Characterization of infinite LSP words and endomorphisms preserving the LSP property

Gwenaël Richomme

Characterization of infinite LSP words
and endomorphisms preserving the LSP property

Gwenaël Richomme

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

August 8, 2018

Abstract

Answering a question of G. Fici, we give an S-adic characterization of the family of infinite LSP words, that is, the family of infinite words having all their left special factors as prefixes. More precisely we provide a finite set of morphisms S and an automaton $\mathcal{A}$ such that an infinite word is LSP if and only if it is S-adic and one of its directive words is recognizable by $\mathcal{A}$. Then we characterize the endomorphisms that preserve the property of being LSP for infinite words. This allows us to prove that there exists no set S' of endomorphisms for which the set of infinite LSP words corresponds to the set of S'-adic words. This implies that an automaton is required no matter which set of morphisms is used.

Keywords: generalizations of Sturmian words, morphisms, S-adicity.

1   Introduction

Free monoid morphisms, also sometimes called substitutions, are basic tools to study finite and infinite words. They are used in various fields as Combinatorics on Words, Formal Languages or Dynamical Systems (see, *e.g.*, [6, 15, 16, 17, 20]). Literature contains many examples of interesting infinite words that are fixed points of endomorphisms. Their success and interest are mainly due to the simplicity of their definitions and the ability given by endomorphisms to prove their properties. Let $\mathbf{w}$ be such a fixed point of morphism and let f be a morphism such that $\mathbf{w} = f(\mathbf{w})$. This word $\mathbf{w}$ can be seen as the limit $\lim_{n\to\infty} f^n(a)$ where a is the first letter of $\mathbf{w}$ (with some extra needed conditions on f to ensure that this limit exists and is an infinite word). This word $\mathbf{w}$ can also be considered as a word which can be recursively desubstituted by the morphism f: here *desubstituted* means the existence of a word $\mathbf{w}_1$ such that $\mathbf{w} = f(\mathbf{w}_1)$ and *recursively desubstituted* means that $\mathbf{w}_1$ can be itself desubstituted by f and so on).

Paraphrasing [3], if one wants to go beyond the morphic case, and thus, get more flexibility in the hierarchical structure, one might want to change the morphism at each desubstitution step (the point of view in [3] is somewhat dual). Thus considering a set S of morphisms instead of a single morphism f, we may consider the S-adic words, that is, words that can be recursively desubstituted over S. A more precise definition is given in Section 3. For more information on S-adic words, readers can consult, *e.g.*, papers [2, 3], Chapter 12 in [17] and their references. Note that we will sometimes write S-adic for substitutive-adic word without direct reference to a set S of morphisms.

S-adicity arises naturally in various studies as, for instance, those of Sturmian words [16, Chap.2][17, Chap. 5] or of Arnoux-Rauzy words [1]. In [9] (see also [12, 13]) Ferenczi proved that any word with factor complexity bounded by some affine function is S-adic for some finite set S of morphisms that depends on the bound of the first difference of the factor complexity (remember that, for any word with factor complexity bounded by some affine function, the first difference of the factor complexity is bounded by a constant [7]). For each of the previous examples, the S-adic properties are not characteristic of the

considered families. Episturmian words that generalize Sturmian and Arnoux-Rauzy words are the S_E-adic words where S_E is the finitely generated set of morphisms that preserve episturmian words [8, 11, 18]. Over the binary alphabet, it is an exercise to show that the set of balanced words (Sturmian words and ultimately periodic balanced words) are the S_S-adic words where S_S is the finitely generated set of morphisms that preserve Sturmian words. As far as the author knows it, the kind of characterizations of the two previous examples is very rare.

S-adic characterizations of families of words is generally twofold. In addition to the induced set S of morphisms, a characterization of allowed infinite sequences of desubstitutions is provided. For instance there exist two disjoint sets S_a and S_b of morphisms such that a word is Sturmian if and only it can be recursively decomposed over $S_a \cup S_b$ using infinitely often elements from S_a and infinitely often elements from S_b [4]. This condition can be described using infinite paths with prohibited segments in an automaton or a graph. These kinds of conditions are also used, for instance, in the characterization of words for which the first difference of factor complexity is bounded by 2 [12, 13] or in the characterization of sequences arising from the study of the Arnoux-Rauzy-Poincaré multidimensional continued fraction algorithm [5]. For the last example, all infinite paths in the graphs are allowed.

Extending an initial work by M. Sciortino and L.Q. Zamboni [21], G. Fici investigated relations between the structure of the suffix automaton built from a finite word w and the combinatorics of this word [10]. He proved that the words having their associated automaton with a minimal number of states (with respect to the length of w) are the words having all their left special factors as prefixes. G. Fici asked in the conclusion of his paper for a characterization of the set of words having the previous property, that he called the LSP property, both in the finite and the infinite case. In this paper, we provide an S_{bLSP}-adic characterization of LSP infinite words (for a suitable S_{bLSP} set of morphisms) using an automaton recognizing allowed infinite desubstitutions over S_{bLSP}. We prove that there exists no set of morphisms S such that the family of LSP words is the family of S-adic words.

Our S-adic characterization is a refinement of the one presented at Conference DLT 2017 [19]. Main ideas of Sections 2 to 7 were already presented and used in [19]. But the set of morphisms considered here allows to provide a smaller automaton (in particular it can be drawn for a three-letter alphabet while this was not possible in [19]) even if this set is larger as it is the set of morphisms considered in [19] plus their restrictions to smaller alphabets. The proof that one cannot have a characterization without restrictions on allowed desubstitutions (Section 8) is new.

The paper is organized as follows. After introducing in Section 2 our basis of morphisms $S_{\mathrm{R-bLSP}}$, in Section 3, we show that all infinite LSP words are $S_{\mathrm{R-bLSP}}$-adic. Section 4 introduces a property of infinite LSP words and a property of morphisms in S_{bLSP} that together allow to explain why the LSP property is lost when applying an LSP morphism to an infinite LSP word. Section 5 allows to trace the origin of the previous property of infinite LSP words. Based on this information, Section 6 defines our automaton and Section 7 proves our characterization of infinite LSP words. In Section 8 we characterize endomorphisms preserving LSP words and deduce that the set of LSP words cannot be characterized as a set of S-adic words, whatever S is.

2 Some basic morphisms

We assume that readers are familiar with combinatorics on words; for omitted definitions (as for instance, factor, prefix, ...) see, *e.g.*, [6, 15, 16]. Given an alphabet A, A^* is the set of all finite words over A, including the empty word ε, and A^ω is the set of all infinite words over A. For a non-empty word u, let $\mathrm{first}(u)$ denote its first letter, $\mathrm{last}(u)$ its last letter and $\mathrm{alph}(u)$ its set of letters. Notations first and alph are similarly defined for infinite words.

A finite word u is a *left special factor* of a finite or infinite word w if there exist at least two distinct letters a and b such that both words au and bu occur in w. Following G. Fici [10], a finite or infinite word $\mathbf{w}$ is *LSP* if all its left special factors are prefixes of $\mathbf{w}$. We also say say $\mathbf{w}$ has *the LSP property*. We study these words using morphisms.

Given two alphabets A and B, a *morphism* (*endomorphism* when $A = B$) f is a map from A^* to B^* such that for all words u and v over A, $f(uv) = f(u)f(v)$. Morphisms are entirely defined by images of letters. Morphisms extend naturally to infinite words. We consider only nonerasing morphisms, that is, morphisms f such that $f(x) = \varepsilon$ implies $x = \varepsilon$. When X and Y are two sets of morphisms, let XY denote the set of morphisms $f \circ g$ with $f \in X$, $g \in Y$ such that if f is from A^* to B^* and g is from C^* to D^* then $D \subseteq A$.

We call *basic LSP morphism* on an alphabet A, or *bLSP* in short, any endomorphism f of A^* verifying:

- there exists a letter α such that $f(\alpha) = \alpha$, and

- for all letters $\beta \neq \alpha$, there exists a letter γ such that $f(\beta) = f(\gamma)\beta$

From this definition one can naturally associate the rooted tree (all labels are on vertices and distinct vertices have distinct labels) whose vertices are elements of A, whose root is α and whose (oriented) edges are pairs (β, γ) of letters such that $f(\beta) = f(\gamma)\beta$. Conversely, given a labeled rooted tree $T = (A, E)$, let f_T be the morphism defined by: for all letters β, $f_T(\beta)$ is the word obtained concatenating vertices on the path in T from the root of T to β. This morphism is bLSP.

Let $S_{\mathrm{bLSP}}(A)$ (or shortly S_{bLSP} when A is clear) denote the set of all bLSP morphisms over the alphabet A. We have just seen that there is a bijection between this set and the set of labeled rooted trees with label in A. Thus, denoting by $\#X$ the cardinality of a set X, there are $(\#A)^{\#A-1}$ elements in $S_{\mathrm{bLSP}}(A)$ (see Sequence A000169 in The On-Line Encyclopedia of Integer Sequences: this sequence enumerates rooted trees; its first values are 1, 2, 9, 64, 625, 7776, 117649, 2097152).

When f is a morphism defined on an ordered alphabet $A = \{a_1, \ldots, a_k\}$, we let $[u_1, u_2, \ldots, u_k]$ denote the morphism defined by $a_1 \mapsto u_1$, $a_2 \mapsto u_2$, $\ldots$, $a_k \mapsto u_k$. For instance $[a, ab, ac, acd, ace]$ defines the bLSP morphism f such that $f(a) = a$, $f(b) = ab$, $f(c) = ac$, $f(d) = acd$, $f(e) = ace$. The rooted tree associated with f is given in Figure 1.
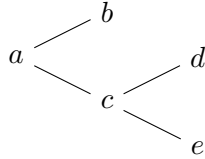

Figure 1: Rooted tree associated with the morphism $[a, ab, ac, acd, ace]$

Example 2.1. Some examples of bLSP morphisms are the standard episturmian morphisms L_α defined, for α a letter, by: $L_\alpha(\alpha) = \alpha$ and $L_\alpha(\beta) = \alpha\beta$ for any letter $\beta \neq \alpha$ (see for instance [11, 18] for example of uses of these morphisms). The associated rooted tree has its root connected to all other vertices. Observe that $S_{\mathrm{bLSP}}(\{a, b\}) = \{L_a, L_b\} = \{[a, ab], [ba, b]\}$. These morphisms are well-known in the context of Sturmian words. They are denoted τ_a and τ_b in [4] from which it can be seen that standard Sturmian words are the non-periodic $\{\tau_a, \tau_b\}$-adic words (see also [14]).

Example 2.2. For any alphabet $A_k = \{a_1, \ldots, a_k\}$, let $\lambda_{a_1 \cdots a_k}$ be the morphism defined by $\lambda_{a_1 \cdots a_k}(a_i) = a_1 \cdots a_i$. The associated rooted tree is a path. For instance $\lambda_{acb} = [a, acb, ac]$. Note that $S_{\mathrm{bLSP}}(\{a, b, c\}) = \{L_a, L_b, L_c, \lambda_{abc}, \lambda_{acb}, \lambda_{bac}, \lambda_{bca}, \lambda_{cab}, \lambda_{cba}\}$. Observe that these morphisms are the mirror morphisms of Arnoux-Rauzy and Poincaré morphisms (here f is a *mirror morphism* of g if $f(a)$ is the mirror image or reversal of $g(a)$ for all letters a) used by V. Berthé and S. Labbé [5].

Let A, B and A' be three alphabets with $A' \subseteq A$ and let f be a morphism from A^* to B^*. The restriction of f to A' is the morphism from A'^* to $\mathrm{alph}(f(A'))^*$ denoted $f_{|A'}$ or $f|A'$ and defined by $f_{|A'}(\alpha) = f(\alpha)$ for each letter α of A'. An R-bLSP morphism is any restriction to a subalphabet of a

morphism in S_{bLSP}. Let $S_{\mathrm{R-bLSP}}(A)$ or simply $S_{\mathrm{R-bLSP}}$ denote the set of R-bLSP morphisms. This means that if f belongs to $S_{\mathrm{R-bLSP}}(A)$, there exists a subalphabet B of A and an element g of S_{bLSP} such that $f = g_{|B}$.

By construction of an R-bLSP morphism f, all images of letters by f begin with the same letter: let $\mathrm{first}(f)$ denote it. Next properties are also direct consequences of the definition of R-bLSP morphisms. They will often be used without explicit mention.

