



# On generalized Lyndon words

Francesco Dolce<sup>a</sup>, Antonio Restivo<sup>b,\*</sup>, Christophe Reutenauer<sup>c</sup>

<sup>a</sup> IRIF, Université Paris Diderot, France

<sup>b</sup> Dipartimento di Matematica e Informatica, Università degli Studi di Palermo, Italy

<sup>c</sup> LaCIM, Université du Québec à Montréal, Québec, Canada



## ARTICLE INFO

### Article history:

Received 18 June 2018

Received in revised form 23 November 2018

Accepted 14 December 2018

Available online 21 December 2018

### Keywords:

Generalized Lyndon words

Nonincreasing Lyndon factorization

Alternating lexicographical order

## ABSTRACT

A generalized lexicographical order on infinite words is defined by choosing for each position a total order on the alphabet. This allows to define generalized Lyndon words. Every word in the free monoid can be factorized in a unique way as a nonincreasing factorization of generalized Lyndon words. We give new characterizations of the first and the last factor in this factorization as well as new characterization of generalized Lyndon words. We also give more specific results on two special cases: the classical one and the one arising from the alternating lexicographical order.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

Let  $A$  be a totally ordered alphabet. A word  $w$  is called a *Lyndon word* if for each nontrivial factorization  $w = uv$ , one has  $w < vu$  (here  $<$  is the lexicographical order). Lyndon words were introduced in [16]. It is easy to see that this property can be expressed in an equivalent way using infinite words, namely  $w^\omega < (vu)^\omega$  (where  $w^\omega = www \dots$ ) for each nontrivial factorization  $w = uv$ .

A well-known theorem of Lyndon states that every finite word  $w$  can be decomposed in a unique way as a nonincreasing product  $w = \ell_1 \ell_2 \dots \ell_n$  of Lyndon words. This theorem, which is a combinatorial counterpart of the famous theorem of Poincaré–Birkhoff–Witt, provides an example of a factorization of the free monoid (see [14]). It has also many algorithmic applications and it may be computed in an efficient way. Indeed, Duval proposed in [11] linear-time algorithms to compute it, while Apostolico and Crochemore proposed in [1] a  $O(\lg n)$ -time parallel algorithm. This factorization is also used in string processing algorithms (see [5]) and for the computation of runs in a word (see, e.g., [10]).

In this paper we consider a variant of this family of words: generalized Lyndon words. These words were first introduced by the third author in [19]. Given a generalized order  $<$ , i.e., an order in which the comparison between two words depends on the length of their common prefix (see Section 3 for the formal definition), a finite word  $w$  is called a *generalized Lyndon word* if for each nontrivial factorization  $w = uv$  we have  $w^\omega < (vu)^\omega$ .

In this paper we present both new results and new proofs of already published results concerning this family of words. In [19] it is proved that the family of generalized Lyndon words is a Hall set, and thus that they provide a factorization of the free monoid. As a consequence, the associated Lie polynomials form a basis of the free Lie algebra (see [18,19]). In the present paper, we give a new proof of this factorization theorem (Theorem 16) using only combinatorial techniques instead of the heavy machinery of Hall set theory.

\* Corresponding author.

E-mail addresses: [dolce@irif.fr](mailto:dolce@irif.fr) (F. Dolce), [antonio.restivo@unipa.it](mailto:antonio.restivo@unipa.it) (A. Restivo), [reutenauer.christophe@uqam.ca](mailto:reutenauer.christophe@uqam.ca) (C. Reutenauer).

Note that Nyldon words, introduced by Grinberg in [12], also provide a factorization of the free monoid (see [8]), but they are not generalized Lyndon words. Inverse Lyndon words introduced in [5] are not generalized Lyndon words neither, while anti-Lyndon words (introduced in the same paper) with respect to a lexicographical order  $<$  can be viewed as classical Lyndon words with respect to the order  $\prec$  (see also Example 9 later).

With our new combinatorial approach we are able, on one hand, to simplify several of the proofs from [19] and, on the other hand, to obtain new interesting results. In particular, we deduce a new characterization of the last factor of the unique nonincreasing factorization in Lyndon words (Corollary 18). We also simplify a result of [19], stating that generalized Lyndon words are characterised by their suffixes and show a new characterization by the prefixes (Theorem 14). This last result is new even for classical Lyndon words.

Next, we focus on two particular cases of generalized orders: the classical and the alternating one.

In Theorem 20 we give two new characterizations of the first factor of the nonincreasing factorization into classical Lyndon words. For a different characterization of the first factor see also [11] and [18, Lemma 7.14 (iii)].

From Theorem 20 we deduce a new proof of a result from Ufnarovskij (Corollary 22) which characterizes Lyndon words by their prefixes.

The second case we focus on, related to continued fractions, is given by Galois words. These are generalized Lyndon words with respect to the *alternating lexicographical order*  $<_{alt}$ , that is the order comparing two words in an opposite way depending on the parity of the length of the common prefix (see Section 6 for the formal definition). The link with continued fractions is that we have that  $a_1a_2a_3 \cdots <_{alt} b_1b_2b_3 \cdots$  if and only if

$$a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{\ddots}}} < b_1 + \frac{1}{b_2 + \frac{1}{b_3 + \frac{1}{\ddots}}}$$

where  $a_i, b_i \in \mathbb{N} \setminus \{0\}$  for all  $i$  (see also [19]).

In Theorem 32 we give a characterization which generalized Ufnarovskij’s Theorem to Galois words. Moreover, we also characterize the first factor of the nonincreasing factorization in Galois words (Theorem 33). This is the analogue for Galois words of Theorem 20.

We conclude in Section 7 with some remarks and open problems.

**Acknowledgment.** We thank the anonymous referee for his/her very detailed report with copious and useful suggestions.

**Dedication.** Maurice Nivat has been one of the main figures of French school of Theoretical Computer Science. Some of his contributions are in the field of combinatorics on words, as for instance the ones related to discrete geometry, where paths are coded by words (see [2]). Lyndon words, whose generalization we focus in this paper, have a significant role in discrete geometry (see, e.g., [7]).

We want to dedicate this paper, in this very international journal that he founded, to his memory.

## 2. Definitions and notations

For undefined notation we refer to [13] and [14]. We denote by  $A$  a finite alphabet, by  $A^*$  the free monoid and by  $A^+$  the free semigroup. Elements of  $A^*$  are called *words* and the identity element, denoted by  $1$  is called the *empty word*. We say that a word  $u$  is a *factor* of the word  $w$  if  $w = xuy$  for some words  $x, y$ ;  $u$  is a *prefix* (resp. *suffix*) if  $x = 1$  (resp.  $y = 1$ ); it is *nontrivial* if  $u \neq 1$  and *proper* if  $u \neq w$ . We say that  $w = ps$  is a *nontrivial factorization* of  $w$  whenever  $p, s$  are both nonempty. The length of a word  $w = a_1a_2 \cdots a_n$ , where  $a_i \in A$  for all  $i$ , is equal to  $n$  and it is denoted by  $|w|$ .

A *period* of a word  $a_1a_2 \cdots a_n$  is a natural integer  $p$  such that  $a_i = a_{i+p}$  for any  $i$  such that  $i, i + p \in \{1, \dots, n\}$ ; it is called a *nontrivial period* if  $0 < p < n$ . A word having a nontrivial period is called *periodic*.

