

# A characterization of binary morphisms generating Lyndon infinite words

Gwenaël Richomme, Patrice Séébold

► **To cite this version:**

Gwenaël Richomme, Patrice Séébold. A characterization of binary morphisms generating Lyndon infinite words. 2021. lirmm-03213221

**HAL Id: lirmm-03213221**

**<https://hal-lirmm.ccsd.cnrs.fr/lirmm-03213221>**

Preprint submitted on 30 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A characterization of binary morphisms generating Lyndon infinite words

Gwenaël Richomme, Patrice Séébold  
LIRMM, Université Paul-Valéry Montpellier 3,  
Université de Montpellier, CNRS, Montpellier, France

April 30, 2021

## Abstract

An infinite word is an infinite Lyndon word if it is smaller, with respect to the lexicographic order, than all its proper suffixes, or equivalently if it has infinitely many finite Lyndon words as prefixes. A characterization of binary endomorphisms generating Lyndon infinite words is provided.

## 1 Introduction

Finite Lyndon words are the non-empty words which are smaller, w.r.t. (with respect to) the lexicographic order, than all their proper suffixes. They are important tools in many studies (see, *e.g.*, [2, 9, 10, 13]). Infinite Lyndon words are defined similarly. They are also the words that have infinitely many finite Lyndon words as prefixes. They occur in many context (see, *e.g.*, [1, 3, 4, 7, 8, 11, 12]).

The aim of the current paper is to provide a characterization, in the binary case, of endomorphisms that generate infinite Lyndon words. This paper continues the study of links between morphisms and Lyndon words done by the first author. In [14] he studied and characterized the morphisms that preserve Lyndon words, calling them *Lyndon morphisms*: these morphisms are those that map any Lyndon word to another Lyndon word. This study was extended to morphisms that preserve infinite Lyndon words in [15].

Note that being a morphism that preserves finite Lyndon words is a sufficient condition to generate an infinite Lyndon word (if the morphism generates an infinite word). Indeed if  $f$  is a morphism that preserves finite Lyndon words and  $u$  is a Lyndon word, then, for any  $n \geq 0$ ,  $f^n(u)$  is a Lyndon word. Applying this process when  $u = a$  with a morphism  $f$  that generates from  $a$  an infinite word  $\mathbf{w}$ , we see that  $\mathbf{w}$  has infinitely many finite Lyndon words as prefixes: it is an infinite Lyndon word. But the condition is not necessary. For instance, the morphism defined by  $f(a) = aba$  and  $f(b) = bb$  generates an infinite Lyndon word (the proof can be done using Proposition 5.1) but it does not preserve finite Lyndon words since  $f(a)$  is not a Lyndon word.

Our main characterization is Theorem 6.1: Over  $\{a, b\}$  with  $a \prec b$ , a non-periodic word generated by a morphism  $f$  prolongable on  $a$  is an infinite Lyndon word if and only if  $f$  preserves the lexicographic order on finite words and  $f^3(a)$  is a prefix of Lyndon words. The proof needs to consider separately the case where  $aa$  is a prefix of  $f^\omega(a)$  and the case where  $ab$  is a prefix of  $f^\omega(a)$ . After some needed preliminaries in Section 2, we prove the following general necessary condition: a binary endomorphism that generates an infinite Lyndon word must preserve the lexicographic order on finite words. In Section 4, we characterize morphisms that generate an infinite Lyndon word beginning with  $aa$  (Proposition 4.1).

In Section 5, we characterize morphisms that generate an infinite Lyndon word beginning with  $ab$  (Proposition 5.1). In Section 6, we prove our main result. We conclude with a few words on what happens on larger alphabets.

## 2 About Lyndon words and morphisms

We assume that readers are familiar with combinatorics on words and morphisms (see, e.g., [9, 10]). We specify our notation and recall useful results.

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*  $\varepsilon$  is the empty sequence. Equipped with the concatenation operation, the set  $A^*$  of finite words over  $A$  is a free monoid with neutral element  $\varepsilon$  and set of generators  $A$ . We let  $A^\omega$  denote the set of infinite words over  $A$ . As usually, for a finite word  $u$  and an integer  $n$ , the  $n^{\text{th}}$  power of  $u$ , denoted  $u^n$ , is the word  $\varepsilon$  if  $n = 0$  and the word  $u^{n-1}u$  otherwise. If  $u$  is not the empty word,  $u^\omega$  denotes the infinite word obtained by infinitely repeating  $u$ . Such a word is called *periodic*. A finite word  $w$  is said *primitive* if for any word  $u$ , the equality  $w = u^n$  (with  $n$  an integer) implies  $n = 1$ .

Given a non-empty word  $u = a_1 \cdots a_n$  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  $u$ .

Let us recall two basic results.

**Proposition 2.1** (see, e.g., [9, Prop. 1.3.2]) *For any words  $u$  and  $v$ ,  $uv = vu$  if and only if there exist a word  $w$  and integers  $k, \ell$  such that  $u = w^k$  and  $v = w^\ell$ .*

**Theorem 2.2** (Fine and Wilf's Theorem, see, e.g., [9, Prop. 1.3.5])

*Let  $x, y \in A^*$ ,  $n = |x|$ ,  $m = |y|$ ,  $d = \gcd(n, m)$ . Assume there exist integers  $p$  and  $q$  such that  $x^p$  and  $y^q$  have a common prefix of length at least equal to  $n + m - d$ . Then  $x$  and  $y$  are powers of the same word.*

### 2.1 Lyndon words

From now on we consider ordered alphabets. We let  $A_n = \{a_1 \prec \dots \prec a_n\}$  denote the  $n$ -letter alphabet  $A_n = \{a_1, \dots, a_n\}$  with order  $a_1 \prec \dots \prec a_n$ . Given an ordered alphabet  $A$ , we let also  $\preceq$  denote the lexicographic order whenever used on  $A^*$  or on  $A^\omega$ . Let us recall that for two different (finite or infinite) words  $u$  and  $v$ ,  $u \prec v$  if and only if  $u = x\alpha y$ ,  $v = x\beta z$  with  $\alpha, \beta \in A$ ,  $\alpha \prec \beta$ ,  $x \in A^*$ ,  $y, z \in A^* \cup A^\omega$ , or if (when  $u$  is finite)  $u$  is a proper prefix of  $v$ . For any finite words  $u, v, w$ , if  $u \prec v$ , then  $wu \prec wv$ . Moreover if  $u$  is not a prefix of  $v$  and  $u \prec v$ , then  $ux \prec vy$  for any words  $x$  and  $y$ .

