On morphisms preserving infinite Lyndon words

(extended abstract)

G. Richomme
LaRIA, Université de Picardie Jules Verne
33, Rue Saint Leu,
F-80039 Amiens cedex 1
(gwenael.richomme@u-picardie.fr)

June 25, 2004

Abstract

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 military order.

1 Introduction

Finite (or infinite) Lyndon words can be encountered in many studies (see for instance [9, 10, 11]). 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 in a product of decreasing (in lexicographic order) Lyndon words. This result was extended to infinite words [18] (In such a case, the decomposition can end with an infinite Lyndon word). Thus some works concern the decomposition in Lyndon words of some infinite words (see for instance [6, 12, 13, 17] for such results).

To obtain the decomposition in Lyndon words of the (infinite) Thue-Morse word, Ido and Melançon [6] give some morphisms that preserve finite Lyndon words. Another use of such morphisms can be found for instance in [10, chapter 4]. In [16], these morphisms are called Lyndon morphisms and studied. In particular, two given uses of these morphisms are new proofs of the decomposition [12, 13] of the Fibonacci word $\underline{F}$ and of the fact that $a\underline{F}$ is an infinite Lyndon word [2]).

Let us recall other results of [16]. Lyndon morphisms are related to morphisms preserving the lexicographic order since 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. More precise characterizations provide algorithms to test if a morphism preserves Lyndon words or the order on finite words. We also note that a morphism preserve the order on finite words if and only if does it strictly.

In [1], the fact that the Thue-Morse morphism μ ($\mu(a) = ab$, $\mu(b) = ba$) preserves the order on infinite word is observed and used. The initial idea of the study presented here is to answer the natural question: what can be say about morphisms preserving infinite Lyndon words or morphisms preserving (possibly strictly) the lexicographic order in infinite words? Our main results are characterization of these new families of morphisms. Before presenting them (Sections 4 and 5), we recall basic notions (Section 2) and present relations between all families of morphisms (Section 3) studied here and in [16].

2 Lyndon words

We assume the reader is familiar with combinatorics on words and morphisms (see [9, 10] for instance). Here we recall only central definitions and notations.

We consider ordered alphabets. We denote $A_n = \{a_1 < \ldots < a_n\}$ the n-letter alphabet $A_n = \{a_1, \ldots, a_n\}$ with order $a_1 < \ldots < a_n$. Given an ordered alphabet A, we denote by $\preceq$ the lexicographic order whenever used on the set of finite words A^* or on the set of infinite words A^ω. Let 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 prefix of v.

The empty word is denoted ε. 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 [3, 9], 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 (that is not a power of another 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. A useful result of Melançon [13] states that 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 a word having no occurrence of a^{n+1}. Thus $a^{n+1}\underline{w}$ is an infinite Lyndon word. See for instance [8] 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 A 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. 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 [16], 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 $f(u) \prec f(v)$.

Lyndon morphisms and morphisms preserving the order on finite words are studied in [16]. 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 infinite property of 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 (stricly) 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 proofs of the inclusion are consequence of the definitions except one: a characterization stated in [16] of morphisms preserving the order on finite words is useful to state that any morphism that preserves infinite Lyndon words also preserves the order on finite words. This proof also needs the fact that no word on the form xyx^ω is an infinite Lyndon word.

The next examples shows that the inclusion are strict.

Example 3.2 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 4.1). 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 3.3 The Thue-Morse morphism μ ($\mu(a) = ab$, $\mu(b) = ba$) is a morphism preserving the order on finite words [1]. On other part, the image of the Lyndon word $abab^\omega$ is not a Lyndon word: $\mu(abab^\omega) = abbaab(ba)^\omega$.

Example 3.4 The morphism f defined on $\{a < b\}$ by $f(a) = ba$, $f(b) = b$ is an example of 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 proper prefix of $f(a)$).

Example 3.5 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 $z \neq \varepsilon$, f preserves the order on infinite words (all infinite words have z^ω has image) but does not did it strictly.