We say that  $v$  is a *fractional power* of  $u$  if  $u = u_1u_2$  and  $v = u^k u_1$  for some nonnegative integer  $k$ . In this case, one writes  $v = u^r$ , where  $r = k + |u_1|/|u|$  is a positive rational. Note that for  $k = 0$  (or  $r < 1$ ) this means that  $v$  is a prefix of  $u$ . Fractional powers are also known as *sesquipowers* (see, e.g., [17]).

We say that the  $v$  is a *strict fractional power* of  $u$  if  $v$  is a fractional power of  $u$  and, with the notations above,  $k \geq 1$  or, equivalently, that  $r \geq 1$ . In this case  $u$  is a prefix of  $v$ .

**Example 1.** Let  $u = abcdef$ . Then  $u^{2/3} = abcd$  and  $u^{5/3} = abcdefabcd$ . The last one is, in particular, a strict fractional power of  $u$ .

We denote by  $A^\omega$  the set of sequences over  $A$ , also called *infinite words*; such a sequence  $(a_n)_{n \geq 1}$  is also written  $a_1a_2a_3 \cdots$ . If  $w$  is a (finite) word of length  $n \geq 1$ ,  $w^\omega$  denotes the infinite word having  $w$  as a prefix and of period  $n$ .

We denote by  $A^\infty = A^* \cup A^\omega$  the set of finite and infinite words.

A *border* of a word  $w$  of length  $n$  is a word which is simultaneously a nontrivial proper prefix and suffix of  $w$ . A word is called *unbordered* if it has no border. It is well-known that a word has a border if and only if it is periodic.

Suppose that  $u, v$  are finite nonempty words. The following fact is well-known: one has  $u^\omega = v^\omega$  if and only if  $u, v$  are power of a common word, and this is true if and only if  $u$  and  $v$  commute (see, for instance, [14, Corollary 6.2.5]).

Observe also that if for two nonempty words  $u, v$ , one has  $u^\omega \neq v^\omega$ , then by Fine and Wilf theorem, their prefixes of length  $|u| + |v| - \gcd(|u|, |v|)$  differ (see, for instance, [13]).

Given an order  $<$  on the alphabet  $A$ , we can define the *lexicographic order*  $<_{lex}$  (or simply  $<$  when it is clear from the context) on  $A^\infty$  in the following way :  $u <_{lex} v$  if either  $u$  is a proper prefix of  $v$  (in which case  $u$  must be in  $A^*$ ) or we may write  $u = pau', v = pbv'$  for some words  $p \in A^*, u', v' \in A^\infty$  and some letters  $a, b \in A$  such that  $a < b$ .

**Definition 2.** Let  $s, t$  be two distinct elements of  $A^\omega$  such that we have a factorization  $s = u_1 \cdots u_k s_0$  with  $u_1, \dots, u_k$  finite nonempty words and  $s_0$  is an infinite word. We say that *the comparison between  $s$  and  $t$  takes place within  $u_k$*  if  $u_1 \cdots u_{k-1}$  is a prefix of  $t$ , but  $u_1 \cdots u_k$  is not. If moreover  $u_1, u_2, \dots, u_k$  are letters we say that the comparison takes place *at position  $k$* .

Note that, when the comparison takes place within  $u_k$ , one may write  $t = u_1 \cdots u_{k-1} u'_k t'$ , for some  $t' \in A^\omega$  and  $u'_k \neq u_k$  such that  $|u'_k| = |u_k|$ .

**Lemma 3.** Let  $u, v$  be nonempty words such that  $u^\omega \neq v^\omega$ . Then the comparison between  $u^\omega$  and  $v^\omega$  takes place within the first factor  $v$  of  $v^\omega$  if and only if  $v$  is not a fractional power of  $u$ .

**Proof.** The comparison between the two infinite words takes place within the first  $v$  if and only if the two prefixes of length  $|v|$  of  $u^\omega$  and  $v^\omega$  are different. The conclusion follows from the fact that  $v$  is a fractional power of  $u$  if and only if  $v$  is a prefix of  $u^\omega$ .  $\square$

### 3. Generalized lexicographical order

**Definition 4.** For each  $n \in \mathbb{N} \setminus \{0\}$ , let  $<_n$  be a total order on  $A$ . To the sequence  $(<_n)_{n \geq 1}$  we associate a total order on  $A^\infty$ , that we still denote by  $<$  when it is clear from the context, called *generalized lexicographical order*, as follows:  $u < v$  if either  $u$  is a proper prefix of  $v$  (in which case  $u$  must be in  $A^*$ ) or we can write  $u = pas, v = pbt$  for some  $p \in A^*$ , some  $s, t \in A^\infty$  and some letters  $a, b \in A$  such that  $a <_{|p|+1} b$ .

**Example 5.** Let  $<$  be the generalized order on  $\{a, b\}^\infty$  defined by  $b <_n a$  if  $n$  is a prime number and  $a <_n b$  otherwise. Then we have  $aba < abaa < aab < bab < baab$  and  $(ab)^\omega < a^\omega < b^\omega < (ba)^\omega$ .

Note that, as for the classical order, when  $u$  is a prefix of  $v$  we could have  $u^\omega \not\leq v^\omega$ , as shown in the next example.

**Example 6.** Let  $A = \{a, b\}$ , and  $<$  as in Example 5. The word  $ab$  is a prefix of  $aba$  but  $(aba)^\omega < (ab)^\omega$ .

Let us consider a generalized lexicographical order  $<$  on  $A^\infty$ .

**Lemma 7.** Let  $s, t \in A^\omega$  be as in Definition 2 (and the sentence following it). Then  $s < t$  (resp.  $s > t$ ) implies  $u_1 \cdots u'_k s' < u_1 \cdots u_k t'$  (resp.  $u_1 \cdots u'_k s' > u_1 \cdots u_k t'$ ) for any infinite words  $s', t'$ .

**Lemma 8.** Let  $u, v$  be nonempty finite words such that  $u^\omega < v^\omega$  and let  $x, y$  be two finite words. Then

- (i) if neither  $u$  or  $v$  is a prefix of the other, then  $(ux)^\omega < (vy)^\omega$ ;
- (ii) if  $v$  is not a fractional power of  $u$ , then  $(u^{k+1}x)^\omega < (vy)^\omega$ , where  $k$  is the largest integer such that  $u^k$  is a prefix of  $v$ . In particular  $u^\omega < (vy)^\omega$ .

**Proof.** In case (i), the comparison between the two infinite words takes place within the prefix of length  $\min(|u|, |v|)$ . Hence we conclude using Lemma 7.

Suppose now that the hypothesis of (ii) holds. Then we can write  $u = u'au_1$  and  $v = u^k u' b v_1$ , with  $u' \in A^*, a, b \in A$  such that  $a \neq b$ , and  $u_1, v_1 \in A^*$ . Let  $m = |u^k u'|$ . Since  $u^\omega = u^k u' a u_1 u^\omega$  and since  $u^\omega < v^\omega$ , we have that  $a <_{m+1} b$ . The two infinite words  $u^\omega$  and  $(u^{k+1}x)^\omega$  share the same prefix of length  $m + 1$ , and the same do  $v^\omega$  and  $(vy)^\omega$ . Thus the comparison between  $(u^{k+1}x)^\omega$  and  $(vy)^\omega$  takes place at position  $m + 1$ . Since  $a <_{m+1} b$ , we can conclude.  $\square$

We use several times the following observation: the opposite order  $\tilde{<}$  of a generalized order  $<$  is also a generalized lexicographical order, obtained by reversing all the orders  $<_i$ .

