A Characterization of Infinite LSP Words[⋆]

Gwenaël Richomme

Univ. Paul-Valéry Montpellier 3, UFR 6, Dpt MIAp, Case J11,
Rte de Mende, 34199 Montpellier Cedex 5, France
and
LIRMM (CNRS, Univ. Montpellier), UMR 5506 - CC 477,
161 rue Ada, 34095 Montpellier Cedex 5, France
`gwenael.richomme@lirmm.fr`

Abstract. G. Fici proved that a finite word has a minimal suffix automaton if and only if all its left special factors occur as prefixes. He called LSP all finite and infinite words having this latter property. We characterize here infinite LSP words in terms of S-adicity. 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 all its directive words are recognizable by $\mathcal{A}$.

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

1 Introduction

Extending an initial work by M. Sciortino and L.Q. Zamboni [15], G. Fici investigated relations between the structure of the suffix automaton built from a finite word w and the combinatorics of this word [8]. He proved that 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. We provide such a characterization for infinite words in the context of S-adicity.

We assume that readers are familiar with combinatorics on words; for omitted definitions (as for instance, factor, prefix, ...) see, *e.g.*, [13,14,6]. Given an alphabet A, A^* is the set of all finite words over A and A^ω is the set of all infinite words over A. 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 au and bu occur in w. 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 extend naturally to infinite words.

Let S be a set of morphisms. An infinite word $\mathbf{w}$ is said S-adic if there exists a sequence $(f_n)_{n \geq 1}$ of morphisms in S and a sequence of letters $(a_n)_{n \geq 1}$ such that $\lim_{n \to +\infty} |f_1 f_2 \cdots f_n(a_{n+1})| = +\infty$ and $\mathbf{w} = \lim_{n \to +\infty} f_1 f_2 \cdots f_n(a_{n+1}^\omega)$. The sequence $(f_n)_{n \geq 1}$ 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 $(f_n)_{n \geq 1}$ if there exists an infinite sequence of infinite words $(\mathbf{w}_n)_{n \geq 1}$ such that $\mathbf{w}_1 = \mathbf{w}$ and $\mathbf{w}_n = f_n(\mathbf{w}_{n+1})$ for all $n \geq 1$. Denoting $w_k = f_k f_{k+1} \cdots f_n(a_{n+1}^\omega)$ shows that 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 morphism mapping a on a. For more information on S-adic systems, readers can consult, *e.g.*, papers [2,3] and their references.

Let $p_\mathbf{w}$ be the *factor complexity* of the infinite word $\mathbf{w}$, that is the function that counts the number of different factors of $\mathbf{w}$. If $\mathbf{w}$ is an infinite LSP word, by definition, it has at most one

---

left special factor of each length. Thus it is well-known that $p_{\mathbf{w}}(n+1) - p_{\mathbf{w}}(n) \leq \#A - 1$ (where for any set X, $\#X$ denotes the cardinality of X). We let readers verify that all infinite LSP words are uniformly recurrent (all factors occur infinitely many times with bounded gaps). By a result of S. Ferenczi [7] (see also [9,10,11]), there exists a finite set S of morphisms such that all infinite LSP words are S-adic. But this general result does not provide a characterization of infinite LSP words.

Our characterization is twofold. First we exhibit an adapted finite set of morphisms S_{bLSP}. Second we show that there exists an automaton that recognizes the set of directive words of infinite LSP words. In the binary case, our result can be seen as a version for infinite words of a result of M. Sciortino and L.Q. Zamboni [15] (see the conclusion). In the ternary case, morphisms in S_{bLSP} are the mirror morphisms of Arnoux-Rauzy-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). These morphisms were used by V. Berthé and S. Labbé [5] to provide an S-adic system recognizing sequences arising from the study of the Arnoux-Rauzy-Poincaré multidimensional continued fraction algorithm. For alphabets of cardinality at least 4, new morphisms appear.

The paper is organized as follows. After introducing in Section 2 our basis of morphisms S_{bLSP}, in Section 3, we show that all infinite LSP words are S_{bLSP}. Section 4 introduces a property of infinite LSP words and a property of morphisms in S_{bLSP} that allow to explain why the LSP property is lost when applying a 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. We end with a few words on characterizations of finite LSP words.

