Discrete Mathematics and Theoretical Computer Science DMTCS vol. **9:2**, 2007, 89–108

On morphisms preserving infinite Lyndon words

Gwénaël Richomme

Université de Picardie Jules Verne, LaRIA,
33 Rue Saint Leu, 80039 Amiens cedex 1, France,
Gwenael.Richomme@u-picardie.fr

received 15 Nov 2004, revised 23 Apr 2005, accepted 2 May 2005.

In a previous paper, we characterized free monoid morphisms preserving finite Lyndon words. In particular, we proved that such a morphism preserves the order on finite words. Here we study morphisms preserving infinite Lyndon words and morphisms preserving the order on infinite words. We characterize them and show relations with morphisms preserving Lyndon words or the order on finite words. We also briefly study morphisms preserving border-free words and those preserving the radix order.

Keywords: Combinatorics on words, morphisms, Lyndon words, infinite words, lexicographic order

1 Introduction

Finite (or infinite) Lyndon words can be encountered in many studies (see for instance [10, 11, 12]). They are the non-empty words which are smaller in lexicographic order than all their proper suffixes. The Lyndon factorization theorem [3] states that any finite word can be decomposed uniquely as a product of non-increasing (in lexicographic order) Lyndon words. This result was extended to infinite words [21] (in such a case, the decomposition can end with an infinite Lyndon word). There are also studies concerning Lyndon decompositions of some specific infinite words (see for instance [2, 7, 13, 14, 20] for such results).

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 [11, chapter 4] or in [6, 16] where they are used to state that "every unbordered Sturmian word is a Lyndon word". In [18], these morphisms are called Lyndon morphisms. In particular, these morphisms are used to give a new proof of the decomposition [13, 14] of the Fibonacci word $\underline{F}$ and to prove that $a\underline{F}$ is an infinite Lyndon word [2].

Let us recall other results of [18]. A morphism is a Lyndon morphism if and only if it preserves the lexicographic order on finite words and if the image of each letter is a Lyndon word. Others characterizations provide algorithms to test if a morphism preserves Lyndon words or the order on finite words. We also note that if a morphism preserves the order on finite words, then it is injective (we say then that the morphism preserves strictly the order).

In [1], the fact that the Thue-Morse morphism μ ($\mu(a) = ab$, $\mu(b) = ba$) preserves the order on infinite words is observed. The study presented here aims to answer the natural questions: what can be said about morphisms preserving infinite Lyndon words or morphisms preserving (possibly strictly) the lexicographic order on infinite words? Do morphisms preserving infinite Lyndon words or morphisms preserving the lexicographic order on infinite words have testable characterizations? How are the families of these morphisms related to those of Lyndon morphisms and morphisms preserving the order on infinite words?

In Section 3, we prove that the families of morphisms previously introduced form a hierarchy: the set of Lyndon morphisms is included in the set of morphisms preserving infinite Lyndon words which is itself included in the set of non-erasing morphisms preserving the order on finite words which is itself included in the set of morphisms preserving *strictly* the order on infinite words which is itself included in the set of morphisms preserving the order on infinite words. All the inclusions are strict.

In Section 4, we characterize the morphisms that preserve infinite Lyndon words. We show that they are exactly the morphisms preserving finite Lyndon words of length n_0 for any given integer $n_0 \geq 2$. This characterization leads to an algorithm to check the property of the morphisms by firstly verifying that it preserves the order on finite words and by secondly examining the images of a finite number of Lyndon words of length at most 2.

In Section 5, we characterize morphisms preserving (resp. strictly) the order on infinite words. Once again we provide an algorithm to verify these properties of morphisms. We have already said that a morphism that preserves the order on infinite words does not do it necessarily strictly. Still in Section 5, we examine the frontier between these two families. We show that for binary, prefix or uniform morphisms this frontier is empty. This is also the case for the family of episturmian morphisms studied in Section 6. There, we also prove that an episturmian morphism preserves Lyndon words if and only if it preserves infinite Lyndon words. Let us recall that the episturmian morphisms that are Lyndon morphisms are known [18].

Sections 7 and 8 provide characterization of two other kinds of morphisms: the morphisms that preserve border-free words (the idea of this study comes from the fact that any Lyndon word is border-free) and the morphisms that preserve the radix order (another used order on words). In Section 9, we remark that the monoids of morphisms considered in this paper are not finitely generated.

An extended abstract of this paper was presented during the Journées Montoises 2004 [19].

2 Lyndon words

We recall here notions on words (see for instance [10, 11] for more details).

An *alphabet* A is a set of symbols called *letters*. Here we consider only finite alphabets. A *word over* A is a sequence of letters from A. The *empty word* ε is the empty sequence. Equipped with the concatenation operation, the set A^* of finite words over A is a free monoid with neutral element ε and set of generators A. We denote by A^ω the set of infinite words over A. As usually, for a finite word u and an integer n, the n^{th} power of u, denoted u^n, is the word ε if $n = 0$ and the word $u^{n-1}u$ otherwise. If u is not the empty word, u^ω denotes the infinite word obtained by infinitely repeating u. A finite word w is said *primitive* if for any word u, the equality $w = u^n$ (with n an integer) implies $n = 1$. Any word is the power of a unique primitive word called the *primitive root* of w. Two words x and y are *conjugate* if there exist words u and v such that $x = uv$ and $y = vu$. Any conjugate of a primitive word is primitive.

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 $|\varepsilon| = 0$. 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.

From now on 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 $a_1 < \dots < a_n$. Given an ordered alphabet A, we denote by $\preceq$ the lexicographic order whenever used on A^* or on A^ω. Let us recall that for two different (finite or infinite) words u and v, $u \prec v$ if and only if $u = xay$, $v = xbz$ with $a, b \in A$, $a < b$, $x \in A^*$, $y, z \in A^* \cup A^\omega$, or if (when u is finite) u is a proper prefix of v.

A non-empty finite word w is a *Lyndon word* if for all non-empty words u and v, $w = uv$ implies $w \prec vu$. This means that a Lyndon word is a primitive word smaller than all its conjugate. Equivalently [3, 10], a non-empty word w is a Lyndon word if all its non-empty proper 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. The following property will be sometimes used implicitly.

Proposition 2.1 (see [10] for instance) *A non-empty word w is a Lyndon word if and only if $|w| = 1$ or $w = uv$ with u and v two Lyndon words such that $u \prec v$.*

A consequence of this proposition is that, for Lyndon words u and v such that $u \prec v$, and for integers $n \geq 1$, $m \geq 1$, $u^n v^m$ is a Lyndon word.

The second definition of Lyndon words extends to infinite words: An infinite word is an infinite Lyndon word if all its proper suffixes are greater than it for the lexicographic order. The following useful result of Melançon [14] will also be sometimes used implicitly.

Lemma 2.2 *An infinite word is a Lyndon word if and only if it has an infinity of prefixes that are Lyndon words.*

It is easy to construct an infinite Lyndon word. Let A be an ordered alphabet with a its least letter, $n \geq 0$ be an integer and $\underline{w}$ be an infinite word having no occurrence of a^{n+1}. Thus $a^{n+1}\underline{w}$ is an infinite Lyndon word. See for instance [9] for another recent example of infinite Lyndon word.