**Example 9.** Let  $<$  be the usual lexicographical order on  $\{a, b\}$ , that is such that  $a <_i b$  for all  $i \geq 1$ . Then  $\tilde{<}$  is defined by  $b \tilde{<}_i a$  for all  $i \geq 1$ .

Note that for two infinite words of the form  $u^\omega$  and  $v^\omega$ , one has

**Example 10.** Let  $<$  be the generalized order defined in Example 5. Then we have  $(aba)^\omega < (aab)^\omega < (bab)^\omega < (baa)^\omega$  and  $(baa)^\omega \gtrsim (bab)^\omega \gtrsim (aab)^\omega \gtrsim (aba)^\omega$ .

Part of the following lemma is stated in [19].

**Lemma 11.** *The following conditions are equivalent for nonempty words  $u, v \in A^*$ :*

- (1)  $u^\omega < v^\omega$ ;
- (2)  $(uv)^\omega < v^\omega$ ;
- (3)  $u^\omega < (vu)^\omega$ ;
- (4)  $(uv)^\omega < (vu)^\omega$ .

**Proof.** The four conditions obtained by replacing in the lemma  $<$  by  $=$  are equivalent (see again [14, Corollary 6.2.5]). We may therefore assume that none of these equalities holds.

Let us assume first that condition (1) holds and prove that the other conditions hold too.

If  $v$  is not a fractional power of  $u$ , then conditions (2), (3) and (4) hold, by point (ii) of Lemma 8, with  $x = u'bv_1$  and  $y = 1$  for case (2),  $x = 1$  and  $y = u$  for case (3), and  $x = u'bv_1$  and  $y = u$  for case (4).

Let us suppose that  $v$  is a strict fractional power of  $u$ . We may therefore write  $v = u^j u_1$ , for some  $j \geq 1$ , and  $u = u_1 u_2$ . Using the observation in the previous section, deduced from the Fine and Wilf theorem, we see that the prefixes of length  $|u| + |v|$  of  $u^\omega$  and  $v^\omega$  are distinct, i.e.,  $u^{j+1}u_1 \neq u^j u_1 u$ . Since both  $u^\omega$  and  $(uv)^\omega$  begin by  $u^{j+1}u_1$  and both  $v^\omega$  and  $(vu)^\omega$  begin by  $u^j u_1 u$ , and since the comparison in all four cases is done in the prefix of length  $|u| + |v|$ , we can conclude that (2), (3) and (4) hold.

Let us now consider the case when  $v$  is a nonstrict fractional power of  $u$ , i.e.,  $v$  is a proper prefix of  $u$ . That implies that either  $u$  is not a fractional power of  $v$  or  $u$  is a strict fractional power of  $v$ . Let us consider  $\gtrsim$  the opposite order of  $<$ . Since  $v^\omega \gtrsim u^\omega$ , from what we have seen above, we have that  $(vu)^\omega \gtrsim u^\omega$ ,  $v^\omega \gtrsim (uv)^\omega$  and  $(vu)^\omega \gtrsim (uv)^\omega$ . Thus, conditions (2), (3) and (4) hold.

Finally, let us suppose that the negation of (1) holds, that is that  $u^\omega \gtrsim v^\omega$  (remember that we supposing  $u^\omega \neq v^\omega$ ). Then, using the same reasoning as above, we have  $(uv)^\omega \gtrsim v^\omega$ ,  $u^\omega \gtrsim (vu)^\omega$  and  $(uv)^\omega \gtrsim (vu)^\omega$ , i.e., the negations of the three last conditions. This shows that each of the condition (2), (3) or (4) implies (1).  $\square$

#### 4. Generalized Lyndon words

In this section we introduce generalized Lyndon words.

**Definition 12.** Given an alphabet  $A$  and a generalized order  $<$  on  $A^\infty$  we say that a finite word  $w \in A^+$  is a *generalized Lyndon word* if for any nontrivial factorization  $w = uv$  one has  $w^\omega < (vu)^\omega$ .

**Example 13.** Let  $A = \{a, b\}$  and  $<$  be the order defined in Example 5. The word  $w = abba$  is a generalized Lyndon word for the order  $<$ . Indeed, one can easily check that  $(abba)^\omega < (aabb)^\omega < (bbaa)^\omega < (baab)^\omega$ .

##### 4.1. Characterization of generalized Lyndon words

In the next theorem we give two characterizations of generalized Lyndon words. Recall that a classical result due to Lyndon states that a word  $w$  is a classical Lyndon word if and only if  $w <_{lex} v$  for any nontrivial proper suffix of  $w$  (see [13, Proposition 5.1.2]).

The second part of the next result has already been proved in [19, Proposition 2.1]. We give here a shorter proof.

**Theorem 14.** *Let us consider a generalized lexicographical order  $<$  on  $A^\infty$ .*

1. *A word  $w$  is a generalized Lyndon word if and only if for any nontrivial factorization  $w = uv$ , one has  $u^\omega < v^\omega$ .<sup>1</sup>*
2. *A word  $w$  is a generalized Lyndon word if and only if for any nontrivial factorization  $w = uv$ , one has  $w^\omega < v^\omega$ .*

**Proof.** By definition,  $w$  is a generalized Lyndon words if and only if for each nontrivial factorization  $w = uv$ , one has  $(uv)^\omega < (vu)^\omega$ . By Lemma 11, this is equivalent both to  $u^\omega < v^\omega$  and to  $(uv)^\omega < v^\omega$ , i.e.  $w^\omega < v^\omega$ .  $\square$

<sup>1</sup> One may find on Wikipedia the following characterization (without proof nor references):  $w$  is a classical Lyndon word if and only if for each nontrivial factorization  $w = uv$  one has  $u < v$ . Our condition is not an extension of this condition to generalized Lyndon words. Indeed, if one takes the usual order  $a < b$ , one has  $b < ba$  but  $b^\omega > (ba)^\omega$ .

**Example 15.** Let  $w, A$  and  $<$  as in Example 13. Let us consider the nontrivial factorization  $w = uv$  with  $u = abb$  and  $v = a$ . One has  $(abb)^\omega < a^\omega$  and  $(abba)^\omega < a^\omega$ .

4.2. Factorization into generalized Lyndon words

The following result is already proved in [19, Theorem 2.1] using the theory of Hall sets. We give here an independent proof, especially for the uniqueness part, using only combinatorial arguments.

**Theorem 16.** Each word in  $A^*$  can be factorized in a unique way as a nonincreasing product of generalized Lyndon words.

**Proof.** Let us consider a generalized lexicographical order  $<$  on  $A^\infty$ .

To prove the existence of such a nonincreasing factorization we follow the proof of [19, Corollary 2.2]. Let  $w \in A^+$  (if  $w = 1$  there is nothing to prove). We define  $z$  as the shortest among all nontrivial suffixes  $s$  of  $w$  such that  $s^\omega$  is minimum. By Theorem 14,  $z$  is a generalized Lyndon word. If  $w = z$ , we have found our factorization. Otherwise, we can write  $w = uz$  and, by induction on the length of  $u$ , we may assume that  $u = \ell_1 \ell_2 \cdots \ell_n$ , where the  $\ell_i$  are generalized Lyndon words with  $\ell_1^\omega \geq \ell_2^\omega \geq \dots \geq \ell_n^\omega$ . Moreover, we have  $\ell_n^\omega \geq z^\omega$ . Indeed, by construction of  $z$  we have  $(\ell_n z)^\omega \geq z^\omega$ , and thus, using Lemma 11,  $\ell_n^\omega \geq z^\omega$ .