Property 2.3. *Let f be an R-bLSP morphism over the alphabet A.*

1. *for all $\beta \in A$, $\mathrm{last}(f(\beta)) = \beta$;*

2. *$f(A)$ is a suffix code (no word of $f(A)$ is a suffix of another word in $f(A)$);*

3. *f is injective both on the set of finite words and the set of infinite words;*

4. *for all $\beta \in A$, x, $y \in A^*$, if $|x| = |y|$ and if $x\beta$ and $y\beta$ are factors of words in $f(A)$, then $x = y$;*

5. *for all letters β, γ, $|f(\beta)|_\gamma \le 1$.*

3 Substitutive-adicity of infinite LSP words

Let S be a set of morphisms. Usually an infinite word $\mathbf{w}$ is said to be *S-adic* if there exist a sequence $(f_n)_{n\ge1}$ of morphisms in S^ω and a sequence of letters $(a_n)_{n\ge1}$ such that $\lim_{n\to+\infty} |f_1 f_2 \cdots f_n(a_{n+1})| = +\infty$ and $\mathbf{w} = \lim_{n\to+\infty} f_1 f_2 \cdots f_n(a_{n+1})$. The sequence $(f_n)_{n\ge1}$ is called the *directive word* of $\mathbf{w}$. We consider here S-adicity in a rather larger way: a word $\mathbf{w}$ is *S-adic* with directive word $(f_n)_{n\ge1}$ if there exists an infinite sequence of infinite words $(\mathbf{w}_n)_{n\ge1}$ such that $\mathbf{w}_1 = \mathbf{w}$ and $\mathbf{w}_n = f_n(\mathbf{w}_{n+1})$ for all $n \ge 1$. If the former definition is verified, the latter is also verified. This second definition may include degenerated cases as, for instance, the word a^ω that is $\{Id\}$-adic with Id the identity morphism. To understand S-adicity, it may be better to see $\mathbf{w}_{n+1}$ as the inverse image of $\mathbf{w}_n$, rather than seing $\mathbf{w}_n$ as the image of $\mathbf{w}_{n+1}$. As morphisms are sometimes called substitutions, we say that $\mathbf{w}_{n+1}$ is a desubstituted word from $\mathbf{w}_n$ and so saying that $\mathbf{w}$ is S-adic means that $\mathbf{w}$ can be recursively desubstituted (using elements of S). Sometimes f_n can be defined on a larger alphabet than $\mathrm{alph}(\mathbf{w}_{n+1})$. In this case f_n can be replaced with its restriction to $\mathrm{alph}(\mathbf{w}_{n+1})$. We will say that $(f_n)_{n\ge1}$ is a *fitted directive* word when for all $n \ge 1$, f_n is defined on $\mathrm{alph}(\mathbf{w}_{n+1})$.

Proposition 3.1. *Any infinite LSP word is $S_{\mathrm{R-bLSP}}$-adic and so is S_{bLSP}-adic. Moreover any LSP word has exactly one fitted directive word over $S_{\mathrm{R-bLSP}}$.*

The proof of this proposition follows the following general scheme of proof and so is a direct consequence of the next two lemmas.

Given a set S of morphisms, in order to prove that infinite words verifying a property P are S-adic, it suffices to prove that for all infinite words $\mathbf{w}$ verifying P,

1. there exist $f \in S$ and an infinite word $\mathbf{w}'$ such that $\mathbf{w} = f(\mathbf{w}')$, and

2. if $\mathbf{w} = f(\mathbf{w}')$ with $f \in S$, then $\mathbf{w}'$ verifies Property P.

To prove the uniqueness of directive words one has also to prove the uniqueness of f and $\mathbf{w}'$ in the first item above.

Lemma 3.2. *Given any finite or infinite LSP word $\mathbf{w}$, there exist a unique infinite word $\mathbf{w}'$ and a unique morphism f in $S_{\mathrm{R-bLSP}}(\mathrm{alph}(\mathbf{w}'))$ such that $\mathbf{w} = f(\mathbf{w}')$.*

Proof. Let $\mathbf{w}$ be a non-empty finite or infinite LSP word and let α be its first letter. Let X be the set of words over $\mathrm{alph}(\mathbf{w}) \setminus \{\alpha\}$ such that $\mathbf{w}$ can be factorized over $\{\alpha\} \cup X$. Let G be the graph $(\mathrm{alph}(\mathbf{w}), E)$ with E the set of edges (β, γ) such that $\beta\gamma$ is a factor of a word αu with $u \in X$. By LSP Property of $\mathbf{w}$, each letter occurring in a word of X is not left special in $\mathbf{w}$. Hence G is a rooted tree with α as root, that is, for any letter β in $\mathrm{alph}(\mathbf{w})$, there exists a unique path from α to β. Let u_β denote the word obtained by concatenating the letters occurring in the path. Let g be the morphism defined by $g(\beta) = u_\beta$ for all β in $\mathrm{alph}(\mathbf{w}) \setminus \{\alpha\}$, and by $g(\alpha) = \alpha$. By construction, g is bLSP.

As α does not occur in words of X and as $\mathbf{w}$ begins with α, $\mathbf{w}$ has a unique decomposition over $\alpha X \cup \{\alpha\}$. If this decomposition holds over αX, set $Y = \alpha X$. Otherwise set $Y = \alpha X \cup \{\alpha\}$. Observe that $Y \subseteq \{u_\beta \mid \beta \in \mathrm{alph}(\mathbf{w})\}$. Let f be the restriction of g to the set $\{\beta \mid u_\beta \in Y\}$: $f \in S_{\mathrm{R-bLSP}}$. By construction $\mathbf{w} = f(\mathbf{w}')$ for a word $\mathbf{w}'$ and $f \in S_{\mathrm{R-bLSP}}(\mathrm{alph}(\mathbf{w}'))$.

Assume $\mathbf{w} = h(\mathbf{w}'')$ with h an R-bLSP morphism and $\mathbf{w}''$ an infinite word. Let $A' = \mathrm{alph}(\mathbf{w}')$ and $A'' = \mathrm{alph}(\mathbf{w}'')$. By the definition of R-bLSP morphisms, all images of letters by f or h begin with the same letter α (which is also the first letter of $\mathbf{w}$). Still by the definition of LSP morphisms, this letter α occurs only once in each image of letters by f and h. Positions of the letter α in $\mathbf{w}$ determine the beginnings of images of letters by f and h. This implies that $f(A') = h(A'')$. By Property 2.3(1), for letters β and γ, $\mathrm{last}(f(\beta)) = \gamma$ or $\mathrm{last}(h(\beta)) = \gamma$ implies $\beta = \gamma$. Hence $f = h$ and $\mathbf{w}' = \mathbf{w}''$. $\qquad \square$

Lemma 3.3. *For any R-bLSP morphism f and any infinite word $\mathbf{w}$, if $f(\mathbf{w})$ is LSP then $\mathbf{w}$ is LSP.*

Proof. Assume by contradiction that $\mathbf{w}$ is not LSP. This means that $\mathbf{w}$ has (at least) one left special factor that is not one of its prefixes. Considering such a factor of minimal length, there exist a word u and letters a, b, β, γ such that $a \neq b$, $\beta \neq \gamma$, ua is a prefix of $\mathbf{w}$, βub and γub are factors of $\mathbf{w}$. As f is an R-bLSP morphism, there exists a unique letter α which is the first letter of all non-empty images by f. The word $f(u)f(a)\alpha$ is a prefix of $f(\mathbf{w})$. Moreover by Property 2.3(1), the words $\beta f(u)f(b)\alpha$ and $\gamma f(u)f(b)\alpha$ are factors of $\mathbf{w}$ (here the fact that $\mathbf{w}$ is infinite is useful: each factor is followed by a letter whose image begins with α). As $f(a) \neq f(b)$ and as the letter α occurs only as a prefix in $f(a)$ and $f(b)$, $f(a)\alpha$ is not a prefix of $f(b)\alpha$ and, conversely, $f(b)\alpha$ is not a prefix of $f(a)\alpha$. Hence there exist a word v and letters a', b' such that $a' \neq b'$, va' and vb' are respectively prefixes of $f(a)\alpha$ and $f(b)\alpha$. It follows that $f(u)va'$ is a prefix of $f(\mathbf{w})$ while $\beta f(u)vb'$ and $\gamma f(u)vb'$ are factors of $f(\mathbf{w})$: $f(\mathbf{w})$ is not LSP, a contradiction. $\qquad \square$

Observe that Lemma 3.3 does not hold for finite words. For instance the word baa is not LSP while its image $abaa$ by the morphism $[a, ab]$ is LSP.

4 Fragilities of infinite LSP words

The converse of Lemma 3.3 is false: there exist an infinite LSP word $\mathbf{w}$ and a bLSP morphism f such that $f(\mathbf{w})$ is not LSP.

Example 4.1. Let $\mathbf{F}$ be the well-known Fibonacci word (the fixed point of the endomorphism $[ab, a]$), and let $g = \lambda_{acb} = [a, acb, ac]$. The word $g^2(\mathbf{F})$ begins with the word $g^2(ab) = g(aacb) = aaacacb$ that contains the factor ac which is left special but not a prefix of the word. Hence the word $g^2(\mathbf{F})$ is not LSP while $\mathbf{F}$ is LSP and g is bLSP (actually one can prove, using Lemma 4.4 below, that $g(\mathbf{F})$ is LSP).

In what follows, we introduce some properties of LSP words and morphisms that explain in which context a (breaking) R-bLSP morphism can map a (fragile) infinite LSP word on a non-LSP word.

Definition 4.2. Let a, b, c be three pairwise distinct letters. An infinite word $\mathbf{w}$ is (a, b, c)-*fragile* if there exist a word u and distinct letters β and γ such that the word ua is a prefix of $\mathbf{w}$ and the words αub and βuc are factors of $\mathbf{w}$. The word u is called an (a, b, c)-*fragility* of $\mathbf{w}$. We will also say that, when we need letters α and β, $\mathbf{w}$ is (a, b, c, β, γ)-fragile and the word u is an (a, b, c, β, γ)-*fragility* of $\mathbf{w}$.

For instance, the empty word ε is an (a, b, c, c, a)-fragility of $g(\mathbf{F})$: εa is a prefix of $g(\mathbf{F}) = aacb\cdots$ while $c\varepsilon b$ and $a\varepsilon c$ are factors of $g(\mathbf{F})$. More generally any factor abc or acb in an infinite word produces an (a, b, c)-fragility. One can also observe that, by symmetry of the definition, any (a, b, c)-fragile word is also (a, c, b)-fragile. Finally let us note that no fragility exists in words over two letters (as the definition needs three pairwise different letters).

The main idea of introducing the previous notion is that for any (a, b, c)-fragile LSP word $\mathbf{w}$, there exists a bLSP morphism such that $f(\mathbf{w})$ is not LSP. For instance, if $u, \alpha, \beta, \mathbf{w}$ are as in Definition 4.2 and if $g = \lambda_{acb} = [a, acb, ac]$, the word $g(u)aa$ is a prefix of $g(\mathbf{w})$ whereas the words $\alpha g(u)acb$ and $\beta g(u)ac$ are factors of $g(\mathbf{w})$, so that $g(\mathbf{w})$ is not LSP since $g(u)ac$ is left special but not a prefix of $g(\mathbf{w})$.

Definition 4.3. Let a, b, c be three pairwise distinct letters. A morphism f is LSP (a, b, c)-*breaking*, if for all (a, b, c)-fragile LSP word w, $f(w)$ is not LSP.

For instance, the morphism $\lambda_{acb} = [a, acb, ac]$ is (a, b, c)-breaking.

