Theoretical Computer Science

# Episturmian words and some constructions of de Luca and Rauzy 

Xavier Droubay ${ }^{\text {a }}$, Jacques Justin ${ }^{\text {b,* }}$, Giuseppe Pirillo ${ }^{\text {c,d }}$<br>${ }^{\text {a }}$ LaRIA, Université de Picardie Jules Verne, CURI, 5, rue du Moulin Neuf, 80000 Amiens, France<br>${ }^{\mathrm{b}}$ LIAFA, 2 place Jussieu, F-75251 Paris, Cedex 05, France<br>${ }^{\mathrm{c}}$ IAMI CNR, Viale Morgagni 67/A, 50134 Firenze, Italy<br>${ }^{\mathrm{d}}$ Université de Marne-la-Vallée, 5, boulevard Descartes, Champs sur Marne, 77454 Marne-la-Vallée Cedex 2, France

Received May 1999; revised June 1999
Communicated by D. Perrin


#### Abstract

In this paper we study infinite episturmian words which are a natural generalization of Sturmian words to an arbitrary alphabet. A characteristic property is: they are closed under reversal and have at most one right special factor of each length. They are first obtained by a construction due to de LUCA which utilizes the palindrome closure. They can also be obtained by the way of extended RAUZY rules. (C) 2001 Elsevier Science B.V. All rights reserved.


Keywords: Sturmian words; Palindrome factors; Rauzy rules; Episturmian words

## 0. Introduction

In [5] de Luca makes a deep study of finite Sturmian words and gives a characterization of standard Sturmian (infinite) words, namely if $u_{n}$ is a palindrome prefix of the infinite word, then $u_{n+1}=\left(u_{n} x_{n}\right)^{(+)}$is also a palindrome prefix for some $x_{n} \in\{a, b\}$, where $v^{(+)}$denotes the palindrome right closure of $v$, i.e. the shortest palindrome having prefix $v$. As we shall see this property is equivalent to the following one, $A l$ : an infinite word $s$ has $A l$ if for any prefix $v$ of $s, v^{(+)}$is also a prefix of $s$.

On the other hand, we have observed another characteristic property of standard Sturmian words, $P i$ : an infinite word $s$ has $P i$ if any leftmost occurrence of a palindrome

[^0]in $s$ is a central factor of a prefix palindrome of $s$. Finite Sturmian words have also the following property, $J u$ : a word has $J u$ if it has a palindrome suffix with no other occurrence in it.

In Section 2, we study the relations between these properties in the case of an arbitrary alphabet. In particular, we prove the non-trivial equivalence between Al and Pi. In Section 3, we apply this to a two letter alphabet and obtain a simple and short proof of de Luca's characterization.

The infinite words considered in Section 2 are a natural and promising generalization to an arbitrary alphabet of the standard Sturmian words. So in Section 4 we call them standard episturmian and study some of their properties. In particular (Theorem 5), an infinite word is episturmian if and only if it is closed under reversal and has at most one right special factor of each length. It appears that, at least in the case that we call strict episturmian and with a three letter alphabet, they have been evoked by Rauzy [11] and, together with Arnoux, studied in [1], so it could perhaps be better to call them Rauzy words. We prove the equivalence of their various definitions, we begin the study of morphisms preserving these words and we also mention some open problems.

Let us also recall that generalizations of Sturmian words in other directions have been considered by several authors (see notes of Berstel and Seebold [2] for a bibliography), in particular the billiard sequences which generalize the definition of Sturmian words by cutting sequences [12].

## 1. Preliminaries

### 1.1. Words

Given a set $A$ (alphabet) whose elements are called letters, the free monoid $A^{*}$ generated by $A$ is the set of the (finite) words on $A$. The empty word is denoted by $\varepsilon$. If $u=u(1) u(2) \cdots u(m), u(i) \in A$, is a word, its length is $|u|=m$. Also $|u|_{x}, x \in A$, is the number of occurrences of $x$ in $u$. Lastly $\tilde{u}$ denotes the reversal of $u$, i.e. the word $u(m) u(m-1) \cdots u(2) u(1)$. A word equal to its reversal is a palindrome.
In the same way, an infinite word $s$ is an infinite sequence $s=s(1) s(2) \cdots s(i) \cdots, s(i)$ $\in A, i \in \mathbb{N}_{+}=\mathbb{N} \backslash\{0\}$. The set of infinite words on $A$ is $A^{\omega}$. For a finite or infinite word $t$, the word $w=t(i) t(i+1) \cdots t(j)$ is a factor of $t$ (proper factor if $w \neq t$ ). It is a prefix of $t$ if $i=1$. If $t \in A^{*}$ and $|t|=m$, then $w$ is a suffix of $t$ if $j=m$. Also we speak of an occurrence of $w$ in $t$ if we consider $w$ together with its position in $t$ defined, for instance, by $i$ or by a prefix $u w, u \in A^{*}$ of $t$. The factor $w$ of $t$ is unioccurrent in $t$ if it has exactly one occurrence in $t$. Lastly, given two occurrences of $w$ in $t$ given by the prefixes $u w$ and $u^{\prime} w$ of $t$, the shift between them is $\| u\left|-\left|u^{\prime}\right|\right|$. Let $u, v$ be two palindromes, then $u$ is a central factor of $v$ if $v=w u \tilde{w}$ for some $w \in A^{*}$.
Let $m \in A^{*} \cup A^{\omega}$. Then $F(m)$ denotes the set of the factors of $m$ and the alphabet of $m$ is $\operatorname{Alph}(m)=F(m) \cap A$. Also $m$ is closed under reversal if $\tilde{u} \in F(m)$ whenever $u \in F(m)$.

Given, $u \in A^{*}$, its palindrome right- (resp. left-) closure is the (unique) shortest palindrome $w=u^{(+)}$(resp. $w=u^{(-)}$) which has prefix (resp. suffix) $u$ [5].

A word is primitive if it is not a power of some shorter word. An infinite word $s$ is ultimately periodic if it can be written $s=u v v v \cdots=u v^{\omega}$ for $u \in A^{*}$ and $v$ a primitive word (periodic if $u=\varepsilon$ ).

Let $s$ be an infinite word, then a factor $u$ of $s$ is right (resp. left) special in $s$ if there exist $x, y \in A, x \neq y$, such that $u x, u y \in F(s)$ (resp. $x u, y u \in F(s)$ ). Clearly, if $s$ is closed under reversal, its right special factors are exactly the reversals of its left special factors.

Finally, an infinite word $s$ is uniformly recurrent if for each $u \in F(s)$ there exists $k$ such that $u \in F(w)$ for any $w \in F(s)$ with $|w|=k$.

### 1.2. Sturmian words

Sturmian words, whose theory was founded by Morse and Hedlund 60 years ago in the frame of symbolic dynamics, are extensively studied (see [2] for a survey).

Definition 1 (Berstel and Seebold [2]; Crisp et al. [3]; de Luca [5]; Justin and Pirillo [7]). Let $\rho$, $\alpha$ be real numbers in [0,1] with $\alpha$ irrational. A Sturmian word on $\{a, b\}$ is an infinite word $s$ given either by

$$
s(n)= \begin{cases}a & \text { if }\lfloor\rho+(n+1) \alpha\rfloor-\lfloor\rho+n \alpha\rfloor=0, \\ b & \text { otherwise }\end{cases}
$$

or by

$$
s(n)= \begin{cases}a & \text { if }\lceil\rho+(n+1) \alpha\rceil-\lceil\rho+n \alpha\rceil=0, \\ b & \text { otherwise. }\end{cases}
$$

Some interesting geometrical equivalent definitions are known, cutting sequences in particular.