To end this section let us note that Example 3.2 provides examples of morphisms that preserve infinite Lyndon words without preserving primitive words (As in [14], we called *primitive* a morphism that preserves primitive words). This situation does not hold for Lyndon morphisms as shown by the following new property.

Lemma 3.6 *Any Lyndon morphism preserves primitive words.*

4 Morphisms preserving infinite Lyndon words

We now present our characterization of morphisms preserving infinite Lyndon words.

Theorem 4.1 *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 of length at least n_0, $f(u)$ is a Lyndon word.*

Our proof of this result is made of several parts. First we consider necessary conditions for a morphism f to preserve infinite Lyndon words. We already know that such a morphism must preserve the order on finite words. In addition, for any Lyndon word $u \neq a_n$, $f(u)$ is a power of a Lyndon word. The following example shows that it is not necessarily true for $u = a_n$.

Example 4.2 The morphism f defined on $\{a < b\}$ by $f(a) = aaa$ and $f(b) = bab$ preserves infinite Lyndon words but $f(b)$ is not a power of a Lyndon word.

Another necessary condition is that the image of any Lyndon word of length at least 2 is a Lyndon word. Hence we get the proof that, in Theorem 4.1, Assertion 1 implies Assertion 2. The proof that Assertion 2 implies Assertion 3 is a consequence of technical considerations on Lyndon words. Finally the fact that Assertion 3 implies Assertion 1 can be viewed using the property that any infinite Lyndon word has an infinity of prefixes that are Lyndon words.

Let us make some remarks about Theorem 4.1.

First of all, Assertion 2 provides an algorithm to test if a morphism preserves infinite Lyndon words. Indeed in [16] an algorithm is given to test 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 providing by the previous result can be implement in time $O(\max\{|f(ab)| \mid a \neq b\})$.

We observe also that the sufficient and necessary conditions given by Assertion 2 is closed to the one obtained in [16] for Lyndon morphisms:

Proposition 4.3 [16] *Let $n \geq 2$ be an integer. A morphism f on A_n is a Lyndon morphism if and only if it verifies the two following properties:*

1. *$f(a)$ is a Lyndon word for each $a \in A_n$,*
2. *f preserves the order on finite word.*

5 Morphisms preserving order on infinite words

Here we study morphisms that preserve (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 $f(u)$.*

5.1 Characterizations

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

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 $\forall i$, $1 \leq i < n$, $f(a_ia_n^\omega) \preceq f(a_{i+1}a_1^\omega)$.*

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 $\{a \prec b\}$ preserves the lexicographic order on infinite words if and only if $f(ab) \preceq f(ba)$.*

We say here nothing about the technical proof of Theorem 5.2. We observe that Theorem 5.2 does not provide (at least directly) an algorithm to test if a morphism preserves the lexicographic order on infinite words (except in the binary case). The following result which is a consequence of the Fine and Wilf's theorem (see [10] for instance) completes the theorem in order to have an algorithm. Using 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 5.4 *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 $|f(ya_1^j)| \geq m$.*

As for Lyndon morphisms, we get a result very similar to one of [16]:

Proposition 5.5 [16] *Let f be a non-empty morphism on A_n ($n \geq 2$). The morphism f is order-preserving if and only if for each i, $1 \leq i < n$, $f(a_ia_n^{k_i}) \prec f(a_{i+1})$ where k_i is the least integer such that $|f(a_ia_n^{k_i})| \geq |f(a_{i+1})|$.*

To end this part, let us note that, replacing $\preceq$ by $\prec$ in Theorem 5.2, Corollary 5.3 and Proposition 5.4, we get valid results for morphisms preserving strictly the order on infinite words.

5.2 Frontier

Now we consider morphisms preserving the order on infinite words without preserving it strictly. In Example 3.5, we have already mentionned the case of periodic morphisms. We can prove that these morphisms are the only ones which preserve the order on infinite morphisms without being injective. But, injectivity is not a necessary condition.

Example 5.6 The morphism f defined by $f(a) = ab$, $f(b) = abb$, $f(c) = ba$ preserves the order but not strictly. Indeed $f(ac^\omega) = f(ba^\omega) = ab(ba)^\omega$.