Lemma 4.4. *Let $\mathbf{w}$ be an infinite LSP word and let f be an R-bLSP morphism defined on* alph$(\mathbf{w})$. *The following assertions are equivalent:*

1. *The word $f(\mathbf{w})$ is not LSP;*

2. *There exist some pairwise distinct letters a, b, c such that $\mathbf{w}$ is (a, b, c)-fragile and the longest common prefix of $f(b)$ and $f(c)$ is strictly longer than the longest common prefix of $f(a)$ and $f(b)$;*

3. *There exist some pairwise distinct letters a, b, c in $\mathbf{w}$ such that $\mathbf{w}$ is (a, b, c)-fragile and f is LSP (a, b, c)-breaking.*

Proof. $1 \Rightarrow 2$. Assume first that $f(\mathbf{w})$ is not LSP. There exists a left special factor V of $f(\mathbf{w})$ which is not a prefix of $f(\mathbf{w})$. Let v be the longest common prefix of V and $f(\mathbf{w})$. Let a', b' be the letters such that va' is a prefix of $f(\mathbf{w})$ and vb' is a prefix of V: by construction $a' \neq b'$. Let also β, γ be distinct letters such that βV and γV are factors of $f(\mathbf{w})$ (also $\beta vb'$ and $\gamma vb'$ are factors of $f(\mathbf{w})$).

By the definition of R-bLSP morphisms, the letter $\alpha = \text{first}(f)$ is the unique letter that can be left special in $f(\mathbf{w})$. As α is the first letter of $f(\mathbf{w})$, we have $v \neq \varepsilon$ and $\text{first}(v) = \text{first}(f)$. As α occurs exactly at the first position in all images of letters, occurrences of α mark the beginning of images of letters in $f(\mathbf{w})$. Considering the last occurrence of α in v, we can write $v = f(u)\alpha x$ with $|x|_\alpha = 0$. Let a, b, c be letters such that:

- ua is a prefix of $\mathbf{w}$, and, $va' = f(u)\alpha xa'$ is a prefix of $f(ua)$ when $a' \neq \alpha$ or $v = f(ua)$ when $a' = \alpha$;

- βub is a factor of $\mathbf{w}$, and, $\beta vb'$ is a prefix of $\beta f(ub)$ when $b' \neq \alpha$ or $v = f(ub)$ when $b' = \alpha$;

- γuc is a factor of $\mathbf{w}$, and, $\gamma vb'$ is a prefix of $\gamma f(uc)$ when $b' \neq \alpha$ or $v = f(uc)$ when $b' = \alpha$.

As $a' \neq b'$, we have $a \neq b$ and $a \neq c$. Observe that until now we did not use the fact that $\mathbf{w}$ is LSP. This implies $b \neq c$ (and so $b' \neq \alpha$). Indeed otherwise ub would be a left special factor of $\mathbf{w}$ without being one of its prefixes: a contradiction with the fact that $\mathbf{w}$ is an LSP word. Thus $\mathbf{w}$ is (a, b, c)-fragile.

This ends the proof of Part $1 \Rightarrow 2$ as $\alpha xb'$ is a common prefix of $f(b)$ and $f(c)$ and αx is the longest common prefix of $f(a)$ and $f(b)$.

$2 \Rightarrow 3$. By hypothesis, $f(a) = v\delta w_1$, $f(b) = v\gamma w_2$ and $f(c) = v\gamma w_3$ for letters δ, γ and words w_1, w_2 and w_3 with $\delta \neq \gamma$. Let $\mathbf{w}'$ be any LSP (a, b, c)-fragile infinite word. Let u', β' and γ' be the word and letters such that $u'a$ is a prefix of $\mathbf{w}'$ while $\beta'u'b$ and $\gamma'u'c$ are factors of $\mathbf{w}'$ with $\beta' \neq \gamma'$. The word $f(\mathbf{w}')$ has $f(u')v\delta$ as a prefix and words $\beta'f(u')v\gamma$ and $\gamma'f(u')v\gamma$ as factors. As $\delta \neq \gamma$, the word $f(\mathbf{w}')$ is not LSP. The morphism f is LSP (a, b, c)-breaking.

$3 \Rightarrow 1$. This follows the definition of (a, b, c)-fragile words and LSP (a, b, c)-breaking morphisms. $\qquad \square$

Observe that we have also proved the next result.

Corollary 4.5. *An R-bLSP morphism is LSP (a, b, c)-breaking for pairwise distinct letters a, b and c if and only if the longest common prefix of $f(b)$ and $f(c)$ is strictly longer than the longest common prefix of $f(a)$ and $f(b)$.*

To end this section let us mention that in the binary case the converses of Lemma 3.3 and Proposition 3.1 hold. Indeed as shown in Lemma 4.4, if $\mathbf{w}$ is LSP, f is an R-bLSP morphism and $f(\mathbf{w})$ is not LSP, then $\mathbf{w}$ contains at least three distinct letters. Recall that the elements of $S_{\mathrm{R-bLSP}}(\{a, b\})$ are the morphisms L_a, L_b and their restrictions to alphabets $\{a\}$ and $\{b\}$.

Corollary 4.6. *If $\mathbf{w}$ is an LSP infinite word over $\{a, b\}$ and if f is defined over $\mathrm{alph}(\mathbf{w})$ and belongs to the set $S_{\mathrm{R-bLSP}}(\{a, b\}$ then $f(\mathbf{w})$ is also LSP. Consequently an infinite word over $\{a, b\}$ is LSP if and only if it is $\{L_a, L_b\}$-adic.*

5 Origin of fragilities

Before characterizing infinite LSP words, we need to know how fragilities in an LSP word $\mathbf{w}$ can appear. We will see that this depends only on the morphisms in the fitted $S_{\mathrm{R-bLSP}}$-adic directive word of $\mathbf{w}$ (Theorem 7.1). At a first step, let f be an R-bLSP morphism from A^* to B^* with $\alpha = \mathrm{first}(f)$. We examine the fragilities occurring in the image by f of an LSP word.

New fragilities. Assume that for some letters b, c, β, γ with $\alpha \neq b \neq c \neq \alpha$ and $\beta \neq \gamma$, the words βb and γc belong to $f(A)$, that is, are factors of images of letters by f. Then for any word $\mathbf{w}$ containing all letters of A (or at least the letters whose images contain βb and γc), the empty word is an $(\alpha, b, c, \beta, \gamma)$-fragility of $f(\mathbf{w})$. We say that ε is an $(\alpha, b, c, \beta, \gamma)$-fragility (or simply an (α, b, c)-fragility) *associated with* f. For instance, (a, b, c) is a fragility associated with the morphism $\lambda_{abc} = [a, ab, abc]$.

Propagated fragilities. Assume now that an infinite word $\mathbf{w}$ over A contains an $(a', b', c', \beta, \gamma)$-fragility v. Observe that: $f(va')\alpha$ is a prefix of $f(\mathbf{w})$; $\beta f(vb')\alpha$ and $\gamma f(vc')\alpha$ are factors of $f(\mathbf{w})$; the word $f(v)\alpha$ is a prefix of the three words $f(va')\alpha$, $f(vb')\alpha$ and $f(vc')\alpha$. Assume there exist a common prefix u of these words and pairwise distinct letters a, b and c such that ua is a prefix of $f(va')\alpha$, βub is a prefix of $\beta f(vb')\alpha$ and γuc us a prefix of $\gamma f(vc')\alpha$. Then u is an (a, b, c, β, γ)-fragility in $f(w)$. We say that this fragility is *propagated* by f from the $(a', b', c', \beta, \gamma)$-fragility v.

Note that $|u| \geq |f(v)\alpha|$ and more precisely $u = f(v)\alpha u'$ for some word u'. Observe that $\alpha u'a$ is a prefix of $f(a')\alpha$, $\alpha u'b$ is a prefix of $f(b')\alpha$ and $\alpha u'c$ is a prefix of $f(c')\alpha$. This is an important fact as it shows that letters a, b and c depend only on f and on letters a', b' and c': if a word contains an (a', b', c')-fragility then $f(\mathbf{w})$ contains an (a, b, c)-fragility. For instance, if $A = \{a, b, c\}$ and $\mathbf{w}$ contains an (a, b, c)-fragility u, then this fragility is propagated by the morphism $f = [a, ab, ac]$. Note that $|f(v)\alpha| \leq |u|$ implies $|v| < |u|$. Hence the propagation of a fragility makes it strictly longer.

Some fragilities can also be not propagated. For instance, the morphism $[a, ab, abc]$ does not propagate the (b, a, c)-fragilities.

Lemma 5.1. *Let $\mathbf{w}$ be an infinite word (not necessarily LSP) and let f be an R-bLSP morphism defined over $\mathrm{alph}(\mathbf{w})$. Fragilities of $f(\mathbf{w})$ are exactly the fragilities associated with f and the fragilities of $\mathbf{w}$ propagated by f.*

Proof. From the definitions given before the lemma, the fragilities associated with f and the fragilities of $\mathbf{w}$ propagated by f are fragilities of $f(\mathbf{w})$. Let u be an (a, b, c, β, γ)-fragility of $f(\mathbf{w})$.

If $u = \varepsilon$, it follows from the definition of an (a, b, c, β, γ)-fragility that $a = \mathrm{first}(f(\mathbf{w}))$ and βb, γc are factors of $f(\mathbf{w})$. Now observe that, still by the same definition, $a \notin \{b, c\}$. Thus by the definition of R-bLSP morphisms, $a = \mathrm{first}(f)$ and βb, γc belong to $\mathrm{Fact}(f(\mathrm{alph}(\mathbf{w})))$. The fragility u is associated with f.

Assume from now on that u is not empty. Let $\alpha = \mathrm{first}(f)$. Considering the last occurrence of α in u, observe that the word u can be decomposed in a unique way as $u = f(v)\alpha x$ with v, x words such that $|x|_\alpha = 0$. As u is an (a, b, c, β, γ)-fragility of $f(\mathbf{w})$, there exist words w_1, w_2 and w_3 such that:

- $|w_1|_\alpha = |w_2|_\alpha = |w_3|_\alpha = 0$;

- $f(v)\alpha x w_1 \alpha$ is a prefix of $f(\mathbf{w})$ and $a = \mathrm{first}(w_1\alpha)$;

- $\beta f(v)\alpha x w_2 \alpha$ and $\gamma f(v)\alpha x w_3 \alpha$ are factors of $f(\mathbf{w})$ with $b = \mathrm{first}(w_2\alpha)$ and $c = \mathrm{first}(w_3\alpha)$.

By the definition of an R-bLSP morphism, there exist letters a', b', c' such that $f(a') = \alpha x w_1$, $f(b') = \alpha x w_2$, $f(c') = \alpha x w_3$. These letters a', b', c' are pairwise distinct since letters $a = \mathrm{first}(w_1\alpha)$, $b = \mathrm{first}(w_2\alpha)$ and $c = \mathrm{first}(w_3\alpha)$ are pairwise distinct. Moreover va' is a prefix of $\mathbf{w}$ and words $\beta v b'$ and $\gamma v c'$ are factors of $\mathbf{w}$ (remember that α marks the beginning of letters in $f(\mathbf{w})$ as f is an R-bLSP morphism and as, for letters x and y, $\mathrm{last}(f(x)) = y$ implies $x = y$). Hence the word v is an $(a', b', c', \beta, \gamma)$-fragility of $\mathbf{w}$. This fragility is propagated by f. $\qquad\square$

6 A first automaton

Lemma 5.1 shows that bLSP morphisms act locally on fragilities on the words. One can then construct an automaton to store the actions of these morphisms on fragilities of LSP words (one can do this for arbitrary words but this is not needed for our purpose). It is important to note that the exact alphabet on which a morphism is applied has a lot of importance. For instance, if $\mathbf{w}$ is an infinite word over $\{a, b, c\}$ and if $f = \lambda_{abc} = [a, ab, abc]$, $f(\mathbf{w})$ contains an (a, b, c)-fragility if and only if c occurs in $\mathrm{alph}(\mathbf{w})$. It is also important to note that the automaton will not follow the fragilities themselves but only the 3-tuples of letters (a, b, c) for which a fragility occurs. Given an infinite word $\mathbf{w}$, let $\mathrm{Frag}(\mathbf{w})$ be the set of 3-tuples (a, b, c) of letters such that $\mathbf{w}$ contains an (a, b, c)-fragility.

Let $\mathcal{A}_1 = (S_{\mathrm{R-bLSP}}, Q_1, \Delta_1)$ be the automaton without initial and final states defined by:

- The alphabet of $\mathcal{A}_1$ is the set $S_{\mathrm{R-bLSP}}$ of R-bLSP morphisms over the alphabet A.

- The set of states is the set $\{(\mathrm{alph}(\mathbf{w}), \mathrm{Frag}(\mathbf{w})) \mid \mathbf{w} \text{ LSP over } A\}$.

