

# Representations of Circular Words

László Hegedüs\*

Benedek Nagy\*<sup>†</sup>

{hegedus.laszlo, nbenedek}@inf.unideb.hu

\*Department of Computer Science,  
Faculty of Informatics, University of Debrecen

<sup>†</sup>Department of Mathematics, Faculty of Arts and Sciences,  
Eastern Mediterranean University, Famagusta, North Cyprus, Mersin-10, Turkey

In this article we give two different ways of representations of circular words. Representations with tuples are intended as a compact notation, while representations with trees give a way to easily process all conjugates of a word. The latter form can also be used as a graphical representation of periodic properties of finite (in some cases, infinite) words. We also define iterative representations which can be seen as an encoding utilizing the flexible properties of circular words. Every word over the two letter alphabet can be constructed starting from  $ab$  by applying the fractional power and the cyclic shift operators one after the other, iteratively.

## 1 Introduction

One of the most popular areas of research in theoretical computer science is combinatorics on words. This field deals with various properties of finite and infinite sequences or words. Being closely related to mathematics, it has connections to algebra, number theory, game theory and several others. Although it was written decades ago, the books of M. Lothaire are good reads and are recommended for researchers who want to get a deep overview of the subject [8, 9, 10]. Axel Thue contributed the first results to the field [19, 20]. Since then many applications in computer science have been discovered (e.g., in string matching, data compression, bioinformatics, etc.).

We deal with circular words (sometimes called necklaces [18] or cyclic words) that are different from linear ones and lead to some interesting new viewpoints. Similar sequences can appear in nature, for example, the DNA sequences of some bacteria has a similar form to a necklace. In the simplest sense, circular words are strongly periodic discrete functions.

Circular words are not as widely investigated as linear words. We hope that our approach and results may show that interesting facts can be obtained by analyzing these sequences. Dirk Nowotka wrote about unbordered conjugates of words in Chapter 4 of his dissertation [13]. Complementing this, we deal with bordered conjugates that have periods smaller than the length of the word. Another related article is [4], where permutations and cyclic permutations of primitive and non-primitive words were investigated. For an overview of current research about circular words, the reader can consult the following articles. Relations to Weinbaum factorizations are investigated in [3]. Several articles were written about pattern avoidance of circular words, for example, [2, 6, 17] to name a few. Other applications in mathematics, namely integer sequences [14, 15] were also considered.

The notion of weak and strong periods was introduced in [7]. One result about periodic functions is often cited in combinatorics on words, since it is clearly about periodic infinite words too. This result belongs to Fine and Wilf [5]. It can be shown by example that this statement is not true for weak periods of circular words [7]. In this paper, we investigate two kinds of representations of circular words continuing the research line of the paper [7] presented at the WORDS 2013 conference in Turku. The

first one is connected to the property that every linear word has a shortest root, while the other one is related to tries (see e.g., [18]).

The structure of the paper is as follows. Section 2 defines the notation and notions used in the rest of the article. After this, in Section 3 we discuss ways of representing circular words with tuples and an algorithm to construct one of these representations. Section 4 is about representing circular words with trees (or tries) and we present some results related to Fibonacci words. At the end in Section 5 some possible directions of future research is discussed.

## 2 Preliminaries

The following notions and notation are used in the rest of the article. We will call a non-empty set of symbols an *alphabet* and denote it by  $\Sigma$ . *Words* (or *linear words*) over  $\Sigma$  are finite sequences of symbols of  $\Sigma$ . The operation of concatenation is defined by writing two words after each-other. The *empty word*, i.e., the empty sequence is denoted by  $\varepsilon$  and it is the unit element of the monoid  $\Sigma^*$ . We also define  $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$ . The *length* of the word  $w \in \Sigma^*$  (denoted by  $|w|$ ) is the length of  $w$  as a sequence, that is, the number of all the symbols in  $w$ . We will use  $\mathbb{N}$  to denote the set of non-negative integers.

We say, that  $v \in \Sigma^*$  is a *factor* of  $w \in \Sigma^*$  if there exist words  $x, y \in \Sigma^*$  such that  $w = xvy$ . Furthermore, if  $x = \varepsilon$  (resp.  $y = \varepsilon$ ), then  $v$  is a *prefix* (resp. *suffix*) of  $w$ . For any word  $w$  and integer  $0 \leq k \leq |w|$ , we denote the *length  $k$  factors* of  $w$  by  $\mathcal{F}_k(w)$ . For arbitrary positive integers  $p$  and  $q$ , we use  $(p \bmod q)$  to denote the remainder of  $\frac{p}{q}$ . Let  $w \in \Sigma^*$  be a word of length  $n$ , that is,  $w = w_1 \dots w_n$ , where  $w_1, \dots, w_n \in \Sigma$ . Then for any  $p \in \mathbb{N}$ , we have  $w^{\frac{p}{n}} = w^{\lfloor \frac{p}{n} \rfloor} w'$ , where  $w' = w_1 \dots w_{(p \bmod n)}$ . We call  $w^{\frac{p}{n}}$  the *fractional power* of  $w$ . From now on we will always refer to the  $i$ th position of a word  $w \in \Sigma^*$  as  $w_i$ . A word  $w \in \Sigma^+$  is *primitive* if there is no word  $v \in \Sigma^*$  such that  $w = v^p$  where  $p \in \mathbb{N}$ ,  $p > 1$ .