A non-empty finite word  $w$  is a *Lyndon word* if for all non-empty words  $u$  and  $v$ ,  $w = uv$  implies  $w \prec vu$ . Equivalently [5, 9], a non-empty word  $w$  is a Lyndon word if all its non-empty proper suffixes are greater than itself for the lexicographic order. For instance, on the one-letter alphabet  $\{a\}$ , only  $a$  is a Lyndon word. On  $\{a \prec b\}$  the Lyndon words of length 6 are  $aaaaab$ ,  $aaaabb$ ,  $aaabab$ ,  $aaabbb$ ,  $aababb$ ,  $aabbab$ ,  $aabbbb$ ,  $ababbb$ ,  $abbbbb$ . Lyndon words are primitive. Note that Lyndon words have no non-empty border, that is, there is no proper prefix of a Lyndon word  $u$  that is also a suffix of  $u$ . Observe also that if  $u$  is a prefix of a Lyndon word then there cannot exist words  $v$  and  $w$  such that the three following conditions hold:  $v$  is a prefix of  $u$ ;  $w$  is a factor of  $u$  which is not a prefix of  $u$ ;  $w \prec v$ .

**Proposition 2.3** (see, e.g., [9, prop. 5.1.3]) *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  $u \prec v$ .*

Lyndon infinite words were introduced in [16] as the infinite words that have infinitely many prefixes that are Lyndon words. It follows from the definition that an infinite Lyndon word is not periodic. More generally an infinite word is Lyndon if and only if all its proper suffixes are greater than it w.r.t. the lexicographic order [16, Prop. 2.2].

## 2.2 Morphisms

Let  $A$  and  $B$  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 say that  $f$  is a morphism over  $A$  if we don't need to refer to  $B$ . When  $A = B$ ,  $f$  is an *endomorphism* over  $A$ . A morphism is *erasing* if  $f(a) = \varepsilon$  for some letter  $a$ . For  $n \geq 0$  and any word (finite or infinite)  $u$ ,  $f^n(u)$  is  $u$  if  $n = 0$  and  $f^{n-1}(f(u))$  otherwise.

An endomorphism is said *prolongable* on  $a$  if  $f(a) = au$  for some word  $u$  and if  $\lim_{n \rightarrow \infty} |f^n(a)| = \infty$ . For such a morphism, for all  $n \geq 0$ ,  $f^n(a)$  is a prefix of  $f^{n+1}(a)$ . Then the sequence  $(f^n(a))_{n \geq 0}$  defines a unique infinite word, denoted  $f^\omega(a)$ . This word is a fixed point of  $f$ .

A morphism *preserves finite Lyndon words* if and only if the image of any finite Lyndon word is also a Lyndon word. Similarly morphisms that *preserve infinite Lyndon words* can be defined. A morphism *preserves the order on finite words* if, for all words  $u$  and  $v$ ,  $u \prec v$  implies  $f(u) \prec f(v)$ . Such a morphism is injective and so non-erasing. In [14], it is proved that 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. We have also the following characterization.

**Proposition 2.4** ([14, prop. 3.3]) *A morphism  $f$  over  $\{a \prec b\}$  preserves the lexicographic order on finite words if and only if  $f(ab) \prec f(b)$ .*

## 3 A necessary condition

In this section we prove the following result that states a necessary condition for a prolongable binary morphism to generate an infinite Lyndon word.

**Proposition 3.1** *Let  $f$  be an endomorphism over  $\{a \prec b\}$ . Assume that  $f$  is prolongable on  $a$ . If  $f^\omega(a)$  is a Lyndon infinite word then  $f$  preserves the lexicographic order on finite words.*

We will use the basic fact and the following characterization of prefixes of Lyndon words.

**Fact 3.2** *Given any finite Lyndon word  $x$  and any proper non-empty prefix  $p$  of  $x$ ,  $px \prec x$ .*

*Proof.* Let  $q$  be the word such that  $x = pq$ . Since  $x$  is a Lyndon word and since  $x \neq q$  and  $x \neq \varepsilon$ ,  $x \prec q$ . It follows that  $px \prec pq = x$ .  $\square$

**Proposition 3.3** ([6, Prop. 1.7]) *Let  $A$  be an ordered alphabet with maximal letter  $c$ . Let  $P$  be the set of prefixes of Lyndon words. The set  $P \cup \{c^k \mid k \geq 2\}$  is equal to the set of all words on the form  $(uv)^k u$  with  $k \geq 1$  an integer and  $u, v$  some finite words such that  $v \neq \varepsilon$  and  $uv$  is a Lyndon word.*

*Proof of proposition 3.1.*

Assume by contradiction that  $f$  does not preserve the lexicographic order on finite words. By Proposition 2.4,  $f(b) \prec f(ab)$  (the equality cannot hold as  $f(a)$  is not empty). Thus, for any integer  $n \geq 0$ ,  $f(a^n b) \prec f(a^{n+1} b)$ . So, for any integer  $n \geq 0$ ,  $f(b) \prec f(a^n b)$ .

From now on let  $i$  be the integer such that  $a^i b$  is a prefix of  $f^\omega(a)$ . Let also  $\mathbf{w}$  be the word such that  $f^\omega(a) = f(a^i b)\mathbf{w}$ . Note that  $i \geq 1$  since  $b^\omega$  is not an infinite Lyndon word.