Let us now prove the uniqueness of this factorization. Suppose that we have  $w = \ell_1 \ell_2 \cdots \ell_n$ , where the  $\ell_i$  are generalized Lyndon words with  $\ell_1^\omega \geq \ell_2^\omega \geq \dots \geq \ell_n^\omega$ . Let us show that  $\ell_n$  is uniquely determined by the following condition: it is the shortest among all nontrivial suffixes  $s$  of  $w$  such that  $s^\omega$  is minimum. To prove this it is enough to show that if  $s$  is a nontrivial proper suffix of  $\ell_n$ , then  $\ell_n^\omega < s^\omega$ ; and if  $s$  is a suffix of  $w$  longer than  $\ell_n$ , then  $\ell_n^\omega \leq s^\omega$ . The first inequality follows from point 2 of Theorem 14 and the fact that  $\ell_n$  is a generalized Lyndon word. Suppose now that the second one is not true. Thus there exists some  $i$ , with  $2 \leq i \leq n$ , and some factorization  $\ell_{i-1} = uv$  with  $v$  nonempty, such that  $s = v \ell_i \cdots \ell_n$ , and

$$(v \ell_i \cdots \ell_{n-1} \ell_n)^\omega < \ell_n^\omega.$$

From this last inequality and from Lemma 11 we deduce that  $(v \ell_i \cdots \ell_{n-1})^\omega < \ell_n^\omega$ . Since  $\ell_n^\omega \leq \ell_{n-1}^\omega$ , we thus have

$$(v \ell_i \cdots \ell_{n-1})^\omega < \ell_{n-1}^\omega.$$

Continuing recursively, we find that  $(v \ell_i)^\omega < \ell_i^\omega$ , therefore  $(v \ell_i)^\omega < \ell_i^\omega$  and, since  $\ell_i^\omega \leq \ell_{i-1}^\omega$ , that  $v^\omega < \ell_{i-1}^\omega$ . This gives us a contradiction to Theorem 14, since  $\ell_{i-1} = uv$  is a generalized Lyndon word. Thus  $\ell_n$  is uniquely determined and, by proceeding recursively we prove the uniqueness of the factorization.  $\square$

From the proof of the previous theorem we obtain the two following results.

**Corollary 17.** Let  $w = \ell_1 \ell_2 \cdots \ell_n$ , with  $\ell_i$  generalized Lyndon words such that  $\ell_1^\omega \geq \ell_2^\omega \geq \dots \geq \ell_n^\omega$ . Then  $\ell_n$  is the shortest among all nontrivial suffixes  $s$  of  $w$  such that  $s^\omega$  is minimum.

**Corollary 18.** With the same hypothesis as in Corollary 17, we have that  $\ell_n$  is the longest suffix of  $w$  which is a generalized Lyndon word.

Note that this result is known for classical Lyndon words (see [18, Lemma 7.14 (ii)] and [11]).

**Proof of Corollary 18.** Indeed, if there exists a suffix  $s$  longer than  $\ell_n$  which is a generalized Lyndon word, then, since  $s$  has  $\ell_n$  as a proper suffix, we would have  $s^\omega < \ell_n^\omega$  by point 2 of Theorem 14, contradicting Corollary 17.  $\square$

**Example 19.** Let us consider the word  $w = aabaabaabb$  with the order defined in Example 5. The unique nonincreasing factorization of  $w$  in generalized Lyndon words is  $w = (a)(aba)(aba)(abb)$ .

Note that every factor  $\ell$  of the factorization in Lyndon words is primitive, i.e., if  $\ell = w^r$  with  $w$  a finite word and  $r$  an integer, then  $r = 1$  and  $\ell = w$ .

5. Classical Lyndon words

In this section, we take as generalized lexicographical order the usual lexicographical order  $<_{lex}$ , simply denoted by  $<$ . Clearly, a generalized Lyndon word for this order is a usual one, since for two finite words of the same length, one has  $u < v$  if and only if  $u^\omega < v^\omega$ . (see [6, Theorem 8]).

### 5.1. Factorization into Lyndon words

The nonincreasing factorization of a word into Lyndon words, as in Theorem 16, is the usual nonincreasing factorization into Lyndon words (see, for instance, [13]).

While at the end of Section 4 we gave two characterizations of the last element of the factorization, here we focus on the first factor. This result is motivated by point 1 of Theorem 14: the fact that a word  $w$  is not a Lyndon word implies the existence of a prefix  $u$  such that  $u^\omega \geq v^\omega$ , where  $v$  is the corresponding suffix. If one chooses the shortest prefix satisfying this property, this turns out to be the first factor in the Lyndon factorization. In the same vein, it is motivated by Ufnarovskij’s Theorem (Corollary 22 below).

**Theorem 20.** *Let  $w = \ell_1 \ell_2 \cdots \ell_n$  be the nonincreasing factorization into Lyndon words of a finite nonempty word  $w$ .*

1. *The word  $\ell_1$  is the shortest nontrivial prefix  $p$  of  $w$  such that, when writing  $w = ps$ , one has either  $s = 1$  or  $p^\omega \geq s^\omega$ .*
2. *The word  $\ell_1$  is the shortest nontrivial prefix  $p$  of  $w$  such that  $p^\omega \geq w^\omega$ .*

In order to prove Theorem 20 we need a preliminary result which refines Lemma 11 in the case of the usual lexicographical order.

Note that, for any infinite words  $s, t$  such that  $s < t$ , with  $<$  the classical order, and for any finite word  $w$ , one has  $ws < wt$ .

**Lemma 21.** *Let  $u, v$  be two nonempty words. Then each of the two following conditions is equivalent to each of the four conditions in Lemma 11:*

- (5)  $u^\omega < (uv)^\omega$ ;
- (6)  $(vu)^\omega < v^\omega$ .

**Proof.** Condition (3) in Lemma 11 is equivalent to condition (5): indeed  $u^\omega < (vu)^\omega \Leftrightarrow uu^\omega < u(vu)^\omega \Leftrightarrow u^\omega < (uv)^\omega$ . Similarly, condition (2) is equivalent to condition (6): indeed,  $(uv)^\omega < v^\omega \Leftrightarrow v(uv)^\omega < vv^\omega \Leftrightarrow (vu)^\omega < v^\omega$ .  $\square$

Note that the previous lemma implies that if  $u^\omega < v^\omega$ , then

$$u^\omega < (uv)^\omega < (vu)^\omega < v^\omega,$$

a result proved by Bergman in [3, Lemma 5.1] (see also [21, p.34 and pp.101–102]).

The following result is [21, Theorem 2, p.35].

**Corollary 22 (Ufnarovskij).** *A word  $w$  is a Lyndon word if and only if for any nontrivial factorization  $w = ps$ , one has  $p^\omega < w^\omega$ .*

**Proof.** It follows from point 1 of Theorem 14 and from Lemma 21.  $\square$

**Example 23.** The word  $w = aabab$  is a Lyndon word. We have  $a^\omega = (aa)^\omega < (aaba)^\omega < (aab)^\omega < w^\omega$ .