Definition 2. A Sturmian word is standard if $\rho=0$ in Definition 1.
Definition 3. A word is a finite Sturmian word if it is a factor of some (infinite) Sturmian word.

Some useful properties [2] are recalled hereafter.

- An infinite word is Sturmian if and only if for every $n$ it has exactly $n+1$ factors of length $n$.
- A finite or infinite non-ultimately-periodic word $u$ is Sturmian if and only if it is balanced, that is whenever $w, w^{\prime}$ are factors of $u$ with $|w|=\left|w^{\prime}\right|$ we have $\|\left. w\right|_{a}-$ $\left|w^{\prime}\right|_{a} \mid \leqslant 1$.
- An infinite word on two letters is standard Sturmian if and only if its prefixes are exactly its left special factors.


## 2. Palindrome factors

In this section the alphabet $A$ is arbitrary.
Definition 4. A word $w$ has property $J u$ (resp. $L J u$ ) if there exists a palindrome suffix (resp. prefix) of $w$ which is unioccurrent in $w$.

Clearly, if $w$ has $J u$ it has exactly one unioccurrent palindrome suffix $v$ and this one is the longest palindrome suffix of $w$, moreover if $w=u v$, then $w^{(+)}=u v \tilde{u}$.

Now, for $u \in A^{*}$, let $P(u)$ denote the number of different palindrome factors of $u$. We have

Proposition 1. $P(w)$ is the number of prefixes (resp. suffixes) of $w$ which have $J u$ (resp. LJu).

Proof. For any $u \in A^{*}$ and $x \in A$, we have:
$P(u x)=P(u) \quad$ if $u x$ has not $J u$,
$P(u x)=P(u)+1 \quad$ if $u x$ has $J u$.
The result for $J u$ follows by induction as $P(\varepsilon)=1$.
In particular, we have the trivial but seemingly not widely known fact
Proposition 2. A word $w$ has at most $|w|+1$ different palindrome factors.
Also
Proposition 3. A word $w$ has exactly $|w|+1$ palindrome factors if and only if all its prefixes (resp. suffixes) have $J u$ (resp. LJu).

Corollary 1. If $P(w)=|w|+1$ then for any factor $u$ of $w, P(u)=|u|+1$. In other words, the language of all such $w$ is factorial.

Proof. Let $w=w^{\prime} u w^{\prime \prime}$. Then by Proposition 3 all prefixes of $w^{\prime} u$ have $J u$ whence, again by Proposition 3, all suffixes of $u$ have $L J u$, that is $P(u)=|u|+1$.

Definition 5. An infinite word $s$ has property $A l$ if, for any prefix $u$ of $s, u^{(+)}$is also a prefix of $s$.

Definition 6. An infinite word $s$ has property $P i$ if every leftmost occurrence of a palindrome in $s$ is a central factor of a palindrome prefix of $s$.

Theorem 1. For an infinite word $s \in A^{\omega}$ the following conditions are equivalent
(i) $s$ has $A l$,
(ii) $s$ has Pi,
(iii) there exist an infinite sequence $u_{1}=\varepsilon, u_{2}, u_{3}, \ldots$ of palindromes and an infinite word $\Delta(s)=x_{1} x_{2} x_{3} \cdots, x_{i} \in A$, such that $u_{n+1}=\left(u_{n} x_{n}\right)^{(+)}$for all $n \geqslant 1$ and that all the $u_{n}$ are prefixes of $s$.

Remark. When $s$ satisfies (iii), $u_{1}, u_{2}, u_{3}, \ldots$ are all the palindrome prefixes of $s$ in increasing length order, because if some palindrome prefix $u$ of $s$ would satisfy $\left|u_{n}\right|<$ $|u|<\left|u_{n+1}\right|$ we would have $\left|\left(u_{n} x_{n}\right)^{(+)}\right| \leqslant|u|<\left|u_{n+1}\right|$, a contradiction.

Proof. (iii) $\Rightarrow$ (i): With notations as in (iii) let $u$ be any non-empty prefix of $s$. There exists $n$ such that $\left|u_{n}\right|<|u| \leqslant\left|u_{n+1}\right|$. As $u_{n} x_{n}$ is a prefix of $u,\left|\left(u_{n} x_{n}\right)^{(+)}\right| \leqslant\left|u^{(+)}\right|$; as $u_{n+1}$ is a palindrome $\left|u^{(+)}\right| \leqslant\left|u_{n+1}\right|$. So $\left|u^{(+)}\right|=\left|u_{n+1}\right|$ and by the unicity of the palindrome right closure of $u_{n} x_{n}$ we get $u^{(+)}=u_{n+1}$. Consequently $s$ has $A l$.
(i) $\Rightarrow$ (ii): Let $v$ be any palindrome factor of $s$ with its leftmost occurrence given by $s=u v s^{\prime}, u \in A^{*}, s^{\prime} \in A^{\omega}$. If $v$ is not the longest palindrome suffix of $u v$ we have $u v=u^{\prime} w v$ with $w \neq \varepsilon$ and $w v$ a palindrome, whence $u v=u^{\prime} v \tilde{w}$ and this is a contradiction as $\left|u^{\prime}\right|<|u|$. So $v$ is the longest palindrome suffix of $u v$, whence $(u v)^{(+)}=u v \tilde{u}$. As $s$ has $A l, u v \tilde{u}$ is a prefix of $s$ and so $v$ is a central factor of a palindrome prefix of $s$.
For the more difficult proof of (ii) $\Rightarrow$ (iii) we require a lemma
Lemma 1. If an infinite words on an alphabet $A$ has property Pi, then all its prefixes have $J u$.

Proof. By induction on $n=|w|$, where $w$ is prefix of $s$, we exhibit a palindrome $w_{2}$ suffix of $w$ and unioccurrent in it, which is trivial for $n=1$. Let $w=v x$, where $x \in A, v \in A^{n},(|w|=n+1)$ and $s=w t$ with $t \in A^{\omega}$.
The induction allows to set $v_{1} \in A^{*}, v_{2} \in A^{+}$with $v=v_{1} v_{2}$ and $v_{2}$ palindrome unioccurrent in $v$. Then Pi involves that $v_{1} v_{2} \widetilde{v_{1}}$ is a prefix of $s$. If $v_{1} \neq \varepsilon, x$ is prefix of $\widetilde{v_{1}}$ and $w_{2}=x v_{2} x$ is the palindrome suffix of $w$ which is unioccurrent in it, since $v_{2}$ is unioccurrent in $v$.