A positive integer  $p$  is a *period* of  $w = w_1 \dots w_n$  if  $w_i = w_{i+p}$  for all  $i = 1, \dots, n - p$ . As a complementary notion, word  $v \in \Sigma^*$  is a *border* of  $w \in \Sigma^*$  if  $v$  is a prefix and also a suffix of  $w$ . Each word  $w \in \Sigma^*$  has *trivial borders*  $\varepsilon$  and  $w$ . It is clear, that word  $w$  has a border  $b$  if and only if  $w$  has period  $|w| - |b|$ .

Words  $x$  and  $y$  are *conjugates* if there exist words  $u, v \in \Sigma^*$  such that  $x = uv$  and  $y = vu$ . Related to this notion, we define the *shift* operation  $\sigma(w)$  for all  $w \in \Sigma^*$  as follows:

$$\sigma(w) = w_2 \dots w_n w_1.$$

Moreover,  $\sigma^\ell(w) = \sigma^{\ell-1}(\sigma(w)) = w_{1+\ell} \dots w_n w_1 \dots w_\ell$ . Also, we will use  $\sigma^{-\ell}(w)$  that can also be written as  $\sigma^{|w|-\ell}(w)$ .

Lyndon and Schützenberger stated the following, which characterizes the relation between a word and its non-trivial borders [12].

**Lemma 1** (Lyndon and Schützenberger). *Let  $x \in \Sigma^+$ ,  $y, b \in \Sigma^*$  be arbitrary words. Then  $xb = by$  if and only if there exist  $u \in \Sigma^+$ ,  $v \in \Sigma^*$  and  $k \in \mathbb{N}$  such that  $x = uv$ ,  $y = vu$  and  $b = (uv)^k u = u(vu)^k$ .*

A *circular word* is obtained from a linear word  $w \in \Sigma^*$  if we link its first symbol after the last one, as seen on Figure 1.

One can see from the figure that circular words do not have a beginning nor an end. Nor do the notions of suffix and prefix make sense. A circular word  $w_\circ$  can be seen as the set of all conjugates of  $w$ , or all cyclic shifts of  $w$ , that is, the set

$$w_\circ = \{v \mid v \text{ is a conjugate of } w\} = \{\sigma^\ell(w) \mid \ell = 0, \dots, |w| - 1\}.$$

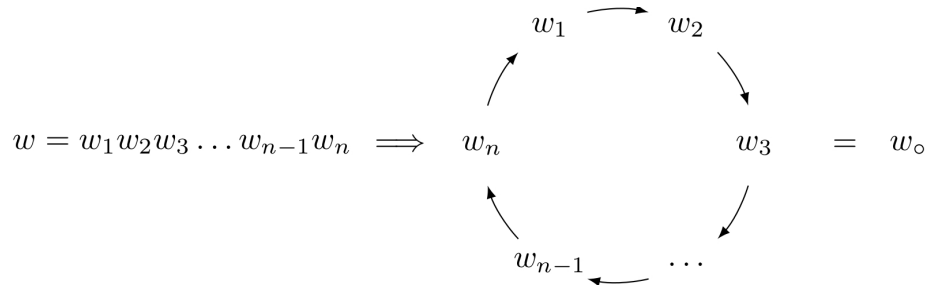


Figure 1: Creating the circular word  $w_\circ$  from the linear word  $w$ .

Note, that  $w_\circ$  consists exactly of the length  $|w|$  factors of  $ww$ . That is,  $w_\circ = \mathcal{F}_{|w|}(ww)$ . The notions of weak- and strong periods were given in [7]. We will only refer to weak periods in this paper and define them as follows.

**Definition 1.** The positive integer  $p$  is a *weak (strong) period* of a circular word  $w_\circ$  if  $p$  is a period of at least one (all) of the conjugates  $v \in w_\circ$ .