Since on a one-letter alphabet there is only one Lyndon word and no infinite Lyndon word, *from now on, we consider only alphabets containing at least two letters.*

3 The hierarchy of morphisms

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 also say 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. A morphism extends naturally on infinite words.

A morphism f *preserves Lyndon words* if the image of any Lyndon word is a Lyndon word. As in [18], we call *Lyndon morphism* such a morphism. We define similarly morphisms that preserve infinite Lyndon words. A morphism f *preserves the lexicographic order* on finite (resp. infinite) words if for any finite (resp. infinite) words u and v, $u \preceq v$ implies $f(u) \preceq f(v)$. We say that f preserves *strictly* the lexicographic order if $u \prec v$ implies $f(u) \prec f(v)$. This means that f preserves the order and f is injective. Note that a necessary condition for preserving the order is that, for any letters a and b with $a \prec b$, $f(a) \preceq f(b)$. This is not sufficient: The morphism f defined by $f(a) = b$, $f(b) = ba$ does satisfy $f(a) \preceq f(b)$, but it does not preserve order, since, for instance, $f(b) = ba \preceq bb = f(aa)$.

Lyndon morphisms and morphisms preserving the order on finite words are studied in [18]. In particular, Lyndon morphisms are injective. This implies that they are non-erasing morphisms (an *erasing* morphism f is a morphism such that $f(a) = \varepsilon$ for at least one letter a). Morphisms preserving the order on finite words are also injective except for the empty morphism ϵ ($\epsilon(a) = \varepsilon$, for each letter a). When considering morphisms that preserve infinite Lyndon words or morphisms preserving the order on infinite words, *we* should also *consider non-erasing morphisms* to preserve the properties of infinite words.

The aim of this section is to show that the families of morphisms defined above form a hierarchy:

Proposition 3.1 *Each of the following sets is* strictly *included in the next one in the list:*

- *set of Lyndon morphisms,*
- *set of morphisms preserving infinite Lyndon words,*
- *set of non-erasing morphisms preserving (strictly) the order on finite words,*
- *set of morphisms preserving* strictly *the order on infinite words,*
- *set of morphisms preserving the order on infinite words.*

The next examples show that the inclusions are strict.

Example 1 Let $n \geq 2$ and $(m_i)_{1 \leq i \leq n}$ be non-zero integers. The morphism f defined on A_n by $f(a_i) = a_i^{m_i}$ is an example of morphism preserving infinite Lyndon words (see Theorem 4.5). But when at least one of the m_i is different from 1, f is not a Lyndon morphism since the image of the Lyndon word a_i is not a Lyndon word.

Example 2 The Thue-Morse morphism μ defined on $\{a < b\}$ by $\mu(a) = ab$, $\mu(b) = ba$ is a morphism preserving the order on finite words [1]. On the other hand, the image of the Lyndon word $abab^\omega$ is not a Lyndon word: $\mu(abab^\omega) = abbaab(ba)^\omega$.

Example 3 The morphism f defined on $\{a < b\}$ by $f(a) = ba$, $f(b) = b$ is an example of a morphism that preserves the lexicographic order on infinite words (see Corollary 5.3) but that does not preserve it on finite words: $a \preceq b$ but $f(b) \prec f(a)$ ($f(b)$ is a prefix of $f(a)$).

Example 4 A morphism f is *periodic* if there exists a word z such that $f(x)$ is a power of z for any word x. When f is not erasing, f preserves the order on infinite words (all infinite words have z^ω as image) but does not do it strictly.

Now we have to prove the inclusions claimed by Proposition 3.1. For this, we recall the following characterization:

Proposition 3.2 [18] *Let f be a non-empty morphism on A_n ($n \geq 2$). The morphism f preserves the lexicographic order on finite words 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 $|f(a_i a_n^{k_i})| \geq |f(a_{i+1})|$.*

We also need the following result:

Lemma 3.3 *Let $x, y \in A^*$ with $x \neq \varepsilon$ and let $n \geq 1$ be an integer. The infinite word $x^n y x^\omega$ is not an infinite Lyndon word.*

Proof of Lemma 3.3: Assume $x^n y x^\omega$ is an infinite Lyndon word. By Lemma 2.2, it has infinitely many prefixes which are (finite) Lyndon words. Consequently there exists a non-empty prefix p of x (possibly $p = x$) such that $x^n y x^m p$ is a Lyndon word for infinitely many integers m. But as p is a prefix and a suffix of $x^n y x^m p$ this one is not a Lyndon word: contradicion. $\square$

Proofs of the inclusions in Proposition 3.1

Inclusion 1: Any Lyndon morphism preserves infinite Lyndon words. This is an immediate consequence of the definition of Lyndon morphisms and of Lemma 2.2.

Inclusion 2: Any (non-erasing) morphism preserving infinite Lyndon words preserves the order on finite words. Let f be a non-erasing morphism that does not preserve the order on finite words. Let us show that f does not preserve the infinite Lyndon words. By Proposition 3.2, there exist two integers i and k such that $f(a_{i+1}) \preceq f(a_i a_n^k)$ and $1 \leq i < n$. Let us consider the infinite Lyndon word $a_i a_n^k a_{i+1}^\omega$.

If $f(a_{i+1})$ is not a prefix of $f(a_i a_n^k)$ then $f(a_i a_n^k a_{i+1}^\omega)$ is not an infinite Lyndon word. Indeed for any prefix v of $f(a_{i+1}^\omega)$, $f(a_i a_n^k a_{i+1})v$ is a prefix of $f(a_i a_n^k a_{i+1}^\omega)$ without being a Lyndon word since $f(a_{i+1})v \preceq f(a_i a_n^k a_{i+1})v$.

If $f(a_{i+1})$ is a prefix of $f(a_i a_n^k)$ then there exists a word y such that $f(a_i a_n^k) = f(a_{i+1})y$. Then $f(a_i a_n^k a_{i+1}^\omega) = xyx^\omega$ with $x = f(a_{i+1})$. Since f is not erasing, $x \neq \varepsilon$. By Lemma 3.3, $f(a_i a_n^k a_{i+1}^\omega)$ is not an infinite Lyndon word.

Inclusion 3: Any non-erasing morphism that preserves the order on finite words also preserves strictly the order on infinite words. Let f be a non-erasing morphism that preserves the order on finite words. Let $\underline{u}$ and $\underline{v}$ be two infinite words such that $\underline{u} \prec \underline{v}$. Let p be the longest common prefix of $\underline{u}$ and $\underline{v}$. Let x (resp. y) be the letter such that px (resp. py) is a prefix of $\underline{u}$ (resp. of $\underline{v}$). There exists a word w such that pxw is a prefix of u and $|f(pxw)| \geq |f(py)|$. From $pxw \prec py$, we get $f(pxw) \prec f(py)$. Since $|f(pxw)| \geq |f(py)|$, for any words u' and v' (finite or infinite), $f(pxw)u' \prec f(py)v'$. It follows $f(\underline{u}) \prec f(\underline{v})$.