Let us suppose now that $v_{1}=\varepsilon$, so $v=v_{2}$ is a palindrome. Either $x$ is unioccurrent in $w$ and $w_{2}=x$ is obviously the unique palindrome suffix of $w$, or, let $u x$ be the prefix of $v$ such that $x$ is unioccurrent in $u x$, and $|u|<|v x|=n+1$. We first prove that $u$ is a palindrome: the case $x$ prefix of $v$ being clear since $u=\varepsilon$ is clearly a palindrome, we now suppose $u \neq \varepsilon$. By induction we can set $u_{1} \in A^{*}, u_{2} \in A^{+}$with $u=u_{1} u_{2}$ and $u_{2}$ a palindrome unioccurrent in $u$. Again, the assumption $u_{1} \neq \varepsilon$ with $P i$ would involve that $u_{1} u_{2} \widetilde{u}_{1}$ is a prefix of $s$, and would contradict the unioccurrence of $x$ in $u x$. So $u_{1}=\varepsilon$, and $u=u_{2}$ is a palindrome. Now we can choose $q x$, the longest prefix of $v$ with $q$ a palindrome. As $v$ is a palindrome, $x q x$ is clearly a palindrome suffix of $w=v x$, and the proof is complete if we show that $w_{2}=x q x$ is unioccurrent in $w$. If it is not true, we set $v=\operatorname{oxqxp}$ with $o, p \in A^{*}$ such that $x q x$ is unioccurrent in oxqx. Let us prove that $l=o x q$ is a palindrome. Otherwise there exists by induction $l_{1}, l_{2} \in A^{+}$with $l=l_{1} l_{2}$ and $l_{2}$ a palindrome unioccurrent in $l$. Moreover $l_{2} \neq q(q$ is a prefix of $l$ since it is a prefix of $v$ and thus $q$ is not unioccurrent in $l$ ), so $l_{1} l_{2} \widetilde{l}_{1}$ is, from $P i$, a prefix of $s$
and $x$ is a prefix of $\tilde{l}_{1}$. Thus $x l_{2} x$ and $x q x$ would be two distinct palindrome suffixes of $l x$, both unioccurrent in it! So $l$ is a palindrome with $l x$ a prefix of $v$, but then, as $|l|>|q|$, this contradicts the choice of $q$, and thus $w_{2}$ is unioccurrent in $w$.
(ii) $\Rightarrow$ (iii): Now $s$ has Pi. Let us build the sequences $\left(u_{n}\right)_{n \geqslant 1}$ and $\Delta(s)$ as denoted in (iii). Clearly, $u_{1}=\varepsilon$ and $x_{1}$ is the letter prefix of $s$. Then we can set $u_{2}=\left(u_{1} x_{1}\right)^{(+)}=x_{1}$ and $x_{2}$ is the letter such that $u_{2} x_{2}$ is a prefix of $s$, with $\left|u_{2}\right| \geqslant\left|u_{1}\right|$. By induction on $n$, the above Lemma 1 ensures the existence of a unioccurrent palindrome suffix $v_{n}$ of $u_{n} x_{n}$, and the leftmost occurrence of $v_{n}$ in $s$ is given by $s=w v_{n} s^{\prime}=u_{n} x_{n} s^{\prime}$, for some $w \in A^{*}, s^{\prime} \in A^{\omega}$. So with $P i, v_{n}$ is a central factor of a palindrome prefix of $s$, say $u^{\prime}$. Moreover, $v_{n}$ is the longest palindrome suffix of $u_{n} x_{n}$, so that it is exactly a central factor of $u^{\prime}=\left(u_{n} x_{n}\right)^{(+)}$. Setting $u_{n+1}=\left(u_{n} x_{n}\right)^{(+)}$which is a palindrome prefix of $s$, and $x_{n+1}$ the letter such that $u_{n+1} x_{n+1}$ is a prefix of $s$, the proof is over.

From now on, $\Delta(s)$ will be called the directive word of $s$.
Corollary 2. If an infinite word satisfies the (equivalent) conditions of Theorem 1, then any factor $w$ of it has exactly $|w|+1$ palindrome factors.

Proof. By Lemma 1 all prefixes have $J u$, so by Proposition 3 any factor $w$ has exactly $|w|+1$ palindrome factors.

Remark. The infinite words having $A l$ are, by Corollary 2, "rich" in palindrome factors. For example, by Corollary 4 hereafter, the well-known [2] Fibonacci word abaababa... has this property. On the contrary, it is easy to find infinite words which are "poor" in palindrome factors, for example $(a b c)^{\omega}$ whose unique palindromes are $a, b, c$, and for a non-periodic example, the image of the Fibonacci word by the morphism $a \mapsto$ $a c, b \mapsto b$, whose unique palindromes are $a, b, c$ and $a c a$.

Proposition 4. If an infinite word s satisfies the (equivalent) conditions of Theorem 1, then it is closed under reversal.

Proof. This is trivial as any factor of $s$ is a factor of some palindrome prefix of $s$.
Theorem 2. If an infinite word s satisfies the (equivalent) conditions of Theorem 1, then it is uniformly recurrent.

Proof. Following the notations of (iii) of Theorem 1, the palindrome prefixes of $s, u_{1}=\varepsilon, u_{2}, u_{3}, \ldots$ satisfy $u_{i+1}=\left(u_{i} x_{i}\right)^{(+)}, x_{i} \in A$. It suffices to show, given any factor $v$ of $s$, that $v$ has infinitely many occurrences in $s$ : this is trivial, and that the shift between successive occurrences of $v$ is bounded. Without loss of generality, we may suppose that $v$ is a palindrome prefix of $s$, that is $v=u_{m}$ for some $m$. For $n>m$, $v$ is a prefix and a suffix of $u_{n}$. Let $k_{n}$ be the maximum shift between successive occurrences of $v$ in $u_{n}$. As $u_{n}$ is prefix and suffix of $u_{n+1}$ and $\left|u_{n+1}\right| \leqslant 2\left|u_{n}\right|+1$, we get $k_{n+1} \leqslant \max \left(k_{n},|v|+1\right)$. For the same reason $k_{m+1} \leqslant|v|+1$. Hence for all $i \geqslant m$, we have $k_{i} \leqslant|v|+1$.

Theorem 3. An infinite word s having Al is ultimately periodic if and only if its directive word $\Delta(s)$ has the form $g x^{\omega}, g \in A^{*}, x \in A$. In this case, $s$ is even periodic.

Proof. If $s$ is ultimately periodic, let $s=f v^{\omega}, f \in A^{*}$ and $v$ a primitive word. There exists $n$ such that for $i \geqslant n$ the palindrome prefix $u_{i}$ defined in (iii) of Theorem 1 satisfies $\left|u_{i}\right| \geqslant|f v|$. Consider two such palindromes $u_{i}, u_{j}, n \leqslant i<j$. Then $u_{i}=f v^{p} v^{\prime}$, $u_{j}=f v^{q} v^{\prime \prime}, p, q$ positive integers and $v^{\prime}, v^{\prime \prime}$ proper prefixes of $v$. As $u_{i}$ is a suffix of $u_{j}, v^{p} v^{\prime}$ is a suffix of $v^{q} v^{\prime \prime}$ whence, as $v$ is primitive, $v^{\prime}=v^{\prime \prime}$. Consequently, the letters $x_{i}, x_{j}$ following respectively $u_{i}, u_{j}$ are the same, $x$ say, whence $\Delta(s)=x_{1} x_{2} \ldots$ $x_{n-1} x^{\omega}$.

Conversely, let $\Delta(s)=x_{1} x_{2} \cdots x_{n-1} x^{\omega}$. As $s$ satisfies condition (iii) of Theorem 1, $u_{i+1}=\left(u_{i} x\right)^{(+)}$for $i \geqslant n$. Now, for $i \geqslant n, x u_{i} x$ is the longest palindrome suffix of $u_{i+1} x$ because $u_{i}$ is the longest palindrome proper prefix (and suffix) of $u_{i+1}$. Let $u_{n+1}=c x u_{n}$, $c \in A^{*}$. Then $u_{n+2}=\left(c x u_{n} x\right)^{(+)}=c x u_{n} x \tilde{c}=c x u_{n+1}$, and continuing in the same way, $u_{n+j}=c x u_{n+j-1}$, for $j \geqslant 1$. It follows $u_{n+j}=(c x)^{j} u_{n}$, for $j \geqslant 0$ whence $s=(c x)^{\omega}$, that is $s$ is periodic.

