Counting Ordered Patterns in Words Generated by Morphisms - I. Rises, Descents, and Repetitions with Gaps

S. Kitaev, T. Mansour, Patrice Séébold

Counting ordered patterns in words generated by morphisms I. Rises, descents, and repetitions with gaps

Sergey Kitaev
Department of Mathematics
University of Kentucky
Lexington, KY 40506-0027, USA
e-mail: kitaev@ms.uky.edu

Toufik Mansour
Department of Mathematics
Haifa University
31905 Haifa, ISRAEL
e-mail: toufik@math2.haifa.ac.il

Patrice Séébold*
LIRMM, UMR 5506 CNRS
161 rue Ada
34392 Montpellier, FRANCE
e-mail: Patrice.Seebold@lirmm.fr

November 3, 2004

Abstract

We start a general study of counting the number of occurrences of ordered patterns in words generated by morphisms. The case of rises and descents is considered. We give recurrence formulas in the general case. We deduce exact formulas for a particular family of morphisms.

Keywords: Ordered patterns, rises, descents, repetitions, morphisms.

*Département Mathématiques Informatique et Applications, Université Paul Valéry, Route de Mende, 34199 Montpellier Cédex 5, France

1 Introduction

Different notions of *pattern* can be encountered in several domains of combinatorics.

In algebraic combinatorics, an occurrence of a pattern p in a permutation π is a subsequence of π (of the same length as the length of p) whose elements are in the same relative order as those in p. For example, the permutation $\pi = 536241$ contains an occurrence of the pattern $p = 2431$: indeed the elements of the subsequence 3641 of π are in the same relative order as those in p. Examples of results concern permutations avoiding a pattern of length 3 in the symmetric group S_n (see [18, 27]).

Motivated by the study of Mahonian statistics, Babson and Steingrímsson introduced a generalisation where two adjacent elements of the pattern must also be adjacent in the permutation [4]. In Claeson, 2001 [11] this generalisation provides interesting results related to set partitions, Dyck paths, Motzkin paths, or involutions.

In combinatorics on words, an occurrence of a pattern p in a word u is a factor of u having the same shape as p, i.e., for which there exists a nonerasing morphism transforming p in this factor. For example the word $u = abaabaaabab$ contains an occurrence of the pattern $p = \alpha\alpha\beta\alpha\alpha\beta$: indeed the morphism $f(\alpha) = a$, $f(\beta) = ba$ transforms the pattern p in $aabaaaba$ which is a factor of u. The main question is to determine whether or not a given pattern is unavoidable, that is if it is possible to construct an infinite word containing no occurrence of the pattern. The interested reader should refer to Chapter 3 of Lothaire, 2002 [20].

In Burstein, 1998 [7], and Burstein and Mansour, 2002, 2003 [8, 9, 10] the authors realised a "mixing" of these two notions. They consider ordered alphabets. Here, an occurrence of a pattern in a word is a factor or a subsequence having the same shape, and in which the relative order of the letters is the same as in the pattern. For example, on the alphabet $\{a, b\}$ with $a < b$, the word $u = abaaabab$ contains an occurrence of the pattern 2111 (the factor $baaa$) but not of the pattern 1222 ($abbb$ is not a factor of u). However, the word u contains an occurrence of the pattern with gaps 1#2#2#2 because $abbb$ is a subsequence of u (here # means that the letters corresponding to 1 and 2 may be not consecutive). To avoid confusion with previous notions we call these patterns *ordered patterns* (with gaps if there is at least one #, without gaps if there is no #).

In Kitaev, Mansour and Séébold, 2004 [17] we computed the number of occurrences of a lot of ordered patterns in the Peano words (words corresponding to finite approximations of the Peano space filling curve). An interesting property of these words is that they are generated by a tag-system, i.e., by applying two morphisms. A motivation for this choice is the interest in studying classes of words defined by iterative schemes, in particular with morphisms that are a fundamental tool of formal languages [2, 20, 24].

In the present paper we start a general study of counting the number of occurrences of ordered patterns in words generated by morphisms. After some preliminaries (Section 2), we give in Section 3 some general results (recurrence formulas) on counting elementary ordered patterns with gaps in words generated by morphisms, and applications to two well known binary words. Section 4 is dedicated to more precise results (exact formulas) in the case of a particular family of morphisms, and in Section 5 we give many examples of morphisms belonging to this family. Further investigations regarding ordered patterns without gaps are presented in section 6.