**Corollary 24.** *If  $\ell_1, \ell_2, \dots, \ell_n$ , with  $n \geq 2$ , are Lyndon words such that  $\ell_1^\omega \geq \ell_2^\omega \geq \dots \geq \ell_n^\omega$ , then  $\ell_1^\omega \geq (\ell_2 \cdots \ell_n)^\omega$ .*

**Proof.** The case  $n = 2$ , it is trivial. Let consider the case  $n \geq 3$ . By induction hypothesis we have  $\ell_2^\omega \geq (\ell_3 \cdots \ell_n)^\omega$ . From Lemma 21 it follows that  $\ell_2^\omega \geq (\ell_2 \cdots \ell_n)^\omega$ . Hence,  $\ell_1^\omega \geq (\ell_2 \cdots \ell_n)^\omega$ .  $\square$

It is well-known that all (classical) Lyndon words are unbordered (see, for instance, [9]).

**Proof of Theorem 20.** Let us prove the first assertion. When  $n = 1$ , then  $w = \ell_1$  is a Lyndon word and the result is true by point 1 of Theorem 14.

Suppose now that  $n \geq 2$ . Then, by Corollary 24, we have  $\ell_1^\omega \geq (\ell_2 \cdots \ell_n)^\omega$ . Let  $p$  be a nontrivial prefix of  $w$  shorter than  $\ell_1$ . Thus, we have a nontrivial factorization  $\ell_1 = pq$  for some  $q \neq 1$ . By Theorem 14, we know that  $p^\omega < q^\omega$ . Since  $\ell_1$  is unbordered,  $q$  cannot be a fractional power of  $p$ . Thus, by point (ii) of Lemma 8, one has  $p^\omega < (q\ell_2 \cdots \ell_n)^\omega$ , which prove the first part of the theorem.

The second assertion just follows from the first one. Indeed, using Lemma 21, we have that if  $s \neq 1$ , then  $p^\omega \geq s^\omega$  is equivalent to  $p^\omega \geq (ps)^\omega$ .  $\square$

**Example 25.** Let  $w = ababaab$ . Its nonincreasing factorization into Lyndon words is  $w = (ab)(ab)(aab)$ . One can check that  $(ab)^\omega > w^\omega > (abaab)^\omega$  while  $a^\omega < w^\omega < (babaab)^\omega$ .

### 6. Galois words

In this section we consider a particular generalized lexicographical order.

**Definition 26.** Let  $<_1$  be an order on  $A$ . The *alternating lexicographical order*  $<_{alt}$  with respect to  $<_1$  is the generalized lexicographical order defined by the sequence  $(<_n)_{n \geq 1}$  with  $<_n = <_1$  if  $n$  is odd, and  $<_n = \tilde{<}_1$  if  $n$  is even.

**Example 27.** Let us consider  $<_1$  as the usual order on  $\{a, b\}$ . Then one has  $(ab)^\omega <_{alt} a^\omega <_{alt} b^\omega <_{alt} (ba)^\omega$ .

This order is relevant when one orders real numbers through their continued fractions, see for example [19, p.1-2]. The terminology in the following definition is justified in [19, p.2].

**Definition 28.** A *Galois word* is a generalized Lyndon word for an alternating lexicographical order.

**Example 29.** Let us consider  $<_1$  the usual order on  $\{a, b, c\}$ . The following are Galois words:  $b, ac, bc, aba, abb, abaa, acab$ .

#### 6.1. Characterization of Galois words

Similarly to what we saw in Section 5.1 for the classical order, for any infinite words  $s, t$  such that  $s <_{alt} t$ , and any finite word  $w$ , one has  $ws <_{alt} wt$  if  $|w|$  is even, and  $ws >_{alt} wt$  if  $|w|$  is odd.

Symmetrically, when  $ws <_{alt} wt$  one has  $s <_{alt} t$  if  $|w|$  is even and  $s >_{alt} t$  if  $|w|$  is odd.

**Example 30.** Let us consider the order of Example 27. One has  $(ab)a^\omega <_{alt} (ab)b^\omega$  and  $ba^\omega >_{alt} bb^\omega$ .

Using the previous observation we can prove the next lemma using the same techniques as in Lemma 21.

**Lemma 31.** Let  $u, v$  be nonempty words. Then each of the two following conditions is equivalent to each of the four conditions in Lemma 11 when considering the order  $<_{alt}$ :

- (5)  $u^\omega <_{alt} (uv)^\omega$  if  $|u|$  is even and  $u^\omega >_{alt} (uv)^\omega$  if  $|u|$  is odd;
- (6)  $(vu)^\omega <_{alt} v^\omega$  if  $|v|$  is even and  $(vu)^\omega >_{alt} v^\omega$  if  $|v|$  is odd.

**Proof.** Let us first suppose that  $|u|$  is even. By the remark at the beginning of the section, one has  $u^\omega <_{alt} (vu)^\omega \Leftrightarrow u^\omega = uu^\omega <_{alt} u(vu)^\omega = (uv)^\omega$ . Using the same remark we have, in the case  $|u|$  is odd,  $u^\omega <_{alt} (vu)^\omega \Leftrightarrow u^\omega = uu^\omega >_{alt} u(vu)^\omega = (uv)^\omega$ . Thus condition (3) of Lemma 11 is equivalent to condition (5).

Similarly, condition (2) of Lemma 11 is equivalent to condition (6). Indeed, whenever  $|v|$  is even one has  $(uv)^\omega <_{alt} v^\omega \Leftrightarrow (vu)^\omega = v(uv)^\omega <_{alt} vv^\omega = v^\omega$ , and whenever  $|v|$  is odd one has  $(uv)^\omega <_{alt} v^\omega \Leftrightarrow (vu)^\omega = v(uv)^\omega >_{alt} vv^\omega = v^\omega$ .  $\square$

The following characterization of Galois words can be seen as a generalization of Ufnarovskij’s Theorem (Corollary 22).

**Theorem 32.** A word  $w$  is a Galois word if and only if for any nontrivial factorization  $w = ps$ , one has the following condition:  $p^\omega <_{alt} w^\omega$  if  $|p|$  is even and  $p^\omega >_{alt} w^\omega$  if  $|p|$  is odd.

**Proof.** The result immediately follow from point 1 in Theorem 14 and from Lemma 31.  $\square$

#### 6.2. Factorization into Galois words

Suppose that  $w = g_1 g_2 \cdots g_n$  is the nonincreasing factorization of  $w$  in Galois words. We call *multiplicity* of  $g_1$  the number  $m = \text{Card}\{i \mid g_i = g_1\}$ . In other words  $w = g_1^m g_{i+1} \cdots g_n$ , with  $g_1^\omega >_{alt} g_{i+1}^\omega$ .

The following result is a generalization of Theorem 20 to Galois words. This result is motivated by Theorems 14 and 32.

**Theorem 33.** Let  $w = g_1 g_2 \cdots g_n$  with  $g_i$  Galois words satisfying  $g_1^\omega \geq_{alt} g_2^\omega \geq_{alt} \cdots \geq_{alt} g_n^\omega$ . Let  $m$  be the multiplicity of  $g_1$ . Let  $p$  be the shortest nontrivial prefix of  $w$  such that

$$p^\omega \geq_{alt} w^\omega \text{ if } |p| \text{ is even and } p^\omega \leq_{alt} w^\omega \text{ if } |p| \text{ is odd.} \tag{\star}$$

Then

- (i) if  $|g_1|$  is odd,  $m$  is even, and  $m < n$ , then  $p = g_1^2$ ;

(ii) otherwise,  $p = g_1$ .

Note that we can give an equivalent condition on  $p$  in the previous statement.

