Lyndon morphisms

G. Richomme

Abstract

We characterize Lyndon morphisms (that is free monoids morphisms that preserve finite Lyndon words) and morphisms that preserve lexicographic order between finite words. We prove that the monoids of these morphisms are not finitely generated. We end characterizing the episturmian morphisms that are Lyndon morphisms and those that are order-preserving.

Résumé

Nous caractérisons les morphismes de Lyndon (à savoir les morphismes qui préservent les mots de Lyndon) ainsi que les morphismes qui préservent l'ordre lexicographique sur les mots finis. Nous montrons que les monoïdes formés de ces morphismes ne sont pas finiment engendrés. Nous caractérisons les morphismes épisturmiens qui appartiennent à chacune des deux familles de morphismes précédentes.

1 Introduction

Let P be a property of words (finite or infinite). To generate some arbitrarily large words with property P, a very easy way (when possible) is to use morphisms. Although it is not necessary, it is sufficient to consider morphisms that preserve the property P of words (and increase the length of words). Such morphisms exist for instance for primitive words [13], square-free words or overlap-free words [24], Sturmian words (see [1] for a general presentation), balanced binary words [15]. Here we consider Lyndon words.

1991 *Mathematics Subject Classification :* 68R15.

Key words and phrases : Combinatorics on words, morphisms, Lyndon words, Sturmian morphisms, lexicographic order.

A finite Lyndon word is a non-empty word which is smaller in lexicographic order than all its proper suffixes (see for instance [9]). The Lyndon factorization theorem [3] states that any finite word can be decomposed uniquely in a product of decreasing (in lexicographic order) Lyndon words. This result was extended to infinite words [23]. To obtain the decomposition in Lyndon words of the (infinite) Thue-Morse word, Ido and Melançon [7] give some morphisms that preserve finite Lyndon words. Another use of such morphisms can be found for instance in [10, chapter 4]. Here we study these morphisms.

In Section 4, we give two characterizations of Lyndon morphisms. First a morphism f on A is a Lyndon morphism if and only if it preserves lexicographic order between finite words and $f(a)$ is a Lyndon word for each a in A. The second characterization is that for each morphism f there exists a finite set T of Lyndon words such that f is a Lyndon morphism if and only if $f(T)$ is a set of Lyndon words. We show that such a finite set T must depend on the morphism f. However we give some families of morphisms for which there exists a finite set T which is the same for each morphism in the concerned family.

Results in Section 4 depend on a characterization of order-preserving morphisms given in Section 3. This result allows to check whether a morphism preserves or does not preserve the lexicographic order. In fact, to determine whether a morphism preserves order, it is sufficient to look at a finite set of couples of words depending necessarily (except for 2-letter alphabets) on the morphism. We also give some families of morphisms having the same set of couples of words to test.

In Section 5, we study the structure of the monoid of Lyndon endomorphisms and the structure of the monoid of endomorphisms that preserve the lexicographic order. In particular, we show that these monoids are not finitely generated. Similar results exist for other monoids [4, 18, 20, 26]. Let us mention the rare situation of the monoid of binary overlap-free morphisms [24] and of the monoid of binary invertible morphisms [25] that are finitely generated. When studying cancellativity and unitarity in the above mentioned monoids, we state a result on equality sets between two Lyndon morphisms defined on a binary alphabet.

Due to the numerous properties they have, Sturmian words have been very well studied (see for instance [1]). A tool in this context is the set of Sturmian morphisms that are the morphisms that preserve Sturmian words. Note that Sturmian morphisms are the invertible binary morphisms [25] (see also [1]). Sturmian words are defined on two-letter alphabets. When considering arbitrary sized alphabets, episturmian words [5, 8] are one generalization of Sturmian words. Episturmian morphisms generalize Sturmian morphisms. In Section 6 we characterize the episturmian morphisms that are Lyndon morphisms and those that preserve lexicographic order.

In Section 7, we give examples of use of Lyndon morphisms. These examples concern the Fibonacci morphism.

2 Preliminaries

We assume the reader is familiar with combinatorics on words and morphisms (see for instance [9, 10]). We precise our notations.

Given a set X we denote by $\mathrm{Card}(X)$ its cardinality.

A *monoid* M is a set equipped with an associative internal operation and a neutral element e for this operation. We denote the internal operation by juxtaposition. Given a subset X of M, X^+ is the subset of M constituted of all the elements $x_1 \dots x_n$ with $n \geq 1$, $x_1, \dots, x_n \in X$. By definition, X^* is the set $X^+ \cup \{e\}$. Given an element x of M, we inductively define the integer powers of x by $x^0 = e$ and $x^n = xx^{n-1}$. We denote x^* the set $\{x\}^*$ and x^+ the set $\{x\}^+$.

A monoid M is *left cancellative* (resp. *right cancellative*) when, for any elements x, y and z in M, if $xy = xz$ (resp. if $yx = zx$) then $y = z$ß. Let $S \subseteq M$ be a monoid. The monoid S is said *left regular* (resp. *right regular*) when, for any elements x and y in M, if x and xy are in S then $y \in S$ (resp. if y and xy are in S then $x \in S$).

A set of *generators* of M is a subset G of M such that any element of M can be decomposed in elements of G. The monoid M is said *finitely generated* if it has a finite set of generators. A *presentation* of the monoid M (based on a set of generators G) is a set S of equalities between elements of G such that given any elements $f_1, \dots, f_n, g_1, \dots, g_p$ in G, the equality $f_1 \dots f_n = g_1 \dots g_p$ holds if and only if it can be stated using the equalities of S (and the trivial relations $em = me = m$ for any element m of M). When there is no relation between generators, the monoid is said *free*.

An *alphabet* A is a set of symbols called *letters*. Here we consider only finite alphabets. A *word over* A is a finite sequence of letters from A. The *empty word* ε is the empty sequence of letters. Equipped with the concatenation operation, the set A^* of words over A is a free monoid with neutral element ε and set of generators A. Given a non-empty word $u = a_1 \dots a_n$ with $a_i \in A$, the *length* $|u|$ of u is the integer n. One has $|\varepsilon| = 0$. For a word u and a letter a, $|u|_a$ is the number of occurrences of a in u. If for some words u, v, p, s (possibly empty), $u = pvs$, then v is a *factor* of u, p is a *prefix* of u and s is a *suffix* of u. When $p \neq u$ (resp. $s \neq u$), we say that p is a *proper prefix* (resp. s is a *proper suffix*) of u. A word w is said *primitive* if for any word u, the equality $w = u^n$ (with n an integer) implies $n = 1$.

Here we consider ordered alphabets. We denote $A_n = \{a_1 < \dots < a_n\}$ the n-letter alphabet $A_n = \{a_1, \dots, a_n\}$ with order $a_1 < \dots < a_n$. Recall [9] that, given an ordered alphabet, the *lexicographic order* on A^* is defined by $u \preceq v$ if and only if either u is a prefix of v or $u = xay$, $v = xbz$ with $a, b \in A$, $a < b$, $x, y, z \in A^*$. We write $u \prec v$ when $u \preceq v$ and $u \neq v$. Of course the lexicographic order is an order and so is reflexive, transitive and antisymmetric. Other properties of the lexicographic order ($u, v \in A^*$) are:

- $\forall w \in A^*$, $u \preceq v \Leftrightarrow wu \preceq wv$.
- if u is not a prefix of v, $\forall w, z \in A^*$, $u \prec v \Rightarrow uw \prec vz$.

A non-empty word w is a *Lyndon word* if for all non-empty words u and v, $w = uv$ implies $w \prec vu$. Equivalently [3, 9], a non-empty word w is a Lyndon word

if all its proper non-empty suffixes are greater than it for the lexicographic order. For instance on the one-letter alphabet $\{a\}$, only a is a Lyndon word. On $\{a < b\}$ the Lyndon words of length at most 5 are a, b, ab, aab, abb, $aaab$, $aabb$, $abbb$, $aaaab$, $aaabb$, $aabab$, $aabbb$, $abbbb$. Lyndon words are primitive.

From now on, since there is only one Lyndon word on a one-letter alphabet, we *consider only alphabets containing at least two letters.*

Let A, B be two alphabets. A *morphism* f from A^* to B^* is a mapping from A^* to B^* such that for all words u, v over A, $f(uv) = f(u)f(v)$. We will say also that f is a morphism on A or that f is defined on A (without any other precision when B has no importance). A morphism on A is entirely known by the images of the letters of A. When $B = A$, f is called an *endomorphism* (on A). Given an integer L, f is *L-uniform* if for each letter a in A we have $|f(a)| = L$. The morphism f is *uniform* if it is L-uniform for some integer $L \geq 0$.

We denote by Id_A or simply Id the identity endomorphism on A. An *erasing* morphism is a morphism for which $f(a) = \varepsilon$ for at least one a in A. The *empty morphism* ϵ on an alphabet A is defined by $\epsilon(a) = \varepsilon$, $\forall a \in A$. A *permutation* on A is an endomorphism f on A such that for all a in A, $|f(a)| = 1$, and such that $\mathrm{Card}(\{f(a) \mid a \in A\}) = \mathrm{Card}(A)$. For $x, y \in A$, let E_{xy} be the morphism on A such that $E_{xy}(x) = y$, $E_{xy}(y) = x$ and $E_{xy}(z) = z$ for $z \in A \backslash \{x, y\}$. Such a morphism is called an *exchange morphism*.

A morphism f on A is called *prefix* (resp. *suffix*) if for all a, b in A, $f(a)$ is not a prefix (resp. not a suffix) of $f(b)$. A prefix or suffix morphism is not erasing. It is also *injective* ($\forall u, v \in A^*$, $u \neq v \Rightarrow f(u) \neq f(v)$).

Let f be a morphism defined on an alphabet A such that there exists a letter a such that $f(a)$ starts with a. Then for each integer $n \geq 0$, $f^n(a)$ is a prefix of $f^{n+1}(a)$. If $\lim_{n\to\infty} |f^n(a)| = \infty$, one says that f generates the infinite words $f^\omega(a) = \lim_{n\to\infty} f^n(a)$.

3 Order-preserving morphisms

In Section 4 we will show that Lyndon morphisms are order-preserving morphisms. Here we start studying these particular morphisms.

Definition 3.1. A morphism f, defined on an ordered alphabet A, is an *order-preserving morphism* if for all words u, v over A, $u \preceq v$ implies $f(u) \preceq f(v)$.

The following lemma will be used very often (even without noticing it).

Lemma 3.2. *Let f be an order-preserving morphism. If $f \neq \epsilon$ then for all words u, v over A, $u \prec v$ implies $f(u) \prec f(v)$.*

Proof. Assume there exist words x and y over A such that $x \prec y$ and $f(x) = f(y)$. We have to prove that $f = \epsilon$.

Let a be the least letter in A. Since $x \neq y$, we have $x \preceq xa \preceq y$. Since f is order-preserving, $f(x) \preceq f(xa) \preceq f(y)$. It follows $f(x) = f(xa)$ and so $f(a) = \varepsilon$. Now let $b \in A \backslash \{a\}$. From $abb \prec b \prec bb$ and f order-preserving, we deduce $f(abb) \preceq f(b) \preceq f(bb)$. But $f(abb) = f(bb)$. So $f(b) = f(bb)$ which implies $f(b) = \varepsilon$. Thus $f = \epsilon$. ∎

One consequence of Lemma 3.2 is that any order-preserving morphism different from the empty morphism is an injective morphism (and in particular a non-erasing morphism). Note also that, given an ordered alphabet A, the unique permutation on A which is order-preserving is the identity morphism on A.