Observe that  $f(b)$  is a prefix of  $f(a^i b)$ . Otherwise, from  $f(b) \prec f(a^i b)$ , we deduce that  $f(b)\mathbf{w} \prec f^\omega(a)$  which contradicts the fact that  $f^\omega(a)$  is an infinite Lyndon word since  $f(b)\mathbf{w}$  is a proper suffix of  $f^\omega(a)$ .

As  $f^\omega(a)$  is an infinite Lyndon word, it has infinitely many prefixes that are Lyndon words. Thus its prefix  $f(a^i b)$  is a prefix of a Lyndon word. Hence by Proposition 3.3, there exist an integer  $k \geq 1$  and words  $u$  and  $v$  such that  $f(a^i b) = (uv)^k u$ ,  $v \neq \varepsilon$  and  $uv$  is a Lyndon word. Consequently  $f(b) = (uv)^j u'$  for some  $j \geq 0$  and some proper prefix  $u'$  of  $uv$ .

Observe that  $ba$  is a factor of  $f^\omega(a)$ . Indeed otherwise  $f^\omega(a) = a^i b^\omega$  which implies  $i = 1$  and  $f(b) \in b^+$ , and so, a contradiction with  $f(b) \prec f(ab)$ .

Assume that  $u' \neq \varepsilon$ . Since  $ba$  is a factor of  $f^\omega(a)$ , the word  $u'uv$  is a factor of  $f(ba)$  and so of  $f^\omega(a)$ . By Fact 3.2,  $u'uv \prec uv$ : since  $uv$  is a prefix of  $f^\omega(a)$ , this contradicts the fact that  $f^\omega(a)$  is an infinite Lyndon word.

Thus  $u' = \varepsilon$ . This means that  $f(b) = (uv)^j$  with  $j \geq 0$ . If  $j = 0$ ,  $f(b) = \varepsilon$  and  $f^\omega(a) = f(a)^\omega$  is a periodic word: a contradiction with the fact it is an infinite Lyndon word. Thus  $j \geq 1$ . Since  $f(b)$  is a suffix of  $f(a^i b) = (uv)^k u$ , we get  $uv = vu$ . Remember that  $v \neq \varepsilon$ . If  $u \neq \varepsilon$ , by Proposition 2.1, the word  $uv$  is not primitive: a contradiction with the primitivity of the Lyndon word  $uv$ . So  $u = \varepsilon$ .

This implies that both  $f(a)$  and  $f(b)$  are powers of  $v$ . So  $f^\omega(a) = v^\omega$ . This is a final contradiction with the fact that an infinite Lyndon word cannot be periodic. The morphism  $f$  preserves the order on finite words over  $\{a \prec b\}$ .  $\square$

Note that the converse of Proposition 3.1 does not hold. Consider, for instance, the morphism  $f$  defined by  $f(a) = abb$  and  $f(b) = baa$ . This morphism preserves the lexicographic order on infinite word but the word  $f^\omega(a)$  is not an infinite Lyndon word.

One could expect a stronger necessary condition as, for instance, a preservation of infinite Lyndon words. The next example shows that this stronger condition is not necessary.

Let  $f$  be defined by  $f(a) = aab$  and  $f(b) = abaabab$ . The word  $\mathbf{w} = abbabb^\omega$  is an infinite Lyndon word. Its image by  $f$  begins with  $ubua$  where  $u = aababaaba$ . Hence  $f$  does not preserve infinite Lyndon words. Nevertheless using Proposition 4.1, one can verify that  $f$  generates an infinite Lyndon word.

## 4 Generating infinite Lyndon words beginning with $aa$

We consider here the case of generated words beginning with  $aa$ .

**Proposition 4.1** *Let  $f$  be an endomorphism over  $\{a \prec b\}$  prolongable on  $a$  such that  $f^\omega(a)$  begins with  $a^i b$  for some integer  $i \geq 2$ .*

*The word  $f^\omega(a)$  is an infinite Lyndon word if and only if*

1.  *$f$  preserves the lexicographic order on finite words, and,*
2.  *$f(a^i b)$  is a Lyndon word.*

The proof of this proposition is based on the next lemmas.

**Lemma 4.2** *Let  $f$  be a morphism that preserves the order on finite words. Let  $i \geq 2$ . Assume that  $f(a^i b)$  is a Lyndon word. For any word  $v$  such that  $a^i b v$  is a Lyndon word, the word  $f(a^i b v)$  is also a Lyndon word.*

*Proof.* We act by induction on  $|v|$ .

By hypothesis the result holds when  $|v| = 0$ . Assume that  $|v| \geq 1$ . By Proposition 2.3, there exist Lyndon words  $\ell$  and  $m$  such that  $a^i b v = \ell m$  and  $\ell \prec m$ . Let us choose  $m$  with the smallest length as possible.

Let us prove that  $a^i b$  is a prefix of  $\ell$ . Assume that this does not hold. Then  $\ell = a$  and  $m = a^{i-1} b v$ . Consequently, since  $m$  is a Lyndon word,  $a^i$  is not a factor of  $a^{i-1} b v$ . Let  $m'$  be the suffix of  $m$  such that  $b m' \in b a^+ b^+$ . If such a factor does not exist (that is if  $m \in a^{i-1} b^+$ ), let  $m' = b$ . In all cases,  $m'$  is a Lyndon word. Let  $\ell'$  be the word such that  $\ell m = \ell' m'$ . The word  $a^i b$  is a prefix of  $\ell'$  (when  $m = b$ , remember that  $|v| \geq 1$ ). Observe that  $\ell' \prec m'$ . The last letter of  $\ell'$  is  $b$ . Indeed, by construction, it could be the letter  $a$  only if  $m \in a^{i-1} b^+$ , that is if  $m = a^{i-1} b^k$  for some  $k \geq 1$ . But then  $m' = a$  and  $\ell' = a^{i-1} b^{k-1}$ . As  $|v| \geq 1$ , we have  $k \geq 2$ , and so, the last letter of  $\ell'$  is  $b$ . Let  $s$  be a proper non-empty suffix of  $\ell'$ . Let  $j \geq 0$  be the integer such that  $s$  begins with  $a^j b$ . Since  $a^i b$  is not a factor of  $a^{i-1} b v$  and since  $a^{i-1} b v$  is a Lyndon word, we deduce that  $j < i$ . So  $\ell' \prec s$ . Hence  $\ell'$  is a Lyndon word: this contradicts the choice made on  $m$  and proves that  $a^i b$  is a prefix of  $\ell$ .