2 Preliminaries

2.1 Definitions and notations

The terminology and notations are mainly those of Lothaire, 2002 [20].

Let A be a finite set called *alphabet* and A^* the free monoid generated by A. The elements of A are called *letters* and those of A^* are called *words*. The *empty word* ε is the neutral element of A^* for the concatenation of words (the *concatenation* of two words u and v is the word uv), and we denote by A^+ the semigroup $A^* \setminus \{\varepsilon\}$.

The *length* of a word u, denoted by $|u|$, is the number of occurrences of letters in u. In particular $|\varepsilon| = 0$. If n is a nonnegative integer, u^n is the word obtained by concatenating n occurrences of the word u. Of course, $|u^n| = n \cdot |u|$. The cases $n = 2$, $n = 3$, and $n = 4$ deserve a particular attention in what follows. A word u^2 (resp. u^3, u^4), with $u \neq \varepsilon$, is called a *square* (resp. a *cube*, a 4-*power*).

A word w is called a *factor* (resp. a *prefix*, resp. a *suffix*) of u if there exist words x, y such that $u = xwy$ (resp. $u = wy$, resp. $u = xw$). The factor (resp. the prefix, resp. the suffix) is *proper* if $xy \neq \varepsilon$ (resp. $y \neq \varepsilon$, resp. $x \neq \varepsilon$). The number of distinct occurrences of w in u is denoted by $|u|_w$. A word u is a *subsequence* of the word v if there exist words $u_1, \ldots, u_n, v_1, \ldots, v_n, v_{n+1}$ such that $u = u_1 \cdots u_n$ and $v = v_1 u_1 v_2 u_2 \cdots v_n u_n v_{n+1}$.

An *infinite word* (or *sequence*) over A is an application $\mathbf{a} : \mathbb{N} \to A$. It is written $\mathbf{a} = a_0 a_1 \cdots a_i \cdots, i \in \mathbb{N}, a_i \in A$.

The notion of factor is extended to infinite words as follows: a (finite) word u is a *factor* (resp. *prefix*) of an infinite word $\mathbf{a}$ over A if there exist $n \in \mathbb{N}$ (resp. $n = 0$) and $m \in \mathbb{N}$ ($m = |u|$) such that $u = a_n \cdots a_{n+m-1}$ (by convention $a_n \cdots a_{n-1} = \varepsilon$).

In what follows, we will consider morphisms on A. Let B be an alphabet (often, $B = A$).

A *morphism on A* (in short *morphism*) is an application $f : A^* \to B^*$ such that $f(uv) = f(u)f(v)$ for all $u, v \in A^*$. It is uniquely determined by its value on the alphabet A. A morphism f on A is a *literal morphism* if $|f(a)| = 1$ for all $a \in A$.

Now, $A = B$. A morphism is *nonerasing* if $f(a) \neq \varepsilon$ for all $a \in A$. It is *prolongable on* x_0, $x_0 \in A^+$, if there exists $u \in A^+$ such that $f(x_0) = x_0 u$. In this case, for all $n \in \mathbb{N}$ the word $f^n(x_0)$ is a proper prefix of the word $f^{n+1}(x_0)$ and this defines a unique infinite word

$$\mathbf{x} = x_0 u f(u) f^2(u) \cdots f^n(u) \cdots$$

which is the limit of the sequence $(f^n(x_0))_{n \geq 0}$. We write $\mathbf{x} = f^\omega(x_0)$ and say that $\mathbf{x}$ is *generated* by f.

A (finite or infinite) word u over A is *square-free* (resp. *cube-free*, *4-power-free*) if none of its factors is a square (resp. a cube, a 4-power). A morphism f on A is *square-free* if the word $f(u)$ is square-free whenever u is a square-free word. The morphism f is *weakly square-free* if f generates a square-free infinite word.

A *D0L-system* is a triple $G = (A, f, u)$ where A is an alphabet, f a morphism on A and $u \in A^*$. An infinite word $\mathbf{x}$ is generated by G if $\mathbf{x} = (f^k)^\omega(u)$ for some $k \in \mathbb{N}$.

A *tag-system* is a quintuple $T = (A, u, f, g, B)$ where A and B are alphabets, $u \in A^+$, f is a nonerasing morphism on A, prolongable on u, and g is a morphism from A onto B. An infinite word $\mathbf{y}$ is generated by G if $\mathbf{y} = g((f^k)^\omega(u))$ for some $k \in \mathbb{N}$.