- A 3-tuple $((A_1, F_1), f, (A_2, F_2))$ is a transition of $\mathcal{A}_1$ if and only if the following condition holds:

  1. $f_{|A_1} = f$ (f is exactly defined on A_1; this point is important as it prevents from having a morphism in the transition that creates a fragility which is not related to A_1; we need to control new and propagated fragilities);

  2. $A_2 = \mathrm{alph}(f(A_1))$ (each letter of A_2 occurs in at least one image of a letter in A_1).

  3. if $(a, b, c) \in A_1$ then f is not LSP (a, b, c)-breaking;

  4. F_2 is the union of the set of 3-tuples (a, b, c) for which an (a, b, c)-fragility is associated with f and the set of 3-tuples (a, b, c) propagated by f from (a', b', c')-fragilities with (a', b', c') in F_1.

One can observe that for each state q of the automaton and each R-bLSP morphism f, there exists at most one state q' such that (q, f, q') is a transition. There is no transition when $q = (B, F)$, $(a, b, c) \in F$ and f is (a, b, c)-breaking.

By the definition of the automaton, Lemma 5.1 has the following immediate corollary.

Corollary 6.1. *Let $\mathbf{w}$ be an LSP word, let $A_0, A_1, \ldots, A_k$ be alphabets and let $f_1, f_2, \ldots, f_k$ be R-bLSP morphisms such that:*

- $A_0 = \mathrm{alph}(\mathbf{w})$;

- *for all i, $1 \le i \le k$, $A_i = \mathrm{alph}(f_i \circ \cdots \circ f_1(\mathbf{w}))$;*

- *For all i, $1 \le i \le k$, $f_i \circ \cdots \circ f_1(\mathbf{w})$ is LSP.*

Then in $\mathcal{A}_1$ there is a unique path labeled by $f_1 \cdots f_k$ from the state $(\mathrm{alph}(\mathbf{w}), \mathrm{Frag}(\mathbf{w}))$ to a state (A_k, F). Moreover F is the set of 3-tuples (a, b, c) for which there exist (a, b, c)-fragilities in $f_k \circ \cdots \circ f_1(\mathbf{w})$.

$$a \mapsto a \qquad\qquad L_a, L_b \qquad\qquad b \mapsto b$$

$$\boxed{\{a\}, \emptyset} \xrightarrow{\;L_b|\{a\} : a \mapsto ba\;} \boxed{\{a,b\}, \emptyset} \xleftarrow{\;L_a|\{b\} : b \mapsto ab\;} \boxed{\{b\}, \emptyset}$$
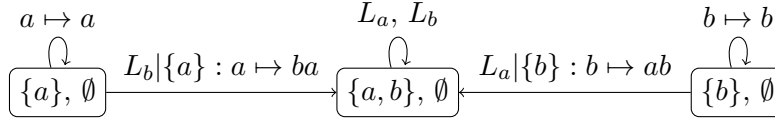
Figure 2: $\mathcal{A}_1$ for the binary alphabet

On the alphabet $\{a, b\}$, as there is no fragility, the automaton $\mathcal{A}_1$ has only three states (see Figure 2): one state for each of the alphabets $\{a\}, \{b\}, \{a, b\}$.

The automaton for the alphabet $\{a, b, c\}$ is provided in Figure 4. Here follow some explanations. We say that $\mathbf{w}$ has fragilities of type a (resp. of type b, of type c) if it has (a, b, c)-fragilities or (a, c, b)-fragilities (resp. (b, a, c)-fragilities or (b, c, a)-fragilities; (c, a, b)-fragilities or (c, b, a)-fragilities). Figure 3 shows the action of R-bLSP morphisms on fragilities. One can observe that, for this ternary case, the type of a propagated fragility is the same then the type of the fragility from which it is propagated.

Properties of the R-bLSP morphism f considered	new fragilities	action on type a fragilities	action on type b fragilities	action on type c fragilities
f is L_a, L_b, L_c or one of their restrictions		propagated	propagated	propagated
$f(c) = abc$ or $f(b) = acb$	a-fragilities	LSP breaking	not propagated	not propagated
$f(c) = bac$ or $f(a) = bca$	b-fragilities	not propagated	LSP breaking	not propagated
$f(b) = cab$ or $f(a) = cba$	c-fragilities	not propagated	not propagated	LSP breaking

Figure 3: Ternary case: action of R-bLSP morphisms on fragilities

Observe that in the last three cases of Figure 3 only one type of fragility is kept. This shows that for ternary LSP words only four sets of fragilities can occur: the emptyset, the set of a-fragilities, the set of b-fragilities and the set of c-fragilities. In other words an LSP word over $\{a, b, c\}$ cannot have simultaneously fragilities of type a and b (nor of type a and c; nor of type b and c).

The presentation of the automaton in Figure 4 is split into two parts. The first part contains the seven states corresponding to words without fragilities. The second part contains the three states corresponding to words with fragilities. In the first part, morphisms with entries "-" denote restrictions of endomorphisms: "-" means "not defined". In the second part, entries "*" denote any possibility of image (defined or not defined as this kind of entries correspond to restriction of $\lambda_{\cdot}$ morphisms). There exist numerous transitions from the first part to the second one. The outgoing side of the transition is presented in the first part and the in-going side of the transition is presented in the second part. For instance there exists a transition from state $(\{b, c\}, \emptyset)$ to state $(\{a, b, c\}, \text{type } a)$ labeled by morphism $[-, ab, abc]$.

For larger alphabets, the number of states explodes. This is due to the increasing number of subalphabets but also to the fact that the set of fragilities may contain more than one fragility. For instance, any LSP morphism f such that $f(a) = abcd$ creates simultaneously several fragilities: in the image of an infinite word containing a, the empty word is an (a, b, c)-fragility, an (a, b, d)-fragility and an (a, c, d)-fragility.

7 The S-adic Characterization of LSP words

Let $\mathbf{w}$ be an LSP word. By Proposition 3.1, $\mathbf{w}$ is $S_{\text{R}-\text{bLSP}}$-adic. Let $\mathbf{f} = (f_k)_{k \geq 0}$ be its fitted directive word over $S_{\text{R}-\text{bLSP}}$. Let also $(\mathbf{w}_k)_{k \geq 1}$ be the sequence of infinite words verifying: $\mathbf{w}_1 = \mathbf{w}$; for all $k \geq 1$, $\mathbf{w}_k = f_k(\mathbf{w}_{k+1})$.
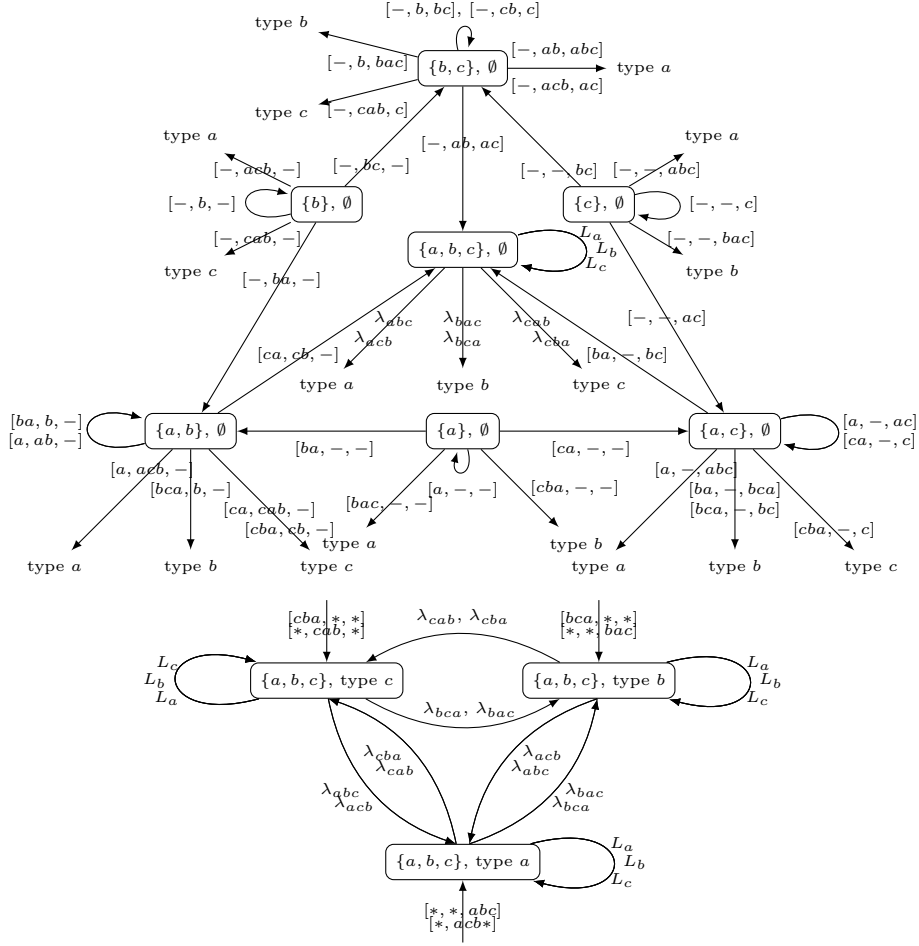
Figure 4: $\mathcal{A}_1$ for the ternary alphabet

Let $\mathcal{A}_{R-bLSP}(A)$, or simply $\mathcal{A}_{R-bLSP}$, be the automaton obtained by reversing transitions in $\mathcal{A}_1$ and considering all states as initial. As $\mathbf{f}$ is fitted, an immediate consequence of Corollary 6.1 is that there exists an infinite path labeled by $\mathbf{f}$ in $\mathcal{A}_{R-bLSP}$ starting on state $(\mathrm{alph}(\mathbf{w}), \mathrm{Frag}(\mathbf{w}))$. The next theorem states that conversely if such an infinite path exists for an $S_{\mathrm{R-bLSP}}$-adic word $\mathbf{w}$, then $\mathbf{w}$ is LSP. If $q = (X, F)$ is a state of $\mathcal{A}_{R-bLSP}$, let $\mathrm{alph}(q) = X$ and $\mathrm{Frag}(q) = F$.

Theorem 7.1. *A word $\mathbf{w}$ is LSP if and only if it is $S_{\mathrm{R-bLSP}}$-adic and there exists an infinite path in $\mathcal{A}_{R-bLSP}$ labeled by the fitted directive word of $\mathbf{w}$.*

Applying this theorem allows to provide particular examples of LSP words. For instance it can be verified that all standard episturmian words are LSP [1, 11]. These words are the $\mathcal{S}$-adic words where S is the set of morphisms containing morphisms L_α and their restrictions. In particular, if $A = \{a_1, \ldots, a_k\}$, the k-bonacci word $\mathbf{B}_{a_1, \ldots, a_k}$ (or Tribonacci word, also denoted $\mathbf{T}$, when $k = 3$) which is the fixed point of the morphism $L_{a_1} \circ \cdots \circ L_{a_k}$ is LSP. Observe that $\mathbf{B}_{a_1, \ldots, a_k}$ contains no fragility as its fitted directive word is recognized using only the state $(\{a_1, \ldots, a_k\}, \emptyset)$ in $\mathcal{A}_{R-bLSP}$. It may be observed that the fitted directive word of an LSP word can be recognized by several paths. For instance the fitted directive word of the Tribonacci word is the word $(L_a L_b L_c)^\omega$ which is recognized by four paths starting respectively from states $(\{a, b, c\}, \emptyset)$, $(\{a, b, c\}, \mathrm{type\ a})$, $(\{a, b, c\}, \mathrm{type\ b})$ and $(\{a, b, c\}, \mathrm{type\ c})$.

Proof of Theorem 7.1. The only if part was already mentioned before the statement of the theorem.

10

Assume, by contradiction, that there exists in $\mathcal{A}_{R-bLSP}$ a path labeled by a directive word $\mathbf{f}$ of a word $\mathbf{w}$ which is S_{R-bLSP}-adic but not LSP (for this part of the proof it is not needed that $\mathbf{f}$ is a fitted directive word). Such a word contains a left special factor u that is not a prefix of $\mathbf{w}$. Among all possible 3-tuples $(\mathbf{f}, \mathbf{w}, u)$, choose one such that $|u|$ is minimal.