If  $a^i b$  is a prefix of  $m$  then, by inductive hypothesis,  $f(\ell)$  and  $f(m)$  are Lyndon words. Since  $\ell \prec m$  and  $f$  preserves the order on finite words,  $f(\ell) \prec f(m)$ . Proposition 2.3 implies that  $f(a^i b v) = f(\ell m)$  is a Lyndon word.

From now on assume that  $a^i b$  is not a prefix of  $m$ . Observe that this implies that  $a^i$  is not a factor of  $m$ . Indeed since  $a^i b v = \ell m$  is a Lyndon word, for any factor  $a^j$  of  $\ell m$ , we have  $j \leq i$ . So  $m$  begins with  $a^k b$  for some integer  $k < i$ . Moreover as  $m$  is a Lyndon word, for any factor  $a^j$  of  $m$ , we have  $j \leq k < i$ . Let  $s$  be a proper non-empty suffix of  $\ell m$ . If  $|s| \leq |f(m)|$  then there exist an integer  $j < i$ , a word  $m'$  and a non-empty suffix  $s'$  of  $f(a^j b)$  such that  $s = s' f(m')$ . The word  $s'$  is a proper non-empty suffix of the Lyndon word  $f(a^j b)$ . So  $f(a^i b) \prec s'$  and  $f(a^i b v) \prec s' \preceq s$ . If  $|s| > |f(m)|$  then  $s = s' f(m)$  with  $s'$  a proper non-empty suffix of  $f(\ell)$ . By inductive hypothesis,  $f(\ell)$  is a Lyndon word. Thus  $f(\ell) \prec s'$  and consequently  $f(a^i b v) = f(\ell m) \prec s' \prec s' f(m) = s$ . The word  $f(a^i b v)$  is a Lyndon word.  $\square$

**Lemma 4.3** *Let  $u$  be a non-empty word. If  $u u$  is a prefix of a Lyndon word, then  $u$  is a power of a Lyndon word.*

*Proof.* Since  $u u$  is a prefix of a Lyndon word, also  $u$  is a prefix of this Lyndon word. By Proposition 3.3, there exist words  $x$  and  $y$  such that  $y \neq \varepsilon$ ,  $x y$  is a Lyndon word and for some integer  $k \geq 1$ ,  $u = (x y)^k x$ . If  $x \neq \varepsilon$ , since  $x y$  is a Lyndon word, we have  $x y \prec y$  and so  $x x y \prec x y$ . Then for any word  $v$ , the word  $x (x y)^k x v$  is a suffix of  $u u v$  and  $x (x y)^k x v \prec u u v$ . This contradicts the fact that  $u u$  is a prefix of a Lyndon word. So  $x = \varepsilon$ . This implies that  $u = y^k$  and  $y$  is a Lyndon word.  $\square$

**Lemma 4.4** *Assume that  $f$  is an endomorphism over  $\{a \prec b\}$  prolongable on  $a$  such that  $f^3(a)$  is a prefix of a Lyndon word,  $f^\omega(a)$  begins with the word  $a^i b$  for some integer  $i \geq 2$  and  $f^\omega(a)$  is not periodic. Then  $f(a^i b)$  is a Lyndon word.*

*Proof.* Let us first observe that  $f(a)$  begins with  $a^i b$ . Indeed otherwise  $f(a)$  is a power of  $a$  contradicting the non-periodicity of  $f^\omega(a)$ .

Observe also that the word  $f(a^i b)$  is a prefix of  $f^2(a)$  which itself is a prefix of  $f^3(a)$ . Hence  $f(a^i b)$  is also a prefix of a Lyndon word. By Proposition 3.3, there exists a Lyndon

word  $v$ , a proper prefix  $p$  of  $v$  ( $p$  may be empty) and an integer  $\ell \geq 1$  such that  $f(a^i b) = v^\ell p$ . Since  $i \geq 2$ , by Lemma 4.3,  $f(a)$  is a power of a Lyndon word  $u$ .

If  $v = u$ , from  $f(a^i b) = v^\ell p$ , we get  $f(b) = v^{\ell'} p$  for some integer  $\ell'$ . In particular  $p$  is a suffix of  $f(b)$ . If  $p = \varepsilon$ , then  $f^\omega(a) = v^\omega$  a contradiction with its non-periodicity. Assume now that  $p \neq \varepsilon$ . Since  $i \geq 2$ , the word  $a^i b$  occurs twice in  $f(a^i b)$  which is a prefix of  $f^2(a)$ . Thus  $ba$  is a factor of  $f^2(a)$  and  $f(ba)$  is a factor of  $f^3(a)$ . Then the word  $pu$  is a factor of  $f^3(a)$ . Note also that  $u$  is a prefix of  $f^3(a)$ . As  $p$  is a proper non-empty prefix of the Lyndon word  $u$ , by Fact 3.2,  $pu \prec u$ . This contradicts the fact that  $f^3(a)$  is a prefix of a Lyndon word. Thus  $v \neq u$ .

Since  $i \geq 2$ ,  $u^2$  is a prefix of  $v^{\ell+1}$ . If  $|u| \geq |v|$ , by Theorem 2.2,  $u$  and  $v$  are powers of the same word. This is not possible as  $u \neq v$  and both words  $u$  and  $v$  are primitive (since they are Lyndon words). Thus  $|v| > |u|$ .

Note that  $v$  is not a factor of  $f(a^i) = f(a)^i$ . Indeed if  $v$  is a factor of  $f(a)^i$  then it is a prefix of a power of  $u$ , and so, a prefix of  $u$  is both a prefix and a suffix of  $v$ : this is impossible since  $v$  is a Lyndon word. It follows that  $p$  is a proper suffix of  $f(b)$ .