Corollary 3. If the infinite word $s$ has $A l$ and is not periodic, then all its prefixes are left special factors.

Proof. Let $u$ be a prefix of $s$. Then $\tilde{u}$ is a suffix of all palindrome prefixes $u_{i}$ of $s$ such that $\left|u_{i}\right| \geqslant|u|$. As $s$ is not periodic, by Theorem 3, two letters at least occur infinitely many times in $\Delta(s)=x_{1} x_{2} \cdots$. So the occurrences of $\tilde{u}$ in $s$ are followed at least by two different letters, and $\tilde{u}$ is right special in $s$.

In order to prove Proposition 5 hereafter which completes Corollary 3, we require a lemma

Lemma 2. Let $s$ be an infinite word having Al and $u$ be a factor of $s$. With the notations of (iii) in Theorem 1 , let $r$ be such that $u \in F\left(u_{r+1}\right) \backslash F\left(u_{r}\right)$, and write $u_{r+1}=$ ded with $u_{r} x_{r}=d e, d, e \in A^{*}$, and e being the unioccurrent palindrome suffix of $u_{r} x_{r}$. Then
(a) the leftmost occurrence of $u$ in the prefix $u_{r+1}$ of $s$ is given by $u_{r+1}=d e \tilde{d}=c u c^{\prime}$ $=c f e f^{\prime} c^{\prime}$, with $c, c^{\prime}, f, f^{\prime} \in A^{*}$. In particular, e is a factor of $u=f e f^{\prime}$
(b) if $u$ is a palindrome, then $\tilde{f}=f^{\prime}$ and $e$ is a central factor of $u$
(c) if $u=x v y$, with $x, y \in A$ and $v$ a palindrome prefix of $s$, then $u$ is either a suffix of $u_{r} x_{r}$ or a prefix of $x_{r} u_{r}$.

Proof. (a) As $u \in F\left(u_{r+1}\right)$, putting into evidence the leftmost occurrence of $u$, we have $u_{r+1}=c u c^{\prime}$ for some $c, c^{\prime} \in A^{*}$. As $u \notin F\left(u_{r}\right)$ and $u_{r+1}=u_{r} x_{r} \tilde{d}$ we have $|c u| \geqslant\left|u_{r} x_{r}\right|$ whence $\left|c^{\prime}\right| \leqslant|d|$. In the same way, $|c| \leqslant|d|$. So writing $d=c f, \tilde{d}=f^{\prime} c^{\prime}, f, f^{\prime} \in A^{*}$, we get $u_{r+1}=c u c^{\prime}=c f e f^{\prime} c^{\prime}$, whence $u=f e f^{\prime}$.
(b) As $u_{r+1}$ and $u$ are palindromes we have $u_{r+1}=c u c^{\prime}=\tilde{c}^{\prime} u \tilde{c}$. So $|c| \leqslant\left|c^{\prime}\right|$ because $u$ is a unioccurrent suffix of $c u$. As $u_{r+1}=c f e f^{\prime} c^{\prime}=c \tilde{f^{\prime}} e \tilde{f} c^{\prime}$ and as $e$ is a unioccurrent suffix of $c f e$, we have $|c f| \leqslant\left|c \tilde{f^{\prime}}\right|$ whence $|f| \leqslant\left|f^{\prime}\right|$. As $|c f|=\left|c^{\prime} f^{\prime}\right|=|d|$ we get $|c|=\left|c^{\prime}\right|,|f|=\left|f^{\prime}\right|$. So $e$ is a central factor of $u$.
(c) We have $u=x v y=f e f^{\prime}$. If $|f| \geqslant 1$ and $\left|f^{\prime}\right| \geqslant 1$, then $e$ is a factor of $v$. As $v$ is a palindrome prefix of $s$, shorter than $u_{r+1}, e$ is a factor of $u_{r}$, a contradiction. So we have, for instance, $f^{\prime}=\varepsilon$ whence $u_{r} x_{r}=c f e=c u$. Similarly if $f=\varepsilon$, then $x_{r} u_{r}=u c^{\prime}$.

Proposition 5. If the infinite word shas $A l$, then all its left special factors are prefixes of it.

Proof. If this is false, let $u$ be a shortest left special factor of $s$ which is not a prefix of $s$. Let $u=v x, x \in A$. Then $v$ is left special, hence is a prefix of $s$. So $v z$ is a prefix of $s$ for some $z \in A, z \neq x$. So $v$ is right special. So $\tilde{v}$ is a prefix of $s$ whence $v=\tilde{v}$. As $u$ is left special we have $y v x, y^{\prime} v x \in F(s)$ for some $y, y^{\prime} \in A, y \neq y^{\prime}$. Appling Lemma 2(c) to $y v x$ we get that $y v x$ or $x v y$ is a prefix of $x_{r} u_{r}$ for some $r$. Clearly $\left|u_{r}\right|>|v|$, so $v z$ is a prefix of $u_{r}$ whence $z \in\{x, y\}$. As $z \neq x$, we get $z=y$. In the same way $z=y^{\prime}$ whence $y=y^{\prime}$, a contradiction.

## 3. Case of a two letter alphabet

Now the alphabet is $A=\{a, b\}$. If $x \in A$ we denote by $\bar{x}$ the other letter of $A$. The infinite words satisfying Theorem 1 can be completely characterized: the equivalence between (iii) and (iv) in Theorem 4 hereafter is due to de Luca [5, Theorem 5 and Proposition 11]. Indeed his formulation of (iii) makes use of palindrome left closure (-) but this amounts to the same thing. Here we prove in a simple way de Luca's result with the help of Section 2.

Theorem 4. For an infinite non-ultimately-periodic words on the two letter alphabet $A=\{a, b\}$, the following conditions are equivalent
(i) $s$ has $A l$,
(ii) $s$ has Pi,
(iii) there exist an infinite sequence of palindromes $u_{1}=\varepsilon, u_{2}, u_{3}, \ldots$ and an infinite word $\Delta(s)=x_{1} x_{2} x_{3} \cdots, x_{i} \in A, \Delta(s) \in A^{\omega} \backslash\left(A^{*} a^{\omega} \cup A^{*} b^{\omega}\right)$ such that the $u_{i}$ are prefixes of $s$ and $u_{i+1}=\left(u_{i} x_{i}\right)^{(+)}$for all $i \geqslant 1$,
(iv) the infinite word $s$ is standard Sturmian.

Proof. The equivalence among (i), (ii) and (iii) follows from Theorems 1 and 3.
(iv) $\Rightarrow$ (ii): Here $s$ is standard Sturmian. Suppose it has not Pi. Then there exist a shortest $u \in A^{*}$, a palindrome $v$ unioccurrent in $u v$ and $s^{\prime} \in A^{\omega}$ such that $s=u v s^{\prime}$ and $\tilde{u}$ is not a prefix of $s^{\prime}$. Clearly $u \neq \varepsilon$ so putting $u=u^{\prime} x, u^{\prime} \in A^{*}, x \in A$
and $s^{\prime}=y s^{\prime \prime}, y \in A, s^{\prime \prime} \in A^{\omega}$ we have $s=u^{\prime} x v y s^{\prime \prime}$. If $x=y, x v x=v^{\prime}$ is unioccurrent in $u^{\prime} v^{\prime}$ and $u^{\prime} v^{\prime} \tilde{u^{\prime}}=u v \tilde{u}$ is not a prefix of $s$. As this contradicts the minimality of $|u|$ we have $x \neq y$. So $s=u^{\prime} x v \bar{x} s^{\prime \prime}$. So $x v, v \bar{x}$ and their reversals $v x, \bar{x} v$ are factors of $s$. So $v$ is left special and is a prefix of $s$, a contradiction with (iv).
(i) $\Rightarrow$ (iv): Now $s$ satisfies Theorem 1. So by Proposition 5 it has at most one left special factor of each length and as it is not ultimately periodic it has exactly one. As by Proposition 5 these left special factors are prefixes, we get that $s$ is standard Sturmian.