The following result characterizes order-preserving morphisms.

Proposition 3.3. *Let f be a non-empty morphism on A_n ($n \geq 2$). The morphism f is order-preserving if and only if for each i, $1 \leq i < n$, $f(a_i a_n^{k_i}) \prec f(a_{i+1})$ where k_i is the least integer such that $|f(a_i a_n^{k_i})| \geq |f(a_{i+1})|$.*

Morphisms f_k ($k \geq 0$) in the proof of Proposition 3.10 explain the necessary and sufficient condition in the previous proposition.

Proof of Proposition 3.3. Let f be a morphism on A_n.

Assume first that f is a non-empty order-preserving morphism. Let i, k be two integers such that $1 \leq i < n$ and $k \geq 0$. Since $a_i a_n^k \prec a_{i+1}$, by Lemma 3.2, $f(a_i a_n^k) \prec f(a_{i+1})$.

From now on, we assume that for each i, $1 \leq i < n$, $f(a_i a_n^{k_i}) \prec f(a_{i+1})$ where k_i is the least integer such that $|f(a_i a_n^{k_i})| \geq |f(a_{i+1})|$. We prove that f is an order-preserving morphism. We first state four successive facts.

Fact 3.4. $\forall i, j,\ 1 \leq i < j \leq n,\ f(a_i) \prec f(a_j)$.

Proof. Let i such that $1 \leq i < n$. The word $f(a_i)$ is a prefix of $f(a_i a_n^{k_i})$. So $f(a_i) \preceq f(a_i a_n^{k_i})$. Since by hypothesis $f(a_i a_n^{k_i}) \prec f(a_{i+1})$, we get $f(a_i) \prec f(a_{i+1})$.

Fact 3.4 follows by induction and by transitivity of $\prec$. ∎

Fact 3.5. *The morphism f is not erasing.*

Proof. For i such that $2 \leq i \leq n$, from Fact 3.4, $f(a_1) \prec f(a_i)$. This implies $f(a_i) \neq \varepsilon$. If $f(a_1) = \varepsilon$, since $|f(a_1 a_n^{k_1})| \geq |f(a_2)|$, we must have $k_1 \neq 0$. Thus $f(a_n)$ is a prefix of $f(a_n^{k_1}) = f(a_1 a_n^{k_1})$. So $f(a_n) \preceq f(a_1 a_n^{k_1})$. By hypotheses $f(a_1 a_n^{k_1}) \prec f(a_2)$. It follows $f(a_n) \prec f(a_2)$: A contradiction with Fact 3.4. ∎

Fact 3.6. *Given any integer i with $1 \leq i < n$, and given any integer $k \geq 0$, $f(a_i a_n^k) \prec f(a_{i+1})$.*

Proof. If $k \leq k_i$, $f(a_i a_n^k) \preceq f(a_i a_n^{k_i})$ since $f(a_i a_n^k)$ is a prefix of $f(a_i a_n^{k_i})$. By hypothesis $f(a_i a_n^{k_i}) \prec f(a_{i+1})$. Thus $f(a_i a_n^k) \prec f(a_{i+1})$

If $k > k_i$, we know by hypothesis that $f(a_i a_n^{k_i}) \prec f(a_{i+1})$ and $|f(a_i a_n^{k_i})| \geq |f(a_{i+1})|$. Thus there exist words u_1, u_2, u_3 and letters c, d such that $f(a_i a_n^{k_i}) = u_1 c u_2$, $f(a_{i+1}) = u_1 d u_3$ and $c \prec d$. From $f(a_i a_n^k) = f(a_i a_n^{k_i}) f(a_n^{k-k_i})$, we deduce that $f(a_i a_n^k) = u_1 c u_2 f(a_n^{k-k_i}) \prec u_1 d u_3 = f(a_{i+1})$. ∎

Fact 3.7. *Given any word u, there exists an integer $k \geq 1$ such that $f(u) \prec f(a_n^k)$.*

Proof. The proof acts by induction on $|u|$. First note that $f(\varepsilon) \prec f(a_n)$ since, by Fact 3.5, f is not erasing. Assume that for a word u and an integer k, $f(u) \prec f(a_n^k)$. We have $f(a_n u) \prec f(a_n) f(a_n^k) = f(a_n^{k+1})$. For i, $1 \leq i < n$, by Fact 3.6, $f(a_i u) \prec f(a_{i+1})$, and thus by Fact 3.4, $f(a_i u) \prec f(a_n)$. ∎

Now we can prove that f is an order-preserving morphism. Let u, v be two words over A_n such that $u \prec v$. If u is a proper prefix of v, there exists a non-empty word w such that $v = uw$. By Fact 3.5, $f(w) \neq \varepsilon$. So $f(u)$ is a proper prefix of $f(v)$: $f(u) \prec f(v)$.

Assume now that u is not a prefix of v. There exists words u_1, u_2, u_3 and integers i, j such that $u = u_1 a_i u_2$, $v = u_1 a_j u_3$, $i < j$. By Fact 3.7, there exists an integer $k \geq 0$ such that $f(u_2) \prec f(a_n^k)$. Thus $f(u) \prec f(u_1 a_i a_n^k)$. By Fact 3.6, $f(u) \prec f(u_1 a_{i+1})$. By Fact 3.4, since $i + 1 \leq j$, $f(a_{i+1}) \preceq f(a_j)$. Thus $f(u) \prec f(u_1 a_j)$. And consequently $f(u) \prec f(v)$. ∎

As example of use of Proposition 3.3, one can check that the rather classical morphism (see, e.g., [14]) recalled in the next example is order-preserving ($k_i = 1$ for each i such that $1 \leq i < n$).

Example 3.8.

$$\begin{array}{lcl} A_n^* & \to & \{a < b\}^* \\ a_i & \mapsto & ab^{i-1} \text{ for each } i,\ 1 \leq i \leq n. \end{array}$$

This example also shows that *given any ordered alphabets A and B containing at least two letters, there exists a non-trivial order-preserving morphism from A^* to B^**. One can also note that the lexicographical order over the code $\{ab^i \mid i \geq 0\}$ induced on $\{a, b\}^*$ coincides with the lexicographical order over $\{a, b\}^*$ (with $a < b$).

Proposition 3.3 states an effective way to determine whether a morphism is order-preserving. Indeed given a morphism f, if $f \neq \epsilon$, we have to verify that the order is verified on $n - 1$ couples of words. These words depend on f. A natural question is: does there exist a finite set to check order preservation that does not depend on the morphism? The following definition introduces formally this kind of set. Proposition 3.10 gives a negative answer to the question.

Definition 3.9. Let $\mathcal{F}$ be a family of non-empty morphisms on an ordered alphabet A. A subset S of $A^* \times A^*$ is a *test-set* for order-preserving morphisms in $\mathcal{F}$ if for each morphism f in $\mathcal{F}$, f is order-preserving if and only if for any u, v in S, $u \prec v$ implies $f(u) \prec f(v)$.

Note that similar notions of test-sets have already been used for other properties of morphisms: see for instance [15, 19, 20, 27]. Note also that, although we check preservation of order, Lemma 3.2 allows to take "$u \prec v$ implies $f(u) \prec f(v)$" in place of "$u \preceq v$ implies $f(u) \preceq f(v)$" in the previous definition.

Despite the fact that Proposition 3.3 does not give a finite test-set for non-empty order-preserving morphisms, it is in some way optimal as shown by the following Proposition.

Proposition 3.10. *Given two ordered alphabets A and B such that* $\mathrm{Card}(A) \geq 3$ *and* $\mathrm{Card}(B) \geq 2$*, there is no finite test-set for non-empty order-preserving morphisms from A^* to B^*.*

Proof. We first prove the proposition in case $\mathrm{Card}(B) = \mathrm{Card}(A)$. Without loss of generality we can consider that $A = B$. Let $k \geq 0$ be an integer. Let f_k be the morphism from A_n^* to A_n^* defined by:

$$\begin{cases} f_k(a_1) & = a_1, \\ f_k(a_2) & = a_1 a_n^k a_2, \\ f_k(a_i) & = a_i \text{ for } 3 \leq i \leq n. \end{cases}$$

This morphism f_k is not an order-preserving morphism since $f_k(a_2) \prec f_k(a_1 a_n^{k+1})$ (since $n \geq 3$, $a_n \neq a_2$ and so $f_k(a_n) = a_n$).

We prove that for any words u and v such that $u \prec v$ and $f_k(v) \preceq f_k(u)$, we have $|u| \geq k+2$. This is sufficient to prove that there exists no finite test-set for order-preserving morphisms on A_n since any test-set contains a couple (u, v) with $|u| \geq k+2$ for each $k \geq 0$.

Since $u \prec v$, v is not a prefix of u. If u is a prefix of v, there exists a non-empty word w such that $v = uw$. From $f_k(v) \preceq f_k(u)$ we get $f_k(w) = \varepsilon$, a contradiction with f_k non-erasing. Thus there exist integers $i \neq j$ $(1 \leq i, j \leq n)$ and words x, u', v' such that $u = x a_i u'$, $v = x a_j v'$. Since $u \prec v$, we have $i < j$. Since $f_k(v) \preceq f_k(u)$, we must have $i = 1$ and $j = 2$. Thus $f_k(u) = f_k(x) a_1 f_k(u')$ and $f_k(v) = f_k(x) a_1 a_n^k a_2 f_k(v')$. Since a_n is the greatest letter of A_n, and since $a_n^k a_2 f_k(v') \preceq f_k(u')$, $f_k(u')$ must start with $a_n^k a_l$ with $a_2 \prec a_l$ which implies that u' starts with $a_n^k a_l$. So $|u| \geq k+2$.

We end the proof of the proposition showing that this inexistence of test-sets does not depend on the co-domain of the morphisms. Indeed consider the order-preserving morphism f defined in Example 3.8. The morphism $f f_k$ is a morphism from A_n^* to $\{a < b\}^*$, and for u, v words, $f f_k(u) \prec f f_k(v)$ if and only if $f_k(u) \prec f_k(v)$ (see also Lemma 5.2). ∎

Now, as corollaries of Proposition 3.3, we consider particular classes of morphisms for which there exist finite test-sets for non-empty overlap-free morphisms: the set $\{(a, b) \in A_n \times A_n \mid a \prec b\}$ is a test-set for non-empty order-preserving prefix (resp. uniform) morphisms on A_n.

Corollary 3.11. *A prefix morphism on A_n is an order-preserving morphism if and only if for each i, $1 \leq i < n$, $f(a_i) \prec f(a_{i+1})$.*

Proof. If f is an order-preserving morphism then for each i, $1 \leq i < n$, since $a_i \prec a_{i+1}$, we have $f(a_i) \prec f(a_{i+1})$.