### 3 Representations with tuples

If not stated otherwise, we assume that alphabet  $\Sigma$  can be arbitrary. Every word  $w \in \Sigma^*$  can be represented by a power of a (possibly shorter) word  $u \in \Sigma^*$  and a positive integer that is the length of  $w$ . In other words, for all  $w \in \Sigma^*$ , there exists a word  $u \in \Sigma^*$  such that  $u^{\frac{|w|}{|u|}} = w$ . We will call such a  $u$  a *root* of  $w$ , while the shortest root is called the *primitive root* of  $w$  (see e.g., pages 10–11 of [18]). In this section we discuss analogous representations of circular words that take advantage of their lack of strictly specified endpoints.

**Definition 2.** A pair  $(u, n) \in \Sigma^* \times \mathbb{N}$  is a *representation* of the circular word  $w_\circ$  over  $\Sigma$  if  $|u| \leq n$ ,  $n = |w_\circ|$  and  $u^{\frac{n}{|u|}} \in w_\circ$ .

**Definition 3.** A *minimal representation* of a circular word  $w_\circ$  over  $\Sigma$  is a representation  $(u, n)$  of  $w_\circ$ , such that  $|u| \leq |u'|$  for any other representation  $(u', n)$  of  $w_\circ$ .

It is clear, that every circular word has a minimal representation, since all of them have a smallest weak period. Trivially, that not all pairs  $(u, n)$  are minimal representations of some circular word. For example, consider the representation  $(baa, 5)$  of the circular word  $(baaba)_\circ$ . This circular word also has a representation  $(ab, 5)$  which is in fact a minimal representation.

It is also true, that a circular word can have more than one minimal representations. For example,  $(ababa, 12)$ ,  $(babaa, 12)$ ,  $(abaab, 12)$  and  $(baaba, 12)$  are all minimal representations of the circular word  $(ababaababaa)_\circ$ . Note, that  $(aabab, 12)$  is not a minimal representation of this circular word, since it represents  $(aababaababaa)_\circ$ .

Clearly, if  $n = k \cdot |u|$  for some  $k \in \mathbb{N}$  in a minimal representation  $(u, n)$ , then  $(\sigma^\ell(u), n)$  is also a minimal representation of the same circular word for all  $\ell = 0, \dots, |u| - 1$ .

Suppose, that  $w = u^m u'$  for some  $u \in \Sigma^*$  where  $u'$  is a non empty prefix of  $u$  and  $m \in \mathbb{N} \setminus \{0\}$ . Then for every  $k \in \mathbb{N}$ , the word  $w' = wu^k$  has a cyclic shift  $\sigma^{|w'|}(w') = u^{k+m} u'$ . Thus the circular word  $w'_\circ$  has a representation  $(u, |w'| + k \cdot |u|)$ .

**Theorem 1.** *Let  $(u, n)$  be a representation of  $w_\circ$ . Suppose, that  $u$  has border  $s$ , that is,  $u = sx = ys$ , and  $n = 2 \cdot |u| - |s|$ . Then  $(y, n)$  is also a representation of  $w_\circ$ . Moreover, if  $s$  is the longest non-trivial border of  $u$ , then  $(y, n)$  is a minimal representation of  $w_\circ$ .*

*Proof.* Let us have a representation  $(u, n)$  of  $w_\circ$  that satisfies the assumption, that is,  $u$  has border  $s$  and  $n = 2 \cdot |u| - |s|$ . Then  $u$  is in the form  $u = sx = ys$  for some  $x, y \in \Sigma^*$  and  $w_\circ = (uy)_\circ = (ysy)_\circ$ . By Lemma 1,  $ys$  has period  $|y|$ , thus  $w_\circ$  has weak period  $|y|$  and a representation  $(y, n)$ .

If  $s$  is the longest non-trivial border of  $u$ , then  $y$  is the primitive root of  $u$ , thus  $(y, n)$  is a minimal representation of  $w_\circ$ .  $\square$

Suppose that we have a representation  $w_\circ = (u, n)$ , where  $u \in \Sigma^*$  and  $n \in \mathbb{N}$ . If  $|u| \geq 2$ , then  $u$  may be compressed further. In other words, we can take a minimal representation  $(u', |u|)$  with an additional parameter  $k \in \mathbb{N}$ , such that  $\sigma^k(u)$  has primitive root  $u'$ . This method of compression can be done finitely many times, until reaching a word  $u_0$  which we will refer to as a *minimal root* of  $w_\circ$ . We will call these representations *iterative representations*, defined formally in Definition 4. Of course, if a minimal root of a word  $w_\circ$  has only one letter, then it is in the form  $(a^{|w|})_\circ$  for some  $a \in \Sigma$ . In this case, this letter is unique and we can refer to it as the minimal root of  $w_\circ$ . Thus words in these forms have trivial representations and we will no longer deal with them.