Example 5.7 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 $f(\underline{u}) \prec f(\underline{v})$.

Next result shows that for some families of morphisms, there is no difference between preserving the order on infinite words and preserving it strictly. Let 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 (resp. not a suffix) of $f(b)$.

Proposition 5.8 *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.*

Proposition 5.8 leads to consider the hierarchy of Proposition 3.1 for cases treated by it. Examples 3.2 and 3.3 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 3.4 shows that a binary morphism can preserve the order on infinite words without preserving it on finite words. The reader can verify that this situation does not hold for prefix morphisms and for non-periodic uniform morphisms.

There is (at least) one another family of morphisms for which we can ask what becomes the hierarchy. Indeed in [16], we have characterized the episturmian morphisms that are Lyndon morphisms. Episturmian morphisms were introduced in [4, 7] in order to study a generalization of Sturmian words (see also [15]). 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 $z \in A \setminus \{x, y\}$. An endomorphism on A is an *episturmian morphism* if it belongs to

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

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

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

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

Let us recall from [16] 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 [16], it is shown that there exist episturmian morphisms that preserve the order on finite words without preserving Lyndon words. For instance, this is the case of the morphism $\overline{\Psi}_a$.

Example 3.4 provides an example of episturmian morphism (Ψ_b with b the greatest letter in A) that preserves the order on infinite words without preserving it on finite words. In fact, one can verify that each morphism $\Psi_a, \overline{\Psi}_a$ (with a a letter) preserves the order on infinite words. In [16], it is proved that any episturmian morphism f can be decomposed $f = p\Psi$ with $p \in \mathrm{Exch}(A)$ and $\Psi \in \{\Psi_\alpha, \overline{\Psi}_\alpha \mid \alpha \in A\}$. Consequently:

Proposition 5.10 *A morphism preserves the order on infinite words if and only if it belongs to* $\{\Psi_\alpha, \overline{\Psi}_\alpha \mid \alpha \in A\}^*$ *if and only if it preserves strictly the order on infinite words.*

6 Conclusion

The study given in this paper comes from a natural question (after [16]). We present answers to other related questions.

A Lyndon word is known to be border-free (that is, no proper suffix is also a prefix) which is a particular case of primitive word. We have already said that a Lyndon morphism is a primitive morphism (Lemma 3.6). So some questions are: does a Lyndon morphism preserve border-free words? Does a morphism preserving border-free words (shortly border-free morphism) also preserves primitive words? The answer to the second question is yes but Lyndon morphisms and border-free morphisms are not related as shown by the following examples. The morphism f defined by $f(a) = abbababbbb$ and $f(b) = abbb$ is a border-free morphism that preserves the order without being a Lyndon morphism. The morphism g defined by $g(a) = a$, $g(b) = ab$ is not border-free, but it is a Lyndon morphism.

Note that the following characterization of border-free morphisms provides an algorithm to test these morphisms:

Proposition 6.1 *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* $a \neq b$.

Another natural question concerns the order. Indeed another order on finite words sometimes used (see for instance [10]) is the *radix order* defined by $x \leq_r y$ if $|x| < |y|$, or $|x| = |y|$ and $x = uax'$ and $y = uby'$ with a, b letters such that $a < b$, or $x = y$. 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. $u <_r v$). The following characterization allows to see that these morphisms have no general relation with morphisms of our hierarchy.

Proposition 6.2 *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* $a < b$, $f(a) <_r f(b)$.

Finally let us recall from [16] that the monoid of Lyndon endomorphisms on A_n (with $n \geq 2$) is not finitely generated. The proof of this result 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. Hence monoids of other endomorphisms in the hierarchy are also not finitely generated.

Acknowledgment. I would like to thank J. Berstel and V. Villain for giving me the idea to consider border-free morphisms and morphisms that preserve radix order respectively.

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 groupes 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] 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.

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

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

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

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

[11] M. Lothaire. *Applied Combinatorics on Words*. To appear. (see http://www-igm.univ-mlv.fr/~berstel).

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

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

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

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

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

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

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