Remark that what we call here a tag-system is sometimes called a *HD0L-system*. The terminology of tag-system comes from the fundamental study of Cobham [12]. Chapter 5 of [23] is dedicated to a deep study of D0L-systems.

2.2 Ordered patterns

Let A be a totally ordered alphabet and let $\aleph$ be the ordered alphabet whose letters are the first n positive integers in the usual order (thus $\aleph = \{1, 2, \ldots, n\}$).

An *ordered pattern* is any word[1] over $\aleph \cup \{\#\}$ where $\# \notin \aleph$. If the pattern contains at least one # it is an *ordered pattern with gaps*; otherwise it is an *ordered pattern without gaps*[2].

A word v over A *contains an occurrence of the ordered pattern* u (or, equivalently the ordered pattern u *occurs in* v) if, for some integer $n \in \mathbb{N}$, $u = u_1 \# u_2 \# \cdots \# u_n$ ($u_i \in \aleph^*$), $v = w_0 v_1 w_1 v_2 w_2 \cdots w_{n-1} v_n w_n$ and there exists a literal morphism $f : \aleph^* \to A^*$ such that $f(u_i) = v_i$, $1 \leq i \leq n$, and if $x, y \in \aleph$, $x < y \Rightarrow f(x) < f(y)$. This means that the word v contains an occurrence of the ordered pattern u if v contains a subsequence v' which is equal to $f(u')$ where u' is obtained from u by deleting all the occurrences of #, with the additional condition that two adjacent letters in u must be adjacent in v. The number of different occurrences of u as an ordered pattern in v is denoted by $|v|_u$.

Example. Let $A = \{a, b, c, d, e, f\}$ with $a < b < c < d < e < f$.

The word $v = eafdbc$ contains one occurrence of the ordered pattern 2#31, namely the subsequence efd ($|e\,a\,f\,d\,b\,c|_{2\#31} = 1$). In v, the ordered pattern 2#3#1 occurs in three occurrences: efd, efb, and efc ($|e\,a\,f\,d\,b\,c|_{2\#3\#1} = 3$); the ordered pattern 231 does not occur in v ($|e\,a\,f\,d\,b\,c|_{231} = 0$).

Of course, since # can correspond to anything, the ordered patterns u, $\#u$, $u\#$, and $\#u\#$ are equal. In particular, if x is a word over $\aleph$, we will write $(x\#)^\ell$ or $(\#x)^\ell$ to represent the ordered pattern $x\#x\#\cdots\#x$ containing l occurrences of the word x.

[1] In algebraic combinatorics when defining a pattern it is claimed that each letter from the interval $[k]$ must occur at least once. This requirement is not useful here, what is important is the relative value of each letter because this gives the order. However it will be often implicite that these letters (which are only formal representations of the pattern) are taken in the order from 1.

[2] Our choice here is to use terminology of combinatorics on words. For example our notion of *pattern without gaps* is often refered to as *pattern without internal dashes* in the literature about algebraic combinatorics (see, e.g., Kitaev, 2003 [16]). However this terminology does not seem to be solid since Burstein and Mansour used *subword pattern without hyphens* [10], and *segmented pattern* is also encountered

3 Ordered patterns with gaps and morphisms

Let k be an integer ($k \geq 2$) and A the k-letter ordered alphabet $A = \{a_1 < a_2 < \cdots < a_k\}$.
Let f be any morphism on A: for $1 \leq i \leq k$, $f(a_i) = a_{i_1} \ldots a_{i_{p_i}}$ with $p_i \geq 0$ ($p_i = 0$ if and only if $f(a_i) = \varepsilon$).

3.1 The incidence matrix of f^n

Let n be a non-negative integer. The *incidence matrix* of f^n is the $k \times k$ matrix

$$M(f^n) = (m_{n,i,j})_{1 \leq i,j \leq k}$$

where $m_{n,i,j}$ is the number of occurrences of the letter a_i in the word $f^n(a_j)$, i.e., $m_{n,i,j} = |f^n(a_j)|_{a_i}$. For details on the incidence matrix of a morphism see, e.g., [2], chapter 8, in which is given the following.

Property 1 *For every $n \in \mathbb{N}$, $M(f)^n = M(f^n)$.*

3.2 Rises, descents, and repetitions with gaps of f^n

In what follows we are interested in some particular forms of ordered patterns. The *rises or descents with gaps* (ordered patterns 1#2 or 2#1), the *repetitions with gaps of one letter* (ordered patterns $(1\#)^p$, $p \geq 1$).