**Lemma 34.** *Let  $w = ps$  be a finite word, with  $p \neq 1$ . Then  $p$  satisfies condition  $(\star)$  if and only if  $p$  is such that one has either  $s = 1$  or  $p^\omega \geq_{alt} s^\omega$ .*

**Proof.** From Lemma 31 it follows that one has  $p^\omega \geq_{alt} s^\omega \Leftrightarrow p^\omega \geq_{alt} w^\omega$  when  $|p|$  is even and  $p^\omega \geq_{alt} s^\omega \Leftrightarrow p^\omega \leq_{alt} w^\omega$  when  $|p|$  is odd.  $\square$

In order to prove Theorem 33 we need several lemmata.

Recall, from Section 5.1 that classical Lyndon words have no border. This is no more true for Galois words, as shown in the next lemma.

**Lemma 35.** *([19, Proposition 3.1]) If a Galois word has a border, then it has odd length.*

**Lemma 36.** *Let  $g, h$  be Galois words with  $g^\omega <_{alt} h^\omega$  and  $g$  a prefix of  $h$ . Then  $|g|$  is even.*

**Proof.** Let  $g$  be a prefix of  $h$  with  $|g|$  odd. Then by Theorem 32, one has  $g^\omega >_{alt} h^\omega$ .  $\square$

**Lemma 37.** *Let  $g, h$  be Galois words with  $g^\omega <_{alt} h^\omega$ . If  $g$  is a strict fractional power of  $h$  then  $|g|$  is even.*

**Proof.** Let  $k \geq 1$  and  $h'$  a prefix of  $h$  such that  $g = h^k h'$ . The factorization  $h = h' h''$  is not trivial since  $h^\omega \neq g^\omega$ . Both  $g^\omega$  and  $h^\omega$  have  $g$  as a prefix, and since  $g^\omega <_{alt} h^\omega$ , we have  $gh' h'' s <_{alt} gh'' h' t$ , where  $s = h^{k-1} h' g^\omega$  and  $t = h'' h^\omega$ .

Let us suppose by contradiction that  $|g|$  is odd. By the remark at the beginning of Section 6.1, we have  $h' h'' s >_{alt} h'' h' t$ . Since  $|h' h''| = |h'' h'| = |h|$  but  $h' h'' \neq h'' h'$ , the comparison of the last inequality takes place within the prefix of length  $|h|$ . Thus, by Lemma 7, one has  $h^\omega >_{alt} (h'' h')^\omega$ , which is impossible since  $h$  is a Galois word.  $\square$

**Lemma 38.** *Let  $w = g_1 g_2 \cdots g_n$  with  $g_i$  Galois words satisfying  $g_1^\omega \geq_{alt} g_2^\omega \geq_{alt} \cdots \geq_{alt} g_n^\omega$ . Let  $m$  be the multiplicity of  $g_1$  and assume that  $m < n$  and that  $n \geq 2$ . Then:*

- (i) if  $|g_1|$  is odd and  $m$  is even, then  $g_1^\omega <_{alt} (g_2 \cdots g_n)^\omega$ ;
- (ii) otherwise,  $g_1^\omega >_{alt} (g_2 \cdots g_n)^\omega$ .

**Proof.** If  $n = 2$  then we have  $m = 1$  and thus  $g_1^\omega >_{alt} g_2^\omega$ , i.e.,  $m$  is odd and condition (ii) holds.

Suppose now that  $n \geq 3$ . Let us first suppose that  $|g_2|$  is even. If  $g_2 = \cdots = g_n$ , then necessarily we have  $g_1 \neq g_2$ , since  $m < n$ . Therefore,  $g_1^\omega >_{alt} g_2^\omega = (g_2 \cdots g_n)^\omega$  and  $m = 1$  is odd, so we are in case (ii). If  $g_2, \dots, g_n$  are not all equal, we can argue by induction on  $n$ . Thus since  $|g_2|$  is even we have  $g_2^\omega >_{alt} (g_3 \cdots g_n)^\omega$ . By applying Lemmata 11 and 31 we find that  $g_2^\omega >_{alt} (g_2 g_3 \cdots g_n)^\omega$ . Finally, since  $g_1^\omega \geq_{alt} g_2^\omega$ , we have  $g_1^\omega >_{alt} (g_2 \cdots g_n)^\omega$ ; moreover, either  $m$  is odd, or  $m$  is even and then  $g_1 = g_2$  and  $|g_1|$  is even, so that we are in case (ii).

Suppose now that  $|g_2|$  is odd. We assume first that  $g_1 \neq g_2$ , i.e.,  $m = 1$ .

We have  $g_1^\omega >_{alt} g_2^\omega$ . We show that  $g_2$  is not a fractional power of  $g_1$ . Indeed,  $g_2$  is not a prefix of  $g_1$  by Lemma 36; moreover, by Lemma 37,  $g_2$  is not a strict fractional power of  $g_1$ ; since being a fractional power is equivalent to be a prefix, or a strict fractional power, we are done.

Hence, by Lemma 8 (ii) (applied to the opposite order), it follows that  $g_1^\omega >_{alt} (g_2 \cdots g_n)^\omega$ .

Finally, let us consider the case  $g_1 = g_2$  and  $|g_2|$  odd. We have  $m \geq 2$  and  $w = g_1^m g_{m+1} \cdots g_n$ , with  $g_1^\omega >_{alt} g_{m+1}^\omega$ . By induction applied to  $g_1 g_{m+1} \cdots g_n$  we have  $g_1^\omega >_{alt} (g_{m+1} \cdots g_n)^\omega$ , i.e., condition (ii) holds. Since  $|g_1| = |g_2|$  is odd, one deduces, by using Lemmata 11 and 31, that  $g_1^\omega <_{alt} (g_1^{m-1} g_{m+1} \cdots g_n)^\omega$  when  $m$  is even and  $g_1^\omega >_{alt} (g_1^{m-1} g_{m+1} \cdots g_n)^\omega$  when  $m$  is odd.  $\square$

An interesting consequence of the previous lemma is the following.

**Corollary 39.** *Let  $w = g_1 g_2 \cdots g_n$  with  $g_i$  Galois words satisfying  $g_1^\omega \geq_{alt} g_2^\omega \geq_{alt} \cdots \geq_{alt} g_n^\omega$ . Let  $m$  be the multiplicity of  $g_1$ . One has  $g_1^\omega >_{alt} w^\omega$  if  $|g_1^m|$  is even and  $g_1^\omega <_{alt} w^\omega$  if  $|g_1^m|$  is odd.*

**Proof.** This follows from Lemma 31 with  $v = g_1^m$  and  $u = (g_{m+1} \cdots g_n)$ .  $\square$

We can now prove the main result of the section.



**Proof of Theorem 33.** Let us prove first that the two prefixes,  $g_1$  for the case (ii) and  $g_1^2$  for the case (i) satisfy condition  $(\star)$ . If we are under the hypotheses of case (ii), i.e., if  $|g_1|$  is even or  $m$  is odd (the two conditions are not mutually exclusive), then  $g_1^\omega >_{alt} (g_2 \cdots g_n)^\omega$  by Lemma 38. If we are under the hypotheses of case (i), then we have  $n - 1 \geq 2$  and the multiplicity of  $g_1$  in  $g_1 g_3 \cdots g_n$  is odd. Hence, applying Lemma 38 we find that  $(g_1)^\omega = (g_1^2)^\omega >_{alt} (g_3 \cdots g_n)^\omega$ . In both cases the result follows from Lemma 34.