2 Some Basic Morphisms

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

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

We let $S_{\mathrm{bLSP}}(A)$ (or shortly S_{bLSP} when A is clear) denote the set of all bLSP morphisms over the alphabet A. Observe that for any bLSP morphism f, there exists a unique letter α such that $f(\alpha) = \alpha$. We let $\mathrm{first}(f)$ denote this letter as it is also the first letter of $f(\beta)$ for any letter β. We also let $[u_1, u_2, \dots]$ denote the morphism defined by $a \mapsto u_1$, $b \mapsto u_2$, $\dots$. For instance, $[a, ab, abc, abcd, abcde]$ defines the morphism f such that $f(a) = a$, $f(b) = ab$, $f(c) = abc$, $f(d) = abcd$, $f(e) = abcde$.

Remark 1. By definition of bLSP morphisms, given an alphabet A, there is a bijection between $S_{\mathrm{bLSP}}(A)$ and the set of labeled rooted trees with label in A (all labels are on vertices and distinct vertices have distinct labels). Given a labeled rooted tree $T = (A, E)$, the associated bLSP morphism f is the one such that, for all letters β, $f(\beta)$ is the word obtained concatenating vertices on the path in T from the root to β. For instance, the rooted trees associated with morphisms $[a, ab, abc, abcd]$, $[a, ab, abc, abd]$, $[a, ab, abc, ad]$ and $[a, ab, ac, ad]$ are given in Figure 1.

The previous remark allows to enumerate bLSP morphisms (see Sequence A000169 in The On-Line Encyclopedia of Integer Sequences whose first values are 1, 2, 9, 64, 625, 7776, 117649, 2097152).

Here follows some examples of bLSP morphisms.
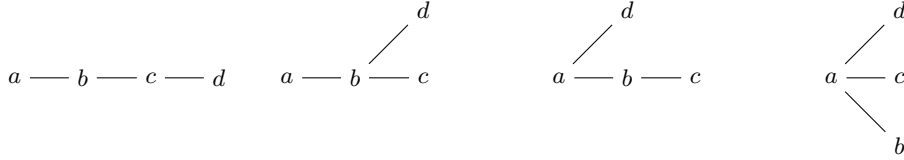
Fig. 1. Rooted trees associated with bLSP morphisms

- $S_{\mathrm{bLSP}}(\{a,b\}) = \{[a,ab],[ba,b]\}$. 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 non-periodic $\{\tau_a, \tau_b\}$-adic words (see also [12]).
- $S_{\mathrm{bLSP}}(\{a,b,c\}) = \{[a,ab,abc],[a,ab,ac],[a,acb,ac],[ba,b,bac],[ba,b,bc],[bca,b,bc],[ca,cb,c],$ $[ca,cab,c],[cba,ca,c]\} = \{p^{-1}\circ[a,ab,abc]\circ p, p^{-1}\circ[a,ab,ac]\circ p \mid p \in perm(A)\}$ where $perm(A)$ is the set of all endomorphisms of A^* whose restriction to the set of letters is a permutation of the alphabet. As mentioned in the introduction, these sets $S_{\mathrm{bLSP}}(\{a,b,c\})$ is also the set of mirror morphisms considered in [5], that is mirrors of the Poincaré substitutions (defined for $\{i,j,k\} = \{a,b,c\}$ by $i \mapsto ijk$, $j \mapsto jk$, $k \mapsto k$) and the Arnoux-Rauzy substitutions (defined for $\{i,j,k\} = \{a,b,c\}$ by $i \mapsto ik$, $j \mapsto jk$, $k \mapsto k$).
- The set $S_{\mathrm{bLSP}}(\{a,b,c,d\})$ is the set of all morphisms on the form $p^{-1}\circ f \circ p$ for $p \in perm(A)$, and f being one of the following morphisms: $[a,ab,abc,abcd]$, $[a,ab,abc,abd]$, $[a,ab,abc,ad]$ and $[a,ab,ac,ad]$.

We end this section with some basic properties of bLSP morphisms that follow directly from the definition. 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.

Property 2. Let f be a bLSP morphism over the alphabet A.

1. there exists a unique letter $\alpha \in A$ such that for all $\beta \in A$, $\mathrm{first}(f(\beta)) = \alpha$;
2. for all $\beta \in A$, $\mathrm{last}(f(\beta)) = \beta$;
3. there exists a unique letter $\alpha \in A$ such that $f(\alpha) = \alpha$: $\alpha = \mathrm{first}(f)$;
4. $f(A)$ is a suffix code (no word of $f(A)$ is a suffix of another word in $f(A)$);
5. f is injective both on the set of finite words and the set of infinite words;
6. 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 $x = y$;
7. for all letters β, γ, $|f(\beta)|_\gamma \leq 1$.