**Definition 4.** Let  $u \in \Sigma^*$ ,  $m \in \mathbb{N} \setminus \{0\}$  and  $\ell_1, \ell_2, \dots, \ell_{m-1}, \ell_m, k_1, k_2, \dots, k_{m-1} \in \mathbb{N}$ . The  $2m$ -tuple

$$(u, \ell_1, k_1, \ell_2, k_2, \dots, \ell_{m-1}, k_{m-1}, \ell_m)$$

is an *iterative representation* of the circular word  $w_\circ = (u_{\frac{\ell_m}{\ell_{m-1}}})_\circ$  over the two letter alphabet  $\{a, b\}$ , where  $u_0 = u$ ,  $u_1 = \sigma^{k_1}(u_0^{\frac{\ell_1}{|u_0|}})$  and  $u_i = \sigma^{k_i}(u_{i-1}^{\frac{\ell_i}{\ell_{i-1}}})$  for all  $i = 2, \dots, m-1$ .

**Example 1.** Consider the circular word  $w_\circ = (bababaabbabaab)_\circ$ . One of its iterative representations is

$$(baa, 4, 0, 6, 4, 14).$$

By using the previous definition of the words  $u_i$ , the following words are obtained during the reconstruction of the circular word:  $u_0 = baa$ ,  $u_1 = baab$ ,  $u_2 = babaab$  and finally,  $w_\circ = (babaabbabaabba)_\circ$ . Note, that no shifting is required in the last step, because  $w_\circ = v_\circ$  for all  $v \in w_\circ$ .

Of course, every circular word has an iterative representation of the form above that can be constructed with the greedy algorithm in Figure 2. Moreover, the algorithm halts if only if it has found a minimal root.

Note, that by using this algorithm, we can process the iterative representation in Example 1 further to obtain  $(ab, 3, 1, 4, 0, 6, 4, 14)$ . In fact, the following can be stated about the iterative representations of circular words over the two letter alphabet  $\{a, b\}$ .

**Theorem 2.** *Let  $w \in \{a, b\}^*$ . If  $(u, \ell_1, k_1, \dots, \ell_{m-1}, k_{m-1}, |w|)$  is a minimal iterative representation of  $w_\circ$ , then  $|u| \leq 2$ .*

*Proof.* It follows from the fact that every word  $u \in \{a, b\}$ ,  $|u| \geq 3$  has a conjugate that has a border of length at least one, thus in this case  $u_\circ$  has a representation  $(v, |u|)$  such that  $|v| < |u|$ .  $\square$

Let  $(u, \ell_1, k_1, \dots, \ell_{m-1}, k_{m-1}, |w|)$  be an iterative representation of  $w_\circ$ . It is *optimal* if for all iterative representations  $(u', \ell'_1, k'_1, \dots, \ell'_{m'-1}, k'_{m'-1}, |w|)$  of  $w_\circ$ ,  $|u| \leq |u'|$  and if  $|u| = |u'|$ , then  $m \leq m'$ . In other words an optimal iterative representation of  $w_\circ$  is one with the shortest possible minimal root, such that

```

construct_iterative_representation( $w_\circ$ )
1.  $u \leftarrow w$ 
2.  $v \leftarrow$  find  $v$  such that  $(v, |w|)$  is a minimal representation of  $w_\circ$ 
3.  $rep \leftarrow [|w|]$  #  $rep$  is a vector of integers
4. while true do
5.    $u \leftarrow v$ 
6.    $v \leftarrow$  find  $v$  such that  $(v, |u|)$  is a minimal representation of  $u_\circ$ 
7.   if  $|u| = |v|$  then # if we have found a minimal root,
8.     break # then the algorithm breaks the loop
9.   endif
10.   $k \leftarrow$  find  $k$  such that  $\sigma^{-k}(u)$  has root  $v$ 
11.   $rep \leftarrow |u| : k : rep$  # append  $|u|$  and  $k$  to  $rep$  from the left
12. endwhile
13. return  $v : rep$ 

```

Figure 2: Algorithm for constructing the iterative representation of  $w_\circ$ .

$w_\circ$  can be reconstructed from it with the least amount of fractional power operations (regardless of the amount of shift operations required).

The algorithm may not provide an optimal solution for all inputs  $w_\circ$ . For example, consider the circular word  $(ababaa)_\circ$ . The algorithm would construct the iterative representation  $(ab, 3, 0, 4, 0, 6)$ , while an optimal solution would be  $(ab, 5, 0, 6)$ . One of the directions of future research is to look for an efficient algorithm that always finds an optimal iterative representation of any circular word  $w_\circ$  (see Section 5).