Let us now prove that any nontrivial proper prefix of  $g_1$  in case (ii) and of  $g_1^2$  in case (i) does not satisfy condition  $(\star)$ .

Let  $g_1 = pt$  be a non trivial factorization of  $g_1$  and  $s = tg_2 \cdots g_n$ . By Theorem 14, we have  $p^\omega <_{alt} t^\omega$ . If the comparison in the last inequality takes place within the first  $t$  of  $t^\omega$ , then  $p^\omega <_{alt} s^\omega$ , and we can conclude by using Lemma 34.

Otherwise, by Lemma 3,  $t$  is a fractional power of  $p$ , i.e., we can write  $t = p^r$ ,  $r \in \mathbb{Q} \setminus \mathbb{N}$  (since  $g_1 = pt$  is primitive). We claim that  $|t|$  is odd. Indeed, if we suppose that  $r < 1$ , then  $t$  is a prefix of  $p$ , hence of  $g_1$ , so that  $|t|$  is odd by Lemma 35. If we suppose that  $r > 1$ , then we can write  $t = p^h p'$ , with  $h \geq 1$  and  $p = p' p''$ . Thus  $g_1 = (p' p'')^{h+1} p'$  and  $p', p p'$  are both borders of  $g_1$ . This implies by Lemma 35 that  $|p'|$  is odd and  $|p|$  even, hence  $|t|$  is odd. This implies that  $|p|$  is even if and only if  $|g_1|$  is odd.

Let us suppose first that  $|g_1|$  is odd. Then  $p^\omega <_{alt} g_1^\omega$  by Theorem 32. If we are in case (ii), we have  $g_1^\omega \leq_{alt} w^\omega$  by Corollary 39, hence  $p^\omega <_{alt} w^\omega$ , as we wanted to prove. If we are in case (i), then  $g_1 = g_2$ . If the comparison between  $p^\omega$  and  $g_1^\omega$  takes place within the first  $g_1$  of  $g_1^\omega$ , we conclude that  $p^\omega <_{alt} (g_1 g_2 \cdots g_n)^\omega = w^\omega$ . Otherwise, by Lemma 3, we can write  $g_1 = p^k p_1$ , with  $k \geq 1$ , and  $p = p_1 p_2$ . Since  $g_1 = (p_1 p_2)^k p_1$  is a Galois word, we have  $((p_1 p_2)^k p_1)^\omega <_{alt} ((p_2 p_1)^k p_1)^\omega$ . Since  $|(p_1 p_2)^k p_1| = |(p_2 p_1)^k p_1|$ , the comparison in the previous (strict) inequality takes place in the prefix of their common length, hence in their prefix of length  $|p| = |p_1 p_2| = |p_2 p_1|$ . Thus,  $(p_1 p_2) s_0 <_{alt} (p_2 p_1) t_0$  for all infinite words  $s_0, t_0$ .

Since  $g_2$  has  $p_1 p_2$  as a prefix, we deduce that  $(g_2 \cdots g_n g_1)^\omega <_{alt} (p_2 p_1)^\omega$ . Therefore, since  $|g_1|$  is odd, we have

$$p^\omega = (p_1 p_2)^k p_1 (p_2 p_1)^\omega = g_1 (p_2 p_1)^\omega <_{alt} g_1 (g_2 \cdots g_n g_1)^\omega = w^\omega.$$

Let us suppose now that  $|g_1|$  is even (and thus  $|p|$  is odd). From Theorem 32 it follows that  $p^\omega >_{alt} g_1^\omega$  and from Corollary 39 it follows that  $g_1^\omega \geq_{alt} w^\omega$ . Thus  $p^\omega >_{alt} w^\omega$ .

We have proved that no nontrivial proper prefix of  $g_1$  satisfy condition  $(\star)$ . It remains to prove that, under the hypotheses of case (i), each proper prefix of  $g_1^2$  of length at least  $|g_1|$  does not satisfy condition  $(\star)$ .

Since we are in case (i),  $|g_1|$  is odd,  $m$  is even and  $g_1 \neq g_n$ . Using Lemma 38 we have that  $g_1^\omega <_{alt} (g_2 \cdots g_n)^\omega$ . Thus it follows from Lemma 34 that  $g_1$  does not satisfy condition  $(\star)$ .

Let now consider  $p = g_1 q$  with  $g_1 = qt$  a nontrivial factorization of  $g_1$ . If  $|q|$  is even, and thus  $|p|$  is odd, we have  $q^\omega <_{alt} g_1^\omega$  by Theorem 32. Using Lemma 31 and Corollary 39 we find

$$p^\omega = (g_1 q)^\omega >_{alt} g_1^\omega \geq_{alt} w^\omega.$$

Finally, let us suppose that  $|q|$  is odd, and thus  $|p|$  and  $|t|$  are even. By Theorem 14 we have  $(qt)^\omega = g_1^\omega <_{alt} t^\omega$ . Since  $t$  is not a prefix of  $g_1$  (being of even length, it cannot be a border of  $g_1$ ), the comparison is within the first  $t$  of  $t^\omega$ . Thus we have  $(qtq)^\omega <_{alt} (tg_3 \cdots g_n qtq)^\omega$ . Since  $|qtq|$  is even and  $g_1 = g_2 = qt$ , we deduce that

$$p^\omega = (g_1 q)^\omega = (qtq)^\omega <_{alt} (qtqtg_3 \cdots g_n)^\omega = w^\omega.$$

Therefore, in both cases  $p$  does not satisfy condition  $(\star)$ .  $\square$

**Example 40.** Let us consider the word  $w = ababbabaa$  using the alternating order of Example 25. The nonincreasing factorization of  $w$  in Galois words is  $w = (abb)(abb)(abaa)$ . One can check that  $((abb)^2)^\omega >_{alt} w^\omega$ , and that each nontrivial proper prefix of  $(abb)^2$  does not satisfy condition  $(\star)$  of Theorem 33.

## 7. Remarks and open problems

Generalized Lyndon words, defined by using orders different than the usual lexicographical one, have different behaviors than classical Lyndon words.

For instance, we have seen that generalized Lyndon words are not, in general, unbordered, as is the case for classical Lyndon words. Moreover, it is known that for each primitive word  $w$ , its unique conjugate which is a classical Lyndon word, is one of the elements in the nonincreasing factorization of  $ww$  into Lyndon words (see, e.g. [18, Section 7.4.1] and [15] where are given algorithms to compute this unique conjugate). This is no more true for generalized Lyndon words, as shown in the next example.

**Example 41.** Let us consider the primitive word  $w = baa$ . Let us consider the alternating order on  $\{a, b\}^\infty$  with  $a <_1 b$ . The unique Galois word conjugate to  $w$  is  $u = aba$ . The nonincreasing factorization of  $ww$  into Galois words is  $ww = (b)(a)(abaa)$ .

In the classical case it is easy to show that we have a symmetric result of Corollary 17, namely that  $\ell_1^\omega$  is maximum along all  $p^\omega$ , with  $p$  a nontrivial prefix of  $w$ . This is not true for general orders, as shown in the next example.

**Example 42.** Let us consider the word  $w = abab$  and the alternating order of Example 41. Its nonincreasing factorization in Galois word is  $w = (ab)(ab)$ . If we consider the nontrivial proper prefix  $a$ , we have  $a^\omega >_{alt} (ab)^\omega$ .