For $n \geq 1$, we denote by f_n the nth letter of $\mathbf{f}$ and $\mathbf{w}_n$ the word directed by $(f_k)_{k \geq n}$ ($\mathbf{w}_1 = \mathbf{w}$; $\mathbf{w}_2$ is the word directed by $f_2 f_3 \cdots$; $\mathbf{w}_n = f_n(\mathbf{w}_{n+1})$ for all $n \geq 1$).

The first three steps do not depend on the automaton.

Step 1: $\mathbf{w}_2$ *contains a fragility*

First observe that $|u| \geq 2$. Indeed we have $|u| \neq 0$ as the empty word is a prefix of $\mathbf{w}$. Moreover, by the structure of images of the R-bLSP morphism f_1, only the letter $\mathrm{first}(f_1)$ can be left special, whence $|u| \neq 1$.

Let $\alpha = \mathrm{first}(\mathbf{w}) = \mathrm{first}(f_1)$. Considering the last occurrence of α in u, the word u can be decomposed in a unique way $u = f_1(v)\alpha x$ with v, x words such that $|x|_\alpha = 0$.

As u is left special, there exist distinct letters β and γ such that βu and γu are factors of $\mathbf{w}$. As the letter α marks the beginning of images of letters in $\mathbf{w}$ and as for all letters δ, $f_1(\delta)$ ends with δ, we deduce that βv and γv are factors of $\mathbf{w}_2$. As $|v| < |u|$ and by the choice of the 3-tuple $(\mathbf{f}, \mathbf{w}, u)$, the word v is a prefix of $\mathbf{w}_2$. Consequently $f_1(v)\alpha$ is a prefix of $\mathbf{w}$ and so $x \neq \varepsilon$.

Assume there exists a unique letter b such that βvb is a factor of $\mathbf{w}_2$ and u is a prefix of $f(vb)$. Assume also that b is the unique letter c such that γvc is a factor of $\mathbf{w}_2$ and u is a prefix of $f(vc)$. As u is not a prefix of $\mathbf{w} = f_1(\mathbf{w}_2)$ and as u is a prefix of $f_1(vb)$, the word vb is not a prefix of $\mathbf{w}_2$. By the choice of the 3-tuple $(\mathbf{f}, \mathbf{w}, u)$, $|vb| \geq |u|$. As $|v| < |u|$, we get $|vb| = |u| = |f_1(v)\alpha x|$. As $|f_1(v)| \geq |v|$, it follows $x = \varepsilon$: a contradiction.

From what precedes, we deduce the existence of two distinct letters b and c such that βvb and γvc are factors of $\mathbf{w}_2$ with u a prefix of $f_1(vb)$ and $f_1(vc)$. As u is not a prefix of $\mathbf{w} = f_1(\mathbf{w}_2)$, the letter a that follows the prefix v of $\mathbf{w}_2$ is different from b and c. Hence the word $\mathbf{w}_2$ is (a, b, c, β, γ)-fragile and v is such a fragility.

Step 2: f_1 *is LSP* (a, b, c)-*breaking*

By the definition of letters b and c at Step 1, the word αx is a common prefix of $f_1(b)$ and $f_1(c)$. Also as $u = f_1(v)\alpha x$ is not a prefix of $\mathbf{w}$ while $f_1(v)a$ is a prefix of $\mathbf{w}$, the word αx is not a prefix of $f_1(a)$. By Corollary 4.5, f_1 is (a, b, c)-breaking.

Step 3: *origin of fragilities of* $\mathbf{w}_2$

Applying iteratively Lemma 5.1, as the lengths of propagated fragilities decrease with the propagation, we deduce the existence of an integer $n \geq 2$ and a sequence of 3-tuples of pairwise different letters $(a_i, b_i, c_i)_{i \in \{2, \cdots, n\}}$, a sequence $(v_i)_{i \in \{2, \cdots, n\}}$ of words such that:

- v_i is an $(a_i, b_i, c_i, \beta, \gamma)$-fragility of $\mathbf{w}$ for all $i \in \{2, \cdots, n\}$;

- $(a_2, b_2, c_2) = (a, b, c)$ and $v_2 = v$;

- $|v_{i+1}| < |v_i|$ for all $i \in \{2, \cdots, n-1\}$ (v_i is a fragility propagated by f_i from v_{i+1});

- $v_n = \varepsilon$ (origin of the fragilities);

- the words $v_i a_i$, $\beta v_i b_i$, $\gamma v_i c_i$ are respectively prefixes of $f_i(v_{i+1} a_{i+1})\alpha_i$, $\beta f_i(v_{i+1} b_{i+1})\alpha_i$, $\gamma f_i(v_{i+1} c_{i+1})\alpha_i$ where $\alpha_i = \mathrm{first}(f_i)$ for $i \in \{2, \cdots, n-1\}$ (by the definition of propagated fragilities);

- $a_n = \mathrm{first}(f_n)$;

- βb_n, γc_n belong to $f_n(\mathrm{alph}(\mathbf{w}_{n+1}))$ (by the definition of new fragilities).

Step 4: conclusion.

Let $(q_i)_{i \geq 1}$ be the sequence of states along the path labeled by $\mathbf{f}$: for all $n \geq 1$, (q_n, f_n, q_{n+1}) is a transition of $\mathcal{A}_{R-bLSP}$. At the end of Step 3, we learn that there exists an (a_n, b_n, c_n)-fragility in $f_n(\mathbf{w}_{n+1})$. Hence a_n, b_n, c_n are pairwise distinct letters. Especially as $a_n = \text{first}(f_n) \notin \{b_n, c_n\}$ by properties of R-bLSP morphisms, the words βb_n and γc_n are factors of images of some letters, say b'_n and c'_n. By the definition of the automaton $\mathcal{A}_{R-bLSP}$, $\text{alph}(q_n) = \text{alph}(f_n(\text{alph}(q_{n+1})))$ and f_n is defined exactly on $\text{alph}(q_{n+1})$. This implies that b'_n and c'_n belong to $\text{alph}(q_{n+1})$ and a_n, b_n and c_n belong to $\text{alph}(q_n)$. Moreover, as βb_n, γc_n are factors of words in $f_n(\text{alph}(q_{n+1}))$, we deduce that $(a_n, b_n, c_n) \in \text{Frag}(q_n)$.

Using a backward induction and the definition of the automaton $\mathcal{A}_{R-bLSP}$, we can show that for all i, $2 \leq i \leq n$, $(a_i, b_i, c_i) \in \text{Frag}(q_i)$. Especially $(a_2, b_2, c_2) \in \text{Frag}(q_2)$. As (q_1, f_1, q_2) is a transition of $\mathcal{A}_{R-bLSP}$, by the definition of transitions, f_1 is not LSP (a_2, b_2, c_2)-breaking: a contradiction with Step 2 as $(a, b, c) = (a_2, b_2, c_2)$. $\qquad\square$

8 Morphisms preserving LSP words

As explained at the end of [19], one can ask for a simpler characterization for LSP words. In the context of substitutive-adicity, the next theorem shows that this is not possible: given any set S of morphisms such that LSP words are S-adic, one needs a way to distinguish directive words of LSP words.

Theorem 8.1. *Let A be an alphabet containing at least three letters. There exists no set S of morphisms such that the set of LSP infinite words over A is the set of S-adic words.*

The main idea of the proof lies on the fact that, given any set S of morphisms, these morphisms preserve the property of being an S-adic word: the image of any S-adic word by an element of S is still S-adic. Thus in this section, we first characterize the set of endomorphisms preserving the LSP property (the image of any LSP word is also LSP) and then show the existence of an LSP word that cannot be decomposed over this set of morphisms.

8.1 Some Morphisms

Let $\mathcal{L}(A) = \{L_a \mid a \in A\}$. When the context will be clear, we will just write $\mathcal{L}$ instead of $\mathcal{L}(A)$. Observe that, in the binary case, $\mathcal{S}_{bLSP} = \mathcal{L}$.

Lemma 8.2. *A bLSP morphism preserves the LSP property if and only if it belongs to $\mathcal{L}$.*

Proof. Let f be a bLSP morphism that does not belong to $\mathcal{L}$. There exist three pairwise distinct letters a, b and c such that $f(a) = a$, $f(b) = ab$, $f(c) = abc$. As previously mentioned (see Example 4.1) there exists an LSP word $\mathbf{w}$ that contains an (a, b, c)-fragility (see the beginning of Section 4). That is, for a word u and distinct letters β, γ, ua is a prefix of $\mathbf{w}$ while βub and γuc are factors of $f(\mathbf{w})$. Then $f(u)aa$ is a prefix of $f(\mathbf{w})$ while $\beta f(u)ab$ and $\gamma f(u)ab$ are factors of $\mathbf{w}$: $f(\mathbf{w})$ is not LSP and so f does not preserve the LSP property.

Conversely for any LSP word $\mathbf{w}$ and any letter α, let us prove that $L_\alpha(\mathbf{w})$ is also LSP. Assume by contradiction that $L_\alpha(\mathbf{w})$ is not LSP and consider a word u and two distinct letters a and b, such that ub is a left special factor of $L_\alpha(\mathbf{w})$ while ua is a prefix of $L_\alpha(\mathbf{w})$. Let β and γ be distinct letters such that βub and γub are factors of $L_\alpha(\mathbf{w})$. Observe that at least one of the letters a and b is different from α, and also at least one of the letters β and γ is different from α. The structure of L_α (any letter different from α is preceded by α in the image of any word) implies that u ends and begins with the letter α (in particular it is not empty) and so can be decomposed $u = L_\alpha(v)\alpha$ for a word v. One can deduce that va is a prefix of $\mathbf{w}$ while βvb and γvb are factors of $\mathbf{w}$: a contradiction with $\mathbf{w}$ LSP. $\qquad\square$

Observe that Lemma 8.2 does not extend to R-bLSP morphisms. For instance $[-, -, abc, ad]$ preserves the LSP property for words over $\{c, d\}$.

It is important also to observe that the first part of the proof of Lemma 8.2 also shows the next result.

Lemma 8.3. *If a bLSP morphism does not belong to $\mathcal{L}$ then it is (a, b, c)-breaking for some pairwise distinct letters a, b and c.*

We will also need the following result.

Lemma 8.4. *Let α be a letter, a, b and c be three distinct letters and $\mathbf{w}$ be a word. The word $\mathbf{w}$ contains an (a, b, c)-fragility if and only if $L_\alpha(\mathbf{w})$ contains an (a, b, c)-fragility.*

Proof. The proof of the if part is similar to the converse part of the proof of Lemma 8.2 (if u is a word such that ua is a prefix of $L_\alpha(\mathbf{w})$ and the words βub and γuc are factors of $L_\alpha(\mathbf{w})$ with $\beta \neq \gamma$ then $u = L_\alpha(v)\alpha$ for a word v such that va is a prefix of $\mathbf{w}$ and the words βwb and γwc are factors of $\mathbf{w}$). The proof of the only if part is straightforward and is left to readers. $\qquad\square$

There exist morphisms that preserve the LSP property without being bLSP morphisms. We introduce some of them. First, we denote by $\mathcal{U}(A, B)$ (or simply $\mathcal{U}$ when the context is clear), the set of all nonerasing morphisms from A^* to B^* such that for each letter a in B, $\sum_{\alpha \in A} |f(\alpha)|_a \leq 1$ (each letter of B occurs at most once in the set of images of letters by f). Here $\mathcal{U}$ stands for uniqueness. *Renaming morphisms* are nonerasing morphisms such that, for all a in A, $|f(a)| = 1$ and such that $\#\{f(a) \mid a \in A\} = A$. Renaming morphisms are elements of $\mathcal{U}(A, B)$. *Permutation morphisms* (or simply *permutations*), that is, morphisms such that $\{f(a) \mid a \in A\} = A$, are particular renaming morphisms. Let $Perm(A)$ (or simply $Perm$) denote the set of permutation morphisms.

Second, we denote by $\mathcal{P}_{\text{LSP}}(A, B)$ (or simply $\mathcal{P}_{\text{LSP}}$ when the context is clear), the set of all nonerasing morphisms from A^* to B^* for which there exists a word u over B with $f(\alpha) \subseteq u^+$ for all letters α in A and u^ω an LSP infinite word. Here $\mathcal{P}$ stands for periodic as any morphism in $\mathcal{P}_{\text{LSP}}$ maps any infinite word to a periodic word. The proof of the next lemma follows quite immediately from the definitions and is left to the reader.