Corollary 4. Any finite Sturmian word $u$ has $J u$ and has exactly $|u|+1$ different palindrome factors.

Proof. This follows from Corollary 2 and Theorem 4.

## 4. Episturmian words

In this section the alphabet $A$ is finite.

## 4.1.

The words satisfying Theorem 1 are very similar to standard Sturmian words and as we will see have several similar properties. So we introduce the following definition

Definition 7. An infinite word $s$ is standard episturmian if it satisfies the equivalent conditions of Theorem 1.
An infinite word $t$ is episturmian if $F(t)=F(s)$ for some standard episturmian word $s$.

The second part of this definition is justified by the uniform recurrence property (Theorem 2) and the analogy with Sturmian words.

An episturmian word is periodic (resp. aperiodic) if the corresponding standard episturmian word is periodic (resp. aperiodic).

The case of periodicity is given by Theorem 3. The next theorem gives a characterization of episturmian words.

Theorem 5. Let s be an infinite word on the finite alphabet $A$, then the following conditions are equivalent:
(i) $s$ is episturmian,
(ii) $s$ is closed under reversal and has at most one right special factor of each length.

Proof. (i) $\Rightarrow$ (ii): It suffices to prove this for standard episturmian words, and this follows immediately from Propositions 4 and 5.
(ii) $\Rightarrow$ (i): Remark that as $s$ is closed under reversal,then any factor of $s$ has infinitely many occurrences in $s$. Construct a "standard" infinite word $\Sigma(s)$ as follows:
(a) if $s$ has exactly one left special factor of each length, then $\Sigma(s)$ is the infinite word whose prefixes are those factors;
(b) if not, $s$ is ultimately periodic and even is periodic as each factor occurs infinitely many times. Let $u$ be the longest left special factor of $s$ and for one of its occurrences, write $s=f u s^{\prime}, f \in A^{*}, s^{\prime} \in A^{\omega}$. Then $\Sigma(s)=u s^{\prime}$.
We prove first that $\Sigma(s)$ has $P i$. If not there is a leftmost occurrence of a palindrome $w$ in $\Sigma(s)$ such that $\Sigma(s)=u x v w \tilde{v} y s^{\prime}$ for some $u, v \in A^{*}, s^{\prime} \in A^{\omega}, x, y \in A, x \neq y$. As $y v w \tilde{v} x \in F(s), v w \tilde{v}$ is left special in $s$, whence $\Sigma(s)=v w \tilde{v}^{\prime \prime}, s^{\prime \prime} \in A^{\omega}$ and this contradicts the fact that the occurrence of $w$ considered previously was the leftmost one. Consequently, $\Sigma(s)$ is standard episturmian.
Let us show now that $F(s)=F(\Sigma(s))$. If $s$ is periodic this is evident by the definition of $\Sigma(s)$. If not, let $u$ be any factor of $s$ and let $s(i) s(i+1) \cdots s(i+k-1)$ and $s(j) s(j+1) \cdots s(j+k-1)$ be two occurrences of $\tilde{u}$ in $s$ with $k=|u|$ and $j>i$. If $\tilde{u}$ is not right special we have $s(i+k)=s(j+k)$. Continuing this way we see that either $s(i+q)=s(j+q)$ for all $q \geqslant 0$ or $\tilde{u}$ is a prefix of some right special factor of $s$. The first case is impossible as $s$ is not periodic. So $u$ is a factor of some left special factor of $s$, whence $u \in F(\Sigma(s))$. So $F(s)=F(\Sigma(s))$ and $s$ is episturmian.

Remark. The condition of closure under reversal cannot be deleted from (ii). Indeed replace in the Fibonacci word $a b a a b \ldots$ each $b$ by $b c$, then we get the infinite word $a b c a a b c a b c \ldots$ which has exactly one left (resp. right) special factor of each length but is not episturmian because it is not closed under reversal.

### 4.2. Complexity of episturmian words

Recall that the complexity function $p$ of an infinite word $s$ is given by $p(n)=\mid A^{n} \cap$ $F(s) \mid$ for all $n \in \mathbb{N}$.

Theorem 6. Let $s$ be a standard episturmian word with sequence of palindrome prefixes $u_{1}=\varepsilon, u_{2}, \ldots$ and $\Delta(s)=x_{1} x_{2} \cdots$ as in (iii) of Theorem 1. Then for $n \in \mathbb{N}_{+}$ and $x \in A, u_{n} x \in F(s)$ if and only if $x$ occurs in $x_{n} x_{n+1} \ldots$.

Proof. If part: If $x=x_{q}$ for some $q \geqslant n$, then $u_{q} x_{q}$ is a prefix of $s$, hence its suffix $u_{n} x_{q}$ is a factor of $s$.

Only if part: Let $u_{n} x \in F(s), x \in A$, and consider a factor $y u_{n} x, y \in A$, of $s$. By Lemma 2(c) $y u_{n} x$ or $x u_{n} y$ is a suffix of $u_{r} x_{r}$ for some $r$. So $\left|y u_{n} x\right| \leqslant u_{r} x_{r}$ whence $n<r$. So $x_{n} u_{n}$ is a suffix of $u_{r}$. So $x=x_{r}$ or $x=x_{n}$, so $x$ occurs in $x_{n} x_{n+1} \ldots$.

Theorem 7. Let $t$ be an aperiodic episturmian word, se the corresponding standard episturmian word, with directive word $\Delta(s)=x_{1} x_{2} x_{3} \cdots$. Then, for $n$ large enough, $t$ has complexity $p(n)=(h-1) n+q$ for some $q \in \mathbb{N}_{+}$and with $h$ the number of letters that occur infinitely many times in $\Delta(s)$.

Proof. For any right special factor $u$ of $t$ let $h_{u}$ be its order, that is the number of letters $x$ such that $u x \in F(t)$. Then $h_{u}$ is a decreasing function of $|u|$ and for $|u|$ large enough $h_{u}=h$ where, by Theorem $6, h$ is the number of the letters occurring infinitely many times in $\Delta(s)$. As $p(|u|+1)-p(|u|)=h_{u}-1$, the result follows.

The case where any right special factor of $t$ has order $k=|A|$ gives the maximum complexity $p(n)=(k-1) n+1$. Such episturmian words will be called $A$-strict according to the following definition.

Definition 8. Let $t \in A^{\omega}$ be an episturmian word and $B=\operatorname{Alph}(t)$. Then $t$ is strict if for any right special factor $u$ of $t, u B \subset F(t)$. Also, $t$ is $A$-strict if $t$ is strict and $B=A$.