Moreover, using the same notation as before, it is not true in general that  $\ell_1$  is the longest among all prefixes of  $w$  which are generalized Lyndon words (this is true for classical Lyndon words, see, e.g., [18, Lemma 7.14 (iii)]).

**Example 43.** Let  $w = abab$  as in Example 42. The prefix  $aba$  is longer than  $ab$  and it is also a Galois word.

For classical Lyndon words, it is known that the unique nonincreasing factorization of a word in Lyndon words is also the factorization into Lyndon words which has the less number of factors.<sup>2</sup> This is no more true for generalized Lyndon words, as shown in the next example.

**Example 44.** Let us consider the alternating lexicographical order of Example 41. Then  $w = (ab)(ab)(ab)$  is a word with its nonincreasing factorization in Galois words. The word  $w$  admits a shorter factorization into Galois words, namely  $w = (ababa)(b)$ .

In [11], Duval shows that given a finite word  $w$ , it is possible to compute in linear time its nonincreasing factorization into classical Lyndon words.

**Open Problem 1.** Generalize Duval's algorithm to generalized Lyndon words.

In the same paper, Duval also proposed a second algorithm generating all Lyndon words of length  $\leq n$ . A consequence of this algorithm is that the number of Lyndon words of length at most  $n$  is equal to the number of words of length  $n$  that are prefixes of a Lyndon word plus 1. This property is no more true for a generalized order, as shown in the next example.

**Example 45.** Let us consider  $A = \{a, b, c\}$  with the usual order. The only 6 Lyndon words of length at most 2 are  $a, b, c, ab, ac$  and  $bc$ . It is easy to check that there are exactly 5 words of length 2 which are prefixes of a Lyndon word, namely  $aa, ab, ac, bb$  and  $bc$ .

Let us now consider  $A$  with the alternating order given by  $a <_1 b <_1 c$ . There are 6 Galois words of length at most 2 (namely  $a, b, c, ab, ac$  and  $bc$ ) but only 3 words of length 2 which are prefixes of Galois words (namely  $ab, ac$  and  $bc$ ).

Finally, all along the paper we only considered finite Lyndon words. In [20] are defined infinite Lyndon words: these are the infinite words which have infinitely many prefixes that are (finite) Lyndon words. Then the authors of [20] prove that  $x$  is an infinite Lyndon word if and only if  $x$  is smaller than any of its nontrivial proper suffixes (Proposition 2.2). They prove also that each infinite word  $x$  is equal to a nondecreasing product of finite Lyndon words and perhaps one infinite one (Proposition 2.3). This means that either  $x = \ell_1 \ell_2 \cdots \ell_n$ , with  $\ell_1, \dots, \ell_{n-1}$  finite Lyndon words and  $\ell_n$  an infinite one, or  $x = \ell_1 \ell_2 \cdots$  is an infinite product of finite Lyndon words; in both cases,  $\ell_1 \geq \ell_2 \geq \dots$ .

Thus, following [4] and [20], we say that an infinite word  $x$  is a *generalized infinite Lyndon word* if  $x$  is smaller than any of its nontrivial proper suffixes. It would be interesting to generalize this result to Galois words and other generalized Lyndon words.

**Open Problem 2.** Prove that each infinite word can be factorized in a unique way as a nonincreasing product of finite and infinite generalized Lyndon words.

## References

- [1] Alberto Apostolico, Maxime Crochemore, Fast parallel Lyndon factorization with applications, *Math. Syst. Theory* 28 (2) (1995) 89–108.
- [2] Danièle Beauquier, Maurice Nivat, On translating one polyomino to tile the plane, *Discrete Comput. Geom.* 6 (6) (1991) 575–592.
- [3] George M. Bergman, Centralizers in free associative algebras, *Trans. Amer. Math. Soc.* 137 (1969) 327–344.
- [4] Luc Boasson, Olivier Carton, Transfinite Lyndon words, in: *Developments in Language Theory*, in: *Lecture Notes in Comput. Sci.*, vol. 9168, Springer, Cham, 2015, pp. 179–190.
- [5] Paola Bonizzoni, Clelia De Felice, Rocco Zaccagnino, Rosalba Zizza, Inverse Lyndon words and inverse Lyndon factorizations of words, *Adv. in Appl. Math.* 101 (2018) 281–319.
- [6] Silvia Bonomo, Sabrina Mantaci, Antonio Restivo, Giovanna Rosone, Marinella Sciortino, Sorting conjugates and suffixes of words in a multiset, *Internat. J. Found. Comput. Sci.* 25 (8) (2014) 1161–1175.
- [7] Srećko Brlek, Jacques-Olivier Lachaud, Xavier Provençal, Christophe Reutenauer, Lyndon + Christoffel = digitally convex, *Pattern Recognit.* 42 (10) (2009) 2239–2246.
- [8] Émilie Charlier, Manon Philibert, Manon Stipulanti, Nyldon Words, 2018, URL: <https://arxiv.org/pdf/1804.09735.pdf>, arXiv:1804.09735.
- [9] Christian Choffrut, Juhani Karhumäki, *Combinatorics of Words, Handbook of Formal Languages*, vol. 1, Springer, Berlin, 1997, pp. 329–438.

<sup>2</sup> This follows easily from the property: if  $u, v$  are Lyndon words and  $u <_{lex} v$ , then  $uv$  is a Lyndon word, see [13, Proposition 5.1.3].

- [10] Maxime Crochemore, Luís M.S. Russo, Cartesian trees and Lyndon trees, CoRR, arXiv:1712.08749, 2017.
- [11] Jean-Pierre Duval, Factorizing words over an ordered alphabet, *J. Algorithms* 4 (4) (1983) 363–381.
- [12] Darij Grinberg, “Nyldon words”: understanding a class of words factorizing the free monoid increasingly, URL, <https://mathoverflow.net/questions/187451/>, 2014.
- [13] M. Lothaire, *Combinatorics on Words*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1997.
- [14] M. Lothaire, *Algebraic Combinatorics on Words*, Encyclopedia of Mathematics and Its Applications, vol. 90, Cambridge University Press, Cambridge, 2002.
- [15] M. Lothaire, *Applied Combinatorics on Words*, Encyclopedia of Mathematics and Its Applications, vol. 105, Cambridge University Press, Cambridge, 2005.
- [16] Roger C. Lyndon, On Burnside’s problem, *Trans. Amer. Math. Soc.* 77 (1954) 202–215.
- [17] Dominique Perrin, Antonio Restivo, Words, in: *Handbook of Enumerative Combinatorics*, Discrete Math. Appl. (Boca Raton), CRC Press, Boca Raton, FL, 2015, pp. 485–539.
- [18] Christophe Reutenauer, *Free Lie Algebras*, London Mathematical Society Monographs. New Series, vol. 7, The Clarendon Press, Oxford University Press, New York, 1993.
- [19] Christophe Reutenauer, Mots de Lyndon généralisés, *Sém. Lothar. Combin.* B 54h (16) (2005) 54.
- [20] Rani Siromoney, Lisa Mathew, V.R. Dare, K.G. Subramanian, Infinite Lyndon words, *Inform. Process. Lett.* 50 (2) (1994) 101–104.
- [21] V.A. Ufnarovskij, Combinatorial and asymptotic methods in algebra, in: *Algebra*, VI, in: *Encyclopaedia Math. Sci.*, vol. 57, Springer, Berlin, 1995, pp. 1–196 [MR1060321 (92h:16024)].