3 S_{bLSP}-Adicity of Infinite LSP Words

Proposition 3. *Any infinite LSP word is S_{bLSP}-adic.*

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 that:

1. there exists $f \in S$ 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.

Hence Proposition 3 is a direct consequence of the next two lemmas.

Lemma 4. *Given any finite or infinite LSP word $\mathbf{w}$, there exist a bLSP morphism f on $\mathrm{alph}(\mathbf{w})$ and an infinite word $\mathbf{w}'$ such that $\mathbf{w} = f(\mathbf{w}')$.*

Proof. Let **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 $\mathrm{alph}(w) \setminus \{\alpha\}$ such that w can be factorized over $\{\alpha\} \cup X$. Let G be the graph $(\mathrm{alph}(w), E)$ 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 **w**, each letter occurring in a word of X is not left special in **w**. Hence G is a rooted tree with α as root, that is, for any letter β, there exists a unique path from α to β. We let u_β denote the word obtained by concatenating the letters occurring in the path. Let f be the morphism defined by $f(\beta) = u_\beta$. By construction, f is bLSP and $\mathbf{w} = f(\mathbf{w}')$ for a word $\mathbf{w}'$. □

Remark 5. The word $\mathbf{w}'$ in Lemma 4 is unique. The morphism f is not unique but its restriction to $alph(\mathbf{w}')$ is. It can also be observed that this restriction is entirely defined by the first letter of **w** and the factors of length two of **w**.

Lemma 6. *For any 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 **w** is not LSP. This means that **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 **w**, βub and γub are factors of **w**. Recall that f is a bLSP morphism: let $\alpha = \mathrm{first}(f)$. The word $f(u)f(a)\alpha$ is a prefix of $f(\mathbf{w})$. Moreover by Property 2(2), words $\beta f(u)f(b)\alpha$ and $\gamma f(u)f(b)\alpha$ are factors of **w** (here the fact that **w** is infinite is needed: 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 α', β' such that $\alpha' \neq \beta'$, $v\alpha'$ and $v\beta'$ are respectively prefixes of $f(a)\alpha$ and $f(b)\alpha$. It follows that $f(u)v\alpha'$ is a prefix of $f(\mathbf{w})$ while $\beta f(u)v\beta'$ and $\gamma f(u)v\beta'$ are factors of $f(\mathbf{w})$: $f(\mathbf{w})$ is not LSP. □

Observe that Lemma 6 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.

To end this section let us mention that in the binary case the converses of Lemma 6 and Proposition 3 hold.

Proposition 7. *If **w** is a binary LSP infinite word and if f belongs to the set $\{[a, ab], [ba, b]\}$ then $f(\mathbf{w})$ is also LSP. Consequently a binary word is LSP if and only if it is $\{[a, ab], [ba, b]\}$-adic.*

Proof. The second part of the proposition is a direct consequence of the first part and of Lemma 4. We prove the first part for $f = [a, ab]$ (exchanging the roles of a and b, the proof is similar for $f = [ba, b]$). Assume **w** is an infinite LSP word over $\{a, b\}$ such that $f(\mathbf{w})$ is not LSP. There exists a left special factor v of $f(\mathbf{w})$ which is not a prefix of $f(\mathbf{w})$. Choose v of minimal length. Hence $v = v'\beta$ and $v'\alpha$ is a prefix of $f(\mathbf{w})$ for a word v' and two letters α and β such that $\{\alpha, \beta\} = \{a, b\}$. As the letter b is always preceded in $f(\mathbf{w}')$ by the letter a, the word v' is a non-empty word ending with a: hence $v' = f(u)a$ for some word u. As $f(u)a\alpha$ is a prefix of $f(\mathbf{w})$, by the structure of f, $u\alpha$ is a prefix of **w**. Also as $f(u)a\beta$ is a left special factor of $f(\mathbf{w})$, as $\mathrm{last}(f(a)) = a$ and $\mathrm{last}(f(b)) = b$, the word $u\beta$ is also a left special factor of **w**. As $\alpha \neq \beta$, this contradicts the fact that **w** is LSP. □

4 Fragility of Infinite LSP Words

For alphabets of cardinality at least 3, the converse of Lemma 6 is false: there exist an infinite LSP word **w** and a bLSP morphism f such that $f(\mathbf{w})$ is not LSP. For instance, let **F** be the

well-known Fibonacci word (the fixed point of the endomorphism $[ab, a]$), and let g be the bLSP morphism $[a, acb, ac]$. 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 10 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) bLSP morphism can map a (fragile) infinite LSP word on a non-LSP word.