Lemma 8.5. *All morphisms in $\mathcal{U}(A, B) \cup \mathcal{P}_{\text{LSP}}(A, B)$ preserve the LSP property.*

8.2 A first necessary condition

Proposition 8.6. *Any nonerasing morphism f that preserves the LSP property for infinite words belongs to $S_{\text{R-bLSP}}(B)^* \mathcal{U}(A, B) \cup \mathcal{P}_{LSP}(A, B)$.*

The converse of this proposition is false as some R-bLSP morphisms do not preserve the LSP property (see the notion of LSP breaking morphisms).

The rest of the section is devoted to the proof of Proposition 8.6. We first need a technical lemma on factors of LSP infinite words.

Lemma 8.7. *Let $\mathbf{w}$ be an LSP word, u be a prefix of $\mathbf{w}$ and $n \geq 1$ be an integer. If u^{n+1} is a factor of $\mathbf{w}$, then u^n is a prefix of $\mathbf{w}$.*

Proof. Let us consider an occurrence of the word u^{n+1}. Observing its extension on the left, we deduce that $u = u_1 u_2$ for two words u_1 and u_2 such that one of the following two possibilities hold:

- $u_2(u_1 u_2)^{n+1}$ is a prefix of $\mathbf{w}$,

- $\alpha u_2(u_1 u_2)^{n+1}$ is a factor of $\mathbf{w}$ with α a letter different from $\text{last}(u_2 u_1)$.

In the second case, the word $(u_2 u_1)^n$ is a left special factor of $\mathbf{w}$. As $\mathbf{w}$ is LSP, $(u_2 u_1)^n$ is a prefix of $\mathbf{w}$. This also holds in the first case. By hypothesis, u is a prefix of $\mathbf{w}$. As $n \geq 1$, both u and $u_2 u_1$ are prefixes of w: $u = u_2 u_1$ and consequently u^n is a prefix of $\mathbf{w}$. $\qquad\square$

For a non-empty finite word w, let $\sqrt{w}$ denote its primitive root, that is, the shortest word u such that $w = u^k$ for some integer k. We now show when the second condition of Proposition 8.6 happens.

Lemma 8.8. *Let f be a nonerasing morphism that preserves the LSP property for infinite words. Assume there exist two distinct letters a and b such that $\sqrt{f(a)} = \sqrt{f(b)}$. Then for all letters α, $f(\alpha) \in \sqrt{f(a)}^+$, that is, $f \in \mathcal{P}_{\mathrm{LSP}}$.*

Proof. Assume by contradiction that there exists a third letter c such that $\sqrt{f(c)} \neq \sqrt{f(a)} = \sqrt{f(b)}$. Let $u = \sqrt{f(a)}$. Consider the Tribonacci word $\mathbf{T}$ over $\{a, b, c\}$. By induction, one can prove that, for all $n \geq 1$, $L_a^{n-1}(\mathbf{T})$ begins with the prefix $a^n b a^n c a$ and contains the word $a^n b a^{n+1} b a^n$. As $L_a^{n-1}(\mathbf{T})$ is an standard episturmian word, it is LSP.

Choose an integer n such that $n \geq 1$ and $|f(a^n b a^n c)u| \leq |f(a^n b a^{n+1} b a^n)| - |u|$. Let m, p be the integers such that $f(a^n b a^n) = u^p$ and $f(a^n b a^{n+1} b a^n) = u^{m+1}$. As f preserves the LSP property, $f(L_a^{n-1}(\mathbf{T}))$ is LSP. As u^{m+1} is a factor of $f(L_a^{n-1}(\mathbf{T}))$ and u is a prefix of $f(L_a^{n-1}(\mathbf{T}))$, by Lemma 8.7, u^m is a prefix of $f(L_a^{n-1}(\mathbf{T}))$. Also $u^p f(c)u$ is a prefix of $f(L_a^{n-1}(\mathbf{T}))$ and $|u^p f(c)u| \leq |u^m|$. Thus $f(c)$ is a power of u or the last occurrence of u in $u^p f(c)u$ is an internal factor of u^2 (i.e. $uu = pus$ with p and s non-empty). The former case is not possible as by hypothesis $\sqrt{f(c)} \neq u$. The latter case contradicts the primitivity of u (as it implies $u = ps = sp$, a well-known equation which implies that u is not primitive – see, e.g., [15]). $\qquad\square$

Lemma 8.8 implies that Proposition 8.6 is a corollary of the next result.

Proposition 8.9. *Let $f : A^* \to B^*$ be a nonerasing morphism that preserves the LSP property for infinite words. Assume that for all letters a and b, the words $f(a)$ and $f(b)$ do not have the same primitive root. Then $f \in S_{\mathrm{R-bLSP}}^* \mathcal{U}$.*

Proof. We proceed by steps. Each step provides a stronger property on f.

Step 1. For all distinct letters a and b in A, $f(a)$ is not a suffix of $f(b)$.

Proof. Assume by contradiction that $f(a)$ is a suffix of $f(b)$ for two distinct letters a and b. Let $\ell \geq 0$ be the greatest integer such that $f(a)^\ell$ is a prefix of $f(b)$. The word $f(a^{3+\ell})$ is a factor of $f(baab)$. Let $\mathbf{w}$ be any LSP word containing the factor $baab$ and having aba as a prefix (for instance the Fibonacci word). By Lemma 8.7, $f(a)^{2+\ell}$ is a prefix of $f(\mathbf{w})$. Also $f(\mathbf{w})$ begins with $f(aba)$. By the definition of ℓ, this is possible only if $f(b) = f(a)^\ell u_1$, $f(a) = u_1 u_2$ for words u_1 and u_2 with $u_2 \neq \varepsilon$ a prefix of $f(a)$. But then, as $f(a)$ is a suffix of $f(b)$, we have $\ell \geq 1$ and $f(a) = u_2 u_1$. Hence $u_1 u_2 = u_2 u_1$. Thus $u_1 = \varepsilon$ or the words u_1 and u_2 have the same primitive root by [15, Prop. 1.3.2]. It follows that $f(a)$ and $f(b)$ have also the same primitive root: a contradiction with our hypotheses. $\qquad\square$

Step 2. For all distinct letters a and b in A, $\mathrm{last}(f(a)) \neq \mathrm{last}(f(b))$.

Proof. Assume by contradiction that $\mathrm{last}(f(a)) = \mathrm{last}(f(b))$ for two distinct letters a and b. By Step 1, $f(a)$ is not a suffix of $f(b)$ and $f(b)$ is not a suffix of $f(a)$. Hence there exist words u, u_1, u_2 and distinct letters α, β such that $f(a) = u_1 \alpha u$, $f(b) = u_2 \beta u$ with $u \neq \varepsilon$.

We observe that for any LSP word $\mathbf{w}$ containing a and b, the word u is a prefix of $f(\mathbf{w})$ (remember that f preserves the LSP property, hence $f(\mathbf{w})$ is LSP and its left special factors are prefixes of it). As $\mathbf{w}$ can be chosen with a as first letter or with b as first letter, this implies that u is a prefix of $f(a)$ and of $f(b)$.

Let $\mathbf{w}$ be any LSP word beginning with aab and containing the word $baaab$ (for instance the word $L_a(\mathbf{F})$ with $\mathbf{F}$ the Fibonacci word). Let ℓ be the greatest integer such that u^ℓ is a prefix of $f(a)$. Observe $\ell \geq 1$. The word $\alpha u^{\ell+1}$ is a factor of $f(aa)$ and the word $\beta u^{\ell+1}$ is a factor of $f(ba)$. So $u^{\ell+1}$ is a left special factor of $f(\mathbf{w})$ and so it is one of its prefixes. By the definition of ℓ, this implies that $f(a) = (u_1 u_2)^\ell u_1$ with $u = u_1 u_2$ and u_2 a prefix of $f(a)$. Now using the fact that $u^{\ell+1}$ is a prefix of $f(aa)$, we deduce that $\alpha u^{\ell+2}$ and $\beta u^{\ell+2}$ are factors of $f(baaa)$ and so of $f(\mathbf{w})$: $u^{\ell+2}$ is a left special factor of $f(\mathbf{w})$. Thus $f(\mathbf{w})$ begins with $u^\ell u_1 u_2 u_1 u_2$ and with $f(aa) = u^\ell u_1 u_1 u_2 u^{\ell-1} u_1$ as $\ell \geq 1$. Hence $u_1 u_2 = u_2 u_1$. By [15, Prop. 1.3.2], $\sqrt{u_1} = \sqrt{u_2}$ which implies that $\sqrt{f(a)} = \sqrt{u}$.

14

Exchanging the roles of a and b, we deduce similarly that $\sqrt{f(b)} = \sqrt{u}$. Hence $\sqrt{f(b)} = \sqrt{f(a)}$: a contradiction with our hypotheses. $\qquad\square$

From Step 2, we know that $f = g \circ r$ for some renaming morphism r and some morphism g such that $\mathrm{last}(g(\alpha)) = \alpha$ for all letters $\alpha \in r(A)$. Observe that f preserves the LSP property for words over A if and only if g preserves the LSP property for words over $r(A)$. Moreover the composition of any element of $\mathcal{U}$ with r is, when defined, an element of $\mathcal{U}$. Thus if Proposition 8.9 holds for g, it also holds for f. Hence replacing f by g, *from now on, we assume that* $\mathrm{last}(f(\alpha)) = \alpha$ *for all letters* α.

Step 3. If the images of two distinct letters begin with the same letter, then all images of letters begin with the same letter.

Proof. Assume $k = \#A \geq 3$ and that for two distinct letters α, β, $\mathrm{first}(f(\alpha)) = \mathrm{first}(f(\beta))$. Recall that we assumed after Step 2 that for all δ in A, $\mathrm{last}(f(\delta)) = \delta$.

Now consider the morphism $g = \lambda_{a_1 \ldots a_k}$ ($g(a_i) = a_1 \ldots a_i$ for all $i = 1, \ldots, k$; $a_1, \ldots, a_k$ are pairwise distinct letters). This morphism is bLSP. As said just after Theorem 7.1 the k-bonacci word $\mathbf{B}_{a_1, a_2, a_3, \ldots, a_k}$ contains no fragility. Hence $g(\mathbf{B}_{a_1, a_2, a_3, \ldots, a_k})$ is LSP. As f preserves the LSP property, $f(g(\mathbf{B}_{a_1, a_2, a_3, \ldots, a_k}))$ is LSP: it begins with $f(a_1)$ and contains the word $f(a_1 a_2 a_3)$ ($\mathbf{B}_{a_1, a_2, a_3, \ldots, a_k}$ begins with $a_1 a_2 a_1 a_3$ and $g(a_3) = a_1 a_2 a_3$). When $a_2 = \alpha$ and $a_3 = \beta$, denoting by a the first letter of $f(\alpha)$, we see that $a_1 a$ and αa are factors of $f(g(\mathbf{w}))$ showing that a is a left special factor of $f(g(\mathbf{w}))$, and so a prefix of $f(a_1)$. Consequently, a is the first letter of $f(a_1)$ whatever $a_1 \in A \setminus \{\alpha, \beta\}$. $\qquad\square$

Step 4. If a letter is left special in $f(A)$, then all images of letters begin with the same letter.

Proof. Here we assume that for a letter α, there exist distinct letters β and γ such that $\beta\alpha$ and $\gamma\alpha$ are both factors of images of letters. For any LSP word $\mathbf{w}$ containing two letters whose images contains $\beta\alpha$ and $\gamma\alpha$, these words are factors of $f(\mathbf{w})$. As f preserves the LSP property, $f(\mathbf{w})$ is LSP and so must begin with the letter α. As for any letter x, one can choose $\mathbf{w}$ in such a way that it begins with x, Thus all images of letters must begin with the same letter (the letter α). $\qquad\square$

Step 5. If there exists a letter a such that a letter occurs twice in $f(a)$, then all images of letters begin with the same letter.

Proof. Let α be a letter occurring twice in $f(a)$. Choose α such that its first occurrence is the leftmost as possible. Assume α is not the first letter of $f(a)$. If α is left special in $f(a)$, then the claim is clear by Step 4. So we can assume that all occurrences of α are preceded by the same letter β. Hence we should have chosen β instead of α: a contradiction. So α is the first letter of $f(a)$. Let b be a letter in A different from a. By Step 3 we only have to consider the case where $f(b)$ begins with a letter γ different from α.