Assume now that f is a prefix morphism and for each i, $1 \leq i < n$, $f(a_i) \prec f(a_{i+1})$. Let i such that $1 \leq i < n$. Since $f(a_i) \prec f(a_{i+1})$ and $f(a_i)$ is not a prefix of $f(a_{i+1})$, there exist words u_1, u_2, u_3 and letters c, d such that $f(a_i) = u_1 c u_2$, $f(a_{i+1}) = u_1 d u_3$ and $c \prec d$. It follows that for any integer $k \geq 0$, $f(a_i a_n^k) \prec f(a_{i+1})$. By Proposition 3.3, f is an order-preserving morphism. ∎

Corollary 3.12. *A non-empty uniform morphism on A_n is an order-preserving morphism if and only if, for each i, $1 \leq i < n$, $f(a_i) \prec f(a_{i+1})$.*

The proof of this corollary is similar to that of Corollary 3.11 and is left to the reader.

We end this section considering the case of binary alphabets.

Lemma 3.13. *Let f be a non-empty morphism on $\{a < b\}$. The four following assertions are equivalent:*

1. *f is order-preserving.*
2. *$f(ab) \prec f(b)$.*
3. *There exists an integer $l \geq 1$ such that $f(a^l b) \prec f(b)$.*
4. *There exist an integer $k \geq 0$, words x, y, z and letters c, d with $c \prec d$ such that $f(a) = xcy$, $f(b) = f(a^k)xdz$.*

Proof. $1 \Rightarrow 2$ and $2 \Rightarrow 3$ are immediate.

$3 \Rightarrow 4$. Assume now that $f(a^l b) \prec f(b)$ for an integer $l \geq 1$. We must have $f(a) \neq \varepsilon$. Thus $|f(a^l b)| > |f(b)|$. There exist words x_1, y_1, z and letters c, d such that $f(a^l b) = x_1 c y_1$, $f(b) = x_1 dz$. Let k be the greatest integer such that $f(b) = f(a^k)u$. We get $f(a^{l+k})u = x_1 c y_1$ and $f(a^k)u = x_1 dz$. Since $c \neq d$, $f(a^k)$ is a prefix of x_1. Let x be the word such that $x_1 = f(a^k)x$. We get $u = xdz$ that is $f(b) = f(a^k)xdz$. Don't forget that $f(a^{l+k})u = x_1 c y_1 = f(a^k)xcy_1$. Thus $f(a^l)u = xcy_1$. By definition of k and x, $f(a)$ is not a prefix of x. Thus xc is a prefix of $f(a)$ since $l \geq 1$. There exists a word y such that $f(a) = xcy$.

$4 \Rightarrow 1$. Assume k, x, y, z, c, d are as in Assertion 4. Let u, v be two words such that $u \prec v$. If u is a prefix of v, then since f is not erasing, $f(u)$ is a proper prefix of $f(v)$. If u is not a prefix of v, there exist words α, β, γ such that $u = \alpha a \beta$, $v = \alpha b \gamma$. If $|\beta|_b = 0$ and $|\beta|_a \leq k - 1$ then $f(u)$ is a prefix of $f(\alpha a^k)$ which is a proper prefix of $f(v)$. Else $f(u)$ starts with $f(\alpha)f(a^k)xc$ and $f(v)$ starts with $f(\alpha)f(a^k)xd$. In all cases $f(u) \prec f(v)$. ∎

Note that "$2 \Rightarrow 1$" is also a Corollary of Proposition 3.3. Assertions 3 and 4 sharpen equivalence "$1 \Leftrightarrow 2$".

A consequence of equivalence of Assertions 1 and 2 in Lemma 3.13 is that the set $\{(ab, b)\}$ is a test-set for non-empty order-preserving morphisms on $\{a < b\}$. One can ask whether $\{(a, b)\}$ is a test-set for non-empty order-preserving morphisms on $\{a < b\}$. The answer is no since for the morphism f defined by $f(a) = a$, $f(b) = aa$, we have $f(a) \prec f(b)$ but f is not an order-preserving morphism on $\{a < b\}$.

As shown by Proposition 3.10 and morphisms used in its proof, Lemma 3.13 cannot be generalized to arbitrary alphabets.

4 Lyndon morphisms

In this section, we characterize Lyndon morphisms (see Proposition 4.2).

Definition 4.1. Given two ordered alphabets A and B, a morphism f from A^* to B^* is a *Lyndon morphism* if it preserves the Lyndon words, that is, for all Lyndon words w over A, $f(w)$ is a Lyndon word over B.

Proposition 4.2. *Let $n \geq 2$ be an integer. A morphism f on A_n^* is a Lyndon morphism if and only if it verifies the two following properties:*

1. *$f(a)$ is a Lyndon word for each $a \in A_n$,*
2. *f is an order-preserving morphism.*

Proof. First we show that the conditions are necessary. So let f be a Lyndon morphism. The first property is an immediate consequence of the fact that for any letter a, the word a is a Lyndon word. Note that since the empty word is not a Lyndon word, f is not erasing and in particular $f \neq \epsilon$. To prove that the second property is necessary, let us consider for all integers $k \geq 0$, and i with $1 \leq i < n$, the Lyndon word $a_i a_n^k a_{i+1}$. Since f is a Lyndon morphism, $f(a_i a_n^k a_{i+1})$ is a Lyndon word. In particular $f(a_i a_n^k) \prec f(a_i a_n^k a_{i+1}) \prec f(a_{i+1})$. By Proposition 3.3, f is order-preserving.

Now we show that the conditions are sufficient. Let f be an order-preserving morphism on A_n such that $f(a)$ is a Lyndon word for each a in A_n. Note once again that f is not erasing. Let u be a Lyndon word over A. We prove that $f(u)$ is a Lyndon word. For this, considering any non-empty proper suffix S of $f(u)$, we have to show that $f(u) \prec S$.

A first case that can occur is that, for non-empty words u_1, u_2, we have $u = u_1 u_2$, $S = f(u_2)$. In this case, since u is a Lyndon word, $u \prec u_2$. Since f is order-preserving, $f(u) \prec f(u_2) = S$.

If the previous case does not hold, there exist words $u_1, u_2, p \neq \varepsilon, s \neq \varepsilon$ and a letter a in A_n such that $u = u_1 a u_2$, $f(a) = ps$ and $S = sf(u_2)$. Since $f(a)$ is a Lyndon word, $f(a) \prec s$. Thus $f(a)$ cannot be a prefix of s. Consequently $f(au_2) \prec sf(u_2) = S$. Since u is a Lyndon word, $u \preceq au_2$. Thus since f is an order-preserving morphism $f(u) \preceq f(au_2) \prec S$. ∎

As a consequence of Proposition 4.2, a Lyndon morphism is injective, non-erasing, and non-empty.

Using Proposition 4.2 we can see that the morphism given at Example 3.8 is a Lyndon morphism. So *given two ordered alphabets A and B with* $\mathrm{Card}(A) \geq 2$, $\mathrm{Card}(B) \geq 2$, *there exists a non-trivial Lyndon morphism from A^* to B^*.*

Morphisms f and fg after the proof of Lemma 5.2 are examples of morphisms verifying Condition 2 of Proposition 4.2 but not Condition 1. In [22], one can find morphisms with the same property. These morphisms π_n are defined over the alphabet A_n by $\pi_n(a_i) = a_i a_{i+1} \dots a_{i+n-1}$ for $1 \leq i \leq n$ (indices are computed modulo n).

Since there exists an algorithm to determine whether a word is a Lyndon word [6], Propositions 4.2 and 3.3 give an algorithm to determine whether a morphism

is a Lyndon morphism. This algorithm can be adapted to the case of morphisms defined on a binary alphabet (using part $1 \Leftrightarrow 2$ of Lemma 3.13), to the case of prefix morphisms (using Corollary 3.11), or to the case of uniform morphisms (using Corollary 3.12).

Proposition 4.2 has the following corollary.

Corollary 4.3. *Let $n \geq 2$ be an integer. A morphism f on A_n^* is a Lyndon morphism if and only if it verifies the two following properties:*

1. *$f(a)$ is a Lyndon word for each $a \in A_n$,*
2. *for each i, $1 \leq i < n$, $f(a_i a_n^{k_i} a_{i+1})$ is a Lyndon word where k_i is the least integer such that $|f(a_i a_n^{k_i})| \geq |f(a_{i+1})|$.*

Proof. The conditions are necessary by definition of a Lyndon morphism. Assume the two conditions are verified. For each i, $1 \leq i < n$, since $f(a_i a_n^{k_i} a_{i+1})$ is a Lyndon word, $f(a_i a_n^{k_i}) \prec f(a_i a_n^{k_i} a_{i+1}) \prec f(a_{i+1})$. By Proposition 3.3, f is an order-preserving morphism. By Proposition 4.2, f is a Lyndon morphism. ∎

As for order-preserving morphisms, we can define a notion of test-set for Lyndon morphisms. Let $\mathcal{F}$ be a family of morphisms on an ordered alphabet A. A set S of Lyndon words over A is a *test-set for Lyndon morphisms* in $\mathcal{F}$, if for any morphism in $\mathcal{F}$: f is a Lyndon morphism if and only if for all words w in S, $f(w)$ is a Lyndon word.

In the general case, we have

Proposition 4.4. *Given two ordered alphabets A and B with* $\mathrm{Card}(A) \geq 3$ *and* $\mathrm{Card}(B) \geq 2$*, there is no finite test-sets for Lyndon morphisms from A^* to B^*.*

Proof. The proof is based on that of Proposition 3.10. If we consider the morphism f_k defined in this proposition, we can see that it is not a Lyndon morphism (since it is not order-preserving) and that the least Lyndon word u such that $f_k(u)$ is not a Lyndon word verifies $|u| \geq k+3$ (for example $u = a_1 a_n^k a_3 a_2$). We let the reader end the proof. ∎

Moreover the next lemmas give three situations where such a test-set exists.

Lemma 4.5. *The set $\{a, b, ab\}$ is a test-set for Lyndon morphisms on $\{a < b\}$.*

Proof. If f is a Lyndon morphism, $f(a)$, $f(b)$ and $f(ab)$ are Lyndon words.

Assume conversely that $f(a)$, $f(b)$ and $f(ab)$ are Lyndon words. Then f is not erasing. Since $f(ab)$ is a Lyndon word, $f(ab) \prec f(b)$. By Lemma 3.13, f is an order-preserving morphism. Thus by Proposition 4.2, f is a Lyndon morphism. ∎

Using Corollaries 3.11 and 3.12, we can similarly prove

Lemma 4.6. *For an integer $n \geq 2$, the set $A_n \cup \{a_i a_{i+1} \mid 1 \leq i < n\}$ is a test-set for prefix Lyndon morphisms on A_n and for uniform Lyndon morphisms on A_n.*

One can ask if in all the previous test-sets the letters are necessary. The answer is yes as shown by the endomorphism f on A_n defined by $f(a_i) = a_i a_i$ for each i with $1 \leq i \leq n$, which is an order-preserving morphism but not a Lyndon one.

To end this section, we give a more precise result than Proposition 4.2 in case of binary alphabet. Using Lemma 3.13, in this case, Proposition 4.2 says: A morphism f on $\{a < b\}$ is a Lyndon morphism if and only if firstly $f(a)$ and $f(b)$ are Lyndon words, and secondly $f(ab) \prec f(b)$. The last part can be changed:

Proposition 4.7. *A morphism f on $\{a < b\}$ is a Lyndon morphism if and only if*

1. *$f(a)$ and $f(b)$ are Lyndon words,*
2. *$f(a) \prec f(b)$.*

Proof. Let f be a morphism on $\{a < b\}$.