Note, that we do not have to restrict ourselves to representations of circular words. If we are looking for a linear word, another shift operation has to be applied at the end of the reconstruction.

Let us now turn to another method of representation, which is not intended as an encoding, nor as a compression, but a way of representing the structure of different conjugates of a word and their relation to each-other (e.g., common prefixes).

## 4 Representations with trees

The tree  $\tau$  is the tree of the circular word  $w_\circ$  if and only if for any word  $v = v_1 \dots v_n$  in  $w_\circ$ , there exists a path in  $\tau$  between the root and a leaf node with a series of edges labeled  $v_1, \dots, v_n$ .

This approach is related to tries that are data structures representing associative structures. They are often used to search for suffixes or other factors of words. Quite similarly, our trees represent a set of words that are conjugates of each-other. For more information on the use of tries consult [1].

We remark, that in our figures the letters appear as nodes, but they are to be considered as labels of edges between two (unnamed) nodes. This way, the represented words can be seen more clearly. First,

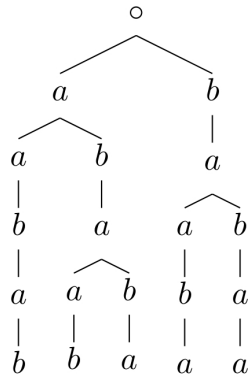


Figure 3: Tree representation of  $(abaab)_{\circ}$ .

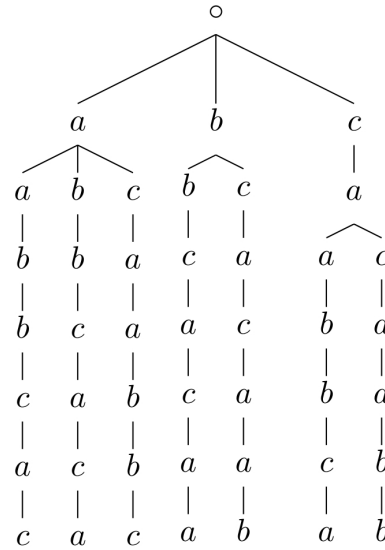


Figure 4: Tree representation of  $(aabbcac)_{\circ}$ .

consider the circular word

$$(abaab)_{\circ} = \{abaab, baaba, aabab, ababa, babaa\}.$$

Its tree representation is shown in Figure 3.

Now, see Figure 4 for the tree of the circular word  $(aabbcac)_{\circ}$  (over the three letter alphabet  $\{a, b, c\}$ ) which is the set

$$(aabbcac)_{\circ} = \{aabbcac, abbcaca, bbcacaa, bcacaab, cacaaab, acaabbc, caabbca\}.$$

Clearly, both trees represent finite-state automata with partially defined, deterministic transition functions. We can distinguish different levels of a tree. Vertex  $\circ$  is on level zero ( $\ell(\circ) = 0$ ) and if there is an edge  $u \rightarrow v$ , then  $\ell(v) = \ell(u) + 1$ .

We can see some branching nodes in both trees. The tree in Figure 4 has two branching nodes on level one while no two branching nodes of the tree in Figure 3 are on the same level.

Examining branching nodes is useful for analyzing trees of circular words and the words themselves. Suppose that tree  $\tau$  has  $u_1, \dots, u_k$  branching nodes such that  $\circ \xrightarrow{a} u_1$  and  $u_i \xrightarrow{a} u_{i+1}$  for all  $i = 1, \dots, k - 1$ . Then there is a letter  $b$  such that  $a^k, a^{k-1}b$ , and thus  $a^{k-2}b, \dots, ab, b$  are all factors of  $w_{\circ}$ . If the level of the leaf nodes is  $k + 1$ , then the represented circular word must be  $(a^k b)_{\circ}$ . Similarly, if there are branching nodes  $u_1, \dots, u_m$  and  $v_1, \dots, v_k$  such that  $\circ \xrightarrow{a} u_1 \xrightarrow{a} \dots \xrightarrow{a} u_m$  and  $\circ \xrightarrow{b} v_1 \xrightarrow{b} \dots \xrightarrow{b} v_k$ , and the level of the leaf nodes is  $m + k$ , then the tree can only represent the circular word  $(a^m b^k)_{\circ}$ . Apart from these simple cases, we can state the following about the relation of circular words and branching nodes in their trees: Let  $w_{\circ}$  be a circular word with tree  $\tau$ . There is a branching node in  $\tau$  on level  $\ell$  if and only if there are two distinct words  $w', w'' \in w_{\circ}$ , such that the longest common prefix of  $w'$  and  $w''$  is a word of length  $\ell$ . Moreover, if there is a branching node in the tree on level  $n > 0$ , then there is a branching node on level  $n - 1$ . These nodes do not necessarily lie on the same path. To verify this, assume that tree  $\tau$  contains the edges  $u \xrightarrow{a} v$  and  $u \xrightarrow{b} s$ , where  $u \neq \circ$ . Then there are words  $xay, xbz \in w_{\circ}$  such that  $x, y, z \in \Sigma^*$  with  $|x| > 0$ , and  $a, b \in \Sigma$ , where  $\Sigma$  is an alphabet of at least two letters. Write  $x = x_1, \dots, x_m$ . Clearly, both