Inclusion 4 is an immediate consequence of the definition of morphisms preserving (resp. strictly) the order on infinite words. □

A consequence of Proposition 3.1 is that known properties of one of the families can be immediately deduced for all the families it contains. For instance, one can see that morphisms preserving infinite Lyndon words are injective. As shown by Example 4, it is not the case of morphisms preserving order on infinite words. But it is the case of (non-erasing) morphisms preserving strictly the order on infinite words. Indeed if f is a non-injective and non-erasing morphism, there exist distinct words u and v such that $u \prec v$ and $f(u) = f(v)$. Since f is not erasing and $u \neq v$, u is not a prefix of v. So $u^\omega \prec v^\omega$ and $f(u^\omega) = f(v^\omega)$: f does not preserve strictly the order on infinite words.

To end this section let us note that Example 1 provides examples of morphisms that preserve infinite Lyndon words without preserving primitive words (As in [15], we call *primitive* a morphism that preserves primitive words). This situation does not hold for Lyndon morphisms as shown by the following new property.

Lemma 3.4 *Any Lyndon morphism preserves primitive words.*

Proof: Assume a morphism f is not primitive. There exist primitive words u, v and an integer $k \geq 2$ such that $f(u) = v^k$. There exist words x, y such that $u = xy$ and yx is a Lyndon word. We have $f(yx) = (v_2v_1)^k$ for some words v_1, v_2 such that $v = v_1v_2$. Since $(v_2v_1)^k$ is not a Lyndon word, f is not a Lyndon morphism. □

4 Morphisms preserving infinite Lyndon words

The aim of this section is to prove Theorem 4.5 which provides characterizations of morphisms preserving infinite Lyndon words. Before we prove several useful results.

Proposition 4.1 *Let f be a non-erasing morphism on A_n that preserves infinite Lyndon words. For any Lyndon word $u \neq a_n$ over A_n, $f(u)$ is a power of a Lyndon word.*

Proof: Let r be the primitive root of $f(u)$ and $m \geq 1$ be the integer such that $f(u) = r^m$. From $u \neq a_n$, we get $u \prec a_n$. Since u and a_n are Lyndon words, by Proposition 2.1, for any integer $p \geq 1$, $u^2a_n^p$ is a Lyndon word. So $u^2a_n^\omega$ is an infinite Lyndon word. If r is not a Lyndon word, there exist two non-empty words x and y such that $r = xy$ with $yx \preceq xy$. Since r is primitive, $yx \prec xy$. The word $f(u^2a_n^\omega)$ starts with xy. Let P be a prefix of $f(a_n^\omega)$. The word $yxyP$ is a suffix of $f(u^2)P$. From $yx \prec xy$, we deduce that $f(u^2)P$ is not a Lyndon word. So the only prefixes of $f(u^2a_n^\omega)$ that can be Lyndon words are the prefixes of $f(u^2)$. Hence f does not preserve the infinite Lyndon words. □

Proposition 4.2 *A morphism that preserves infinite Lyndon words also preserves Lyndon words of length 2.*

Proof: Any Lyndon word of length 2 is of the form ab with a, b letters such that $a \prec b$. Let f be a morphism that preserves infinite Lyndon words. By Proposition 4.1, there exist a Lyndon word w and an integer $k \geq 1$ such that $f(ab) = w^k$. Assume $k \geq 2$. We distinguish two cases.

Case $|f(a)| \geq |w|$. In this case $f(a) = w^i w_1$ for an integer $i \geq 0$ and a non-empty prefix w_1 of w (if $i = 0$, $w_1 = w$). By Proposition 4.1, $f(a) = z^j$ for an integer $j \geq 1$ and a Lyndon word z. Hence $w = z^l z_1$ for an integer l and a non-empty suffix z_1 of z. Since w is a Lyndon word, $w = z_1$ ($l = 0$). So $|z| \geq |w|$. Hence $|z| = |w|$, that is $z = w$, otherwise w_1 is a non-empty prefix and suffix of the Lyndon word z. It follows $f(b) \in w^*$ and so $f(ab^\omega)$ is not a Lyndon word.

Case $|f(a)| < |w|$. There exists a word w_2 such that $w = f(a)w_2$, $f(b) = w_2 w^{k-1}$. Although $abbabb^\omega$ is a Lyndon word, the word $f(abbabb^\omega)$ is not a Lyndon word. Indeed it starts with $w^k w_2 w^{k+1} w^{k-2}$. Since w is a Lyndon word and w_2 is one of its proper suffixes $w \prec w_2$ and $w^{k+1} \prec w^k w_2$. So we cannot have $k \geq 2$. Hence $k = 1$. $\square$

Lemma 4.3 *If u and v are two non-empty words such that uv is a Lyndon word, then uvv is also a Lyndon word.*

Proof of Lemma 4.3: Let s be a non-empty suffix of uvv. If s is a suffix of v then as uv is a Lyndon word $uv \prec s$ whence $uvv \prec s$. Otherwise $s = s'v$ for a non-empty suffix s' of uv. Again $uv \prec s'$ implies $uvv \prec s'v = s$. $\square$

Let us remark that the converse of Lemma 4.3 does not hold (take $u = aba$, $v = b$).

Lemma 4.4 *Let u be a Lyndon word, v be a primitive word different from u, and let $n \geq 1, m \geq 1$ be two integers. If $u^n \prec v^m$, then $u \prec v$.*

Proof of Lemma 4.4: Assume by contradiction that $v \preceq u \preceq u^n \prec v^m$. Then necessarily v is a prefix of u. Let $i \geq 1$ be the greatest integer such that v^i is a prefix of u. Note that $m \geq i+1$. Let x be the word such that $u = v^i x$. Since u is a Lyndon word, and $u \neq v$, we have $x \neq \varepsilon$ and $u \prec x$. In particular x is not a prefix of v. By construction, v is not a prefix of x. So from $v^i x \prec x$, we deduce the existence of words α, β and γ and letters a, b such that $a \prec b$, $v = \alpha a \beta$ and $x = \alpha b \gamma$. We get a contradiction with $u^n \prec v^m$ since u^n starts with $v^i \alpha b \gamma$ and v^m starts with $v^i \alpha a \beta$. $\square$

One can also observe that the converse of Lemma 4.4 does not hold (take $u = ab$, $v = aba$), but as a consequence of Proposition 2.1, the reader can verify that the converse holds if we assume also that v is a Lyndon word.

Theorem 4.5 *Let f be a non-erasing morphism on A_n. The following assertions are equivalent:*

1. *f preserves infinite Lyndon words;*
2. *f preserves the order on finite words and for each i, $1 \leq i < n$,*
   - *$f(a_i)$ is a power of a Lyndon word, and*
   - *$f(a_i a_n)$ is a Lyndon word;*

3. *given an integer $n_0 \geq 2$, for any Lyndon word u of length at least n_0, $f(u)$ is a Lyndon word.*

Proof: $3 \Rightarrow 1$ is a direct consequence of Lemma 2.2.

$1 \Rightarrow 2$ is a direct consequence of Propositions 3.1, 4.1 and 4.2.

Proof of $2 \Rightarrow 3$. Let f be a morphism verifying Assertion 2. Let u be a Lyndon word of length at least 2. We prove by induction on $|u|$ that $f(u)$ is a Lyndon word.