3.2.1 Rises and descents with gaps

Let n be a non-negative integer.

The *vector of rises with gaps* of f^n is the k vector whose i-th entry is the number of occurrences of the ordered pattern 1#2 in the word $f^n(a_i)$, i.e.,

$$RG(f^n) = (|f^n(a_i)|_{1\#2})_{1 \leq i \leq k}.$$

The *vector of descents with gaps* of f^n is the k vector whose i-th entry is the number of occurrences of the ordered pattern 2#1 in the word $f^n(a_i)$, i.e.,

$$DG(f^n) = (|f^n(a_i)|_{2\#1})_{1 \leq i \leq k}.$$

Our goal is to obtain recurrence formulas giving the entries of $RG(f^{n+1})$ and $DG(f^{n+1})$. Since $f^{n+1} = f^n \circ f = f \circ f^n$, we have two different ways to compute $RG(f^{n+1})$ and $DG(f^{n+1})$.

Let ℓ be an integer, $1 \leq \ell \leq k$. Either $|f^{n+1}(a_\ell)|_{1\#2}$ (resp. $|f^{n+1}(a_\ell)|_{2\#1}$) will be obtained from the value of $f(a_\ell)$ and the entries of $RG(f^n)$ (resp. $DG(f^n)$) (see 1. below), or it will be computed from the values of $RG(f)$ (resp. $DG(f)$) and $f^n(a_\ell)$ (see 2. below).

1. From $f^{n+1} = f^n \circ f$.

Since $f(a_\ell) = a_{\ell_1} \ldots a_{\ell_{p_\ell}}$, the number of occurrences of the ordered pattern 1#2 in $f^{n+1}(a_\ell) = f^n(f(a_\ell)) = f^n(a_{\ell_1} \ldots a_{\ell_{p_\ell}})$ is obtained by adding two values:

- the number of occurrences of the ordered pattern 1#2 in each $f^n(a_{\ell_i})$, $1 \leq i \leq p_\ell$. Since the ℓ-th column of the incidence matrix of f indicates which letters appear in $f(a_\ell)$ (and how many times), this number is obtained by multiplying the vector $RG(f^n)$ by the ℓ-th column of $M(f)$, i.e., it is equal to $\sum_{t=1}^{k} |f^n(a_t)|_{1\#2} \cdot m_{1,t,\ell}$,
- the number of occurrences of the ordered pattern 1#2 in each of the $f^n(a_{\ell_i} a_{\ell_j})$, $1 \leq i < j \leq p_\ell$, where the letter corresponding to 1 is in $f^n(a_{\ell_i})$ and the letter corresponding to 2 is in $f^n(a_{\ell_j})$. In what follows we will call such an occurrence of 1#2 in $f^n(a_{\ell_i} a_{\ell_j})$ an *external* occurrence of the ordered pattern 1#2 in $f^n(a_{\ell_i} a_{\ell_j})$, and denote it $|f^n(a_{\ell_i} a_{\ell_j})|^{ext}_{1\#2}$.
The value of $|f^n(a_{\ell_i} a_{\ell_j})|^{ext}_{1\#2}$ is obtained by adding, for all the integers r, $1 \leq r \leq k-1$, the product of the number of occurrences of the letter a_r in $f^n(a_{\ell_i})$ ($|f^n(a_{\ell_i})|_{a_r}$) by the number of occurrences of all the letters of $f^n(a_{\ell_j})$ greater than a_r ($|f^n(a_{\ell_j})|_{a_s}$, $r+1 \leq s \leq k$). This gives $\sum_{r=1}^{k-1}(m_{n,r,\ell_i} \cdot \sum_{s=r+1}^{k} m_{n,s,\ell_j})$.
The number of external occurrences of 1#2 in all the $f^n(a_{\ell_i} a_{\ell_j})$, $1 \leq i < j \leq p_\ell$, is thus given by $\sum_{1 \leq i < j \leq p_\ell} |f^n(a_{\ell_i} a_{\ell_j})|^{ext}_{1\#2} = \sum_{1 \leq i < j \leq p_\ell}(\sum_{r=1}^{k-1}(m_{n,r,\ell_i} \cdot \sum_{s=r+1}^{k} m_{n,s,\ell_j}))$.