Definition 8. 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 α and β such that the word ua is a prefix of $\mathbf{w}$ and the words αub and βuc are factors of $\mathbf{w}$. We will also say that $\mathbf{w}$ is (a, b, c, α, β)-fragile when we need letters α and β. The word u is also called 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 starting with the letter a (and with $a \neq b \neq c \neq a$) 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 distinct 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 8, the word $g(u)aa$ is a prefix of $g(\mathbf{w})$ whereas 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 9. 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 infinite LSP word $\mathbf{w}$, $f(\mathbf{w})$ is not LSP.

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

Lemma 10. *Let $\mathbf{w}$ be an infinite LSP word and let f be a bLSP morphism. 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, 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 Property 2, the letter $\alpha = \mathrm{first}(f)$ is the unique letter that can be left special in $f(\mathbf{w})$. This implies $v \neq \varepsilon$ and $\mathrm{first}(v) = \mathrm{first}(f)$. 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, $va' = f(u)\alpha x a'$ 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', α', β' be the word and letters such that $u'a$ is a prefix of $\mathbf{w}'$ while $\alpha'u'b$ and $\beta'u'c$ are factors of $\mathbf{w}'$ with $\alpha' \neq \beta'$. The word $f(\mathbf{w}')$ has $f(u')v\delta$ as a prefix and words $\alpha'f(u')v\gamma$ and $\beta'f(u')v\gamma$ 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. □

Observe that we have also proved the next result.

Corollary 11. *A 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 $f(b)$.*

5 Origin of Fragilities

Before characterizing infinite LSP words, we need to know how fragilities in an LSP word can appear. This is explained by next result. For a set X of words, we let $\mathrm{Fact}(X)$ denote the set of factors of words in X.

Lemma 12. *Assume a word u is an (a, b, c, β, γ)-fragility of $f(\mathbf{w})$ for an infinite word $\mathbf{w}$ (not necessarily LSP) over an alphabet A and f is a bLSP morphism (by definition of fragilities, a, b, c, β, γ are letters).*

– *(New fragilities) If $u = \varepsilon$, then $a = \mathrm{first}(f)$ and βb, $\gamma c \in \mathrm{Fact}(f(\mathrm{alph}(\mathbf{w})))$.*
– *(Propagated fragilities) If $u \neq \varepsilon$, there exist letters a', b', c' in $\mathrm{alph}(\mathbf{w})$ and an $(a', b', c', \beta, \gamma)$-fragility v of $\mathbf{w}$ such that $|v| < |u|$, $f(v)$ is a proper prefix of u and words ua, βub, γuc are respectively prefixes of $f(va')\alpha$, $\beta f(vb')\alpha$, $\gamma f(vc')\alpha$ with $\alpha = \mathrm{first}(f)$.*

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

(Propagated fragilities) We assume here 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:

– $|w_1|_\alpha = |w_2|_\alpha = |w_3|_\alpha = 0$;
– $f(v)\alpha xw_1\alpha$ is a prefix of $f(\mathbf{w})$ and $a = \mathrm{first}(w_1\alpha)$;
– $\beta f(v)\alpha xw_2\alpha$ and $\gamma f(v)\alpha xw_3\alpha$ are factors of $f(\mathbf{w})$ with $b = \mathrm{first}(w_2\alpha)$ and $c = \mathrm{first}(w_3\alpha)$.

By definition of a 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 $a = \text{first}(w_1\alpha)$, $b = \text{first}(w_2\alpha)$ and $c = \text{first}(w_3\alpha)$ 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 a bLSP morphism). Hence the word v is an $(a', b', c', \beta, \gamma)$-fragility of $\mathbf{w}$. Finally let us observe that $|v| \leq |f(v)| < |u|$. □

6 An Automaton to Follow Fragilities

In this section, we introduce an automaton that allows to recognize all directive words of LSP words viewed as S_{bLSP}-adic words. We will prove the converse in next section. Observe that transitions of the automaton are defined in order to follow fragilities using Lemma 12.