For this, let us first observe that for $1 \leq i < j < n$, $a_i a_j$ is a Lyndon word. Moreover $f(a_i a_j)$ is a Lyndon word. Indeed by hypotheses, there exist Lyndon words u_i and u_j and integers m_i and m_j such that $f(a_i) = u_i^{m_i}$ and $f(a_j) = u_j^{m_j}$. Since f preserves the order on finite words, $f(a_i) \prec f(a_j)$. By Lemma 4.4, $u_i \prec u_j$. By Proposition 2.1, $u_i^{m_i} u_j^{m_j} = f(a_i a_j)$ is a Lyndon word. Hence f preserves Lyndon words of length 2.

Now assume $|u| \geq 3$. Choose w and v such that $u = vw$ and w is the proper suffix of u of greatest length which is a Lyndon word. Proposition 5.1.3 in [10] shows that v is a Lyndon word and moreover $v \prec w$. Of course $v \neq a_n$ and so by hypotheses and by inductive hypothesis, we deduce that $f(v)$ is a power of a Lyndon word. If $w \neq a_n$, $f(w)$ is also a power of a Lyndon word. As for case $|u| = 2$, $f(u) = f(vw)$ is a Lyndon word. If $w = a_n$, we must have $u = a_i a_n^{|u|-1}$. Observe that by hypotheses $f(a_i a_n)$ is a Lyndon word. Hence Lemma 4.3 implies that $f(a_i a_n^k)$ is a Lyndon word for any integer $k \geq 1$. In particular $f(u)$ is a Lyndon word. □

We end this section with some remarks.

First of all, Assertion 2 provides an algorithm to test if a morphism preserves infinite Lyndon words. Indeed by Proposition 3.2, we know how to verify if a morphism preserves the order on finite words. Moreover it is easy to compute the primitive root of a word, and there exists a well-known algorithm to test if a word is a Lyndon one [5]. A more precise analysis can show that the algorithm provided by the previous result can be implemented in time $O(\max\{|f(ab)| \mid a \neq b\})$.

We observe also that the sufficient and necessary conditions given by Assertion 2 are close to the one obtained in [18]:

Proposition 4.6 [18] *Let $n \geq 2$ be an integer. A morphism f on A_n is a Lyndon morphism if and only if it satisfies the following two properties:*

1. *$f(a)$ is a Lyndon word for each $a \in A_n$,*

2. *f preserves the order on finite word.*

Finally let us observe that a_n cannot be considered in Proposition 4.1 as shown in the following example.

Example 5 The morphism f defined on $\{a < b\}$ by $f(a) = aa$ and $f(b) = bab$ preserves the lexicographic order (this is a direct consequence of the fact that $f(a)$ and $f(b)$ start respectively with a and b). Moreover $f(a)$ is a power of the Lyndon word a and $f(ab) = ababb$ is a Lyndon word. Thus by Theorem 4.5, f preserves infinite Lyndon words but $f(b)$ is not a power of a Lyndon word.

5 Morphisms preserving order on infinite words

Here we study morphisms that preserve (resp. strictly) the order on infinite words. We have already noticed that there exist morphisms that preserve the order on infinite words without preserving it on finite words. The following proposition explains the main difference between these two families of morphisms.

Proposition 5.1 *A non-erasing morphism f preserves the order on infinite words if and only if for all* finite *words u and v, if $u \prec v$ then $f(u) \preceq f(v)$ or $f(v)$ is a prefix of $f(u)$.*

Proof: Assume first that f preserves the order on infinite words. Let u and v be two finite words such that $u \prec v$. If u is a prefix of v, then $f(u) \prec f(v)$. Else $u^\omega \prec v^\omega$. Since $f(u^\omega) \preceq f(v^\omega)$, $f(v)$ is a prefix of $f(u)$ or $f(u) \preceq f(v)$.

To prove the converse, assume that f does not preserve the order on infinite words. Let $\underline{x}$, $\underline{y}$ be two infinite words such that $\underline{x} \prec \underline{y}$ and $f(\underline{y}) \prec f(\underline{x})$. There exist a finite word α, letters a, b and infinite words $\underline{X}$, $\underline{Y}$ such that $a \prec b$, $f(\underline{x}) = \alpha b\underline{X}$, $f(\underline{y}) = \alpha a\underline{Y}$. Let u (resp. v) be a prefix of $\underline{x}$ (resp. $\underline{y}$) such that αb (resp. αa) is a prefix of $f(u)$ (resp. $f(v)$). By construction, $f(v) \prec f(u)$ and $f(v)$ is not a prefix of $f(u)$. So u is not a prefix of v and v is not a prefix of u. By definition of u and v and since $\underline{x} \prec \underline{y}$, we have $u \prec v$. $\square$

5.1 Characterizations

The following theorem states our characterization of morphisms preserving the lexicographic order on infinite words.

Theorem 5.2 *A non-erasing morphism f on A_n preserves the lexicographic order on infinite words if and only if $f(a_1a_n) \preceq f(a_na_1)$ and for all i, $1 \le i < n$, $f(a_ia_n^\omega) \preceq f(a_{i+1}a_1^\omega)$.*

For letters a and b, one can observe that $f(ab^\omega) \preceq f(ba^\omega)$ is equivalent to $f(ab) \preceq f(ba)$. Consequently:

Corollary 5.3 *A non-erasing morphism f on $\{a < b\}$ preserves the lexicographic order on infinite words if and only if $f(ab) \preceq f(ba)$.*

To prove Theorem 5.2, we need the next three lemmas.

Lemma 5.4 *Let u and v be two non-empty words. If $uv = vu$ then $u^\omega = v^\omega$. If $uv \prec vu$ then $u^\omega \prec v^\omega$.*

Proof: of Lemma 5.4 If $uv = vu$, then u and v are powers of a same word z (see [10] for instance). Thus $u^\omega = z^\omega = v^\omega$.

When $uv \prec vu$, we can show by induction that, for any integers $i \neq 0$, $j \neq 0$, $u^iv^j \prec v^ju^i$. In particular, $u^{|v|}v^{|u|} \prec v^{|u|}u^{|v|}$. Since $|v^{|u|}| = |u^{|v|}|$, $u^{|v|} \prec v^{|u|}$ and $u^\omega \prec v^\omega$. $\square$

Lemma 5.5 *For non-empty words u and v, if $uv^\omega \preceq wu^\omega$ and $uv \preceq vu$, then $uw \preceq wu$.*

Proof: of Lemma 5.5 Assume by contradiction that $wu \prec uw$. By induction, for any integer i, we can prove $wu^i \prec u^iw$. In particular, $wu^{|v|+1} \prec u^{|v|+1}w$. Since $|wu^{|v|+1}| = |u^{|v|+1}w|$, $wu^\omega = wu^{|v|+1}u^\omega \prec u^{|v|+1}wu^\omega$. From $uv^\omega \preceq wu^\omega$, we get $v^\omega \prec u^{|v|}wu^\omega$. So for length reason, $v^{|u|} \prec u^{|v|}$.