Observe that  $a^i b$  is a prefix of  $f(a)$  and so  $f(a)a^i b$  is a prefix of  $f(aa)$  and so of  $f^2(a)$ . Since  $f^2(a)$  is a prefix of a Lyndon word, it cannot contain  $a^{i+1}$  as a factor and so the last letter of  $f(a)$  must be  $b$ . Hence  $ba^i b$  and  $f(ba^i b)$  are factors of  $f^3(a)$ . This implies that  $pv$  is also a factor of  $f^3(a)$ . By Fact 3.2,  $pv \prec v$  if  $p \neq \varepsilon$ . This contradicts the fact that  $f^3(a)$  is a prefix of a Lyndon word. So  $p = \varepsilon$  and  $f(a^i b) = v^\ell$ . Assume that  $\ell \geq 2$ .

Since  $a^i b$  is a prefix of  $f(a)$  and since  $f(a^i b) = v^\ell$ ,  $a^i b$  is also a prefix of  $v$ . Thus  $v^\ell$  is a prefix of  $f(v)$  itself a prefix of  $f^2(a)$ . Since  $\ell \geq 2$ ,  $f(v)f(v)$  and  $f(v)v^\ell$  are prefixes of  $f^3(a)$ .

Let us prove that  $f(v)$  is not a prefix of  $v^\omega$ . Assume by contradiction that  $f(v) = v^k p'$  for some proper prefix  $p'$  of  $f(v)$  and some integer  $k$ . Since  $v^\ell$  is a prefix of  $f(v)$ , we have  $k \geq \ell \geq 2$ . If  $p' \neq \varepsilon$ , by Fact 3.2,  $p'v \prec v$ . Since  $p'v$  is a factor of  $f(v)f(v)$ , this contradicts the fact that  $f^3(a)$  is a prefix of a Lyndon word. So  $p' = \varepsilon$  and  $f(v) = v^k$ . Hence by induction, for all  $n \geq 0$ ,  $f^n(v) \in v^+$ . Moreover we have  $\lim_{n \rightarrow \infty} |f^n(v)| = \infty$ . So  $f^\omega(a) = v^\omega$ : a contradiction with the non-periodicity of  $f^\omega(a)$ .

So  $f(v)$  is not a prefix of  $v^\omega$ . There exists an integer  $k$ , a proper prefix  $\pi$  of  $v$  and letters  $\alpha, \beta$  such that  $v^k \pi \beta$  is a prefix of  $f(v)$  and  $\pi \alpha$  is a prefix of  $v$ . Since  $f^3(a)$  is a prefix of a Lyndon word,  $\alpha = a$  and  $\beta = b$ . Note that  $v^{k+1} \prec v^k \pi \beta$ .

We have already mentioned that  $v$  is not a factor of  $f(a^i)$ . From  $f(a^i b) = v^\ell$  and  $\ell \geq 2$ , we deduce that  $v$  is a suffix of  $f(b)$ . Moreover since  $v$  is a Lyndon word beginning with  $a^i b$ , the last letter of  $v$  is  $b$ :  $v$  is so a suffix of  $f(v)$ . Since  $f(v)f(v)$  is a factor of  $f^3(a)$ , the word  $v^{k+1}$  is a factor of  $f^3(a)$ , This contradicts the fact that  $f^3(a)$  is a prefix of a Lyndon word.

Thus  $\ell = 1$ :  $f(a^i b)$  is a Lyndon word. □

*Proof of Proposition 4.1.* We first prove that the two conditions are sufficient. First observe that, for any integer  $n \geq 1$ ,  $a^i b$  is a prefix of  $f^n(a^i b)$  (this is a direct consequence of the facts that  $f$  is prolongable on  $a$  and that  $f^\omega(a)$  begins with  $a^i b$ ). Thus by induction, using Lemma 4.2, we get: for any integer  $n \geq 0$ ,  $f^n(a^i b)$  is a Lyndon word. As  $\lim_{n \rightarrow \infty} |f^n(a^i b)| = \infty$ , the word  $f^\omega(a)$  has infinitely many prefixes that are Lyndon words. By definition, it is an infinite Lyndon word.

From now on assume that  $f^\omega(a)$  is an infinite Lyndon word. Proposition 3.1 shows that  $f$  preserves the lexicographic order on finite words. Observe that since it is an infinite Lyndon word,  $f^\omega(a)$  is not periodic and  $f^3(a)$  is a prefix of a Lyndon word. Lemma 4.4 states that  $f(a^i b)$  is a Lyndon word. □

## 5 Generating infinite Lyndon words starting with $ab$

We consider here the case of generated words beginning with  $ab$ . The word  $ab^\omega$  is an infinite Lyndon word. A morphism  $f$  generates it if and only if  $f(a) = ab^i$  for some integer  $i \geq 1$  and if  $f(b) \in b^+$ .

In what follows we only consider the case where  $f^\omega(a)$  begins with  $ab^i a$  for some  $i \geq 1$ .

**Proposition 5.1** *Let  $f$  be an endomorphism over  $\{a \prec b\}$  prolongable on  $a$  such that  $f^\omega(a)$  begins with  $ab^i a$  for some integer  $i \geq 1$ .*

*The word  $f^\omega(a)$  is an infinite Lyndon word if and only if*

1.  *$f$  preserves the lexicographic order on finite words,*
2.  *$f(ab^i)$  is a power of a Lyndon word  $u \neq ab^i$ , and,*
3. *if  $i = 1$ ,  $|u| > |f(b^i)|$ .*

Here follows an example showing that indeed in item 2,  $f(ab^i)$  is not necessarily a Lyndon word.

**Example 5.2** Let  $f$  be defined by  $f(a) = abbab$  and  $f(b) = b$ :  $f(ab) = (abb)^2$  is the square of a Lyndon word. By induction one can verify that all words  $f^n(abb)$  are Lyndon words with the relation  $f^{n+1}(abb) = f^n(abb)f^n(abb)b$ . This confirms that  $f^\omega(a)$  is an infinite Lyndon word.