$x_2 \dots x_m a y x_1$  and  $x_2 \dots x_m b z x_1$  are in  $w_\circ$ , having a common prefix of length  $|x| - 1$ . Thus there must be a node  $u'$  such that the path from  $\circ$  to  $u'$  reads  $x_2 \dots x_m$  and two nodes  $v'$  and  $s'$ , such that  $u' \rightarrow^a v'$  and  $u' \rightarrow^b s'$ .

**Proposition 1.** Consider a circular word  $w_\circ \in \{a, b\}$  with tree  $\tau$ . If  $\tau$  has a branching node on level  $|w| - 2$ , then there is exactly one branching node on all levels  $m = 0, \dots, |w| - 2$  of  $\tau$ .

*Proof.* From the previous argument, it follows that all levels  $k < |w| - 2$  of the tree has at least one branching node. Clearly, the depth of the tree is  $|w|$ . Since the root node is branching, the number of possible paths (words) up to level one is two. Moreover, if level  $k > 0$  has  $m_k \in \mathbb{N}$  branching nodes, then the number of all possible paths up to level  $k + 1$  is equal to the number of all possible paths up to level  $k$ , plus  $m_k$ . Then we get that the number of possible paths on the level of the leaf nodes is  $2 + m_1 + \dots + m_{|w|-1} + m_{|w|} = |w|$ . We have stated,  $m_i > 0$  for all  $i = 1, \dots, |w| - 2$ , thus  $m_{|w|-1} = m_{|w|} = 0$  and  $2 + m_1 + \dots + m_{|w|-2} = |w|$ . If  $m_i > 1$  for any  $i \geq 1$ , then  $m_j = 0$  for some  $j \neq i$ . This is impossible, since all levels under  $|w| - 2$  have at least one branching node, thus  $m_i = 1$  for all  $i = 1, \dots, |w| - 2$ .  $\square$

Now, let us analyze an interesting class of words. Let  $f_1 = b$ ,  $f_2 = a$  and define  $f_n = f_{n-1}f_{n-2}$  for all  $n \geq 3$ . We call  $f_n$  (where  $n \geq 1$ ) the *n*th finite Fibonacci word. The infinite Fibonacci word is the limit of the sequence  $f_1, f_2, \dots$

The following lemma describes a well known property of the infinite Fibonacci words.

**Lemma 2** (see Séébold [16]). If a word  $u^2$  is a factor of the infinite Fibonacci word, then  $u$  is a conjugate of some finite Fibonacci word.  $\square$

Note that the tree in Figure 3 represents the circular word obtained from  $f_5$  which is the fifth Fibonacci word. See the trees of  $(f_6)_\circ$  and  $(f_7)_\circ$  in Figure 5. One can observe that the structure of these trees are very similar. This is strongly related to the definition of Fibonacci words.