The hypothesis $uv \preceq vu$ implies $u^{|v|}v^{|u|} \preceq v^{|u|}u^{|v|}$. Once again for length reason $u^{|v|} \preceq v^{|u|}$: a contradiction. □

Similarly one can prove:

Lemma 5.6 *For non-empty words u and v, if $wv^\omega \preceq vu^\omega$ and $uv \preceq vu$, then $wv \preceq vw$.*

Proof: of Theorem 5.2 Let f be a non-erasing morphism on A_n that preserves the lexicographic order on infinite words. By definition, for all letters x, y with $x \prec y$, $f(xa_n^\omega) \preceq f(ya_1^\omega)$. This is true in particular when $x = a_i$ and $y = a_{i+1}$ for $1 \leq i < n$. This is also true when $x = a_1$ and $y = a_n$. In this last case, $f(a_1a_n^\omega) \preceq f(a_na_1^\omega)$ implies, for length reason, $f(a_1a_n) \preceq f(a_na_1)$.

Assume now that f is a non-erasing morphism on A_n such that $f(a_1a_n) \preceq f(a_na_1)$ and for all i, $1 \leq i < n$, $f(a_ia_n^\omega) \preceq f(a_{i+1}a_1^\omega)$. We first observe that Lemma 5.4 implies $f(a_1^\omega) \preceq f(a_n^\omega)$. Consequently, for any integer i, $1 \leq i < n$, $f(a_{i+1}a_1^\omega) \preceq f(a_{i+1}a_n^\omega)$. So by induction, we can prove that for i and j such that $1 \leq i < j \leq n$, $f(a_ia_n^\omega) \preceq f(a_ja_1^\omega)$. In other words for $x, y \in A_n$ such that $x \prec y$, $f(xa_n^\omega) \preceq f(ya_1^\omega)$.

Let $x \in A_n$. When $x \neq a_1$, from $f(a_1a_n^\omega) \preceq f(xa_1^\omega)$ and $f(a_1a_n) \preceq f(a_na_1)$, Lemma 5.5 implies $f(a_1x) \preceq f(xa_1)$. This also holds when $x = a_1$. Using Lemma 5.6, we also deduce that $f(xa_n) \preceq f(a_nx)$.

To prove that f preserves the lexicographic order on infinite words, it is sufficient to prove that $f(x\underline{u}) \preceq f(y\underline{v})$ when x, y are two letters such that $x \prec y$ and $\underline{u}$ and $\underline{v}$ are infinite words. Since $x\underline{u} \preceq xa_n^\omega \preceq ya_1^\omega \preceq y\underline{v}$ and since $f(xa_n^\omega) \preceq f(ya_1^\omega)$, it is sufficient to prove $f(\underline{u}) \preceq f(a_n^\omega)$ and $f(a_1^\omega) \preceq f(\underline{v})$. We just prove $f(\underline{u}) \preceq f(a_n^\omega)$ (the proof of the other inequality is similar).

Assume by contradiction $f(a_n^\omega) \prec f(\underline{u})$. There exists an integer p, a prefix u_1 of $\underline{u}$, words w_1, w_2, w_3, letters α, β such that $\alpha \prec \beta$, $f(a_n^p) = w_1\alpha w_2$, $f(u_1) = w_1\beta w_3$. Then $f(a_n^p) \prec f(u_1) \preceq f(u_1a_n^p)$. Since for all letters x, $f(xa_n) \preceq f(a_nx)$, we can get $f(u_1a_n^p) \preceq f(a_n^pu_1)$. So $f(a_n^p) \prec f(u_1) \preceq f(a_n^pu_1)$. Since $f(a_n^p)$ is not a prefix of $f(u_1)$, this is impossible. Consequently $f(\underline{u}) \preceq f(a_n^\omega)$. □

One question is: How to test whether a morphism preserves the lexicographic order on infinite words? For binary morphisms, Corollary 5.3 gives an answer. The next proposition treats the general case. With Theorem 5.2, it shows that to check whether or not a morphism preserves the lexicographic order on infinite words it is sufficient to check the preservation of the order only on a finite set of couples of finite words. The couples of words to be tested are (a_1, x) and (x, a_n) for letters x and the couples (xa_n^i, ya_1^j) where x, y, i, j are as in the next proposition.

Proposition 5.7 *Let $n \geq 2$ be an integer, f be a non-erasing morphism on A_n, x, y be letters such that $x \prec y$ and $m = \max(|f(x)|, |f(y)|) + |f(a_1)| + |f(a_n)| - \gcd(|f(a_1)|, |f(a_n)|)$. The following assertions are equivalent:*

1. *$f(xa_n^\omega) \preceq f(ya_1^\omega)$,*

2. *$f(xa_n^i) \preceq f(ya_1^j)$ or $f(ya_1^j)$ is a prefix of $f(xa_n^i)$ where i and j are the least integers such that $|f(xa_n^i)| \geq m$ and $|f(ya_1^j)| \geq m$.*

The proof of this result is a consequence of Fine and Wilf's theorem (see [11] for instance) recalled below. This explains the *gcd* that appears in the statement.

Theorem 5.8 (Fine and Wilf's theorem) *Let u and v be two non-empty words. If a power of u and a power of v have a common prefix of length $|u| + |v| - \gcd(|u|, |v|)$, then u and v are powers of a same word.*

Proof: of Proposition 5.7 If for integers i, j, $f(ya_1^j) \prec f(xa_n^i)$ and $f(ya_1^j)$ is not a prefix of $f(xa_n^i)$, then $f(ya_1^\omega) \prec f(xa_n^\omega)$.

Assume now $f(xa_n^i) \preceq f(ya_1^j)$ or $f(ya_1^j)$ is a prefix of $f(xa_n^i)$ where $m = \max(|f(x)|, |f(y)|) + |f(a_1)| + |f(a_n)| - \gcd(|f(a_1)|, |f(a_n)|)$, i and j are (the least) integers such that $|f(xa_n^i)| \geq m$ and $|f(ya_1^j)| \geq m$.

If $f(xa_n^i) \preceq f(ya_1^j)$ and $f(xa_n^i)$ is not a prefix of $f(ya_1^j)$ then $f(xa_n^\omega) \preceq f(ya_1^\omega)$.

If $f(xa_n^i)$ is a prefix of $f(ya_1^j)$ or if $f(ya_1^j)$ is a prefix of $f(xa_n^i)$, then by hypotheses on i and j, $f(a_1^\omega)$ and $f(a_n^\omega)$ have a common factor of length $|f(a_1)|+|f(a_n)|-\gcd(|f(a_1)|, |f(a_n)|)$. It follows by Fine and Wilf's theorem that $f(a_1)$ and $f(a_n)$ are powers of conjugate words ($f(a_1) = (uv)^\alpha$, $f(a_n) = (vu)^\beta$ for words u and v and integers α and β). Then for any integers k and l such that $k \geq i$ or $l \geq j$, we get $f(xa_n^k)$ is a prefix of $f(ya_1^l)$ or $f(ya_1^l)$ is a prefix of $f(xa_n^k)$. This implies $f(xa_n^\omega) = f(ya_1^\omega)$. □

To end this part, let us note that, when considering morphisms preserving strictly the order on infinite words, we can get similarly as previously:

Theorem 5.9 *A non-erasing morphism f on A_n preserves* strictly *the lexicographic order on infinite words if and only if $f(a_1a_n) \prec f(a_na_1)$ and for all i, $1 \leq i \leq n$, $f(a_ia_n^\omega) \prec f(a_{i+1}a_1^\omega)$.*