We now provide an example showing the necessity of item 3.

**Example 5.3** Let  $f$  be defined by  $f(a) = aba$  and  $f(b) = bbababb$ :  $f(ab) = u^2$  with  $u = ababb$  is the square of a Lyndon word. Condition 3 is not verified and indeed  $f^\omega(a)$  is not a Lyndon word. It could be verified that  $f^\omega(a)$  begins with  $u^4 b b u^5$  and so contains the factor  $u^4 a$  which is smaller than the prefix  $u^4 b$ .

The proof of Proposition 5.1 is based on the next lemmas.

**Lemma 5.4** *Let  $f$  be a morphism that preserves the lexicographic order on finite words over  $\{a \prec b\}$ . Assume that  $i \geq 2$  is an integer and that  $f(ab^i)$  is a power of a Lyndon word  $u$ . Then  $|u| > |f(b)|$ .*

*Proof.* Assume by contradiction that  $|u| \leq |f(b)|$ . Assume first that  $|u| < |f(b)|$ . Since  $f(b)f(b)$  is a suffix of  $f(ab^i)$  so of a power of  $u$ , there exist words  $p$  and  $s$  and an integer  $k \geq 1$  such that  $u = ps$ ,  $f(b)$  ends with  $p$  and  $f(b) = su^k$ . Since  $u$  is a Lyndon word,  $p$  cannot be both a prefix and a suffix of  $u$  except if  $p = \varepsilon$ . When  $p = \varepsilon$ ,  $f(b)$  is a power of  $u$ . If  $|u| = |f(b)|$  then  $f(b) = u$ . In all cases both  $f(a)$  and  $f(b)$  are powers of  $u$ . Hence  $f$  is not injective, a contradiction with the fact that  $f$  preserves the lexicographic order on finite words.  $\square$

**Lemma 5.5** *Let  $f$  be a morphism that preserves the lexicographic order on finite words over  $\{a \prec b\}$ . Assume that  $f(ab^i)$  is a power of a Lyndon word  $u$  for some integer  $i \geq 1$ . Assume also that  $|u| > |f(b)|$  if  $i = 1$ . Then, for any non-empty word  $v$  over  $\{a \prec b\}$  such that  $ab^i v$  is a Lyndon word, the word  $f(ab^i v)$  is also a Lyndon word.*



*Proof.* We act by induction on  $|v|$ . Let us observe that  $f$  is non-erasing and injective since it preserves the lexicographic order on finite words. Let  $n$  be the integer such that  $f(ab^i) = u^n$ . Observe that  $|f(b)| < |u|$  (by hypothesis if  $i = 1$  and by Lemma 5.4 if  $i \geq 2$ ).

We first assume that  $|v| = 1$ . In this case, since  $ab^i v$  is a Lyndon word,  $v = b$ . Any suffix  $s$  of  $f(ab^{i+1})$  with  $|s| \leq |f(b)|$  is also a suffix of the Lyndon word  $u$ . Thus  $u \prec s$  (and for length reason,  $u$  is not a prefix of  $s$ ). Hence  $f(ab^{i+1}) \prec s$ . Consider now a suffix  $s$  of  $f(ab^{i+1})$  such that  $|f(b)| < |s| < |f(ab^{i+1})|$ . We have  $s = s'f(b)$  for some suffix  $s'$  of  $f(ab^i) = u^n$ . If  $s' = s''u^k$  for some proper non-empty suffix  $s''$  of  $u$  and some integer  $k$  then  $u \prec s''$  and  $u$  is not a prefix of  $s''$ . Once again  $f(ab^{i+1}) \prec s$ . If  $s' = u^k$  for some integer  $k$  such that  $1 \leq k < n$ ,  $s = u^k f(b)$ . As  $f(b)$  is a proper non-empty suffix of  $u$ ,  $u \prec f(b)$ . Hence  $u^{k+1} \prec u^k f(b)$ . Moreover since  $k + 1 \leq n$ ,  $f(ab^i) \prec u^k f(b)$ . So for any proper non-empty suffix  $s$  of  $f(ab^{i+1})$ ,  $f(ab^{i+1}) \prec s$ :  $f(ab^{i+1})$  is a Lyndon word.

From now on assume that  $|v| \geq 2$ . By Proposition 2.3, there exist two Lyndon words  $\ell$  and  $m$  such that  $ab^i v = \ell m$  and  $\ell \prec m$ . Two cases can hold.

Case  $|m| \geq 2$ . As  $m$  cannot begin with the letter  $b$  (as any Lyndon word of length at least 2 over a binary alphabet),  $\ell$  must begin with  $ab^i$ . Moreover as  $ab^i v = \ell m$  is a Lyndon word,  $m$  is on the form  $ab^k$  with  $k > i$  or begins with a factor  $ab^k a$  with  $k \geq i$ . In both cases,  $ab^i$  is a proper prefix of  $m$ , and by inductive hypothesis  $f(m)$  is a Lyndon word. If  $\ell \neq ab^i$ ,  $f(\ell)$  is also a Lyndon word. Moreover, since  $f$  preserves the lexicographic order,  $f(\ell) \prec f(m)$ . By Proposition 2.3,  $f(ab^i v) = f(\ell m)$  is a Lyndon word. If  $\ell = ab^i$ ,  $f(\ell) = u^n$ . Since  $f$  preserves the lexicographical order,  $u \preceq f(\ell) \prec f(m)$ . Using Proposition 2.3, one can prove by induction that  $u^k f(m)$  is a Lyndon word for any  $k \geq m$ . Once again,  $f(ab^i v) = u^n f(m)$  is a Lyndon word.