Definition 13. We let $\mathcal{A}_{bLSP}$ denote the non-deterministic automaton whose elements are described below.

- The alphabet of $\mathcal{A}_{bLSP}$ is the set bLSP of basic LSP morphisms.
- The set of states Q is the set $2^A \times \text{bLSP} \times 2^{A^5}$. Hence a state is the data of a sub-alphabet of A, of a bLSP morphism and of a set of 5-tuples (a, b, c, β, γ) of letters whose aim is to represent the set of fragilities of a word. For a state q, we let $\text{alph}(q)$ denote the sub-alphabet of A, by $\text{bLSP}(q)$ the morphism and by $\text{set}(q)$ the set of 5-tuples.
- The set of transitions Δ is the set of triples (q, f, q') such that
  1. $f = \text{bLSP}(q)$;
  2. $\text{alph}(q) = \text{alph}(f(\text{alph}(q')))$;
  3. if $(a, b, c, \beta, \gamma) \in \text{set}(q')$ then f is not LSP (a, b, c)-breaking;
  4. $\text{set}(q)$ is the set of all 5-tuples (a, b, c, β, γ) such that a, b, c, β, γ are letters of $\text{alph}(q)$, $a \neq b \neq c \neq a$, $\beta \neq \gamma$ and one of the following two conditions holds:
     (a) $a = \text{first}(f)$, βb, γc in $\text{Fact}(f(\text{alph}(q')))$ and $\beta \neq \gamma$;
     (b) there exist a', b', c' such that $(a', b', c', \beta, \gamma) \in \text{set}(q')$ and a word x such that $xa \in \text{pref}(f(a')\alpha)$, $xb \in \text{pref}(f(b')\alpha)$ and $xc \in \text{pref}(f(c')\alpha)$ with $\alpha = \text{first}(f)$.
- All states are initial.

Figure 2 shows this automaton when the alphabet is $\{a, b\}$. In this figure, $\tau_a = [a, ab]$ and $\tau_b = [ba, b]$. States q with $set(q) \neq \emptyset$ are not drawn since binary infinite LSP words contain no fragilities. Moreover states $(\emptyset, \tau_a, \emptyset)$ and $(\emptyset, \tau_b, \emptyset)$, $(\{a\}, \tau_b, \emptyset)$ and $(\{b\}, \tau_a, \emptyset)$ are not drawn as there are no transition leaving them.
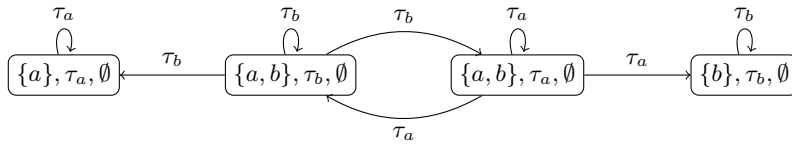
Fig. 2. $\mathcal{A}_{\text{bLSP}}$ for the binary alphabet

For alphabets with at least three letters, automaton $\mathcal{A}_{bLSP}$ is too huge to be drawn even restricting to states q such that $set(q)$ is a set of fragilities of an LSP word.

An infinite word $\mathbf{f}$ over bLSP is said to be *recognized* by $\mathcal{A}_{bLSP}$ if there exists an infinite path in $\mathcal{A}_{bLSP}$ whose label is $\mathbf{f}$. The aim of $\mathcal{A}_{bLSP}$ is to recognize bLSP directive words of infinite LSP words.

Let $\mathbf{w}$ be an LSP word. We associate with $\mathbf{w}$ a state of $\mathcal{A}_{bLSP}$ that we let denote $q(\mathbf{w})$. This state is the state q such that:

- $\mathrm{alph}(q) = \mathrm{alph}(\mathbf{w})$:
- $\mathrm{bLSP}(q)$ is any morphism f such that $\mathbf{w} = f(\mathbf{w}')$ for some word $\mathbf{w}'$ (such a morphism exists by Lemma 4).
- $\mathrm{set}(q)$ is the set of all 5-tuples (a, b, c, β, γ) such that $\mathbf{w}$ is (a, b, c, β, γ)-fragile.