Corollary 5.10 *A morphism f qon $\{a < b\}$ preserves strictly the order on infinite words if and only if $f(ab) \prec f(ba)$.*

Proposition 5.11 *Let $n \geq 2$ be an integer, f be a non-erasing morphism on A_n, x, y be letters such that $x \prec y$ and $m = \max(|f(x)|, |f(y)|) + |f(a_1)| + |f(a_n)| - \gcd(|f(a_1)|, |f(a_n)|)$. The following assertions are equivalent:*

1. *$f(xa_n^\omega) \prec f(ya_1^\omega)$,*
2. *$f(xa_n^i) \prec f(ya_1^j)$ and $f(xa_n^i)$ is not a prefix of $f(ya_1^j)$ where i and j are the least integers such that $|f(xa_n^i)| \geq m$ and $|f(ya_1^j)| \geq m$.*

5.2 Frontier

Now we study the frontier between the two families of morphisms characterized in Section 5.1. In Example 4, we have already mentioned that periodic morphisms preserve the order on infinite words but not strictly. The next proposition shows that these morphisms are the only ones which preserve the order on infinite morphisms without being injective.

Proposition 5.12 *A non-erasing morphism f that preserves the lexicographic order on infinite words is not injective if and only if it is periodic.*

Proof: Of course, if a morphism is periodic, it is not injective (Recall that we consider alphabet with at least two letters).

Now assume that f is a non-erasing morphism preserving the lexicographic order and assume that u and v are two different finite words such that $f(u) = f(v)$. Assume $u \prec v$ (case $v \prec u$ is symmetric to the present case). The word u is not a prefix of v since $f(u) = f(v)$, $u \neq v$ and f is not erasing. Hence for any letters x, y with $x \prec y$, $uy^\omega \prec vx^\omega$. Since f preserves the order on infinite words, $f(x^\omega) \preceq f(y^\omega)$ and $f(uy^\omega) \preceq f(vx^\omega)$. Since $f(u) = f(v)$, we get $f(x^\omega) = f(y^\omega)$ and so (by Fine and Wilf's theorem) $f(x)$ and $f(y)$ are powers of a same word. Since this is true for each pair of letters $\{x, y\}$, f is periodic. $\square$

Now the question is: do there exist morphisms which are not periodic and that preserve the order on infinite words but not strictly? The answer is yes as shown by the following examples.

Example 6 The morphism f defined on $\{a < b\}$ by $f(a) = ab$, $f(b) = abb$, $f(c) = ba$ preserves the order but not strictly. Indeed $f(ac^\omega) = f(ba^\omega) = ab(ba)^\omega$. Note that for any infinite words $\underline{u} \neq ac^\omega$ and $\underline{v} \neq ba^\omega$, if $\underline{u} \prec \underline{v}$, then $f(\underline{u}) \prec f(\underline{v})$.

Example 7 The same property holds for the morphism f defined by $f(a) = ab$, $f(b) = ac$, $f(c) = acb$, $f(d) = ba$. For this morphism, $f(bd^\omega) = f(ca^\omega) = ac(ba)^\omega$. We note also that for any infinite words $\underline{u}$ and $\underline{v}$ such that $\underline{u} \neq bd^\omega$ or $\underline{v} \neq ca^\omega$, if $\underline{u} \prec \underline{v}$, then $f(\underline{u}) \prec f(\underline{v})$.

But in some special cases, the answer to the previous question is no. Let us recall two definitions. A morphism is *uniform* if the images of the letters have all the same length. A morphism f on A is *prefix* if for all a, b in A, $f(a)$ is not a prefix of $f(b)$.

Proposition 5.13 *Let f be a non-periodic morphism defined on an ordered alphabet A. If A contains two letters, or if f is prefix or uniform, then f preserves the order on infinite words if and only if it does it strictly.*

Proof: By Proposition 3.1, we have just to prove the only if part. Let f be a non-periodic morphism defined on an alphabet A. Assume that f preserves the order on infinite words.

Case Card$(A) = 2$. By Theorem 5.2, $f(ab) \preceq f(ba)$ and $f(ab^\omega) \preceq f(ba^\omega)$. Since f is not periodic, $f(ab) \prec f(ba)$ and consequently $f(ab^\omega) \prec f(ba^\omega)$. By Theorem 5.9, f preserves strictly the order on infinite words.

Case f prefix. Let $\underline{u}$ and $\underline{v}$ be two different infinite words. We can write $\underline{u} = px\underline{u'}$ and $\underline{v} = py\underline{v'}$ for a word p, two letters x and y and two infinite words $\underline{u'}$ and $\underline{v'}$. Assume $\underline{u} \prec \underline{v}$, then $x \prec y$. Since $f(px\underline{u'}) = f(\underline{u}) \preceq f(\underline{v}) = f(py\underline{v'})$, since $f(x)$ is not a prefix of $f(y)$ and $f(y)$ is not a prefix of $f(x)$ then $f(x) \prec f(y)$ and $f(\underline{u}) \prec f(\underline{v})$.

Case f uniform. If $f(x) = f(y)$ for two letters x and y, then f is not injective. By Proposition 5.12, f is periodic: a contradiction. So f is prefix, and the result holds by the previous case.

□

Proposition 5.13 leads to consider the hierarchy given in Proposition 3.1 for cases treated by it. Examples 1 and 2 show that in this cases the set of Lyndon morphisms is different from the set of morphisms preserving infinite Lyndon words which is itself different from the set of morphisms preserving the order on finite words. Moreover Example 3 shows that a binary morphism can preserve the order on infinite words but not on finite words. The reader can verify that this situation does not hold for prefix morphisms and for non-periodic uniform morphisms.

6 Episturmian morphisms

There is (at least) another family of morphisms for which hierarchical questions arise. Indeed in [18], we characterized the episturmian morphisms that are Lyndon morphisms. Episturmian morphisms were introduced in [4, 8] in order to study a generalization of Sturmian words (see also [17]). We recall them. Let A be an alphabet. For x, y letters, the exchange morphism E_{xy} is defined on A by $E_{xy}(x) = y$, $E_{xy}(y) = x$ and $E_{xy}(z) = z$ for $z \in A \setminus \{x, y\}$. In what follows, we denote the composition of applications just by juxtaposition and we use regular operations to denote particular subsets of the monoid of all endomorphisms on A. An endomorphism on A is an *episturmian morphism* if it belongs to

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

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

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

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. We first observe:

Proposition 6.1 *An episturmian morphism preserves Lyndon words if and only if it preserves infinite Lyndon words.*

Proof: By Proposition 3.1, a Lyndon morphism preserves infinite Lyndon words. Let f be an episturmian morphism that preserves infinite Lyndon words. By Theorem 4.5, f preserves the order on finite word, and for each i, $1 \leq i < n$, $f(a_i)$ is a power of a Lyndon word. For $\alpha \in A$, Ψ_α is a primitive morphism. (Indeed if $\Psi_\alpha(u) = v^k$ for two primitive words u, v and an integer $k \geq 2$, the word v starts with the letter α. Consequently $v = \Psi_\alpha(x)$ for a word x. Hence $\Psi_\alpha(x^k) = \Psi_\alpha(u)$. Since the morphism Ψ_α is injective, $u = x^k$.) Similarly we can see that exchange morphisms and morphisms $\overline{\Psi}_\alpha$ for $\alpha \in A$ are primitive morphisms. So f is a primitive morphism. In particular for each i, $f(a_i)$ is a Lyndon word. By Proposition 4.6, f is a Lyndon morphism. □

Let us recall from [18] that the episturmian morphisms on A that preserve Lyndon words are the morphisms in $(\{\overline{\Psi}_\alpha \mid \alpha \in A \setminus \{a\}\}^*\Psi_a)^*\overline{\Psi}_b^*$. In [18], it is also shown that there exist episturmian morphisms that preserve the order on finite words without preserving Lyndon words. For instance, this is the case for $\overline{\Psi}_a$.