Case  $|m| = 1$ . In this case,  $m = b$ . Let  $s$  be a proper non-empty suffix of  $f(ab^i v)$ . If  $|s| \leq |f(b)|$ , then  $s$  is a suffix of the Lyndon word  $u$  (remember that  $|f(b)| < |u|$  and  $f(ab^i) = u^n$ ). This implies that  $u \prec s$  and so that  $f(ab^i v) \prec s$ . If  $|f(b)| < |s| < |f(ab^i v)|$ , we have  $s = s'f(b)$  for some proper non-empty suffix  $s'$  of the Lyndon word  $f(\ell)$  (since  $|v| \geq 2$ ,  $|\ell| > |ab^i|$  and the inductive hypothesis can be applied). Thus  $f(\ell) \prec s'$  which implies that  $f(ab^i v) \prec s$ . Hence  $f(ab^i v)$  is a Lyndon word. □

**Lemma 5.6** *Assume that  $f$  is an endomorphism over  $\{a \prec b\}$  prolongable on  $a$  such that  $f^3(a)$  is a prefix of a Lyndon word,  $f^\omega(a)$  begins with the word  $ab^i a$  for some integer  $i \geq 1$  and  $f^\omega(a)$  is not periodic. Then  $f(ab^i)$  is a power of a Lyndon word  $u \neq ab^i$ . Moreover if  $i = 1$ ,  $|u| > |f(b)|$ .*

*Proof.* The word  $ab^i a$  is a prefix of  $f^\omega(a)$ . Let us prove that the word  $f^3(a)$  has a prefix on the form  $ab^i ab^k a$ . If  $f(a)$  has  $ab^i a$  as a prefix, then  $f^2(a)$  (and so  $f^3(a)$ ) contains at least 4 occurrences of  $a$ . Since  $f^2(a)$  is a prefix of a Lyndon word, it cannot contain the factor  $aa$ . Hence we get the result. Assume now that  $f(a) = ab^j$  for some  $j < i$ . Since  $f$  is prolongable on  $a$ ,  $j > 0$ . It follows that  $f(b)$  begins with  $b^{j-i} a$ . Then  $f^3(a)$  contains at least 3 occurrences of  $a$ . And once again  $f^3(a)$  has a prefix on the form  $ab^i ab^k a$ .

Since  $f^3(a)$  is a prefix of a Lyndon word, we have  $k \geq i$  and so  $(ab^i)^2$  is a prefix of  $f^3(a)$ . Lemma 4.3 shows that  $f(ab^i)$  is a power of a Lyndon word  $u$ :  $f(ab^i) = u^n$  for an integer  $n \geq 1$ . If  $u = ab^i$ , we have  $f^\omega(a) = (ab^i)^\omega$  which contradicts the fact that  $f^\omega(a)$  is aperiodic. Thus  $u \neq ab^i$ .

Assume now that  $i = 1$  and  $|f(b)| \geq |u|$ . From  $f(ab) = u^n$  and  $f(a) \neq \varepsilon$ , we get  $n \geq 2$ . Let  $s$  be the proper suffix of  $u$  and let  $j \geq 1$  be the integer such that  $f(b) = su^j$ . If  $s = \varepsilon$ , then both  $f(a)$  and  $(b)$  are powers of  $u$ . This implies that  $f^\omega(a) = u^\omega$ , a contradiction.

Assume now that  $s \neq \varepsilon$ . Let  $p$  be the word such that  $u = ps$ :  $f(a) = u^k p$  for some integer  $k \geq 0$  and  $p \notin \{\varepsilon, u\}$  since  $s \notin \{\varepsilon, u\}$ . Since  $u$  is a Lyndon word different from  $ab$  but beginning with  $aba$ , we deduce that  $u$  begins with  $(ab)^m b$  for some  $m \geq 2$ .

Since  $n \geq 2$ , the word  $(ab)^m b$  has at least one non prefix occurrence in  $f(ab)$  so in  $f^3(a)$ . This occurrence must be preceded by the letter  $b$  since  $aa$  cannot occur in  $f^3(a)$  which is a prefix of a Lyndon word. Hence the word  $f(ab)^m s = u^{nm} s$  is a prefix of  $f((ab)^m b)$  itself a prefix of  $f^3(a)$ , and, the word  $uu^{nm}$  which is a suffix of  $f(b(ab)^m)$  is a factor of  $f^3(a)$ . Since  $u \prec s$ , we have  $u^{nm} u \prec u^{nm} s$ : this contradicts the fact that  $f^3(a)$  is a prefix of a Lyndon word.  $\square$

*Proof of Proposition 5.1.* Let us first show that the three conditions imply that  $f^\omega(a)$  is an infinite Lyndon word. Since  $f$  preserves the order on finite words,  $f$  is not erasing. Let  $u$  be the word occurring in condition 2 and let  $k$  be the integer such that  $f(ab^i) = u^k$ . Observe that  $ab^i$  and the prefix  $u$  of  $f(ab^i)$  are both prefixes of  $f(a)^\omega$ . From  $|f(ab^i)| \geq |ab^i|$ ,  $ab^i$  is a prefix of  $f(ab^i) = u^k$ . Hence  $ab^i$  is a prefix of  $u$ . By hypothesis, we cannot have  $ab^i = u$ . So  $ab^i$  is a proper prefix of  $u$ . For any  $n \geq 0$ ,  $f^n(u)$  is a prefix of  $f^\omega(a)$  and so  $ab^i$  is a proper prefix of  $f^n(u)$ . Due to condition 3, one can apply Lemma 5.5. Thus it follows by induction that  $f^n(u)$  is a Lyndon word for all  $n \geq 0$ :  $f^\omega(a)$  is an infinite Lyndon word.

Let us show that the conditions are necessary. First Proposition 3.1 shows that  $f$  preserves the lexicographic order on finite words. Observe that since it is an infinite Lyndon word,  $f^\omega(a)$  is not periodic and  $f^3(a)$  is a prefix of a Lyndon word. Lemma 5.6 states that  $f(ab^i)$  is a power of a Lyndon word  $u$  and, when  $i = 1$ ,  $|u| > |f(b)|$ .  $\square$

## 6 A general characterization

Let us prove our main characterization.

**Theorem 6.1** *Let  $f$  be an endomorphism over  $\{a \prec b\}$  prolongable en  $a$ . The word  $f^\omega(a)$  is an infinite Lyndon word if and only if*

1.  $f$  preserves the lexicographic order on finite words,
2.  $f^\omega(a)$  is not periodic and
3. the word  $f^3(a)$  is a prefix of a Lyndon word.