There exists an infinite LSP word $\mathbf{w}$ beginning with a and containing the word aba (for instance the Fibonacci word). In $f(aba)$, only the first letter α of $f(\mathbf{w})$ can be left special. But there also exist infinite LSP words beginning with b and containing the word aba (exchange the roles of a and b in the Fibonacci word). As $f(b)$ does not begin with α, this letter α cannot be left special in $f(aba)$.

Let u be the word such that $|u|_\alpha = 0$ and $\alpha u \alpha$ is a prefix of $f(a)$. As no letter is left special in the factor $f(ab)\alpha u\alpha$ of $f(aba)$, each occurrence of α in $f(ab)\alpha u\alpha$ is preceded by αu: αu is a period of $f(ab)$. As αu is a prefix of $f(a)$, $f(ab)$ is a power of αu. For some words x and y and some integers k and ℓ with $k \geq 1$, $\ell \geq 0$, we have $u = x\gamma y$, $f(a) = (\alpha x \gamma y)^k \alpha x$, $f(b) = \gamma y (\alpha x \gamma y)^\ell$.

Observe that $f(ba) = (\gamma y \alpha x)^{k+\ell+1}$. Consider an LSP word $\mathbf{w}$ beginning with *babbabab* (for instance the word obtained from the Fibonacci word exchanging the roles of a and b). The word $f(\mathbf{w})$ contains $(f(ba))^2 = (\gamma y \alpha x)^{2(k+\ell+1)}$ and so by Lemma 8.7, $(\gamma y \alpha x)^{2k+2\ell+1}$ is a prefix of $f(\mathbf{w})$: in particular as $k \geq 1$, $(\gamma y \alpha x)^{\ell+1+k+\ell}\gamma y \alpha$ is a prefix of $f(\mathbf{w})$. But $f(\mathbf{w})$ begins with $f(babb)$ that begins with $(\gamma y \alpha x)^{\ell+1+k+\ell}\gamma y \gamma$. This contradicts the fact that $\alpha \neq \gamma$. $\qquad\square$

Step 6. If a letter occurs at least twice in $f(A)$, then all images of letters begin with the same letter.

Proof. Assume α occurs twice in $f(A)$. After Step 5, we can assume that α occurs in $f(a)$ and $f(b)$ for two distinct letters a and b. For any letter c (possibly a or b), there exists an LSP word $\mathbf{w}$ beginning with c and containing the word aba. Thus if α is left special in $f(aba)$, α is the first letter of $f(c)$ for all letters c.

Hence assume that α is not left special in $f(aba)$ nor in $f(bab)$. Thus all occurrences of α in $f(aba)$ and in $f(bab)$ are preceded by the same letter β. Possibly replacing iteratively α by β, we see that we can choose α as the first letter of $f(a)$ or of $f(b)$. Assume without loss of generality that it is the first letter of $f(a)$. By Step 3, we can assume that α is not the first letter of $f(b)$. As α is not left special in $f(ba)$ and as $f(b)$ ends with b (so $b\alpha$ is a factor of $f(ba)$), b has two occurrences in $f(b)$. By Step 5 all images of letters begin with the same letter. $\qquad\square$

Step 7. End of the proof of Proposition 8.9.

The proof acts by induction on $||f|| = \sum_{a \in A} |f(a)|$. When $||f|| = \#A$, by Step 2, f is a renaming morphism: it belongs to $\mathcal{U}$. Assume $||f|| > \#A$ and that f does not belong to $\mathcal{U}$. This means that a letter occurs at least twice in $f(A)$. By Step 6, all images of letters begin with the same letter. Let a denote this letter. By Step 2, for all distinct letters x and y in A, $\mathrm{last}(f(x)) \neq \mathrm{last}(f(y))$.

Observe that as f preserves the LSP property for infinite words, at most one letter can be left special in $f(A)$: this letter must be the letter a. Thus there exist an R-bLSP morphism g_1 and a morphism h, such that $f = g_1 \circ h$ and g_1 is defined over $\mathrm{alph}(h(A))$. Assume by contradiction that h does not preserve the LSP property for infinite words. This means there exists an infinite LSP word $\mathbf{w}$ such that $h(\mathbf{w})$ is not LSP. By Lemma 3.3, $f(\mathbf{w}) = g(h(\mathbf{w}))$ is neither LSP: a contradiction. Hence h preserves the LSP property for infinite words. By the definition of R-bLSP morphisms, $||g_1|| \geq \#A + 1$ and so $||h|| < ||f||$. By induction, $h \in S^*_{\mathrm{R-bLSP}}\mathcal{U}$: this also holds for f and Proposition 8.9 holds by induction. $\qquad\square$

8.3 Characterization of endomorphisms preserving LSP infinite words

Theorem 8.10. *The set of nonerasing endomorphisms over an alphabet A that preserve the LSP property for infinite words over A is $\mathcal{L}^* Perm \cup \mathcal{P}_{\mathrm{LSP}}$*

Proof. Permutation morphisms (elements of $Perm$) and the elements of $\mathcal{P}_{\mathrm{LSP}}$ preserve the LSP property for infinite words. By Lemma 8.2, the elements of $\mathcal{L}$ also preserve this property. Hence any element of $\mathcal{L}^* Perm \cup \mathcal{P}_{\mathrm{LSP}}$ preserves the LSP property for infinite words.

Conversely let f be a nonerasing endomorphism over A^* that preserves the LSP property for infinite words. Assume $f \notin \mathcal{P}_{\mathrm{LSP}}$. By Lemma 8.8 and the definition of $\mathcal{P}_{\mathrm{LSP}}$, for all letters a and b, the words $f(a)$ and $f(b)$ do not have the same primitive root. By Proposition 8.9, there exist an integer $n \geq 0$, n R-bLSP morphisms $g_1, g_2, \ldots, g_n$ and an element ν of $\mathcal{U}$ such that $f = g_1 \circ \cdots \circ g_n \circ \nu$.

Let $B = \mathrm{alph}(\nu(A))$ and define the alphabets $(A_i)_{i=n,n-1,\ldots 1}$ by $A_n = \mathrm{alph}(g_n(B))$ and for i such that $n-1 \geq i \geq 1$, $A_i = \mathrm{alph}(g_i(A_{i+1}))$. By Property 2.3(1), for any i in $\{n-1, \ldots, 1\}$, $A_{i+1} \subseteq A_i$ and $B \subseteq A_n$. Note that $A_1 = A$. Hence $B \subseteq A$. As ν belongs to $\mathcal{U}$, $\#B \geq \#A$. Hence $A = B$ and ν is a permutation morphism. It follows that f preserves the LSP property for infinite words if and only if $g_1 \circ \cdots \circ g_n$ preserves it. Thus from now on we assume that ν is the identity and $f = g_1 \circ \cdots \circ g_n$. From what precedes, we have $A_i = A$ for all i in $\{n, \ldots, 1\}$. Hence all morphisms g_i are endomorphisms. By definitions, the only R-bLSP morphisms that are endomorphisms are LSP morphisms.

To end we prove by induction that each endomorphism g_i belongs to $\mathcal{L}$. By Lemma 3.3, g_n must preserve the LSP property for infinite words (otherwise morphisms $g_i \circ \cdots \circ g_n$ $(1 \leq i \leq n)$ and so f would not preserve the LSP property for infinite words). By Lemma 8.2, it must belong to $\mathcal{L}$. Assume we have already proved that $g_{m+1}, \ldots, g_n \in \mathcal{L}$ for some integer m, $1 \leq m < n$. By Lemma 8.3 if $g_m \notin \mathcal{L}$, then g_m is (a, b, c)-breaking for some pairwise distinct letters a, b and c. Let $\mathbf{w}$ be an (a, b, c)-fragile LSP word. By Lemma 8.4, $g_{m+1} \circ \cdots \circ g_n(\mathbf{w})$ is (a, b, c)-fragile and so $g_m \circ \cdots \circ g_n(\mathbf{w})$ is not LSP. By Lemma 3.3, we deduce that $g_1 \circ \cdots \circ g_m \circ \cdots g_n(\mathbf{w}) = f(\mathbf{w})$ is not LSP: a contradiction with the fact that

f preserves the LSP property for infinite words. Then for all pairwise distinct letters a, b and c, g_m is not (a, b, c)-breaking: by Lemma 8.2 g_m belongs to $\mathcal{L}$. Hence by induction $g_i \in \mathcal{L}$ for all i, $1 \leq i \leq n$. $\qquad\square$

8.4   A particular word

To end the proof of Theorem 8.1, we need an LSP word defined on three letters that cannot be decomposed on two words. Let ξ_a (resp. ξ_b, ξ_c) be the fixed point of $f_a = \lambda_{abc} \circ \lambda_{bca} \circ \lambda_{cab} = [ababcababca,$ $ababcababcaab, ababc]$ (resp. $f_b = \lambda_{bca} \circ \lambda_{cab} \circ \lambda_{abc}$, $f_c = \lambda_{cab} \circ \lambda_{abc} \circ \lambda_{bca}$). Considering the permutation $\pi = [b, c, a]$, observe that $\pi \circ \lambda_{abc} = \lambda_{bca} \circ \pi$, $\pi \circ \lambda_{bca} = \lambda_{cab} \circ \pi$, and $\pi \circ \lambda_{cab} = \lambda_{abc} \circ \pi$. Hence $\xi_a = \pi(\xi_b)$ and $\xi_b = \pi(\xi_c)$ and $\xi_c = \pi(\xi_a)$.

From Theorem 7.1, ξ_a, ξ_b and ξ_c are LSP words as there exist infinite paths labeled by $(\lambda_{abc}\lambda_{bca}\lambda_{cab})^\omega$, $(\lambda_{bca}\lambda_{cab}\lambda_{abc})^\omega$ and $(\lambda_{cab}\lambda_{abc}\lambda_{bca})^\omega$ in the automaton $\mathcal{A}_{\mathrm{bLSP}}$.

Lemma 8.11. *There does not exist two words u and v such that ξ_a can be decomposed on $\{u, v\}$, that is, for any choice of the words u and v, $\xi \notin \{u, v\}^\omega$.*

Proof. First observe that ξ_a is not periodic (it has infinitely many left special factors). Then if X is a set of words such that $\xi_a \in X^\omega$, the cardinality of X is at least two. Assume by contradiction that there exist two words u and v such that $\xi_a \in u\{u, v\}^\omega$. Choose u and v such that $|uv|$ is minimal. Observe also that u begins with the letter a as it is a prefix of ξ_a. It is important to remark that for any word w beginning by a, there exists a word w' such that $w = \lambda_{abc}(w')$. We distinguish three cases depending on the first letter of v.

Case v begins with the letter a. Then $u = \lambda_{abc}(u')$ and $v = \lambda_{abc(v')}$ and $\xi_b \in \{u', v'\}^\omega$. As abc occurs in u or v, $|u'v'| \leq |uv| - 2$. As $\xi_a = \pi(\xi_b)$, we get $\xi_a \in \{\pi(u'), \pi(v')\}^\omega$ with $|\pi(u')\pi(v')| < |uv|$. This contradicts the choice of u and v.

Case v begins with the letter b. Let I be the set of integers such that $uv^i u$ occurs in ξ_a: $\xi_a \in \{uv^i \mid i \in I\}^\omega$. Note that $\#I \geq 2$ since ξ_a is not periodic. By the definition of λ_{abc}, since uv occurs in $\xi_a = \lambda_{abc}(\xi_b)$ and since v begins with b, u ends with the letter a. This is also the case for v if $I \neq \{0, 1\}$, that is, if vv occurs in ξ_a.

Assume that vv occurs in ξ_a (there exists $j \in I$ with $j \geq 2$). There exist words u_1, v_1 such that $u = \lambda_{abc}(u_1)a$ and $av = \lambda_{abc}(v_1)a$. As $\xi_a = \lambda_{abc}(\xi_b)$, $\xi_b \in \{u_1 v_1^i a \mid i \in I\}^\omega$. As aba and $abca$ must occur at least once each in uv, $|u_1 v_1| \leq |uv| - 4$. Both words u_1 and v_1 are not empty, otherwise $\xi_a \in \{\pi(u_1), b\}^\omega$ or $\xi_a \in \{\pi(v_1), b\}^\omega$: a contradiction with the choice of u and v.