Example 3 provides Ψ_b as an example of episturmian morphism that preserves the order on infinite words without preserving it on finite words. In fact, one can verify that each morphism $\Psi_\alpha, \overline{\Psi}_\alpha$ (with α a letter) preserves strictly the order on infinite words. In [18], it is proved that any episturmian morphism f can be decomposed as $f = p\Psi$ with $p \in \mathrm{Exch}(A)$ and $\Psi \in \{\Psi_\alpha, \overline{\Psi}_\alpha \mid \alpha \in A\}^*$. Consequently:

Proposition 6.2 *The following assertions are equivalent for an episturmian morphism f:*

1. *f preserves the order on infinite words,*
2. *f preserves strictly the order on infinite words,*
3. *$f \in \{\Psi_\alpha, \overline{\Psi}_\alpha \mid \alpha \in A\}^*$.*

7 On morphisms preserving border-free words

A word u is a *border* of a word w if u is a proper prefix and a proper suffix of w. A word w is said *border-free* if it has no border except the empty word ε. Let us recall that a Lyndon word is a special case of border-free word and that a border-free word is primitive. We say that a morphism f *preserves border-free words* or is a *border-free morphism* if $f(w)$ is border-free for each border-free word w.

We have already said that a Lyndon morphism is a primitive morphism (Lemma 3.4). Example 1 shows that a morphism that preserves infinite Lyndon words is not necessarily a primitive morphism. Contrarily to the case of words, Lyndon morphisms and border-free morphisms are not related as shown by the following examples.

To show that there exist border-free morphisms that are not Lyndon morphisms, it suffices to consider any permutation, that is any composition of exchange morphisms. For instance the exchange morphism E defined by $E(a) = b$, $E(b) = a$ is a border-free morphism but not a Lyndon morphism. Note that it does not preserve the lexicographic order. The following example is a border-free morphism that preserves the order without being a Lyndon morphism:

$$f(a) = abbababbbb,$$
$$f(b) = abbb.$$

Now consider the morphism f defined by $f(a) = a$, $f(b) = ab$. This morphism is not a prefix morphism so that it is not border-free. But it is a Lyndon morphism. An example of a Lyndon morphism which is not suffix is $f(a) = ab$, $f(b) = b$ (a *suffix morphism* is a morphism such that for all words u, v, $f(u)$ is a prefix of $f(v)$ if and only if u is a prefix of v). Let us also consider the following example: $f(a) = aab$, $f(b) = ababb$. This morphism is a Lyndon morphism. It is both a prefix morphism and a suffix morphism. But it is not a border-free morphism since $f(ba)$ has the border ab.

Observe now that a border-free morphism is non-erasing except if it is the totally erasing morphism ϵ. Indeed if there exist letters a, b such that $f(a) = \varepsilon$ and $f(b) \neq \varepsilon$ then $f(abb)$ has $f(b)$ as border although abb is border-free.

A non-empty border-free word is a primitive word. This property extends to morphism.

Lemma 7.1 *A border-free morphism different from ϵ preserves primitive words.*

Proof: Key elements of the proof are the well-known facts: a word w is primitive if and only if any of its conjugate is primitive; any primitive word is a conjugate of a Lyndon word.

Let us consider a primitive word w. Let u, v be the words such that $w = uv$ and vu is a Lyndon word. The word vu is border-free. Hence $f(vu)$ is border-free. This implies that $f(vu)$ is a primitive word. Its conjugate $f(w) = f(uv)$ is also primitive.

□

Now we provide a characterization of border-free morphism.

Proposition 7.2 *A non-erasing morphism $f : A^* \to B^*$ preserves border-free words if and only if*

1. *$f(a)$ is border-free for each letter $a \in A$,*
2. *$f(ab)$ is border-free for all $a, b \in A$, with $a \neq b$.*

Proof: The condition is immediately seen to be necessary.

Assume now that f is a non-erasing morphism such that $f(a)$ is border-free for each letter $a \in A$ and $f(ab)$ is border-free for all $a, b \in A$, with $a \neq b$. We first note that f is a prefix morphism. Indeed if $f(a)$ is a prefix of $f(b)$ then $f(ba)$ has $f(a)$ as border: this contradicts the second hypothesis. Similarly f is a suffix morphism.

Now let u be a border-free word such that $f(u)$ has a non-empty border x. Assume first that there exists a word v such that $x = f(v)$. Since f is both a prefix morphism and a suffix morphism, and since x is a proper prefix and a proper suffix of $f(u)$, the word v is a proper prefix and a proper suffix of u. This contradicts the fact that u is border-free. Now assume that for no word v, $x = f(v)$. Then $x = f(v)y$ with y a non-empty prefix of $f(a)$ for a letter a. Moreover $y \neq f(a)$. Note that the word y is a suffix of $f(u)$. Let bt with $b \in A$ and $t \in A^*$ be the least suffix of u such that y is a suffix of $f(bt)$. For a suffix z of $f(b)$, $y = zf(t)$. By definition of bt, $z \neq \varepsilon$. If $z = f(b)$ then $x = f(vbt)$ a contradiction with the case hypothesis. So z is a proper suffix of $f(b)$ and a proper prefix of $f(a)$. Since $f(a)$ is border-free, $a \neq b$. So finally z is a non-empty border of $f(ab)$. This contradicts the second hypothesis. Hence f is border-free. □

A consequence of Proposition 7.2 is that one can test in time $O(\max\{|f(ab)| \mid a, b \in A, a \neq b\})$ if a morphism is border-free. Indeed for a word w, it is now a classical algorithm to compute in time $O(|w|)$ the length of the greatest border of w (see [12] for instance).

Let us end by a remark: the only episturmian morphisms that are border-free are the morphisms in $(\mathrm{Exch}(A))^*$.

8 Morphisms preserving the radix order

Another order on finite words sometimes used (see for instance [11]) is the *radix order* defined by $x \leq_r y$ if $|x| < |y|$, or $x = y$, or $|x| = |y|$ and $x = uax'$ and $y = uby'$ with a, b letters such that $a < b$. A natural question is the study of morphisms preserving (resp. strictly) the radix order, that is morphisms f such that $f(u) \leq_r f(v)$ (resp. $f(u) <_r f(v)$) for all words u, v with $u \leq_r v$ (resp. $u <_r v$). This section is devoted to this study. First we note that these morphisms can be easily characterized (we still work on alphabet with at least two letters).

