions using a set of nonerasing endomorphisms. We have shown that several known sets of words are stable sets: the set of binary balanced words, the set of Sturmian words, the set of Lyndon Sturmian words, the set of standard episturmian words (which corresponds, in the binary case, to the set of LSP words), the set of strict standard episturmian words (which corresponds, in the binary case, to the set of standard words), the set of episturmian words and the set of strict standard episturmian words. Among all these sets, only the set of binary balanced words and the set of standard episturmian words are stable sets of a finite set of substitutions. A first natural question is whether there exist other sets defined by combinatorial properties that are stable sets of a (finite) set of substitutions.

A characterization of a set of words as the stable set of an infinite set of substitutions may be more difficult to understand and to use than a characterization as a subset of a stable set of a finite set of substitutions using conditions on directive sequences. For instance, it is probably more interesting to know that standard Sturmian words are the elements of $\mathrm{Stab}(\mathcal{L} \cup \mathcal{R})$ whose directive sequences contain infinitely many occurrences of L_a and infinitely many occurrences of L_b than to know that they are the elements of $\mathrm{Stab}(\mathcal{S}_{stand})$, even if this latter formulation states the same result in a more compact form. Similarly, it may be more interesting to know that episturmian words are the recurrent elements of $\mathrm{Stab}(\mathcal{L} \cup \mathcal{R})$ than to know that they are the elements of $\mathrm{Stab}(\mathcal{R}^* \mathcal{L})$. More generally one can search for a characterization of a set of words as a subset of a stable set whose directive sequence of elements verify a particular condition. Such a result was obtained for the set of LSP words [21]. Moreover this approach is often done w.r.t. the concept of S-adicity instead of stable sets (see, *e.g.* the case of Sturmian words or the paper [2] and its references).

Some extensions of the notion of stable set could also be studied. For instance it should be quite natural to search for sets of words that are images by a morphism or by a set of morphisms (not necessarily endomorphisms) of a stable set. But except the morphic words, the author knows no example among classical sets of words.

Another direction of study could be to have a better formalization of the possible changes of alphabets. Indeed, remember that in the definitions of S-adicity, the considered morphisms are not necessarily endomorphisms. Also in the examples of stable sets one can observe that some elements

of stable sets have desubstituted words written on alphabets whose cardinalities may decrease (see for instance the word $L_c L_b(a^\omega)$).

Acknowledgements

References

[1] P. Arnoux, M. Mizutani, and T. Sellami. Random product of substitutions with the same incidence matrix. *Theor. Comput. Sci.*, 543:68–78, 2014.

[2] V. Berthé and V. Delecroix. Beyond substitutive dynamical systems: S-adic expansions. In S. Akiyama, editor, *Numeration and Substitution 2012*, volume B46 of *RIMS Kôkyûroku Bessatsu*, pages 81–123, 2014.

[3] V. Berthé, C. Holton, and L. Q. Zamboni. Initial powers of Sturmian sequences. *Acta Arith.*, 122:315–347, 2006.

[4] V. Berthé and M. Rigo, editors. *Combinatorics, Automata and Number Theory*, volume 135 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2010.

[5] X. Droubay, J. Justin, and G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255:539–553, 2001.

[6] S. Ferenczi. Rank and symbolic complexity. *Ergodic Theory Dynam. Systems*, 16:663–682, 1996.

[7] G. Fici. Special factors and the combinatorics of suffix and factor automata. *Theoret. Comput. Sci.*, 412:3604–3615, 2011.

[8] Amy Glen, Florence Levé, and Gwénaël Richomme. Directive words of episturmian words: equivalences and normalization. *Theor. Inform. Appl.*, 43(2):299–319, 2009.

[9] E. Godelle. The stable set of a self-map. *Adv. in Appl. Math.*, 45:438–448, 2010.

[10] J. Justin and G. Pirillo. Episturmian words and episturmian morphisms. *Theoret. Comput. Sci.*, 276(1-2):281–313, 2002.

[11] J. Leroy. Contribution à la résolution de la conjecture S-adique. Doctoral Thesis, Université de Picardie Jules Verne, 2012.

[12] J. Leroy. Some improvements of the S-adic conjecture. *Adv. App. Math.*, 48:79–98, 2012.

[13] J. Leroy. An S-adic characterization of minimal subshifts with first difference of complexity $p(n+1)-p(n) \leq 2$. *Discrete Math. Theor. Comput. Sci.*, 16(1)(1):233–286, 2014.

[14] J. Leroy and G. Richomme. A combinatorial proof of S-adicity for sequences with linear complexity. *Integers*, 13(Article #A5), 2013.

[15] F. Levé and G. Richomme. Quasiperiodic Sturmian words and morphisms. *Theor. Comput. Sci.*, 372(1):15–25, 2007.

[16] M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, 1983. Reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, UK, 1997.

[17] M. Lothaire. *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2002.

[18] M. Morse and G. A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.*, 62:1–42, 1940.

[19] N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer, 2002. (V. Berthé, S. Ferenczi, C. Mauduit, A. Siegel, editors).

[20] G. Richomme. Lyndon morphisms. *Bulletin of the Belgian Mathematical Society*, 10(5):761–786, 2003.

[21] G. Richomme. Characterization of infinite LSP words and endomorphisms preserving the LSP property. *Internat. J. Found. Comput. Sci.*, 30(1):171–196, 2019.

[22] A. Thue. Uber die gegenseitige Lage gleigher Teile gewisser Zeichenreihen. *Kristiania Videnskapsselskapets Skrifter Klasse I. Mat.-naturv*, 1:1–67, 1912.