If f is a Lyndon morphism, by Proposition 4.2, $f(a)$ and $f(b)$ are Lyndon words. Moreover f is an order-preserving morphism. Since $a \prec b$, $f(a) \prec f(b)$.

Assume conversely that $f(a)$ and $f(b)$ are Lyndon words, and $f(a) \prec f(b)$. By Lemma 3.13 and Proposition 4.2, to prove that f is a Lyndon morphism, we just have to state that $f(ab) \prec f(b)$. If $f(a)$ is not a prefix of $f(b)$, $f(a) \prec f(b)$ implies that $f(a) = xay$, $f(b) = xbz$ for words x, y, z over $\{a < b\}$. Then $f(ab) \prec f(b)$. If $f(a)$ is a prefix of $f(b)$, $f(b) = f(a)u$ for a non-empty word u. Since $f(b)$ is a Lyndon word, $f(a)u \prec u$. Consequently $f(a)f(a)u \prec f(a)u$, that is, $f(ab) \prec f(b)$. ∎

5 Non finitely generated monoids

Let f and g be two composable morphisms. If both f and g are Lyndon (resp. order-preserving, uniform) morphisms, then fg is a Lyndon (resp. order-preserving, uniform) morphism. On the other hand, the identity morphism (on any alphabet) is a Lyndon (resp. order-preserving, uniform) morphism. The aim of this section is, given an ordered alphabet A, to study the monoid of (uniform) Lyndon endomorphisms on A, and the monoid of (uniform) order-preserving endomorphisms on A. We first prove:

Proposition 5.1. *Given an integer $n \geq 2$, the monoid of Lyndon endomorphisms on A_n, the monoid of uniform Lyndon endomorphisms on A_n, the monoid of order-preserving endomorphisms on A_n and the monoid of uniform order-preserving endomorphisms on A_n are not finitely generated.*

Proof. For any integer $p \geq 2$, we define a uniform endomorphism f_p on A_n by:

$$\begin{cases} f_p(a_1) & = a_1^p a_2, \\ f_p(a_i) & = a_{i-1} a_i^{p-1} a_{i+1} \text{ for } i \text{ such that } 2 \leq i \leq n-1, \\ f_p(a_n) & = a_{n-1} a_n^p. \end{cases}$$

By Corollary 3.12, the morphism f_p is order-preserving. By Proposition 4.2, it is a Lyndon morphism.

We prove that for each $p \geq 2$, f_p belongs to any set of generators of the monoid of (uniform) Lyndon endomorphisms on A_n, or, of the monoid of (uniform) order-preserving endomorphisms on A_n: So this set is not finite. Let g and h be two

order-preserving endomorphisms on A_n such that $f_p = gh$. We state that g or h is the identity morphism on A_n.

Let $G = \{g(a) \mid a \in A_n\}$. Since $g \neq \epsilon$ and g is order-preserving, for each $(a, b) \in A \times A$, $g(a) \neq g(b)$: $\mathrm{Card}(G) = n$. Since $f_p = gh$, for each i between 1 and n, $f_p(a_i) \in G^+$. Thus for each j, $1 \leq j \leq n-1$, there is at least one word in G starting with a_j. We consider two cases.

Case 1: G contains one word starting with a_n. Since g is order-preserving, for each i, $1 \leq i \leq n$, $g(a_i)$ is the unique word in G starting with a_i. Since $p \geq 2$, $f_p(a_1)$ starts with a_1a_1, thus with $g(a_1)$. Moreover $f_p(a_2)$ starts with a_1a_2, thus with $g(a_1)$. It follows $g(a_1) = a_1$. Since $f_p(a_1) = a_1^p a_2 \in G^*$, and since $g(a_1)$ is the unique word in G starting with a_1, we have $a_2 \in G$: $g(a_2) = a_2$. In the same manner, by induction, we can show that $g(a_i) = a_i$ for $3 \leq i \leq n$ using $f_p(a_{i-1}) = a_{i-2}a_{i-1}^{p-1}a_i$. So $g = Id_{A_n}$.

Case 2: G contains no word starting with a_n. There exists an integer m, $1 \leq m \leq n-1$ such that G contains two words starting with a_m and such that for each integer $i \neq m$ with $1 \leq i < n$, G contains only one word starting with a_i.

Assume first $m > 1$. Since g is order-preserving, for each j with $1 \leq j < m$, $g(a_j)$ is the word in G starting with a_j, $g(a_m)$ and $g(a_{m+1})$ start with a_m, and, for each j with $m+2 \leq j \leq n$, $g(a_j)$ is the word in G starting with a_{j-1}. As in Case 1 (using values $f_p(a_j)$, $1 \leq j \leq m-1$), we can see that for each i with $1 \leq i \leq m$, $g(a_i) = a_i$. Since no word in G starts with a_n, and since $f_p(a_n) = a_{n-1}a_n^p \in G^*$, $g(a_n) = f_p(a_n)$. If $m \neq n-1$ (possible if $n \geq 3$ since $n-1 \geq m \geq 2$), since $f_p(a_{n-1}) = a_{n-2}a_{n-1}^{p-1}a_n \in G^*$, since $g(a_n)$ is the only word in G starting with a_{n-1}, and since no word in G starts with a_n, $g(a_{n-1}) = f_p(a_{n-1})$. By induction, we can see for each i with $m+1 \leq i \leq n$, $g(a_i) = f_p(a_i)$. In these conditions, we cannot have $f_p(a_m) = a_{m-1}a_m^p a_{m+1} \in G^*$. A contradiction.

Thus $m = 1$. As in the last part of the case $m > 1$, we can see for each i with $2 \leq i \leq n$, $g(a_i) = f_p(a_i)$. Since none of these $n-1$ words of G is a factor of $f_p(a_1) = a_1^p a_2$, we have $g(a_1) = f_p(a_1)$. Consequently $g = f_p$, and so h is a permutation of A_n. Since h is order-preserving, $h = Id_{A_n}$.

■

Note that the previous proposition can be stated (with the same proof) for the monoid of prefix Lyndon (resp. prefix order-preserving) endomorphisms on A_n.

Now we study unitarity in the previous monoids. For immediate reasons, we exclude the empty morphism from the discussion. The following lemma states the left unitarity:

Lemma 5.2. *Let f and g be two composable non-empty morphisms (defined on alphabets of cardinality at least two) where f is an order-preserving morphism.*

1. *fg is an order-preserving morphism if and only if g is an order-preserving morphism.*

2. if fg is a Lyndon morphism then g is a Lyndon morphism.

Note that the converse of Assertion 2 does not hold. For instance take $g = Id_{A_n}$ and define f by $f(a_i) = a_i a_i$ for $1 \leq i \leq n$: f is order-preserving, g is a Lyndon morphism but fg is not a Lyndon morphism.

Proof of Lemma 5.2. Assume that g is not an order-preserving morphism. Then there exist words u and v such that $u \prec v$ and $g(v) \preceq g(u)$. Since f is an order-preserving morphism, $fg(v) \preceq fg(u)$. By Lemma 3.2, the morphism f is not erasing. Since $g \neq \epsilon$, we also have $fg \neq \epsilon$. Once again by Lemma 3.2, fg is not an order-preserving morphism. This ends the proof of the first part.

Assume f is order-preserving and fg is a Lyndon morphism on A. From Proposition 4.2, fg is order-preserving and for all a in A, $fg(a)$ is a Lyndon word (so fg is not erasing). From the first part of Lemma 5.2, g is order-preserving. Let a be a letter and let S be a non-empty proper suffix of $g(a)$. Since $f \neq \epsilon$, by Lemma 3.2, f is not erasing and so $f(S) \neq \varepsilon$. Since $fg(a)$ is a Lyndon word, $fg(a) \prec S$. Since f is order-preserving, $g(a) \prec S$. Thus $g(a)$ is a Lyndon word. By Proposition 4.2 g is a Lyndon morphism. ∎

The monoid of order-preserving non-empty endomorphisms on A_n is not right unitary: there exist non-empty endomorphisms f, g on A_n such that fg and g are order-preserving but not f. This is the case for instance for the morphisms f and g defined by:

$$\begin{cases} f(a_1) & = a_2 a_1, \\ f(a_2) & = a_2, \\ f(a_i) & = a_i a_i \text{ for } 3 \leq i \leq n. \end{cases} \qquad \begin{cases} g(a_1) & = a_1, \\ g(a_2) & = a_2 a_2, \\ g(a_i) & = a_i \text{ for } 3 \leq i \leq n. \end{cases}$$

We have $fg(a_1) = a_2 a_1$, and $fg(a_i) = a_i a_i$ for $2 \leq i \leq n$. By Corollary 3.11, fg and g are order-preserving. Since $f(a_2) \prec f(a_1)$, f is not order-preserving.

The situation is different when we consider only uniform endomorphisms on A_2.

Lemma 5.3. *Let f and g be two non-empty uniform endomorphisms on $\{a < b\}$. If fg and g are order-preserving, then f is order-preserving.*

Proof. Let f and g be two non-empty uniform endomorphisms on $\{a < b\}$ such that fg and g are order-preserving. Since g is order-preserving, $g(a) \prec g(b)$. Since g is uniform, there exist words x, y, z such that $g(a) = xay$, $g(b) = xbz$ with $|y| = |z|$. Since fg is order-preserving and $fg \neq \epsilon$, by Lemma 3.2, $fg(a) \prec fg(b)$ which implies $f(a)f(y) \prec f(b)f(z)$. Since $y, z \in \{a, b\}^*$ and $|y| = |z|$, $f(a) = f(b)$ implies $fg(a) = fg(b)$. Thus $f(a) \neq f(b)$. Since f is uniform, $f(a)$ is not a prefix of $f(b)$, and $f(b)$ is not a prefix of $f(a)$. From $f(a)f(y) \prec f(b)f(z)$, we get $f(a) \prec f(b)$. ∎

When $n \geq 3$, the monoid of uniform order-preserving endomorphisms on A_n is not right unitary (this is essentially due to the existence of morphisms from A_n^* to A_{n-1}^*). To prove this, it is sufficient to consider for instance the morphisms f and g defined by

$$\begin{cases} f(a_1) & = a_1, \\ f(a_2) & = a_2, \\ f(a_3) & = a_1, \\ f(a_i) & = a_i \text{ for } 4 \leq i \leq n. \end{cases} \qquad \begin{cases} g(a_1) & = a_1 a_1 a_1 a_2, \\ g(a_2) & = a_1 a_1 a_2 a_2, \\ g(a_3) & = a_1 a_2 a_2 a_2, \\ g(a_i) & = a_1 a_i a_i a_i \text{ for } 4 \leq i \leq n. \end{cases}$$

The situation depicted in Lemma 5.3 does not hold in case of Lyndon morphisms: if, for non-empty endomorphisms f and g, fg and g are Lyndon morphisms, then f is not necessary a Lyndon morphism. Actually, for any integer $n \geq 2$, the monoid of Lyndon morphisms on A_n, and the monoid of uniform Lyndon morphisms on A_n are not right unitary. This can be shown using the endomorphisms on A_n defined by

$$\begin{cases} f(a_1) & = a_1a_1, \\ f(a_2) & = a_2a_2, \\ f(a_i) & = a_ia_i \text{ for } 3 \leq i \leq n. \end{cases} \qquad \begin{cases} g(a_1) & = a_1a_1a_2, \\ g(a_2) & = a_1a_2a_2, \\ g(a_i) & = a_1a_ia_i \text{ for } 3 \leq i \leq n. \end{cases}$$

We now study cancellativity in the considered monoids. Once again we exclude the empty morphism from the discussion.