Proposition 8.1 *A morphism f preserves the radix order if and only if f is uniform and one of the two following assertions is verified*

- *for all letters a, b, $f(a) = f(b)$,*
- *for all letters a, b with $a < b$, $f(a) <_r f(b)$.*

Proof: Let f be a morphism that preserves the radix order. We first prove that f is uniform. Assume by contradiction that there exist two letters a, b such that $a < b$ and $|f(a)| \neq |f(b)|$. From $f(a) \leq_r f(b)$, we get $|f(a)| < |f(b)|$. But then, although $abb <_r baa$, we have $|f(abb)| > |f(baa)|$ and so $f(abb) >_r f(baa)$. This contradicts the fact that f preserves the radix order. Hence f is uniform.

Now we prove that if there exist two letters a and b such that $f(a) = f(b)$ then for all letters c, $f(c) = f(a)$. Assume by contradiction that c is a letter with $f(c) \neq f(a)$. First we note that we cannot have $a < c$ and $a < b$. Indeed in this case $f(a) <_r f(c)$ and $f(ac) \leq_r f(ba)$ which implies $f(c) \leq_r f(a)$. Similarly cases $b < a < c$ (consider words bac and aba in place of ac and ba) and $c < a$ are not possible. Hence for all letters c, $f(c) = f(a)$.

If f is uniform with $f(a) = f(b)$ for all letters a, b then immediately f preserves the radix order.

Now let f be a uniform morphism such that $f(a) <_r f(b)$ for all letters a, b with $a < b$. Let u, v be words with $u <_r v$. If $|u| < |v|$ then $|f(u)| < |f(v)|$ and so $f(u) <_r f(v)$. If $|u| = |v|$, $u = wax'$, $v = wby'$ for words w, x', y' and a, b letters such that $a < b$, since $f(a) <_r f(b)$, $f(u) <_r f(v)$. □

Similarly, we have:

Proposition 8.2 *A morphism f preserves strictly the radix order if and only if f is uniform and for all letters a, b with $a < b$, $f(a) <_r f(b)$.*

We end this section with several remarks:

- Of course any constant morphism on the alphabet preserves the radix order and so not all morphisms preserving the radix order also preserve strictly this order (the converse is trivially true).
- From all characterizations, one can observe that morphisms that preserve (even strictly) the radix order have no general relation with Lyndon morphisms or with morphisms that preserve (even strictly) the lexicographic order or with border-free morphisms.
- It follows from the previous characterization and other results stated or recalled in this paper that if f is a uniform morphism, f preserves (resp. strictly) the radix order if and only if it preserves (resp. strictly) the lexicographic order.
- Except morphisms in $\mathrm{Exch}(A)$, no episturmian morphism is uniform and so no episturmian morphism preserves the radix order.

9 A remark on monoids of morphisms

Let A be an alphabet. For any family of endomorphisms on A considered in this paper (morphisms preserving infinite Lyndon words, morphisms preserving (strictly) the order on infinite words, border-free morphisms, morphisms preserving the radix order), the family has naturally a structure of monoid for the composition of applications. We end this paper with a remark about these monoids.

In [18], it is shown that the monoid of Lyndon endomorphisms on A_n (with $n \geq 2$) is not finitely generated. This is done using an infinite family of Lyndon endomorphisms f_p on A_n ($f_p(a_1) = a_1^p a_2$, $f_p(a_i) = a_{i-1} a_i^{p-1} a_{i+1}$ for $2 \leq i \leq n-1$, and $f_p(a_n) = a_{n-1} a_n^p$) that cannot be written $f_p = gh$ with g and h two endomorphisms on A_n preserving the order on finite words and different from the identity. It follows from the hierarchy that this family can also be used to show that the monoid of endomorphisms preserving infinite Lyndon words is not finitely generated. In fact morphisms f_p cannot be decomposed using morphisms preserving the order on infinite words. This shows that monoids of other endomorphisms in Proposition 3.1 are also not finitely generated.

The monoid of border-free endomorphisms on an alphabet A (containing at least two letters) is also not finitely generated. This can be viewed considering, for $p \geq 0$, the morphisms defined by

$$\begin{aligned} &a \mapsto ab \\ &b \mapsto aabbb^p \\ &c \mapsto c \text{ for } c \notin \{a, b\} \end{aligned}$$

Finally let $n \geq 2$ be an integer. Considering for $p \geq n+1$, the morphisms g_p defined on A_n by $g_p(a_i) = a_1^{p+i} a_2^{p-i}$, we can see that the monoid of endomorphisms on A_n that preserve the radix order is not finitely generated.

10 Conclusion

There are other questions of interest concerning morphisms and (infinite) Lyndon words. The first one is to find more applications of the morphisms studied in this paper and in [18] for instance in order to decompose in Lyndon words some infinite words (as done for instance for the Fibonacci word in [18]). A second question is: which are the morphisms that generate an infinite Lyndon word? Let A be an ordered alphabet with a its least letter. Let f be a morphism such that $f^\omega(a) = \lim_{n\to\infty} f^n(a)$ exists and is an infinite word. One can verify that if f is a Lyndon morphism or if f preserves infinite Lyndon words then the previous limit is an infinite Lyndon word. But there exist morphisms that generate an infinite Lyndon word without preserving finite or infinite Lyndon words. This is the case of the morphism f defined on $\{a < b\}$ by $f(a) = aba$ and $f(b) = bb$: for $n \geq 0$ the words $f^n(ab)$ are Lyndon words that are prefixes of $f^\omega(a)$; $f(a^2b^\omega)$ is not a Lyndon word.

Acknowledgements

The author would like to thank J. Berstel and V. Villain for giving him the idea to consider border-free morphisms and morphisms that preserve radix order respectively. He thanks also the referees for their remarks.

References

[1] J.-P. Allouche, J. Currie, and J. Shallit. Extremal infinite overlap-free binary words. *Electronic J. Combinatorics*, 5(1):paper R27, 1998.

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

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

[6] T. Harju and D. Nowotka. Minimal Duval extensions. *Inter. J. Foundations of Comput. Science*, 15(2):349–354, 2004.

[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] J. Justin and G. Pirillo. On a characteristic property of Arnoux-Rauzy sequences. *RAIRO Theoretical Informatics and Applications*, 36:385–388, 2002.

[10] 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.

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

[12] M. Lothaire. *Applied Combinatorics on Words*, volume 105 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, UK, 2005.

[13] G. Melançon. Lyndon factorization of infinite words. In *STACS'96*, volume 1046 of *Lect. Notes in Comp. Sci.*, pages 147–154, 1996.

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

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

[16] D. Nowotka. *Periodicity and unbordered factors of words.* PhD thesis, Turku Centre for Computer Science, 2004. Dissertation 50.

[17] G. Richomme. Conjugacy and episturmian morphisms. *Theoretical Computer Science*, 302:1–34, 2003.

[18] G. Richomme. Lyndon morphisms. *Bulletin of the Belgian Mathematical Society*, 10(3):761–786, 2003.

[19] G. Richomme. On morphisms preserving infinite Lyndon words. In *10th International Conference "Journées Montoises d'Informatique Théorique",* prépublication 04.006 de l'Institut de Mathématiques, Université de Liège, pages 325–333, 2004.

[20] P. Séébold. Lyndon factorization of the Prouhet words. *Theoretical Computer Science*, 307:179–197, 2003.

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