In [1] uniformly recurrent infinite words of complexity $p(n)=(k-1) n+1$ with exactly one right and one left special factor of each length have been studied (with emphasis on the case $k=3$ ). We claim that they are exactly the strict episturmian words. This follows from Section 4.3 hereafter but we can sketch a more direct proof. Let $T$ be such an infinite word with alphabet $A$. In [1] it is studied by the mean of its factor graphs $\Gamma_{n}$. The directed graph $\Gamma_{n}$ has for vertice set $A^{n} \cap F(T)$ and has an arrow from $u$ to $v$ if $u y=x v \in F(T)$ for some $x, y \in A$. It consists of $k$ simple paths from $D_{n}$ to $G_{n}$ and one simple path from $G_{n}$ to $D_{n}$, with $G_{n}$ (resp. $D_{n}$ ) the left (resp. right) special factor of length $n$. The evolution of these graphs when $n$ grows shows that $T$ has infinitely many factors $w_{1}, w_{2}, \ldots$ which are both left and right special and that $\left|w_{n+1}\right| \leqslant 2\left|w_{n}\right|+1$. Consequently if $w_{n}$ is a palindrome, then $w_{n+1}$ is also a palindrome. Consequently, $T$ has infinitely many palindrome factors hence is closed under reversal. So, by Theorem 5, $T$ is episturmian.

### 4.3. Rauzy rules

The construction of standard Sturmian words by Rauzy pairs and rules is well known [8]. The extension to alphabets of at least three letters is evoked in [11] and implicitly made in [1].
Let $A=\left\{a_{1}, a_{2}, \ldots, a_{k}\right\}$ be a $k$-letter alphabet, $k \geqslant 2$. We construct a sequence $\left(R_{n}\right)_{n \in \mathbb{N}}$ of Rauzy $k$-uples $R_{n}=\left(A_{n}^{(1)}, A_{n}^{(2)}, \ldots, A_{n}^{(k)}\right)$ as follows: $R_{0}=\left(a_{1}, a_{2}, \ldots, a_{k}\right), R_{n+1}$ is obtained from $R_{n}$ by applying one of the Rauzy rules, labelled $1,2, \ldots, k$, with the rule $i \in[1, k]$ defined by

$$
\begin{aligned}
& A_{n+1}^{(i)}=A_{n}^{(i)}, \\
& A_{n+1}^{(j)}=A_{n}^{(i)} A_{n}^{(j)} \quad \text { for } j \in[1, k] \backslash\{i\} .
\end{aligned}
$$

We impose no restriction on the choice of the rule at each stage. There exists a unique infinite word $t$ such that every prefix of $t$ is a prefix of infinitely many of the $A_{n}^{(q)}, n \in \mathbb{N}, q \in[1, k]$. So any Rauzy sequence $\left(R_{n}\right)_{n \in \mathbb{N}}$ defines an infinite word.

Theorem 8. An infinite word $s$ on the finite alphabet $A=\left\{a_{i} ; 1 \leqslant i \leqslant k\right\}$ is standard episturmian if and only if it can be obtained by the Rauzy rules for $A$. Moreover, if $\Delta(s)=a_{i_{1}} a_{i_{2}} a_{i_{3}} \cdots$ then the sequence of the applied Rauzy rules is $i_{1}, i_{2}, i_{3}, \ldots$.

Proof. (Only if ): Let $s$ be the standard episturmian with directive word $\Delta(s)=x_{1} x_{2} \ldots$ and corresponding palindrome prefixes $u_{1}=\varepsilon, u_{2}, \ldots$. For any $n$ define $v_{n}, v_{n}^{\prime}$ and $w_{n}$ by $u_{n+1}=v_{n} u_{n}=u_{n} \widetilde{u_{n}}, v_{n}=v_{n}^{\prime} x_{n}, u_{n} x_{n}=v_{n}^{\prime} w_{n}$. Then $w_{n}$ is the unioccurrent palindrome suffix of $u_{n} x_{n}$. We have also $u_{n+1}=v_{n} v_{n-1} \cdots v_{p} u_{p}$ for any $p \in[1, n]$. Now define $k$ uples $R_{n}=\left(A_{n}^{(1)}, A_{n}^{(2)}, \ldots, A_{n}^{(k)}\right)$ by $R_{0}=\left(a_{1}, a_{2}, \ldots, a_{k}\right)$ and, for $i \in[1, k]$,
(a) $A_{n}^{(i)}=v_{n-1} v_{n-2} \cdots v_{1} a_{i}=u_{n} a_{i}$ if $a_{i}$ does not occur in $x_{1} x_{2} \cdots x_{n-1}$
(b) $A_{n}^{(i)}=v_{n-1} v_{n-2} \cdots v_{p}$ with $p$ maximal in $[1, n-1]$ such that $x_{p}=a_{i}$ if $a_{i}$ occurs in $x_{1} x_{2} \cdots x_{n-1}$.
We will show that $R_{n+1}$ is obtained from $R_{n}$ by applying a Rauzy rule and more precisely the rule $i$ such that $x_{n}=a_{i}$.
(a) Suppose we are in the first case for the definition of $A_{n}^{(i)}$, that is $A_{n}^{(i)}=u_{n} a_{i}$. Then as $a_{i}$ does not occur in $u_{n}$, we have $u_{n+1}=u_{n} a_{i} u_{n}$ whence $v_{n}=u_{n} a_{i}=A_{n}^{(i)}$. We have also $A_{n+1}^{(i)}=v_{n}$ because $x_{n}=a_{i}$, so $A_{n+1}^{(i)}=A_{n}^{(i)}$.
(b) Suppose we are in the second case for the definition of $A_{n}^{(i)}$, that is $A_{n}^{(i)}=v_{n-1} \cdots v_{p}$ with $x_{p}=a_{i}$ and $x_{n-1}, \ldots, x_{p+1} \neq a_{i}$. Then we have $u_{n}=v_{n-1} \cdots v_{p+1} v_{p}^{\prime} a_{i} u_{p}$, hence $u_{p}$ is the longest palindrome suffix of $u_{n}$ which is preceded by $a_{i}$ in $u_{n}$. So $a_{i} u_{p} a_{i}=w_{n}$. It follows that $u_{n} x_{n}=v_{n-1} \cdots v_{p+1} v_{p}^{\prime} w_{n}$ whence $v_{n-1} \cdots v_{p+1} v_{p}^{\prime}=v_{n}^{\prime}$. As $A_{n+1}^{(i)}=v_{n}$ we have again $A_{n+1}^{(i)}=A_{n}^{(i)}$.
So in both cases $A_{n+1}^{(i)}=A_{n}^{(i)}=v_{n}$.
Now for $j \neq i$, by the definition of $A_{n}^{(j)}$ and $A_{n+1}^{(j)}$ we have $A_{n+1}^{(j)}=v_{n} A_{n}^{(j)}$ because the occurrences, if any, of $a_{j}$ in $x_{1} x_{2} \cdots x_{n-1}$ and in $x_{1} x_{2} \cdots x_{n}$ are the same. Consequently $A_{n+1}^{(j)}=A_{n}^{(i)} A_{n}^{(j)}$. So $R_{n+1}$ can be obtained from $R_{n}$ by applying the Rauzy rule $i$.
(If ): It follows from the only if part that any sequence of Rauzy rules can be obtained with a suitable standard episturmian word, and so generates it.