2. From $f^{n+1} = f \circ f^n$.

Let $q_\ell = |f^n(a_\ell)|$: $f^{n+1}(a_\ell) = f(f^n(a_\ell)) = f(a_{\ell'_1} \dots a_{\ell'_{q_\ell}})$. Here the number of occurrences of the ordered pattern 1#2 in $f^{n+1}(a_\ell)$ is obtained by adding

- the number of occurrences of the ordered pattern 1#2 in each $f(a_{\ell'_i})$, $1 \leq i \leq q_\ell$. Since the ℓ-th column of the incidence matrix of f^n indicates which letters appear in $f^n(a_\ell)$ (and how many times), this number is obtained by multiplying the vector $RG(f)$ by the ℓ-th column of $M(f^n)$, i.e., it is equal to $\sum_{t=1}^{k} |f(a_t)|_{1\#2} \cdot m_{n,t,\ell}$,
- the number of external occurrences of the ordered pattern 1#2 in each of the $f(a_{\ell'_i} a_{\ell'_j})$, $1 \leq i < j \leq q_\ell$. This number is obtained by adding, for all the integers r, $1 \leq r \leq k-1$, the product of the number of occurrences of the letter a_r in $f(a_{\ell'_i})$ ($|f(a_{\ell'_i})|_{a_r}$) by the number of occurrences of all the letters of $f(a_{\ell'_j})$ greater than a_r ($|f(a_{\ell'_j})|_{a_s}$, $r+1 \leq s \leq k$). This gives $\sum_{r=1}^{k-1}(m_{1,r,\ell'_i} \cdot \sum_{s=r+1}^{k} m_{1,s,\ell'_j})$.
  The number of external occurrences of 1#2 in all the $f(a_{\ell'_i} a_{\ell'_j})$, $1 \leq i < j \leq q_\ell$, is thus given by $\sum_{1\leq i<j\leq q_\ell} |f(a_{\ell'_i} a_{\ell'_j})|^{ext}_{1\#2} = \sum_{1\leq i<j\leq q_\ell}(\sum_{r=1}^{k-1}(m_{1,r,\ell'_i} \cdot \sum_{s=r+1}^{k} m_{1,s,\ell'_j}))$.

The same reasoning applies for calculating the entries of $DG(f^{n+1})$, replacing 1#2 by 2#1 and "greater" by "smaller".

Thus we have the following.

Proposition 1 *For each letter $a_\ell \in A$, let p_ℓ and q_ℓ be such that $f(a_\ell) = a_{\ell_1} \dots a_{\ell_{p_\ell}}$ and $f^n(a_\ell) = a_{\ell'_1} \dots a_{\ell'_{q_\ell}}$. Then, for all $n \in \mathbb{N}$,*

$$|f^{n+1}(a_\ell)|_{1\#2} = \sum_{1\leq i<j\leq p_\ell} (\sum_{r=1}^{k-1}(m_{n,r,\ell_i} \cdot \sum_{s=r+1}^{k} m_{n,s,\ell_j})) + \sum_{t=1}^{k} |f^n(a_t)|_{1\#2} \cdot m_{1,t,\ell} \,, \tag{1}$$

$$= \sum_{1\leq i<j\leq q_\ell} (\sum_{r=1}^{k-1}(m_{1,r,\ell'_i} \cdot \sum_{s=r+1}^{k} m_{1,s,\ell'_j})) + \sum_{t=1}^{k} |f(a_t)|_{1\#2} \cdot m_{n,t,\ell} \,, \tag{2}$$

$$|f^{n+1}(a_\ell)|_{2\#1} = \sum_{1\leq i<j\leq p_\ell} (\sum_{r=2}^{k}(m_{n,r,\ell_i} \cdot \sum_{s=1}^{r-1} m_{n,s,\ell_j})) + \sum_{t=1}^{k} |f^n(a_t)|_{2\#1} \cdot m_{1,t,\ell} \,, \tag{3}$$

$$= \sum_{1\leq i<j\leq q_\ell} (\sum_{r=2}^{k}(m_{1,r,\ell'_i} \cdot \sum_{s=1}^{r-1} m_{1,s,\ell'_j})) + \sum_{t=1}^{k} |f(a_t)|_{2\#1} \cdot m_{n,t,\ell} \,. \tag{4}$$

Of course, the analysis we have done above could be realized to compute more complex ordered patterns with gaps, as 1#23, 1#2#3, $\cdots$ The only difficulty is to adapt the computation of external rises or descents.