The fact that, for any LSP word $\mathbf{w}$, any of its directive word is recognized by $\mathbf{w}$, is a direct consequence of next lemma.

Lemma 14. *Let $\mathbf{w}$, $\mathbf{w}'$ be LSP words such that $\mathbf{w} = f(\mathbf{w}')$ with $f = \mathrm{bLSP}(q(\mathbf{w}))$. The transition $(q(\mathbf{w}), \mathrm{bLSP}(q(\mathbf{w})), q(\mathbf{w}'))$ exists in $\mathcal{A}_{\mathrm{bLSP}}$.*

Proof. Let $f = \mathrm{bLSP}(q(\mathbf{w}))$. Observe that $\mathrm{alph}(q(\mathbf{w})) = \mathrm{alph}(\mathbf{w})$, $\mathbf{w} = f(\mathbf{w}')$ and $\mathrm{alph}(q(\mathbf{w}')) = \mathrm{alph}(\mathbf{w}')$. Whence we have $\mathrm{alph}(q(\mathbf{w})) = \mathrm{alph}(f(\mathrm{alph}(q(\mathbf{w}'))))$.

By Lemma 12 and the definition of $q(\mathbf{w})$, a 5-tuple (a, b, c, β, γ) belongs to set $q(\mathbf{w})$ if and only if one of the following two conditions holds:

- $a = \mathrm{first}(f)$, βb, γc belong to $\mathrm{Fact}(f(\mathrm{alph}(\mathbf{w}')))$
- there exist $a', b', c', \beta, \gamma$ in $\mathrm{alph}(\mathbf{w}')$ and an (a, b, c, β, γ)-fragility u of $\mathbf{w}$ and an $(a', b', c', \beta, \gamma)$-fragility v of $\mathbf{w}'$ such that $|v| < |u|$, $f(v)$ is a proper prefix of u and words ua, βub, γuc are respectively prefixes of $f(va')\alpha$, $\beta f(vb')\alpha$, $\gamma f(vc')\alpha$ with $\alpha = \mathrm{first}(f)$.

For the second case, $u = f(v)x$ for a word x. The word xa is a prefix of $f(a')\alpha$, xb is a prefix of $f(b')\alpha$ and xc is a prefix of $f(c')\alpha$. As $\mathbf{w}'$ is $(a', b', c', \beta, \gamma)$-fragile, $(a', b', c', \beta, \gamma) \in \mathrm{set}(q(\mathbf{w}'))$. Thus in both cases, Condition 4 for $(q(\mathbf{w}), \mathrm{bLSP}(q(\mathbf{w}), q(\mathbf{w}'))$ to be a transition of $\mathcal{A}_{\mathrm{bLSP}}$ is verified.

To end the proof we have to check Property 3 of transitions of $\mathcal{A}_{\mathrm{bLSP}}$. Assume there exists an (a, b, c, β, γ)-fragility in $\mathrm{set}(q(\mathbf{w}'))$. By definition of $q(\mathbf{w}')$, this implies that $\mathbf{w}'$ has an (a, b, c, β, γ)-fragility. As $\mathbf{w} = f(\mathbf{w}')$ is LSP, $f = \mathrm{bLSP}(q(\mathbf{w}))$ is not LSP (a, b, c)-breaking. □

7 A Characterization of LSP Words

Theorem 15. *A word $\mathbf{w}$ is LSP if and only if it is S_{bLSP}-adic and all of its directive word are recognized by the automaton $\mathcal{A}_{\mathrm{bLSP}}$.*

Proof. Proposition 3 and Lemma 14 prove the only if part of Theorem 15. Let us prove the if part of Theorem 15.

Assume, by contradiction, that $\mathcal{A}_{\mathrm{bLSP}}$ recognizes a directive word $\mathbf{f}$ of a word $\mathbf{w}$ which is S_{bLSP}-adic but not LSP. Such a word contains a left special factor u that is not a prefix of $\mathbf{w}$. Among all possible triples $(\mathbf{f}, \mathbf{w}, u)$, choose one such that $|u|$ is minimal.

For $n \geq 1$, we let f_n denote 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 directed by $f_2 f_3 \cdots$; $\mathbf{w}_n = f_n(\mathbf{w}_{n+1})$ for $n \geq 1$).