### 4.4. Morphisms

The theory of Sturmian morphisms [2] on a 2-letter alphabet has a natural extension for any $k$-letter alphabet $A=\left\{a_{1}, \ldots, a_{k}\right\}$ with $k \geqslant 2$. Hereafter we sketch some results and proofs, at least for standard episturmian words.
We define endomorphisms $\psi_{a}, \theta_{a b}$ of $A^{*}, a, b \in A, a \neq b$ by

$$
\begin{aligned}
& \psi_{a}(a)=a, \quad \psi_{a}(x)=a x, \quad x \in A, \quad x \neq a, \\
& \theta_{a b}(a)=b, \quad \theta_{a b}(b)=a, \quad \theta_{a b}(x)=x, \quad x \in A, \quad x \neq a, b .
\end{aligned}
$$

These morphisms satisfy relations such as $\theta_{a b} \circ \psi_{a}=\psi_{b} \circ \theta_{a b}$. Endomorphisms of $A^{*}$ are naturally extended to $A^{\omega}$. The following lemmas have easy proofs.

Lemma 3. Let $u \in A^{*}$ and $a \in A$. Then
(a) $\psi_{a}(\tilde{u}) a$ is the reversal of $\psi_{a}(u) a$,
(b) $u$ is a palindrome if and only if $\psi_{a}(u) a$ is a palindrome,
(c) $u$ is a right (resp. left) special factor of an infinite word $t$ if and only if $\psi_{a}(u) a$ is a right (resp. left) special factor of $\psi_{a}(t)$.

For any $m \in A^{*} \cup A^{\omega}$ let us say that $a \in A$ is separating for $m$ if for any $x, y \in A$, $x y \in F(m)$ implies $a \in\{x, y\}$.

Lemma 4. (a) If $s$ is a standard episturmian word with first letter $a$, then a is separating for $s$ and its factors.
(b) If two different letters $a, b$ are separating for $m \in A^{*} \cup A^{\omega}$, then $m$ is $(a b)^{\omega}$ or $(b a)^{\omega}$ or $m \in F\left((a b)^{\omega}\right)$.

Theorem 9. The infinite word $s \in A^{\omega}$ is standard episturmian if and only if there exist a standard episturmian word $t$ and $a \in A$ such that $s=\psi_{a}(t)$. Moreover the first letter of $s$ is $a, t$ is unique, and the directive words satisfy $\Delta(s)=a \Delta(t)$.

Proof. (Only if): If $s$ is standard episturmian its first letter, $a$ say, is separating, so $s=\psi_{a}(t)$ for some $t \in A^{\omega}$. By Lemma 3 and easy arguments, $t$ is closed under reversal and has at most one left special factor of each length, and this one is a prefix of $t$, hence by Theorem $5, t$ is standard episturmian. Clearly $t$ is unique. At last if $u_{1}^{\prime}=\varepsilon, u_{2}^{\prime}, \ldots$ are the palindrome prefixes of $t$ and $\Delta(t)=x_{1}^{\prime} x_{2}^{\prime} \cdots, x_{n}^{\prime} \in A$ then by Lemma $3, \psi_{a}\left(u_{n}^{\prime}\right) a$ is a palindrome prefix of $s$ and even is $u_{n+1}$ from the remark after Theorem 1, with $u_{1}=\varepsilon, u_{2}, \ldots$ being the sequence of all palindrome prefixes of $s$. Moreover $x_{n+1}=x_{n}^{\prime}$, so $\Delta(s)=x_{1} \Delta(t)=a \Delta(t)$.
(If ): Follows from Lemma 3 in the same way.

Now let $M$ be the monoid of morphisms generated by the $\psi_{a}$ and the $\theta_{a b}, a, b \in A$. We have

Theorem 10. (a) If $\varphi \in M$, then for any standard episturmian word $t \in A^{\omega}, \varphi(t)$ is standard episturmian.
(b) If $\varphi$ is an endomorphism of $A^{*}$ such that for any $A$-strict standard episturmian word the infinite word $\varphi(t)$ is also $A$-strict standard episturmian, then $\varphi \in M$.

Proof. Part (a) follows by induction from Theorem 9. For part (b) suppose by contradiction that there exists a morphism $\varphi \notin M$ such that $\varphi(t)$ is $A$-strict standard episturmian whenever $t$ is. Choose $\varphi$ such that $|\varphi|=\Sigma_{x \in A}|\varphi(x)|$ is minimal. Let $C$ be the set of the first letter of each $\varphi(x) \neq \varepsilon, x \in A$. For any $x \in A$ such that $\varphi(x) \neq \varepsilon$ let $t$ be an $A$-strict standard episturmian word with first letter $x$. Then $\varphi(x)$ is a prefix of $\varphi(t)$, so the first letter of $\varphi(x)$ is separating for $\varphi(t)$ hence for all $\varphi(y) \neq \varepsilon, y \in A$. So if $|C| \geqslant 3$, three different letters are separating for all non-empty $\varphi(y), y \in A$. By Lemma 4(b) this is impossible unless $|\varphi(x)| \leqslant 1$ for all $x \in A$. In this case we must have $A=\{\varphi(x) ; x \in A\}$, so $\varphi$ is a permutation of $A$, hence $\varphi \in M$.

If $|C|=2$, say $C=\{a, b\}, a \neq b$, then by Lemma 4(b), $|\varphi(x)| \geqslant 2, x \in A$ gives $\varphi(x) \in$ $F\left((a b)^{\omega}\right)$. Also $|\varphi(x)|=1$ gives $\varphi(x) \in C$. So only $a$ and $b$ occur in $\varphi(x), x \in A$. So $A=C$ and the $A$-strict standard episturmian words are the standard Sturmian words on $A$. Then $\varphi(x) \in M$ by the theory of standard Sturmian morphisms [2,4].

Finally if $C=\{a\}$ for some $a \in A, a$ is separating for all non-empty $\varphi(x), x \in A$ and we see easily that $\varphi(x)=\psi_{a}\left(u_{x}\right), u_{x} \in A^{*}, x \in A$. Hence $\varphi(x)=\psi_{a} \circ \varphi_{1}$ where $\varphi_{1}$ is defined by $\varphi_{1}(x)=u_{x}, x \in A$. Also if $s=\varphi(t)$ with $t A$-strict standard episturmian, then by Theorem $9, s=\psi_{a}\left(s^{\prime}\right)$ with $s^{\prime} A$-strict standard episturmian. So $\varphi_{1}(t)=s^{\prime}$ whence $\varphi_{1}$ satisfies the hypotheses for $\varphi$. By the minimality of $|\varphi|$ we have either $\varphi_{1} \in M$, whence $\varphi \in M$, or $\left|\varphi_{1}\right| \geqslant|\varphi|$ whence $\varphi(x) \in a^{*}$ for all $x \in A$, which is impossible as $|A| \geqslant 2$.

The next theorem is a stronger version of part (b) of Theorem 10.
Theorem 11. Let $\varphi \in \operatorname{Endo}\left(A^{*}\right)$ be such that $\varphi(t)=s$ for some $A$-strict standard episturmian words $s, t \in A^{\omega}$. Then $\varphi \in M$.

Proof (sketch). If this is false then there exist $\varphi, s, t$ with $|\varphi|$ minimal satisfying the hypotheses and such that $\varphi \notin M$.