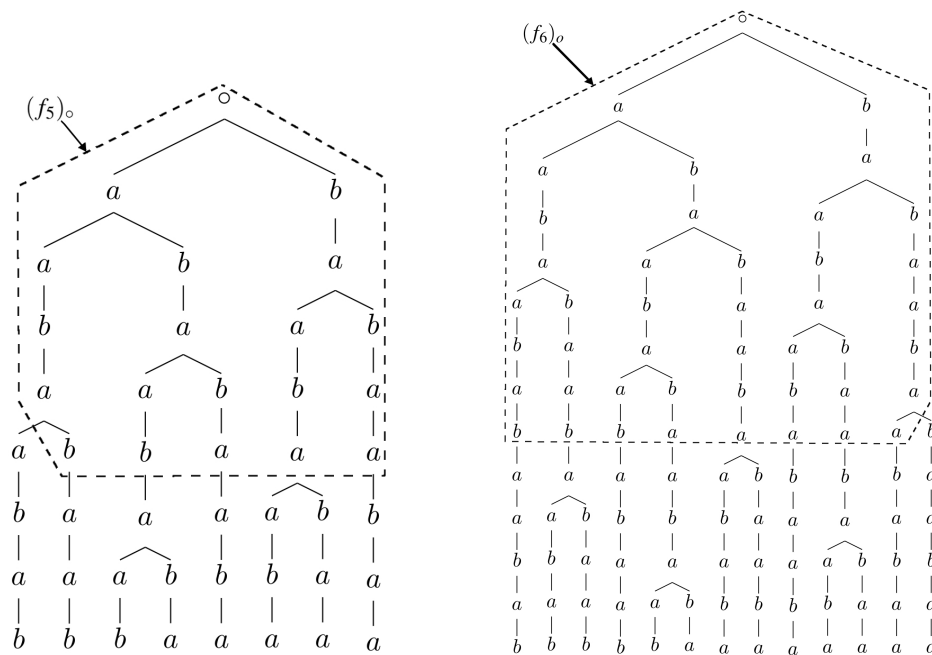


Figure 5: Trees of  $(f_6)_\circ$  and  $(f_7)_\circ$ .

**Theorem 3.** *Let us denote the tree of the finite Fibonacci word  $f_i$  by  $\varphi_i$  for all  $i \in \mathbb{N}$ . Then for all  $i \in \mathbb{N}$ , the tree  $\varphi_i$  has exactly one branching node on all of its levels, except for the last two.*

*Proof.* Consider the tree  $\varphi_i$  of the circular Fibonacci word  $(f_i)_\circ$  and let  $\ell \in \{0, \dots, |f_i|\}$ . The paths from  $\circ$  to nodes on level  $k$  represent the length  $k$  factors of  $(f_i)_\circ$ . By the properties of Fibonacci words (or Sturmian words), we know that the number of distinct factors of length  $k$  in the infinite Fibonacci word is  $k + 1$ . Since all of the length  $k$  words of the tree appear in the infinite Fibonacci word (because it has factor  $f_i^2$ ), their number must not be more than  $k + 1$ . On the other hand, each tree of a primitive word of length  $n$  must contain  $n$  branching nodes. Thus in  $\varphi_i$  all branching nodes must be on different levels.  $\square$

Based on the proof, we can state the following about the trees of circular Fibonacci words.

**Corollary 1.** *For all  $i, j \in \mathbb{N} \setminus \{0\}$ , if  $j > i$ , then  $\varphi_i$  is a subtree of  $\varphi_j$ .*

Thus the trees of Fibonacci words are not only very similar, but they contain recurring subtrees. Notice in Figure 5, that the tree of  $(f_5)_\circ$  appears in the tree of  $(f_6)_\circ$  which also appears in the tree of  $(f_7)_\circ$ , marked by the dashed lines. Thus we can define the tree  $\varphi$  which belongs to the limit of the sequence of Fibonacci words, that is, the infinite Fibonacci word. Each path in the tree  $\varphi$  defines an infinite suffix of the infinite Fibonacci word. This is a consequence of the structure of the trees  $\varphi_i$  ( $i = 1, 2, \dots$ ), since all of their words are factors of the infinite Fibonacci word and an infinite factor must be a suffix.

Let us state another interesting fact about branching nodes of trees of circular Fibonacci words.

**Theorem 4.** *Consider the tree  $\varphi_i$  for any  $i \in \mathbb{N}$ . Let  $u$  and  $u'$  be branching nodes of  $\varphi_i$  such that they lie on the same path and there are no other branching nodes between them. Then  $|\ell(u) - \ell(u')|$  is a Fibonacci number.*

*Proof.* Assume the contrary, that is, there is a Fibonacci word  $f_i$  such that there are two branching nodes  $u, u'$  in tree  $\varphi_i$  that lie on the same path and do not have any other branching nodes between them, but  $|\ell(u) - \ell(u')|$  is not a Fibonacci number. Then, there exists a Fibonacci word  $f_j$  with  $j \geq i$  such that  $(f_j)_\circ$  has square factor  $vv$  where  $v$  is the word constructed from the labels on the path between  $u$  and  $u'$ . Moreover, this will be true for all Fibonacci words  $f_{j'}$  where  $j' \geq j$ . Thus the infinite Fibonacci word must contain the square factor  $vv$ . This contradicts Lemma 2, since  $v$  cannot be a conjugate of any Fibonacci word because its length is not a Fibonacci number. Thus our indirect assumption is false.  $\square$

## 5 Conclusion and future directions

Combinatorics on circular words is a field that still has countless open problems and many possible research directions. We have shown some non-traditional methods of considering (representing) circular words. The following questions are still open and may lead to a better characterization of these sequences.

1. The algorithm presented in Section 3 does not always provide optimal solutions. Is there a way of deciding how to choose the best sequence of roots in the algorithm?
2. Theorem 2 is about the minimal roots of words over the two letter alphabet. What can we say about words over alphabets of more than two letters?
3. One could use the tree  $\varphi$  to deduce some properties of the infinite Fibonacci word.