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

First observe $|u| \geq 2$. Indeed we have $|u| \neq 0$ as the empty word is a prefix of $\mathbf{w}$. Moreover, by the structure of images of the 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 choice of the triple $(\mathbf{f}, \mathbf{w}, u)$, 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 $\mathbf{w}$ and u is a prefix of $f(vb)$. Assume also that b is the unique letter c such that γvc is a factor of $\mathbf{w}$ 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 choice of the triple $(\mathbf{f}, \mathbf{w}, u)$, $|vb| \geq |u|$. 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 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 11, f_1 is (a, b, c)-breaking.

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

Applying iteratively Lemma 12, we deduce the existence of an integer $n \geq 2$, a sequence of triples of pairwise distinct letters $(a_i, b_i, c_i)_{i \in \{2, \cdots, n\}}$, a sequence of $(a_i, b_i, c_i, \beta, \gamma)$-fragilities $(v_i)_{i \in \{2, \cdots, n\}}$ such that:

- v_i occurs in $\mathbf{w}_i$ 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 $i \in \{2, \cdots, n-1\}$;
- $v_n = \varepsilon$.
- words $v_i a_i$, $\beta v_i b_i$, $\gamma v_i c_i$ are respectively prefixes of the words $f_i(v_{i+1}a_{i+1})\alpha_i$, $\beta f_i(v_{i+1}b_{i+1})\alpha_i$, $\gamma f_i(v_{i+1}c_{i+1})\alpha_i$ where $\alpha_i = \mathrm{first}(f_i)$ for $i \in \{2, \cdots, n-1\}$;
- $a_n = \mathrm{first}(f_n)$;
- βb_n, γc_n belong to $\mathrm{Fact}(f_n(\mathrm{alph}(\mathbf{w}_{n+1})))$.

Step 4: Conclusion Let $(q_i)_{i \geq 1}$ be the sequence of states along a path recognizing $\mathbf{f}$: for all $n \geq 1$, (q_n, f_n, q_{n+1}) is a transition of $\mathcal{A}_{\mathrm{bLSP}}$.

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 $a = \mathrm{first}(f_n) \notin \{b_n, c_n\}$ by properties of bLSP morphisms, the words βb_n and γc_n are factors of images of letters, say b'_n and c'_n. As $\mathrm{alph}(q_n) = \mathrm{alph}(f_n(q_{n+1}))$, this implies that b'_n and c'_n belong to $\mathrm{alph}(q_{n+1})$ and a_n, b_n and c_n belong to $\mathrm{alph}(q_n)$. Moreover, as βb_n, γc_n are factors of words in $f_n(\mathrm{alph}(q_{n+1}))$, we deduce that $(a_n, b_n, c_n, \beta, \gamma) \in \mathrm{set}(q_n)$.

By backward induction, we can show that for all i, $2 \leq i \leq n$, $(a_i, b_i, c_i, \beta, \gamma) \in \mathrm{set}(q_n)$. Especially $(a_2, b_2, c_2, \beta, \gamma) \in \mathrm{set}(q_2)$. As, by Step 2, f_1 is LSP (a_2, b_2, c_2)-breaking and (q_1, f_1, q_2) is a transition of $\mathcal{A}_{\mathrm{bLSP}}$, we get our final contradiction. $\square$

8 Conclusion

Recall that G. Fici [8] asked for a characterization of both finite and infinite words. Observe that it can be proved that any non-empty finite LSP word w is right extendable to a longer LSP word (That is there exists a letter a occurring in w such that wa is a LSP). As a consequence one can prove:

Lemma 16. *A finite word is LSP if and only if it is a prefix of an infinite LSP word.*

Lemma 16 is a consequence of next result.

Lemma 17. *Any non-empty finite LSP word w is right extendable to a longer LSP word. That is there exists a letter a occurring in w such that wa is a LSP.*

Proof. The result is immediate for words of length 1. Let w be a finite LSP word of length at least two. Let u be its longest border (that is the longest word distinct from w that is both a prefix and a suffix of w) and let a be the letter such that ua is a prefix of w. From now on we prove that wa is LSP. Consider a left special factor of wa. If it is a factor of w, as w is LSP, it is a prefix of w and so of wa. Assume now that va is both a suffix and a left special factor of wa. As v is then a left special factor of w, v is a prefix of w and so it is a border of w. If $v = u$ then, by definition, va is a prefix of wa. In the remaining case $v \neq u$, va is a suffix of ua which is a prefix of w. Hence va occurs in w, and as a left special factor, it is a prefix of w and so of wa. Hence wa is LSP. $\square$

Proof of Lemma 16. One can observe that if u is the longest border of a word w and if ua is a prefix of w, then ua is the longest border of wa. Hence iterating the proof of the previous lemma, one obtains that: if $w = pu$ with u the longest border of w, then the infinite extension p^ω of w is LSP. As any prefix of an LSP word is also LSP, Lemma 16 follows.