*Proof.* Assume first that  $f^\omega(a)$  is an infinite Lyndon word. Conditions 2 and 3 are direct consequences of this hypothesis. Proposition 3.1 states condition 1.

Assume now that the three conditions hold. If  $f^\omega(a)$  begins with  $aa$ , then it begins with  $a^i b$  for some integer  $i \geq 2$ . Lemma 4.4 states that  $f(a^i b)$  is a Lyndon word. Thus from Proposition 4.1  $f^\omega(a)$  is an infinite Lyndon word.

If  $f^\omega(a) = ab^\omega$ , it is an infinite Lyndon word.

If  $f^\omega(a)$  begins with  $ab^i a$  for some integer  $i \geq 1$ , Lemma 5.6 states that  $f(ab^i)$  is a power of a Lyndon word  $u \neq ab^i$ . Moreover if  $i = 1$  then  $|u| > |f(b)|$ . Thus from Proposition 5.1  $f^\omega(a)$  is an infinite Lyndon word.  $\square$

**Example 6.2** Let  $f$  be the morphism defined by  $f(a) = aba$  and  $f(b) = bab$ . This morphism fulfills conditions 1 and 3 of Theorem 6.1. It generates the periodic word  $(ab)^\omega$ . This shows the importance of the condition  $f^\omega(a)$  is not periodic that does not occur in Propositions 4 and 5,

**Example 6.3** Let  $\varphi$  be the Fibonacci morphism defined by  $\varphi(a) = ab$  and  $\varphi(b) = a$ . We have  $\varphi^2(a) = aba$  and  $\varphi^3(a) = abaab$ . This examples shows the optimality of the exponent 3 in the last condition of Theorem 6.1.

## 7 Conclusion

After Theorem 6.1, a natural problem is to obtain a characterization of morphisms that generate infinite Lyndon words over an alphabet containing at least three letters.

Let us observe that Proposition 3.1 does not extend to morphisms over alphabets with at least three letters. Indeed consider any endomorphism  $f$  such that  $f(a) = au$  with  $u$ ,  $f(b)$  and  $f(c)$  belonging to  $\{b, c\}^*$  (note that one of the two words  $f(b)$  and  $f(c)$  could be the empty word: we just need that  $\lim_{n \rightarrow \infty} |f^n(a)|$  is infinite). Then  $f^\omega(a)$  is an infinite Lyndon word whatever is  $f$  (that may not preserve the lexicographic order). Note that the previous example can include some erasing morphisms. We don't know whether the condition  $f$  preserves the lexicographic order is necessary if  $f$  generates a recurrent word.

Note also that, if an analog of Theorem 6.1 exists for a larger alphabet  $A$ , then the exponent in the last condition would be at least  $\#A + 1$  with  $\#A$  the cardinality of  $A$ . Indeed if  $A_n = \{a_1 \prec \dots \prec a_n\}$ , one can extend Example 6.3 defining the morphism  $f$  by  $f(a_1) = a_1a_2$ ,  $f(a_i) = a_{i+1}$  for  $2 \leq i < n$  and  $f(a_n) = a_1$ . Then for  $1 \leq i \leq n$ ,  $f^i(a_i)$  is a prefix of  $a_1a_2 \cdots a_na_1$  and so a prefix of Lyndon word while  $f^{n+1}(a)$  is not such a prefix since it begins with  $a_1a_2 \cdots a_na_1a_1$ .

## References

- [1] J.-P. Allouche, J. Currie, and J. Shallit. Extremal infinite overlap-free binary words. *Electron. J. Combin.*, 5(1):paper R27, 1998.
- [2] J. Berstel, A. Lauve, C. Reutenauer, and F. Saliola. *Combinatorics on Words: Christoffel Words and Repetitions in Words*, volume 27 of *CRM Monograph Series*. American Mathematical Society, 2008.
- [3] J.P. Borel and F. Laubie. Quelques mots sur la droite projective réelle. *J. Théor. Nombres Bordeaux*, 5:23–51, 1993.
- [4] E. Charlier, T. Kamae, S. Puzynina, and L. Q. Zamponi. Infinite self-shuffling words. *J. Combin. Theory Ser. A*, 128:1–40, 2014.
- [5] K.T. Chen, R.H. Fox, and R.C. Lyndon. Free differential calculus IV – the quotient groups of the lower central series. *Ann. of Math. (2)*, 68:81–95, 1958.
- [6] J.-P. Duval. Factorizing words over an ordered alphabet. *J. Algorithms*, 4(4):363–381, 1983.
- [7] A. Glen, F. Levé, and G. Richomme. Quasiperiodic and Lyndon episturmian words. *Theoret. Comput. Sci.*, 409(3):578–600, 2008.
- [8] F. Levé and G. Richomme. Quasiperiodic Sturmian words and morphisms. *Theoret. Comput. Sci.*, 372(1):15–25, 2007.
- [9] M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, 1983. Reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, UK, 1997.
- [10] M. Lothaire. *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2002.
- [11] G. Paquin. A characterization of infinite smooth lyndon words. *Discret. Math. Theor. Comput. Sci.*, 12(5):25–62, 2010.

- [12] M. Postic and L. Q. Zamboni. Reprint of:  $\omega$ -lyndon words. *Theoret. Comput. Sci.*, 834(60-65), 2020.
- [13] C. Reutenauer. *From Christoffel words to Markoff numbers*. Oxford University Press, 2019.
- [14] G. Richomme. Lyndon morphisms. *Bull. Belg. Math. Soc. Simon Stevin*, 10(5):761–786, 2003.
- [15] G. Richomme. On morphisms preserving infinite Lyndon words. *Discret. Math. Theor. Comput. Sci.*, 9(2):89–108, 2007.
- [16] R. Siromoney, L. Mathew, V. R. Dare, and K. G. Subramanian. Infinite Lyndon words. *Inform. Process. Lett.*, 50:101–104, 1994.