Let $x$ be the first letter of $t$. If $\varphi(x)=\varepsilon$, we have $s=\varphi\left(t^{\prime}\right)$ with $t^{\prime}=\psi_{x}^{-1}(t)$. If $y$ is the first letter of $t^{\prime}$, either $\varphi(y) \neq \varepsilon$ and we consider $t^{\prime}$ instead of $t$, or $\varphi(y)=\varepsilon$ and we repeat the argument. So, without loss of generality, we assume $\varphi(x) \neq \varepsilon$. Let $a$ be the first letter of $\varphi(x)$, hence of $s$. Put $s_{1}=\psi_{a}^{-1}(s)$. If all non-empty $\varphi(z), z \in A$ have $a$ for their first letter we can write $\varphi(z)=\psi_{a}\left(u_{z}\right), u_{z} \in A^{*}, z \in A$. Let $\varphi_{1}: A^{*} \rightarrow A^{*}$ be given by $\varphi_{1}(z)=u_{z}$ for all $z \in A$. Then $\varphi=\psi_{a} \circ \varphi_{1}$ and $s_{1}=\varphi_{1}(t)$. Clearly $\varphi_{1} \notin M$ and $\left|\varphi_{1}\right| \leqslant|\varphi|$ so, by the minimality of $|\varphi|,\left|\varphi_{1}\right|=|\varphi|$ whence $\varphi(z) \in a^{*}$ for all $z \in A$. This is impossible.

So at least two different letters are prefixes of the $\varphi(z), z \in A$. It can be shown then that all $\varphi(z), z \in A$ are palindromes. Let $\varphi^{\prime}=\varphi \circ \psi_{x}$ and $t^{\prime}=\psi_{x}^{-1}(t)$. Then $s=\varphi^{\prime}\left(t^{\prime}\right)$. As all the $\varphi^{\prime}(z), z \in A$ have the same first letter, $a$, and with $s_{1}=\psi_{a}^{-1}(s)$ as previously we have $s_{1}=\varphi_{1}\left(t^{\prime}\right)$ with $\varphi_{1}$ being given by $\varphi^{\prime}=\psi_{a} \circ \varphi_{1}$. In the particular case where $|\varphi(x)|=1$, i.e. $\varphi(x)=a$ it is easy to see that $\left|\varphi_{1}\right| \leqslant|\varphi|$, whence by the minimality of $|\varphi|, \varphi_{1} \in M$ or $\left|\varphi_{1}\right|=|\varphi|$ and then the restriction of $\varphi_{1}$ to $A$ is a permutation of $A$. Hence $\varphi_{1} \in M$, whence $\varphi \in M$. If $|\varphi(x)|>1$, and if $b$ is the first letter of $s_{1}$ we repeat the argument with $s_{2}=\psi_{b}^{-1}\left(s_{1}\right), \varphi_{2}$ given by $\varphi_{1}=\psi_{b} \circ \varphi_{2}$ and $s_{2}=\varphi_{2}\left(t^{\prime}\right)$. Continuing this way as far as possible and with a careful analysis of the process, we arrive to some $\varphi_{n}$ such that $\left|\varphi_{n}\right| \leqslant|\varphi|$ and we conclude that $\varphi \in M$.

From Theorem 11 we deduce a result for standard episturmian words which are fixed points of morphisms.

Theorem 12. Let $s$ be a strict standard episturmian word. Then $s$ is the fixed point of a morphism if and only if its directive word $\Delta(s)$ is periodic.

Proof. (If ): Let $\Delta(s)=\left(x_{1} x_{2} \cdots x_{n}\right)^{\omega}$. It follows from Theorem 9 that $s=\left(\psi_{x_{1}} \circ \psi_{x_{2}} \circ\right.$ $\left.\cdots \circ \psi_{x_{n}}\right)(s)$.
(Only if ): Without loss of generality, assume $s$ is $A$-strict. If $s=\varphi(s)$ then, by Theorem 11, $\varphi \in M$. So $\varphi=\psi_{y_{1}} \circ \psi_{y_{2}} \circ \cdots \circ \psi_{y_{n}} \circ \sigma, y_{i} \in A$, with $\sigma$ a permutation of $A$. It follows by Theorem 9 that $\Delta(s)=y_{1} y_{2} \cdots y_{n} \sigma(\Delta(s))$ whence, easily, $\Delta(s)$ is periodic.

### 4.5. Other problems

Let us mention three among the many problems about episturmian words.
First, do episturmian words satisfy a "balance" condition which extends the one for Sturmian words?
Another problem is: is there a number theoretical definition of episturmian words on $A$ recalling the one for Sturmian words and is there some kind of continued fraction expansion that could play a role there as for Sturmian words? Some particular cases have been examined. In [10] the infinite word on $\{a, b, c\}$ which is the fixed point of the morphism $\varphi$ given by $\varphi(a)=a b, \varphi(b)=a c, \varphi(c)=a$, that is the standard episturmian word with directive word $(a b c)^{\omega}$, is studied in this spirit. In [1] any $A$-strict episturmian word on $\{a, b, c\}$ is characterized by partitioning the circle of length 1 into 3 intervals of lengths $\alpha, \beta, \gamma$ where $\alpha, \beta, \gamma$ are the frequencies of $a, b, c$ in the infinite word, and performing a transformation on those intervals.

Lastly, it has been shown (Corollary 2) that any factor $w$ of an episturmian word has exactly $|w|+1$ palindrome factors. However, the converse is false: if in the Fibonacci word $a b a a b \cdots$ we replace $a$ by $a a$ and $b$ by $b b$, we get $a a b b a a a a b b \cdots$ whose any factor of length $n$ has exactly $n+1$ palindrome factors. So the problem of characterizing such words is open.

## References

[1] P. Arnoux, G. Rauzy, Représentation géometrique des suites de complexité $2 n+1$, Bull. Soc. Math. France 119 (1991) 199-215.
[2] J. Berstel, P. Seebold, Sturmian words, in: M. Lothaire (Ed.), Algebraic Combinatorics on Words, Preliminary Version (February 16, 1999), Institut Gaspard Monge, Univ. Marne-la-Vallée (Chapter 2).
[3] D. Crisp, W. Moran, A. Pollington, P. Shiue, Substitution invariant cutting sequences, J. Théorie Nombres Bordeaux 5 (1993) 123-138.
[4] A. de Luca, On standard Sturmian morphisms, Theoret. Comput. Sci. 178 (1-2) (1997) 205-224.
[5] A. de Luca, Sturmian words: structure, combinatorics and their arithmetics, Theoret. Comput. Sci. 183 (1997) 45-82.
[6] X. Droubay, G. Pirillo, Palindromes and Sturmian words, Theoret. Comput. Sci. 223 (1999) 73-85.
[7] J. Justin, G. Pirillo, Decimations and Sturmian words, Inform. Théor. Appl. 31 (1997) 271-290.
[8] F. Mignosi, P. Seebold, Morphismes Sturmiens et règles de Rauzy, J. Théor. Nombres Bordeaux 5 (1993) 221-233.
[9] G. Rauzy, Suites à termes dans un alphabet fini, Sém. Théor. Nombres Bordeaux 25 (1982-1983) 2501-2516.
[10] G. Rauzy, Nombres algébriques et substitutions, Bull. Soc. Math. France 110 (1982) 147-178.
[11] G. Rauzy, Mots infinis en arithmétique, in: M. Nivat, D. Perrin (Eds.), Automata on Infinite Words, Lecture Notes in Computer Science, Vol. 192, 1985, pp. 167-171.
[12] S. Tabachnikov, Billiards, Panoramas et Synthèses 1, Soc. Math. France, 1995.


[^0]:    * Correspondence address: J. Justin, 19 rue de Bagneux, 92330 Sceaux, France.

    E-mail addresses: droubay@laria.u-picardie.fr (X. Droubay), justin@liafa.jussieu.fr (J. Justin), pirillo@udini.math.unifi.it (G. Pirillo).