Since any non-empty order-preserving morphism is injective, the monoid of (uniform) non-empty order-preserving endomorphisms on A_n is left cancellative: for f, g, h non-empty order-preserving endomorphisms on A_n, if $hf = hg$ then $f = g$. It is also the case for the monoid of (uniform) Lyndon endomorphisms on A_n.

For any integer $n \geq 2$, the monoid of non-empty order-preserving endomorphisms on A_n is not right cancellative. This is shown by the order-preserving endomorphisms f, g, h defined on A_n by $f(a_1) = a_1a_2a_1$, $f(a_2) = a_2$, $g(a_1) = a_1$, $g(a_2) = a_2a_1a_2$, $h(a_1) = a_1a_2$, $h(a_2) = a_2a_1$ and for $3 \leq i \leq n$, $f(a_i) = g(a_i) = h(a_i) = a_i$. We have $fh = gh$ and $f \neq g$.

The following example shows that right cancellativity is not obtained even for uniform order-preserving endomorphisms on A_n when $n \geq 3$: $h(a_1) = a_2a_2a_2a_3$, $h(a_2) = a_2a_2a_3a_3$, $h(a_3) = a_2a_3a_3a_3$, $h(a_i) = a_2a_ia_ia_i$ for $i \geq 4$, $f(a_i) = g(a_i) = a_1a_ia_ia_i$ for $i \geq 2$ and $f(a_1) = a_1a_1a_1a_2$, $g(a_1) = a_1a_1a_2a_2$. But observe

Lemma 5.4. *Let A and B be two ordered alphabets. Let h be a non-erasing morphism from A^* to B^* such that each letter of B is a factor of a word in $h(A)$. Let f, g be two uniform morphisms on B such that $fh = gh$. Then $f = g$.*

Proof. Assume f is L_1-uniform and g is L_2-uniform. We have $L_1|h(a)| = |fh(a)| = |gh(a)| = L_2|h(a)|$. Since h is not erasing, $L_1 = L_2$. The lemma follows immediately from the fact that each letter of B is a factor of a word in $h(A)$. ∎

As a consequence of this lemma, we can see that right cancellativity is verified in the monoid of non-empty uniform order-preserving endomorphisms on A_2. Indeed, it is not difficult to see that if h is a non-empty order-preserving endomorphism on A_2, both a and b are factors of words of $h(A_2)$.

The example before Lemma 5.4 also shows that the monoid of (resp. uniform) Lyndon endomorphisms on A_n is not right cancellative when $n \geq 3$.

Now we prove that the monoid of Lyndon endomorphisms on A_2 is right cancellative. This is a consequence of the following proposition. Let recall [21, chapter 7], that for two morphisms f and g defined on an alphabet A, the equality set of f and g is the set:

$$E(f, g) = \{w \in A^+ \mid f(w) = g(w)\}.$$

Proposition 5.1 in [21, chapter 7] says: *Given two non-periodic morphisms f and g defined on a binary alphabet, there exist words u, v, w such that $E(f, g) = \emptyset$ or*

$E(f,g) = \{u,v\}^+$ *or* $E(f,g) = (uw^*v)^+$ (a morphism on A is *periodic* if there exists a word z such that for all letter a in A, $f(w) \subseteq z^*$: a Lyndon morphism is not periodic). In the particular case of Lyndon morphisms, we have:

Proposition 5.5. *Given two different Lyndon morphisms on* $\{a < b\}$*, we have* $E(f,g) = \emptyset$ *or* $E(f,g) = b^+$ *or* $E(f,g) = (ab^n)^+$ *for an integer* $n \geq 0$*.*

Proof. Let f and g be two Lyndon morphisms on $\{a < b\}$ such that $E(f,g) \neq \emptyset$.

First we consider the case $f(a) = g(a)$. In this case $a^+ \subseteq E(f,g)$ (or equivalently $(ab^0)^+ \subseteq E(f,g)$). Assume there exists a word w in $E(f,g) \setminus a^+$. Then $|w|_b \neq 0$. If $|w|_b = 1$, since $f(w) = g(w)$ and $f(a) = g(a)$, we have $f(b) = g(b)$: a contradiction with $f \neq g$. It follows $|w|_b \geq 2$, that is, $w = a^m bvba^n$ for a word v and integers m and n. From $f(w) = g(w)$ and $f(a) = g(a)$, we deduce $f(bvb) = g(bvb)$. Without loss of generality, we can assume $|f(b)| < |g(b)|$. Thus $f(b)$ is both a prefix and a suffix of $g(b)$: a contradiction with $g(b)$ a Lyndon word because $f(b) \neq \varepsilon$. We conclude in this case $E(f,g) = a^+$.

Similarly, in case $f(b) = g(b)$ we have $E(f,g) = b^+$.

From now on, we assume that $f(a) \neq g(a)$ and $f(b) \neq g(b)$.

Observe $E(f,g) \cap a(a,b)^*a = \emptyset$. Indeed otherwise let $awa \in E(f,g)$. Without loss of generality we can assume $|f(a)| < |g(a)|$. Then $f(awa) = g(awa)$ implies $f(a)$ is both a prefix and a suffix of $g(a)$: a contradiction with $g(a)$ is a Lyndon word.

Similarly $E(f,g) \cap b(a,b)^*b = \emptyset$.

Now we prove $E(f,g) \cap b(a,b)^*a = \emptyset$. Let u be a word in $b(a,b)^*a$ such that $f(u) = g(u)$. Without loss of generality we assume $|f(a)| < |g(a)|$. We have

$$|u|_a(|g(a)| - |f(a)|) = |u|_b(|f(b)| - |g(b)|). \tag{1}$$

So $|f(b)| > |g(b)|$. It follows that $g(b)$ is a proper prefix of $f(b)$ and $f(a)$ is a proper suffix of $g(a)$. Consequently since $g(a)$ is a Lyndon word, $g(a) \prec f(a)$. Let w be the longest word such that wa is a suffix of u and $f(wa)$ is a suffix of $g(a)$. Let v be the word such that $u = vwa$ and let S be the suffix of $f(v)$ such that $g(a) = Sf(wa)$. If $S = \varepsilon$, then since $g(a) \neq f(a)$, $|w| \geq 1$. Let x be the first letter of w. Since $f(x)$ is a proper prefix of $g(a)$, $f(x) \prec g(a)$. But since f is a Lyndon morphism, by Proposition 4.2, $f(a) \prec f(b)$: a contradiction with $g(a) \prec f(a)$. Thus $S \neq \varepsilon$. By construction, S is a proper suffix of $f(a)$ or of $f(b)$. Since $g(a) = Sf(wa)$ is a Lyndon word, S is not a suffix of $f(a)$. Since $f(b)$ is a Lyndon word, $f(b) \prec S$. But once again $f(a) \prec f(b)$. We get $f(b) \prec S \prec g(a) \prec f(a) \prec f(b)$: this is impossible.

From now on let u be a word in $E(f,g) \cap a(a,b)^*b$. Once again without loss of generality we assume $|f(a)| < |g(a)|$. Equation (1) is still valid and so $|f(b)| > |g(b)|$. By definition of u, we deduce that $f(a)$ is a prefix of $g(a)$ and $g(b)$ is a suffix of $f(b)$. Consequently $f(a) \prec g(a)$, and since $f(b)$ is a Lyndon word, $f(b) \prec g(b)$. Moreover since f and g are Lyndon morphisms, by Proposition 4.2, $f(a) \prec f(b)$ and $g(a) \prec g(b)$.

Let w be the longest word such that aw is a prefix of u and $f(aw)$ is a prefix of $g(a)$. Let v be the word such that $u = awv$ and let P be the prefix of $f(v)$ such that $g(a) = f(aw)P$.

Let us prove $g(a) \prec f(b)$. If $P = \varepsilon$, since $g(a) \neq f(a)$, $|w| \geq 1$. Since $g(a) = f(aw)$ is a Lyndon word, w must ends with b and so $g(a) \prec f(b)$. In this case $g(a)$ cannot be a prefix of $f(b)$. If $P \neq \varepsilon$, by construction, P is a proper prefix of $f(a)$ or a proper prefix of $f(b)$. Since $g(a)$ is a Lyndon word and starts with $f(a)$, P must be a prefix of $f(b)$ and so $g(a) \prec P \prec f(b)$. Note that since P is a proper suffix of $g(a)$, $g(a)$ cannot be a prefix of $f(b)$.

So $f(a) \prec g(a) \prec f(b) \prec g(b)$ and $g(a)$ is not a prefix of $f(b)$. It follows:

- $g(a)$ is not a prefix of a word in $\{f(a), f(b), g(b)\}$,
- $g(b)$ is not a prefix of a word in $\{f(a), f(b), g(a)\}$,
- $f(a)$ is not a prefix of a word in $\{f(b), g(b)\}$,
- $f(b)$ is not a prefix of a word in $\{f(a), g(a)\}$.

In particular f and g are prefix morphisms. Moreover since $g(b)$ is a proper suffix of $f(b)$, $f(b)$ is not a prefix of $g(b)$.

We have $P \neq \varepsilon$. Indeed if $P = \varepsilon$, $g(a) = f(aw)$. Since $g(a)$ is a Lyndon word, $|w|_b \neq 0$. There exist an integer $n \geq 1$ and a word w_1 such that $aw = a^n b w_1$. Since $f(u) = g(u)$ and $u = a^n b w_1 v$, $f(v) = g(a^{n-1})g(b)g(w_1 v)$. Since f is a prefix morphism and $f(v)$ starts with $f((ab^n w_1)^{n-1})$, we get v starts with $(a^n b w_1)^{n-1}$. Let v' be the word such that $v = (a^n b w_1)^{n-1} v'$. We have $f(v') = g(b)g(w_1 v)$. But this equality is impossible because g is not erasing (as a Lyndon morphism), $g(b)$ is a prefix neither of $f(a)$ nor of $f(b)$, and $f(a)$ and $f(b)$ are not prefixes of $g(b)$.

From now on $P \neq \varepsilon$. Let recall that P is a prefix of $f(v)$. We have already said that P, by construction, must be a prefix of $f(a)$ or of $f(b)$. But P cannot be a prefix of $f(a)$ since $g(a)$ is a Lyndon word. So v starts with b. Let v' be the word such that $v = bv'$ and let S be the word such that $f(b) = PS$. We have $Sf(v') = g(wbv')$.

Let w_3 be the longest prefix of wbv' such that $g(w_3)$ is a prefix of S. Let S' be the word such that $S = g(w_3)S'$.

Assume $S' \neq \varepsilon$. By construction S' is a proper prefix of $g(a)$ or of $g(b)$. But S' is a suffix of $f(b)$. Since $f(b)$ is a Lyndon word, $f(b) \prec S'$. It follows $g(a) \prec S'$. This implies that S' is a proper prefix of $g(b)$. But S' and $g(b)$ are suffixes of $f(b)$. Thus S' is both a prefix and a suffix of the Lyndon word $g(b)$: a contradiction.

So $S' = \varepsilon$, that is $S = g(w_3)$, $g(a) = f(aw)P$ and $f(v') = g(w_4)$.