By definition, left special factors are prefixes of LSP words. Hence readers can verify that Lemma 17 can not be stated for left extendability. More precisely, if, for a letter a and a word x, both a and x are LSP, then ax has no left special factor except the empty word or $ax = a^n u$ for a word u whose left special factors are words a^i with $i \leq n$. $\square$

Lemma 16 shows that 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). For instance in the binary case, this allows to find back M. Sciortino and L.Q. Zamboni's result [15]: "binary words having suffix automaton with the minimal possible numbers of states are exactly the finite prefixes of standard Sturmian words" (that can be reformulated after G. Fici's work : "finite binary LSP words are exactly the finite prefixes of standard Sturmian words"). For this purpose, one can first see from Theorem 15 and Figure 2 that directive words of binary infinite LSP words are ultimately τ_a or ultimately τ_b or ultimately contain both τ_a and τ_b. By classical results (see, *e.g.*, [1]) it can be deduced that an infinite LSP word is an infinite repetition of a finite standard word or is an infinite standard word. As any power of a finite standard word is a prefix of an infinite standard word (see [14, Chap. 2] for instance), we get M. Sciortino and L.Q. Zamboni's result.

We end this paper mentioning natural questions arising from this work. Can a smaller automaton than $\mathcal{A}_{\mathrm{bLSP}}$ can be found for recognizing directive words of LSP infinite words? Can a similar S-adicity system can be found for infinite words having at most one left special factor? Does there exist a finite or infinite set S of morphisms such that an infinite word is LSP if and only if it S-adic (as it occurs for infinite balanced binary words)?

References

1. Berstel, J., Séébold, P.: Sturmian words. In: Lothaire, M. (ed.) Algebraic Combinatorics on Words, Encyclopedia of Mathematics and its Applications, vol. 90, pp. 45–110. Cambridge University Press (2002)
2. Berthé, V.: S-adic expansions related to continued fractions. In: Akiyama, S. (ed.) Natural extension of arithmetic algorithms and S-adic system. RIMS Kôkyûroku Bessatsu, vol. B58, pp. 61–84 (2016)
3. Berthé, V., Delecroix, V.: Beyond substitutive dynamical systems: S-adic expansions. In: Akiyama, S. (ed.) Numeration and Substitution 2012. RIMS Kôkyûroku Bessatsu, vol. B46, pp. 81–123 (2014)
4. Berthé, V., Holton, C., Zamboni, L.Q.: Initial powers of Sturmian sequences. Acta Arith. 122, 315–347 (2006)
5. Berthé, V., Labbé, S.: Factor complexity of S-adic words generated by the Arnoux–Rauzy–Poincaré algorithm. Adv. App. Math. 63(90 - 130) (2015)

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

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

8. Fici, G.: Special factors and the combinatorics of suffix and factor automata. Theoret. Comput. Sci. 412, 3604–3615 (2011)

9. Leroy, J.: Contribution à la résolution de la conjecture S-adique. Doctoral Thesis, Université de Picardie Jules Verne (2012)

10. Leroy, J.: 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)

11. Leroy, J., Richomme, G.: A combinatorial proof of S-adicity for sequences with linear complexity. Integers 13(Article #A5) (2013)

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

13. Lothaire, M.: Combinatorics on Words, Encyclopedia of Mathematics and its Applications, vol. 17. Addison-Wesley (1983), reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, UK, 1997

14. Lothaire, M.: Algebraic Combinatorics on Words, Encyclopedia of Mathematics and its Applications, vol. 90. Cambridge University Press (2002)

15. Sciortino, M., Zamboni, L.Q.: Suffix automata and standard Sturmian words. In: Harju, T., Karhumäki, J., Lepistö, A. (eds.) Developments in Language Theory: 11th International Conference, DLT 2007, Turku, Finland, July 3-6, 2007. Proceedings. Lecture Notes in Computer Science, vol. 4588, pp. 382–398. Springer Berlin Heidelberg (2007)