4. Or the tree representations can be utilized to prove results about the structure of other (possibly infinite) words.
5. We believe, that Theorem 3 is true for all standard sturmian words (see e.g., [11] for their definition).

## Acknowledgements

The authors would like to thank the reviewers for their valuable and useful comments. The work is supported by the TÁMOP 4.2.2/C-11/1/KONV-2012-0001 and 4.2.2/B-10/1-2010-0024 projects. The projects are implemented through the New Hungary Development Plan, co-financed by the European Social Fund and the European Regional Development Fund.

## References

- [1] Maxime Crochemore & Wojciech Rytter (2002): *Jewels of Stringology*. World Scientific Publishing Company, Incorporated, doi:10.1142/4838.
- [2] James D. Currie & D. Sean Fitzpatrick (2002): *Circular words avoiding patterns*. *Proceedings of the 6th International Conference on Developments in Language Theory. LNCS 2450.*, pp. 319–325, doi:10.1007/3-540-45005-X\_28.
- [3] Volker Diekert, Tero Harju & Dirk Nowotka (2006): *Factorizations of cyclic words*. *Workshop on Words and Automata at CSR 7*.
- [4] Szilárd Zsolt Fazekas & Benedek Nagy (2008): *Scattered Subword Complexity of Non-primitive Words*. *J. Autom. Lang. Comb.* 13(3), pp. 233–247.
- [5] Nathan J. Fine & Herbert S. Wilf (1965): *Uniqueness theorems for periodic functions*. *Proceedings of the American Mathematical Society* 16, pp. 109–114, doi:10.1090/S0002-9939-1965-0174934-9.
- [6] D. Sean Fitzpatrick (2005): *There are binary cube-free circular words of length  $n$  contained within the Thue-Morse word for all positive integers  $n$* . *Ars Combinatorica* 74.
- [7] László Hegedüs & Benedek Nagy (2013): *Periodicity of circular words*. *Local Proceedings of WORDS 2013, TUCS Lecture Notes* 20, pp. 45–56.
- [8] M. Lothaire (1983): *Combinatorics on words*. Addison-Wesley.
- [9] M. Lothaire (2002): *Algebraic Combinatorics on Words*. *Encyclopedia of Mathematics and its Applications* 90, Cambridge University Press, doi:10.1017/CBO9781107326019.
- [10] M. Lothaire (2005): *Applied Combinatorics on Words*. *Encyclopedia of Mathematics and its Applications* 105, Cambridge University Press, doi:10.1017/CBO9781107341005.
- [11] Aldo de Luca & Filippo Mignosi (1994): *Some combinatorial properties of Sturmian words*. *Theoretical Computer Science* 136(2), pp. 361–385, doi:10.1016/0304-3975(94)00035-H.
- [12] Roger C. Lyndon & Marcel-Paul Schützenberger (1962): *The equation  $a^M = b^N c^P$  in a free group*. *Michigan Math. J.* 9(4), pp. 289–298, doi:10.1307/mmj/1028998766.
- [13] Dirk Nowotka (2004): *Periodicity and unbordered factors of words*. TUCS Dissertations No. 50.
- [14] Benoît Rittaud & Laurent Vivier (2011): *Circular words and applications*. *Proceedings of Words 2011, Electronic Proceedings in Theoretical Computer Science* 63, pp. 31–36, doi:10.4204/EPTCS.63.6.
- [15] Benoît Rittaud & Laurent Vivier (2012): *Circular words and three applications: factors of the Fibonacci word,  $\mathcal{F}$ -adic numbers, and the sequence 1, 5, 16, 45, 121, 320, ...* *Funct. Approx. Comment. Math.* 47(2), pp. 207–231, doi:10.7169/facm/2012.47.2.6.

- [16] Patrice Séébold (1985): *Propriétés combinatoires des mots infinis engendrés par certains morphismes*. Thèse de doctorat, Université P. et M. Curie, Institut de Programmation.
- [17] Arseny M. Shur (2010): *On ternary square-free circular words*. *The Electronic Journal of Combinatorics* 17.
- [18] William Smyth (2003): *Computing patterns in strings*. Addison-Wesley.
- [19] Axel Thue (1906): *Über unendliche Zeichenreihen*. *Kra. Vidensk. Selsk. Skrifter, I. Mat. Nat. Kl. 7*, pp. 1–22.
- [20] Axel Thue (1912): *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*. *Kra. Vidensk. Selsk. Skrifter, I. Mat. Nat. Kl. 46*, pp. 1–67.