If $v' \neq \varepsilon$ or $w_4 \neq \varepsilon$, since f and g are not erasing then $v' \neq \varepsilon$ and $w_4 \neq \varepsilon$. Recall that $f(a)$ is not a prefix of $g(b)$, that $f(b)$ is a prefix neither of $g(a)$ nor of $g(b)$, that $g(a)$ is not a prefix of $f(b)$ and that $g(b)$ is a prefix neither of $f(a)$ nor of $f(b)$. In this case both v' and w_4 must start with a. Thus $f(v')$ starts with $g(a) = f(aw)P$. Since f is a prefix morphism, v' starts with aw. Morevoer since P is not a prefix of $f(a)$, v' starts with awb and $g(w_4)$ starts with $f(awb) = g(aw_3)$. Since g is a prefix morphism w_4 must start with aw_3. By induction we deduce that for an integer $m \geq 1$:

$$u = (awb)^m = (aw_3)^m.$$

So $aw_3 = awb$.

Remember that $S = g(w_3)$ is a proper suffix of the Lyndon word $f(b)$, $f(a) \prec f(b)$ and $f(a)$ is not a prefix of $f(b)$. So $f(a)$ cannot be a factor of S. Since $f(a)$ is a prefix of $g(a)$, this implies $|w_3|_a = 0$. Taking $n = |w_3|$, we have $awb = ab^n$ and $u \in (ab^n)^+$.

We have seen that when $(v', w_4) \neq (\varepsilon, \varepsilon)$, v' starts with awb and w_4 starts with aw_3. In the same manner, we can see that words w, w_3 and integer n are the same for any word u such that $f(u) = g(u)$.

Thus in case $f(a) \neq g(a)$, $f(b) \neq g(b)$, $E(f, g) \neq \emptyset$, we have $E(f, g) = (ab^n)^+$ for an integer $n \geq 1$. ∎

Before concluding about cancellativity, we give examples of equality sets for Lyndon morphisms. Let $n \geq 1$ be an integer. Let f_1, f_2 and f_3 be the three Lyndon endomorphisms on $\{a < b\}$ defined by $f_1(a) = aab$, $f_1(b) = b$, $f_2(a) = aab$, $f_2(b) = ab^{n+1}$, $f_3(a) = aab(ab^{n+1})^{n-1}ab$, $f_3(b) = b$. We can see that $E(f_1, f_2) = a^+$, $E(f_1, f_3) = b^+$, $E(f_2, f_3) = (ab^n)^+$. This example shows that integer n in Proposition 5.5 can be arbitrary.

Corollary 5.6. *The monoid of Lyndon endomorphisms on A_2 is right cancellative.*

Proof. Let f, g, h be three Lyndon endomorphisms on A_2 such that $fh = gh$. If f and g are different, then since $h(a)$ and $h(b)$ are Lyndon words, they are primitive words. So by Proposition 5.5, $h(a) = h(b) = b$ or $h(a) = h(b) = ab^n$ for an integer $n \geq 0$. By Proposition 4.2, h is order-preserving: a contradiction with Lemma 3.2. ∎

6 Episturmian morphisms

In this section, we study the episturmian morphisms that are order-preserving morphisms and those that are Lyndon morphisms. Let recall that episturmian morphisms [5, 8, 16, 17] are a generalization of Sturmian morphisms [1].

An endomorphism on an alphabet A is an *episturmian morphism* if it belongs to

$$\mathrm{Episturm}(A) = (\mathrm{Exch}(A) \cup \{\Psi_\alpha, \overline{\Psi}_\alpha / \alpha \in A\})^*.$$

where $\mathrm{Exch}(A)$ is the set of exchange morphisms and for $\alpha \in A$,

$$\Psi_\alpha\colon \begin{cases} \alpha \to \alpha \\ x \to \alpha x, \quad \forall x \in A \backslash \{\alpha\} \end{cases} \qquad \overline{\Psi}_\alpha\colon \begin{cases} \alpha \to \alpha \\ x \to x\alpha, \quad \forall x \in A \backslash \{\alpha\} \end{cases}$$

In the rest of this section, we consider an ordered alphabet A of cardinality at least two. We call a (resp. b) the least (resp. the greater) letter of A for the lexicographic order.

Let $\alpha \in A$. One can note that $\overline{\Psi}_\alpha$ preserves order. But Ψ_α preserves the lexicographic order if and only if $\alpha = a$. Indeed if $\alpha \neq a$, $\Psi_\alpha(\alpha)$ is a prefix of $\Psi_\alpha(a)$. More generally, we have:

Proposition 6.1. *An episturmian morphism is order-preserving if and only if*

1. *it belongs to* $\{\Psi_\alpha, \overline{\Psi}_\alpha/\alpha \in A\}^*$ *and*

2. *for all* $x, y \in A$, $x \prec y$ *implies* $f(y)$ *is not a prefix of* $f(xy)$.

The proof of this proposition needs the following relations and lemma.

Let $z \in A$ and $E \in \mathrm{Exch}(A)$. We have $E\Psi_z(z) = E(z) = \Psi_{E(z)}E(z)$, $E\overline{\Psi}_z(z) = E(z) = \overline{\Psi}_{E(z)}E(z)$, and for x in $A\backslash\{z\}$, $E\Psi_z(x) = E(z)E(x) = \Psi_{E(z)}E(x)$, $E\overline{\Psi}_z(x) = E(x)E(z) = \overline{\Psi}_{E(z)}E(x)$. So

$$E\Psi_z = \Psi_{E(z)}E, \tag{2}$$
$$E\overline{\Psi}_z = \overline{\Psi}_{E(z)}E \tag{3}$$

Lemma 6.2. *Let* $a_1, \ldots, a_k, b_1, \ldots, b_k$ *be* $2k$ *letters* ($k \geq 1$). *Let* $P_1, \ldots, P_k$ *be permutations such that* $b_i = (P_iP_{i+1}\ldots P_k)^{-1}(a_i)$. *For each* i, $1 \leq i \leq n$, *let* $f_i \in \{\Psi_{a_i}, \overline{\Psi}_{a_i}\}$, *and let* $g_i = \Psi_{b_i}$ *if* $f_i = \Psi_{a_i}$, *or* $g_i = \overline{\Psi}_{b_i}$ *if* $f_i = \overline{\Psi}_{a_i}$. *Then the following relation can be stated using Relations (2) and (3).*

$$f_1P_1f_2P_2\ldots f_kP_k = P_1\ldots P_kg_1\ldots g_k. \tag{4}$$

Proof. Relations (2) and (3) can be used to show inductively that for exchange morphisms $E_1, \ldots, E_n$:

$$\Psi_z E_1E_2\ldots E_n = E_1\ldots E_n\Psi_{E_nE_{n-1}\ldots E_1(z)}.$$

This proves Formula (4) when $k = 1$, since $(E_1\ldots E_n)^{-1} = E_n\ldots E_1$ and any permutation can be decomposed as a product of exchange morphisms.

It follows from what precedes that with the hypotheses of the lemma, we have for each i, $1 \leq i \leq k$,

$$f_iP_iP_{i+1}\ldots P_k = P_iP_{i+1}\ldots P_kg_i$$

The proof of lemma follows by induction. ∎

Let observe that if, in the hypothesis of Lemma 6.2, we take $b_i = (P_1\ldots P_i)^{-1}(a_i)$ for each i, $1 \leq i \leq n$, we can state

$$P_1f_1P_2f_2\ldots P_kf_k = g_1\ldots g_kP_1\ldots P_k.$$

Consequently any episturmian morphism can be decomposed as $f = g_1P = Pg_2$ where $g_1, g_2 \in \{\Psi_\alpha, \overline{\Psi}_\alpha/\alpha \in A\}^*$ and P is a permutation (such decompositions are unique [8]).

Proof of Proposition 6.1. Let f be an episturmian morphism. If for letters x, y with $x \prec y$, $f(y)$ is a prefix of $f(xy)$, then f does not preserve the lexicographic order. Thus, in what follows, we suppose f is an episturmian morphism such that for all $x, y \in A$, $x \prec y$ implies $f(y)$ is not a prefix of $f(xy)$. Then we have to prove that f is order-preserving if and only if $f \in \{\Psi_\alpha, \overline{\Psi}_\alpha/\alpha \in A\}^*$.

By Lemma 6.2, there exist a permutation P and $k \geq 0$ morphisms $f_1, \ldots, f_k$ in $\{\Psi_\alpha, \overline{\Psi}_\alpha/\alpha \in A\}$ such that $f = Pf_1\ldots f_k$. So we have to prove that f is order-preserving if and only if P is the identity morphism. Before doing this, we prove the following fact.

Fact 6.3. *Let x, y be letters such that $x \prec y$. For each integer i, $1 \leq i \leq k+1$, there exists a word w_i such that $f_i \dots f_k(xy)$ starts with $w_i x$ and $f_i \dots f_k(y)$ starts with $w_i y$.*

Proof. We prove this fact by induction on k. Denoting $f_{k+1} \dots f_k$ the identity morphism, the property is true for $i = k+1$ with $w_{k+1} = \varepsilon$.

Now assume the existence of w_{i+1} for an integer i, $1 \leq i \leq k$. If $f_i \in \{\overline{\Psi}_\alpha / \alpha \in A\}$, then taking $w_i = f_i(w_{i+1})$, we get the expected property. Assume now $f_i = \Psi_\alpha$ for a letter α. Let $w_i = f_i(w_{i+1})\alpha$. When $\alpha \notin \{x, y\}$, the property is verified. If $\alpha = x$, then $f_i \dots f_k(y)$ starts with $w_i y$. Let z be the word such that $f_{i+1} \dots f_k(xy) = w_{i+1} x z$. We have $f_i \dots f_k(xy) = w_i f_i(z)$. Since $|f_i \dots f_k(xy)| \geq |f_i \dots f_k(y)| \geq |w_i| + 1$, we get $z \neq \varepsilon$ and $f_i \dots f_k(xy)$ starts with $w_i x$. Finally let consider the case $\alpha = y$. Then $f_i \dots f_k(xy)$ starts with $w_i x$. Let z be the word such that $f_{i+1} \dots f_k(y) = w_{i+1} y z$. We have $f_i \dots f_k(y) = w_i f_i(z)$. If $z = \varepsilon$, $f_i \dots f_k(y)$ is a prefix of $f_i \dots f_k(xy)$. It follows $f_1 \dots f_k(y)$ is a prefix of $f_1 \dots f_k(xy)$: a contradiction. Thus $z \neq \varepsilon$ and $f_i \dots f_k(y)$ starts with $w_i \alpha = w_i y$. The property is once again verified. ∎

Now we end the proof of Proposition 6.1.

If P is not the identity morphism, there exist letters x, y such that $x \prec y$ and $P(y) \prec P(x)$. Taking w_1 defined as in Fact 6.3, $f(xy)$ starts with $P(w_1)P(x)$ and $f(y)$ starts with $P(w_1)P(y)$: f does not preserve the lexicographic order.