Remember that $\xi_b = \lambda_{bca}(\xi_c)$. As $v_1 v_1 a$ occurs in ξ_b ($j \geq 2$ belongs to I), the word v_1 ends with bc. and consequently the word v_1 begins with a or b.

Assume that v_1 begins with a. Then $\xi_b \in \{u_1 a(v_1' a)^i \mid i \in I\}^\omega$ where v_1' is the word such that $v_1 = av_1'$. Hence $\xi_a = \pi(\xi_b) \in \{\pi(u_1 a)\pi(v_1' a)^i \mid i \in I\}^\omega$. As $|\pi(u_1 a)\pi(v_1' a)| = |u_1 a v_1| < |uv|$, this contradicts the choice of u and v.

Thus v_1 begins with b. We still have $\xi_b = \lambda_{bca}(\xi_c) \in \{u_1 v_1^i a \mid i \in I\}^\omega$. Recall that v_1 ends with bc. There exists a word v_2 such that $v_1 = \lambda_{bca}(v_2)bc$. If $v_2 = \varepsilon$, as $(bc)^3$ is not a factor of ξ_b, $I \subseteq \{0, 1, 2\}$. Also $0 \in I$ implies that u_1 ends with bc and so $(bc)^3 = bcv_1^2$ is a factor of ξ_b. Thus $I = \{1, 2\}$ and $\xi_b \in \{u_1 bca, u_1 bcbca\}^\omega$. As u_1 cannot ends with bc, it ends with b or bca. Then $bbca$ or $bcabca$ is a factor of ξ_b which implies that ba or aa is a factor of ξ_c. Both are impossible. Hence $v_2 \neq \varepsilon$. As $v_1^2 a = (\lambda_{bca}(v_2)bc)^2 a$ is a factor of ξ_b, the word $v_2 c v_2 a$ is a factor of $\xi_c = \lambda_{cab}(\xi_a)$. So v_2 must end with the letter c. As ccc is not a factor of ξ_c, v_2 does not begin with c: its first letter is a. This implies that each occurrence of v_2 in $\xi_c = \lambda_{cab}(\xi_a)$ is preceded by the letter c. Hence each occurrence of $\lambda_{bca}(v_2)$ in ξ_b is preceded by bc. It follows that the word u_1 must end with bc. This word u_1 also begins with b as it is a prefix of ξ_b. Then there exists a word u_2 such that $u_1 = \lambda_{bca}(u_2)bc$. We get $\xi_c \in \{u_2(cv_2)^i a \mid i \in I\}^\omega$. Recall that v_2 begins with a and ends with c. There exists a word v_3 such that $cv_2 = \lambda_{cab}(v_3)c$. The word u_2 is not empty. Indeed otherwise $\xi_a = \pi^2(\xi_c) \in \{\pi^2(cv_2), \pi^2(a)\}^\omega$ which contradicts the choice of u and v since $|cv_2| + |a| < |uv|$.

- If $0 \in I$, the factor u_2a is a factor of ξ_c and so the word u_2 ends with the letter c. As u_2 also begins with c as it is a prefix of ξ_c, there exists a word u_3 such that $u_2 = \lambda_{cab}(u_3)c$. We get $\xi_a \in \{u_3(cv_3)^i a \mid i \in I\}^\omega$.

- If $0 \notin I$, for a word u_3, $u_2 = \lambda_{cab}(u_3)$ and $\xi_a \in \{u_3(v_3c)^{i-1}v_3a \mid i \in I\}^\omega$.

Assume $v_3 = \varepsilon$. Observe that $0 \notin I$ since I contains an integer greater than or equals to 2 and cc is not a factor of ξ_a Similarly $I = \{1, 2\}$. In this case, $\xi_a \in \{u_3a, u_3ca\}^\omega$. As u_3c is a factor of $\xi_a = \lambda_{abc}(\xi_b)$, the word u_3 must end with ab which implies that u_3a ends with aba. As a is the first letter of u_3, $abaa$ is a factor of ξ_a: a contradiction. Thus $v_3 \neq \varepsilon$. In both cases $0 \in I$ and $0 \notin I$, the factor u_3v_3c of ξ_a ends with abc. As cv_3a is a factor of ξ_a, $v_3 \neq b$. Then v_3 ends with ab. Once again as u_3 begins with a, we get the factor $abaa$ of ξ_a: a contradiction that ends the study of the case "vv occurs in ξ_a".

To continue the study of the case "v begins with b", we have to study case $I = \{0, 1\}$. The word u begins with a and so there exist words u_1 and v_1 such that $u = \lambda_{abc}(u_1)a$, and either $v = b\lambda_{abc}(v_1)$ or $v = bc\lambda_{abc}(v_1)$. It follows that $\xi_b \in \{u_1a, u_1bv_1\}^\omega$ or $\xi_b \in \{u_1a, u_1cv_1\}^\omega$. As $\xi_b = \lambda_{bca}(\xi_c)$, the word u_1 ends with bc. As cc is not a factor of ξ_b, necessarily $\xi_b \in \{u_1a, u_1bv_1\}^\omega$ holds. Consequently there exist words u_2 and v_2 such that $u_1 = \lambda_{bca}(u_2)bc$ and $bv_1 = \lambda_{bca}(v_2)$: $\xi_c \in \{u_2a, u_2cv_2\}^\omega$. Necessarily u_2 begins with c, the first letter of $\xi_c = \lambda_{cab}(\xi_a)$. From the factor u_2a we deduce that u_2 ends also with c: $u_2 = \lambda_{cab}(u_3)c$ and $cv_2 \in \lambda_{cab}(v_3)$. It follows that $\xi_a \in \{u_3a, u_3cv_3\}^\omega$. From the factor u_3c we deduce that u_3 ends with ab. Also u_3 must begin with a, the first letter of ξ_a. Once again we find a factor $abaa$ in ξ_a: a contradiction. This ends the study of the case "v begins with b".

Case v begins with the letter c. Let I be defined as in the previous case: $\xi_a \in \{uv^i \mid i \in I\}^\omega$. Assume first that vv occurs in $\xi_a = \lambda_{abc}(\xi_b)$. Then both words u and v ends with ab. There exist words u_1 and v_1 such that $u = \lambda_{abc}(u_1)ab$ and $v = c\lambda_{abc}(v_1)ab$. Then $\xi_b \in \{u_1(cv_1)^i b \mid i \in I\}^\omega$. As cv_1c is a factor of $\xi_b = \lambda_{bca}(\xi_c)$, the word v_1 is not empty and ends with b. The word u_1c is a prefix of ξ_b and so u_1 begins with b. As v_1bu_1 is a factor of ξ_b, bbb also occurs in ξ_b: a contradiction. Hence vv does not occur in $\xi_a = \lambda_{abc}$. There exist words u_1 and v_1 such that $u = \lambda_{abc}(u_1)ab$ and $v = c\lambda_{abc}(v_1)$. It follows that $\xi_b \in \{u_1b, u_1cv_1\}$. If u is empty, we find a contradiction with the choice of u and v as $|bcv_1| < |uv|$. If u is not empty, it begins with b. The existence of the factor u_1c shows that u_1 also ends with b. As u_1bu_1 is a factor of ξ_b, we deduce that bbb is a factor of ξ_b: a final contradiction. $\qquad\square$

8.5 Proof of Theorem 8.1

Assume by contradiction that there exists a set S of morphisms such that the set of LSP infinite words is the set of S-adic words. Then all morphisms in S preserve the LSP property for infinite words over A.

Let ξ_a be the word studied in Section 8.4. Recall that ξ_a is LSP. So it is S-adic. By Lemma 8.11 ξ_a cannot be decomposed on two words. It follows that any element f_n occurring in a directive word $(f_n)_{n \geq 1} \in S^\omega$ must be a morphism defined on at least three letters and cannot belong to $\mathcal{P}_{\text{LSP}}$. By Proposition 8.6, $f_n \in S^*_{\text{R}-\text{bLSP}}\mathcal{U}$. If f_n is defined from the alphabet A_n to the alphabet B_n, then $\#A_n \leq \#B_n$. As $\#B_1 = 3$, by induction one can see that $\#A_n = \#B_n = 3$ for all $n \geq 1$. This implies that $f_n = g_n \circ \pi_n$ with $g_n \in S^*_{\text{R}-\text{bLSP}}$ and π_n a renaming morphism. As g_n is an endomorphism over B_n, $g_n \in S^*_{\text{bLSP}}$.

The fact that f_n preserves the LSP property for infinite words (as all elements of S) implies that g_n also preserves the LSP property. Hence $g_n \in \mathcal{L}^*Perm$ by Theorem 8.10 (ξ_a is not periodic so g_n cannot belong to $\mathcal{P}_{\text{LSP}}$).

From what precedes $\xi_a = L_\alpha(\mathbf{w})$ for a letter α and an infinite word $\mathbf{w}$. Indeed there must exist an integer k such that $g_k \in \mathcal{L}^+Perm$ and $g_i \in Perm$ for all $1 \leq i \leq k-1$. As one can see that $PermPerm \subseteq Perm$ and $Perm\mathcal{L} \subseteq \mathcal{L}Perm$, relation $\xi_a = L_\alpha(\mathbf{w})$ follows. This is a contradiction with the fact that the word $abca$ is a factor of ξ.

9 Final remarks

In [10] G. Fici asked for a characterization of both finite and infinite words. As explained in [19], any finite LSP word can be extended to a longer LSP word and so: A finite word is LSP if and only if it is a prefix of an infinite LSP word. And thus any characterization of infinite LSP words provides naturally a characterization of finite LSP words (adding "is a prefix of" before the characterization of infinite LSP words).

A natural open question comes from the content of this paper. Does there exist an S-adic characterization of infinite words having at most one left special factor of each length (but not necessarily as a prefix)? Another question comes after Lemma 8.11. The proof of this result is rather technical while the result itself seems to be extendable. Let $\mathbf{w}$ be an infinite LSP word such that for all prefixes p, there exist at least three letters a, b and c such that ap, bp and cp are factors of $\mathbf{w}$. Is it true that $\mathbf{w}$ cannot be decomposed on two words? More generally let $\mathbf{w}$ be an infinite word having infinitely many factors that have at least k left extensions (factors u such that there exist at least k distinct letters a_1, ..., a_k with $a_i u$ factor of $\mathbf{w}$ for all i, $1 \leq i \leq k$). Is it true that $\mathbf{w}$ cannot be decomposed on a set of $k - 1$ or less words?

References

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

[2] V. Berthé. S-adic expansions related to continued fractions. In S. Akiyama, editor, *Natural extension of arithmetic algorithms and S-adic system*, volume B58 of *RIMS Kôkyûroku Bessatsu*, pages 61–84, 2016.

[3] 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.

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

[5] V. Berthé and S. Labbé. Factor complexity of S-adic words generated by the Arnoux–Rauzy–Poincaré algorithm. *Adv. App. Math.*, 63(90 - 130), 2015.

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

[7] J. Cassaigne. Special factors of sequences with linear subword complexity. In J. Dassow, G. Rozenberg, and A. Salomaa, editors, *Developments in Language Theory II*, pages 25–34. World Scientific, 1996.

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

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

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

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

[12] J. Leroy. *Contribution à la résolution de la conjecture S-adique.* Doctoral thesis, Université de Picardie Jules Verne, 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):233–286, 2014.

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

[15] 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.

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

[17] 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).

[18] G. Richomme. Conjugacy and episturmian morphisms. *Theoret. Comput. Sci.*, 302:1–34, 2003.

[19] G. Richomme. A characterization of infinite LSP words. In Émilie Charlier, Julien Leroy, and Michel Rigo, editors, *Developments in Language Theory - 21st International Conference, DLT 2017, Liège, Belgium, August 7-11, 2017, Proceedings*, volume 10396 of *Lect. Notes Comput. Sci.*, pages 320–331. Springer, 2017.

[20] G. Rozenberg and A. Salomaa, editors. *Handbook of Formal Languages.* Springer, 1997.

[21] M. Sciortino and L. Q. Zamboni. Suffix automata and standard Sturmian words. In T. Harju, J. Karhumäki, and A. Lepistö, editors, *Developments in Language Theory: 11th International Conference, DLT 2007, Turku, Finland, July 3-6, 2007. Proceedings*, volume 4588 of *Lect. Notes Comput. Sci.*, pages 382–398. Springer Berlin Heidelberg, 2007.