If P is the identity morphism, we apply Proposition 3.3. Let $n = \mathrm{Card}(A)$: $A = A_n$ with $a_1 = a$ and $a_n = b$. Let i, $1 \leq i < n$ and let m be the least integer such that $|f(a_i a_n^m)| \geq |f(a_{i+1})|$. As done for the previous fact, we can see that for each j, $1 \leq j \leq k+1$, there exists a word w_j such that $f_j \dots f_k(a_i a_n^m)$ starts with $w_j a_i$, $f_j \dots f_k(a_{i+1})$ starts with $w_j a_{i+1}$ (it is useful to prove simultaneously that $f_j \dots f_k(a_i a_{i+1})$ starts with $w_j a_i$). Consequently $f(a_i a_n^m) \prec f(a_{i+1})$. By Proposition 3.3, f is order-preserving. ∎

Let now study connections between episturmian morphisms and Lyndon morphisms. Let $\alpha \in A$. By Proposition 4.2, the morphism Ψ_α is a Lyndon morphism if and only if $\alpha = a$. Moreover the morphism $\overline{\Psi}_\alpha$ is a Lyndon morphism if and only if $\alpha = b$. More generally:

Proposition 6.4. *An episturmian morphism is a Lyndon morphism if and only if it belongs to the set*

$$(\{\overline{\Psi}_\alpha / \alpha \in A \backslash \{a\}\}^* \Psi_a)^* \overline{\Psi}_b^*.$$

Proof. First we prove that the condition is sufficient. We have already said that Ψ_a and $\overline{\Psi}_b$ are Lyndon morphisms.

Let prove that for $\alpha \in A \backslash \{a\}$, $\overline{\Psi}_\alpha$ preserves the Lyndon words starting with a. Let $u \in A^*$ such that au is a Lyndon word. Let $\alpha \in A$ with $\alpha \neq a$ and $P \neq \varepsilon$, $S \neq \varepsilon$ such that $\overline{\Psi}_\alpha(au) = PS$. If S starts with α, since PS starts with $a \prec \alpha$, $\overline{\Psi}_\alpha(au) \prec S$. Otherwise there exist non-empty words p, s such that $P = \overline{\Psi}_\alpha(p)$, $S = \overline{\Psi}_\alpha(s)$ and $au = ps$. Since au is a Lyndon word, $au \prec s$. Since $\overline{\Psi}_\alpha$ is order-preserving, $\overline{\Psi}_\alpha(au) \prec S$. The word $\overline{\Psi}_\alpha(au)$ is a Lyndon word.

For all x in A, $\Psi_a(x)$ is a Lyndon word starting with a. From what precedes, given any element f in $\{\overline{\Psi}_\alpha / \alpha \in A \backslash \{a\}\}^* \Psi_a$, for all x in A, $f(x)$ is a Lyndon word

starting with a. Moreover each morphism in $\{\overline{\Psi}_\alpha/\alpha \in A\backslash\{a\}\} \cup \{\Psi_a\}$ is an order-preserving morphism. So any morphism in $\{\overline{\Psi}_\alpha/\alpha \in A\backslash\{a\}\}^*\Psi_a$ is order-preserving. By Proposition 4.2 such a morphism is a Lyndon morphism.

Consequently any element in $(\{\overline{\Psi}_\alpha/\alpha\backslash\{a\}\}^*\Psi_a)^*\{\overline{\Psi}_b\}^*$ is a Lyndon morphism as a product of Lyndon morphisms.

Now we prove that the condition is necessary. Let f be an episturmian morphism which is a Lyndon morphism. By Proposition 4.2, f is order-preserving. If it is the identity, the result is obvious. Assume $f \neq Id$. By Proposition 6.1, there exist $n \geq 1$ morphisms $f_1, \ldots, f_n \in \{\Psi_\alpha, \overline{\Psi}_\alpha/\alpha \in A\}$ such that $f = f_1 \ldots f_n$.

If for all i, $1 \leq i \leq n$, $f_i \in \{\overline{\Psi}_\alpha/\alpha \in A\}$, then $f(b)$ starts with b. If moreover there exists j, $1 \leq j \leq n$, $f_j \neq \overline{\Psi}_b$ then $\alpha \prec b$ occurs in $f(b)$: a contradiction since $f(b)$ is a Lyndon word. Then $f \in \overline{\Psi}_b^*$.

Now assume there exists an integer i and a letter β in A, $1 \leq i \leq n$, such that $f_i = \Psi_\beta$. Choose i in such a way that for all j, $1 \leq j < i$, $f_j \in \{\overline{\Psi}_\alpha/\alpha \in A\}$. If $\beta \neq a$, the word $f_i \ldots f_n(a)$ is not a Lyndon word since it starts with β and contains the letter a. Since $\{\overline{\Psi}_\alpha/\alpha \in A\}$ is a set of order-preserving morphisms, $f(a)$ is not a Lyndon word. So $\beta = a$. Similarly we can prove that for each j, $1 \leq j < i$, $f_j \neq \overline{\Psi}_a$.

It follows that $f_1 \ldots f_i$ belongs to $\{\overline{\Psi}_\alpha/\alpha \in A\backslash\{a\}\}^*\Psi_a$. From the sufficient condition, it is a Lyndon morphism. By Lemma 5.2(2) and Proposition 4.2, $f_{i+1} \ldots f_n$ is a Lyndon morphism. The proof ends by induction. ∎

Note that when $\mathrm{Card}(A) = 2$ (that is in case of Sturmian morphisms), the monoid of Sturmian morphisms which are Lyndon morphisms is (finitely) generated by $\{\Psi_a, \overline{\Psi}_b\}$. A question is: is it still the case when $\mathrm{Card}(A) \geq 3$ or when considering episturmian order-preserving morphisms? To answer this question, we need a presentation of the monoid of episturmian morphisms with $\mathrm{Exch}(A) \cup \{\Psi_\alpha, \overline{\Psi}_\alpha/\alpha \in A\}$ as a set of generators.

Proposition 6.5. *The monoid* $\mathrm{Episturm}(A)$ *with* $\mathrm{Exch}(A) \cup \{\Psi_\alpha, \overline{\Psi}_\alpha/\alpha \in A\}$ *as a set of generators has the presentation given by Relations (2), (3) and the following relations (x, y, z, t are pairwise different letters):*

$$E_{xy}E_{xy} = Id, \tag{5}$$
$$E_{xy}E_{yz} = E_{yz}E_{zx}, \tag{6}$$
$$E_{xy}E_{zt} = E_{zt}E_{xy}, \tag{7}$$
$$\Psi_1\Psi_2 \ldots \Psi_k\overline{\Psi}_1 = \overline{\Psi}_1\overline{\Psi}_2 \ldots \overline{\Psi}_k\Psi_1 \tag{8}$$

where $k \geq 1$ is an integer and $\Psi_1, \ldots, \Psi_k \in \{\Psi_\alpha/\alpha \in A\}$ with $\Psi_1 \neq \Psi_i$ for all i, $2 \leq i \leq k$.

The proof of this proposition will use Lemma 6.2 and the following presentation proved in [16].

Proposition 6.6. *The monoid* Episturm(A) *with* Exch$(A) \cup \{\Psi_a, \overline{\Psi}_a\}$ *as a set of generators has the presentation given by Relations (5) to (7) and the following relations:*

$$E_{xy}\Psi_a = \Psi_a E_{xy} \quad \text{when } a \notin \{x, y\}, \tag{9}$$

$$E_{xy}\overline{\Psi}_a = \overline{\Psi}_a E_{xy} \quad \text{when } a \notin \{x, y\}, \tag{10}$$

$$\Psi_a E_1 \Psi_a E_2 \dots \Psi_a E_k \overline{\Psi}_a = \overline{\Psi}_a E_1 \overline{\Psi}_a E_2 \dots \overline{\Psi}_a E_k \Psi_a \tag{11}$$

where $k \geq 1$ is an integer and $E_1, \dots, E_k$ are exchange morphisms such that $E_1 \dots E_k(a) = a$, and for each integer i, $2 \leq i \leq k$, $E_i \dots E_k(a) \neq a$.

Proof of Proposition 6.5. We have already said that Relations (2) and (3) are valid. By Proposition 6.6, Relations (5) to (7) are valid.

Particular cases of Lemma 6.2 hold when for each i, $1 \leq i < k$, $f_i = \Psi_a$ and $f_k = \overline{\Psi}_a$, or when for each i, $1 \leq i < k$, $f_i = \overline{\Psi}_a$ and $f_a = \Psi_a$. Consequently Relation (8) can be stated using Relation (5), Relations (2), (3) and (11). This proves the validity or Relation (8) and so the validity of all relations in Proposition 6.5.

Using the same particular cases of Lemma 6.2, we deduce that Relation (11) can be stated using Relation (5), and Relations (2), (3) and (8). Relations (9) and (10) are particular cases of Relations (2) and (3). So all relations used in Proposition 6.6 can be stated using Relations in Proposition 6.6. By Proposition 6.6, any relations between elements of $(\text{Exch}(A) \cup \{\Psi_\alpha, \overline{\Psi}_\alpha / \alpha \in A\})^*$ can be deduced using Relations (5) to (7), and Relations (2), (3) and (8). ∎

We answer the previous question.

Proposition 6.7. *Let A be an alphabet.*

When Card$(A) \geq 2$*, the monoid of episturmian morphisms on A that are order-preserving is not finitely generated.*

When Card$(A) \geq 3$*, the monoid of episturmian morphisms on A that are Lyndon morphisms is not finitely generated.*

Proof. Recall that a (resp. b) is the least (resp. greatest) letter of A.

Let $n \geq 0$ be an integer. Let $f_n = \Psi_b^n \overline{\Psi}_a$: $f_n(a) = b^n a$, $f_n(b) = b^{n+1} a$, $f_n(\alpha) = b^n \alpha b^n a$ for $\alpha \in A \setminus \{a, b\}$. By Proposition 6.1, f_n is order-preserving. Let g, h be two order-preserving episturmian morphisms such that $f = gh$. By Proposition 6.1, $g, h \in \{\Psi_\alpha, \overline{\Psi}_\alpha / \alpha \in A\}^*$. By Proposition 6.5, $g = f$ and $h = Id$, or for an integer i, $0 \leq i \leq n$, $g = \Psi_b^i$ and $h = \Psi_b^{n-i} \overline{\Psi}_a$. We have $\Psi_b^i(a) = b^i a$ and $\Psi_b^i(b) = b$. Consequently since g is order-preserving, $i = 0$.

From what precedes, we deduce that for each $n \geq 0$, f_n must be an element of any set of generators of the monoid of order-preserving episturmian morphisms: this monoid is not finitely generated.

Similarly when Card$(A) \geq 3$, given a letter $\alpha \in A \setminus \{a, b\}$, the morphisms $\overline{\Psi}_\alpha^n \Psi_a$ for all $n \geq 0$ can be used to state that the monoid of Lyndon episturmian monoids is not finitely generated. ∎

Let us give some remarks.

Note that a consequence of Propositions 6.4 and 6.5 is that the monoid of Sturmian morphisms (case $\mathrm{Card}(A) = 2$) that are Lyndon morphisms is a free monoid on two generators: no relation exists between Ψ_a and $\overline{\Psi}_b$.

The monoid of episturmian morphisms is left and right cancellative [8, 16]. Consequently for any alphabet A ($\mathrm{Card}(A) \geq 2$), the monoid of order-preserving episturmian morphisms and the monoid of Lyndon episturmian morphisms are left and right cancellative.

The monoid of episturmian morphisms is left and right unitary [16]. Lemma 5.2 implies that the monoid of order-preserving episturmian morphisms and the monoid of Lyndon episturmian morphisms are left unitary.

When $\mathrm{Card}(A) = 2$, Corollary 5.6 implies that the monoid of Sturmian morphisms that preserve Lyndon words is right unitary.

For larger alphabet, this is not true. Indeed if $\alpha \in A \setminus \{a, b\}$, $\overline{\Psi}_\alpha \Psi_a$ and Ψ_a are Lyndon morphisms, but $\overline{\Psi}_\alpha$ is not a Lyndon morphisms.

When $\mathrm{Card}(A) \geq 2$, the monoid of order-preserving morphisms is not right unitary. Indeed $\Psi_b \overline{\Psi}_a$ and $\overline{\Psi}_a$ are order-preserving, but Ψ_b is not order-preserving.

To end this section, we give a more precise result than Proposition 6.1 when $\mathrm{Card}(A) = 2$.

Corollary 6.8. *Let $A = \{a < b\}$. A Sturmian morphism is order-preserving if and only if*

1. *it belongs to* $\{\Psi_a, \Psi_b, \overline{\Psi}_a, \overline{\Psi}_b\}^*$
2. *it does not belong to* $\mathrm{Episturm}(A)\Psi_b\Psi_a^*$.

Proof. We have to prove that for a Sturmian morphism f, $f(b)$ is a prefix of $f(ab)$ if and only if $f \in \mathrm{Episturm}(A)\Psi_b\Psi_a^*$. First note for $k \geq 0$, $\Psi_b\Psi_a^k(a) = ba$, $\Psi_b\Psi_a^k(b) = (ba)^k b$. Consequently for $f \in \mathrm{Episturm}(A)\Psi_b\Psi_a^*$, $f(b)$ is a prefix of $f(ab)$.

Conversely let f be a Sturmian morphism such that $f(b) = f\overline{\Psi}_b(b)$ is a prefix of $f(ab) = f\overline{\Psi}_b(a)$. There exists a word u such that $f\overline{\Psi}_b(a) = f\overline{\Psi}_b(b)u$. Let g be the morphism defined by $g(a) = u$, $g(b) = f\overline{\Psi}_b(b)$. We have $f\overline{\Psi}_b = g\Psi_b$. Since the monoid of Sturmian morphisms is right cancellative, g is Sturmian. By Proposition 6.5, $f\overline{\Psi}_b \in \mathrm{Episturm}(A)\Psi_b\Psi_a^*\overline{\Psi}_b$ and so $f \in \mathrm{Episturm}(A)\Psi_b\Psi_a^*$. ∎

7 About Fibonacci morphism

The aim of this section is to give examples of use of Lyndon morphisms (see Lemmas 7.2 and 7.3). We work on $A_2 = \{a < b\}$. These examples concern one particular Sturmian morphism which is the so-called Fibonacci morphism. This morphism, denoted φ, is defined by $\varphi(a) = ab$, $\varphi(b) = a$. That is $\varphi = \Psi_a E$ where E is the exchange morphism defined by $E(a) = b$ and $E(b) = a$.

The standard morphisms are the Sturmian morphisms that belong to $\{\varphi, E\}^*$. As a corollary of Corollary 6.8, we have:

Corollary 7.1. *A standard morphism is order-preserving if and only if it is a Lyndon morphism if and only if it belongs to $(\varphi E)^*$.*

In [12], Melançon gives a decomposition in finite Lyndon words of particular Sturmian words, called characteristic Sturmian words. When such a word is generated by a morphism, this morphism is standard. The previous result of Melançon implies that any standard morphism that generates an infinite word cannot be a Lyndon morphism. Corollary 7.1 is then without surprise since morphisms in $(\varphi E)^*$ are the only standard morphisms that do not generate infinite words: $(\varphi E)^n(a) = a$, $(\varphi E)^n(b) = a^n b$.

We give a new proof of the following result stated in [2].

Lemma 7.2. *The word $a\varphi^\omega(a)$ is a Lyndon infinite word.*

Proof. First observe that $\varphi\tilde{\varphi}(a) = aab$, $\varphi\tilde{\varphi}(b) = ab$. We prove by induction that for all $n \geq 0$,

$$a\varphi^{2n}(a) = (\varphi\tilde{\varphi})^n(a)a \tag{12}$$

$$a\varphi^{2n}(ba) = (\varphi\tilde{\varphi})^n(ab)a \tag{13}$$

This is immediate for $n = 0$ (remember $f^0 = Id$ for any morphism f). Assume Equations (12) and (13) hold for an integer n. Then $a\varphi^{2(n+1)}(a) = a\varphi^{2n}(aba) = (\varphi\tilde{\varphi})^n(aab)a = (\varphi\tilde{\varphi})^{n+1}(a)a$. Moreover

$$a\varphi^{2(n+1)}(ba) = a\varphi^{2n}(ababa) = (\varphi\tilde{\varphi})^n(aabab)a = (\varphi\tilde{\varphi})^{n+1}(ab)a.$$

It follows from Equation (12) that

$$a\varphi^\omega(a) = (\varphi\tilde{\varphi})^\omega(a).$$

But by Proposition 4.2, the morphism $\varphi\tilde{\varphi}$ is a Lyndon morphism. So the word $(\varphi\tilde{\varphi})^\omega(a)$ is a Lyndon infinite word ∎

Now we give a slightly new proof of the following result of Melançon.

Lemma 7.3. [11, 12] *The factorization in Lyndon words of the Fibonacci word is*

$$\varphi^\omega(a) = \Pi_{l\geq 0}(\varphi\tilde{\varphi})^l(ab).$$

In particular, for $l \geq 0$ $(\varphi\tilde{\varphi})^l(ab)$ is a Lyndon word and $(\varphi\tilde{\varphi})^l(ab) \prec (\varphi\tilde{\varphi})^{l+1}(ab)$.

Proof. In the proof of Lemma 7.2, we have already proved that $\varphi\tilde{\varphi}$ is a Lyndon morphism. So for $l \geq 0$, $(\varphi\tilde{\varphi})^l(ab)$ is a Lyndon word. For $l \geq 0$, $(\varphi\tilde{\varphi})^l(ab)$ is a prefix of $(\varphi\tilde{\varphi})^{l+1}(ab)$, so $(\varphi\tilde{\varphi})^l(ab) \prec (\varphi\tilde{\varphi})^{l+1}(ab)$.

To end the proof it is sufficient to state:

$$\forall n \geq 0, \quad \varphi^{2n}(a) = \left[\Pi_{l=0}^{n-1}(\varphi\tilde{\varphi})^i(ab)\right] a \tag{14}$$

This equation is immediately true for $n = 0$. For $n \geq 1$,

$$\varphi^{2n}(a) = \varphi^{2(n-1)}(\varphi^2(a)) = \varphi^{2(n-1)}(a)\varphi^{2(n-1)}(ba).$$

Thus by induction

$$\forall n \geq 0, \quad \varphi^{2n}(a) = a\left[\Pi_{l=0}^{n-1}\varphi^{2i}(ba)\right] \tag{15}$$

Using Equation (13), we get Equation (14). ∎

Acknowledgment. The author would like to thank P. Séébold for his remarks and his encouragements, and J. Berstel for his remarks after the presentation of the present work at the conference Journées Montoises 2002. Thanks also to the referees for their comments and to a referee of [17] that gives the idea of Lemma 6.2.

References

[1] J. Berstel and P. Séébold. Sturmian words. Chapter 2 in M. Lothaire, *Algebraic Combinatorics on Words*, volume 90 of Encyclopedia of Mathematics. Cambridge University Press, 2002, p45-110.

[2] J.P. Borel and F. Laubie. Quelques mots sur la droite projective réelle. *Journal de Théorie des Nombres de Bordeaux*, 5:23–51, 1993.

[3] K.T. Chen, R.H. Fox, and R.C. Lyndon. Free differential calculus IV – the quotient groups of the lower central series. *Ann. Math. 68*, 68:81–95, 1958.

[4] M. Crochemore. Régularités évitables (thèse d'état). Technical Report 83-43, LITP, November 1983.

[5] X. Droubay, J. Justin, and G. Pirillo. Episturmian words and some constructions of De Luca and Rauzy. *Theoretical Computer Science*, 255:539–553, 2001.

[6] J.-P. Duval. Factorizing words over an ordered alphabet. *Journal of Algebra*, 4:363–381, 1983.

[7] A. Ido and G. Melançon. Lyndon factorization of the Thue-Morse word and its relatives. *Discret Math. and Theoret. Comput. Sci.*, 1:43–52, 1997.

[8] J Justin and G. Pirillo. Episturmian words and episturmian morphisms. *Theoretical Computer Science*, 276(1-2):281–313, 2002.

[9] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics*. Addison-Wesley, 1983. Reprinted in 1997 by Cambridge University Press in the Cambridge Mathematical Library, Cambridge, UK, 1997.

[10] M. Lothaire. *Algebraic Combinatorics on words*, volume 90 of *Encyclopedia of Mathematics*. Cambridge University Press, Cambridge, UK, 2002.

[11] G. Melançon. Lyndon factorization of infinite words. In *STACS'1996*, volume 1996 of *Lecture Notes in Computer Science*, pages 147–154, 1996.

[12] G. Melançon. Lyndon factorization of sturmian words. *Discrete Mathematics*, 210:137–149, 2000.

[13] V. Mitrana. Primitive morphisms. *Information Processing Letters*, 64:277–281, 1997.

[14] C. Reutenauer, *Free Lie Algebras*, Oxford University Press, Oxford, 1993.

[15] G. Richomme. Test-words for Sturmian morphisms. *Bulletin of the Belgian Mathematical Society*, 6:481–489, 1999.

[16] G. Richomme. Conjugacy and episturmian morphisms. *Theoretical Computer Science*, 302, p1-34, 2003.

[17] G. Richomme. Some algorithms to compute the conjugates of episturmian morphisms. To appear in *RAIRO Theoretical Informatics and Applications* (see Technical Report 2002-10, LaRIA, 2002).

[18] G. Richomme. Some non finitely generated monoids of repetition-free endomorphisms. *Information Processing Letter*, 85(2):61–66, 2003.

[19] G. Richomme and F. Wlazinski. Finite test-sets for overlap-free morphisms. In K. Diks and W. Rytter, editors, *MFCS'2002*, volume 2420 of *Lecture Notes in Computer Science*, pages 605–614, 2002.

[20] G. Richomme and F. Wlazinski. Some results on k-power-free morphisms. *Theoretical Computer Science*, 273:119–142, 2002.

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

[22] P. Séébold. Lyndon factorization of the Prouhet words. Technical report, LaRIA 2002-02, March 2002. To appear in *Theoret. Comput. Sci.*

[23] R. Siromoney, L. Mathew, V.R. Dare, and K.G. Subramanian. Infinite Lyndon words. *Information Processing Letters*, 50:101–104, 1994.

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

[25] Z.-X. Wen and Z.Y. Wen. Local isomorphisms of invertible substitutions. *C. R. Acad. Sci. Paris*, 318, série I:299–304, 1994.

[26] Z.X. Wen and Z. Yiping. Some remarks on invertible substitutions on three letter alphabet. *Chin. Sci. Bulletin*, 44(19):1755–1760, 1999.

[27] F. Wlazinski. A test-set for k-power-free binary morphisms. *Theoretical Informatics and Applications*, 35:437–452, 2001.

LaRIA Université de Picardie Jules Verne
33 rue Saint Leu
F-80039 Amiens cedex 1
richomme@laria.u-picardie.fr