3.2.2 Repetitions of one letter

Let n be a non-negative integer and p a positive integer. The *vector of p-repetitions of one letter* of f^n is the k vector whose i-th entry is the number of occurrences of the ordered pattern $(1\#)^p$ in the word $f^n(a_i)$, i.e.,

$$R_p(f^n) = (|f^n(a_i)|_{(1\#)^p})_{1\leq i\leq k}.$$

The following is obvious.

Proposition 2 *For each letter $a_\ell \in A$ and for all $n \in \mathbb{N}$,*

$$|f^n(a_\ell)|_{(1\#)^p} = \sum_{t=1}^{k} \binom{m_{n,t,\ell}}{p}. \tag{5}$$

3.3 Some examples in the binary case

Since equations (1) to (5) are recurrence formulas they are not always suitable to produce exact formulas giving the entries of $RG(f^n)$, $DG(f^n)$, and $R_p(f^n)$. However, in some particular cases we obtained such exact formulas. This is in particualr the case for the following two classical morphisms on the two-letter ordered alphabet $\{a_1 < a_2\}$.

3.3.1 The Thue-Morse morphism

The *Thue-Morse morphism* μ was introduced in 1912 by Thue [28], although it was hinted at sixty years earlier by Prouhet [22]. It was discovered independently in 1921 by Morse [21]. This morphism is defined by $\mu(a_1) = a_1a_2$, $\mu(a_2) = a_2a_1$. It generates the famous *Thue-Morse sequence* $\mathbf{t} = \mu^\omega(a_1)$ which has been widely studied (see, e.g., Lothaire, 1983 [19], or Allouche and Shallit, 2003 [2], and references therein).

For every positive integers n, the incidence matrix of μ^n is

$$M(\mu^n) = \begin{bmatrix} 2^{n-1} & 2^{n-1} \\ 2^{n-1} & 2^{n-1} \end{bmatrix}.$$

Thus, from equations (1), (3), and (5) we obtain

$$\begin{array}{rcccl} |\mu^{n+1}(a_1)|_{1\#2} & = & |\mu^{n+1}(a_2)|_{1\#2} & = & 2^{2(n-1)} + |\mu^n(a_1)|_{1\#2} + |\mu^n(a_2)|_{1\#2}, \\ |\mu^{n+1}(a_1)|_{2\#1} & = & |\mu^{n+1}(a_2)|_{2\#1} & = & 2^{2(n-1)} + |\mu^n(a_1)|_{2\#1} + |\mu^n(a_2)|_{2\#1}, \\ |\mu^n(a_1)|_{(1\#)^p} & = & |\mu^n(a_2)|_{(1\#)^p} & = & 2 \cdot \binom{2^{n-1}}{p}. \end{array}$$

Since $RG(\mu) = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and $DG(\mu) = \begin{bmatrix} 0 & 1 \end{bmatrix}$, we obtain from Proposition 1 the following well known result.

Corollary 1 *For any integer $n \geq 2$,*

$$RG(\mu^n) = DG(\mu^n) = \begin{bmatrix} 2^{2n-3} & 2^{2n-3} \end{bmatrix} \text{ and } R_p(\mu^n) = \begin{bmatrix} 2 \cdot \binom{2^{n-1}}{p} & 2 \cdot \binom{2^{n-1}}{p} \end{bmatrix}.$$

3.3.2 The Fibonacci morphism

The *Fibonacci morphism* φ is defined by $\varphi(a_1) = a_1a_2$, $\varphi(a_2) = a_1$. It generates the well known *Fibonacci sequence* $\mathbf{f} = \varphi^\omega(a_1)$ which has numerous remarkable properties and is the prototype of a Sturmian word (see, e.g., chapter 2 of Lothaire, 2002 [20]).

Let $(F_n)_{n\geq -1}$ be the sequence of Fibonacci numbers: $F_{-1} = 0$, $F_0 = 1$, $F_n = F_{n-1} + F_{n-2}$ for $n \geq 1$. The following property of Fibonacci numbers will be useful below.

Property 2 *For every positive integer n,*

$$F_n.F_{n-2} = F_{n-1}^2 + \begin{cases} 1 & \text{if } n \text{ is even,} \\ -1 & \text{if } n \text{ is odd.} \end{cases}$$

An easy computation gives that, for every positive integer n, the incidence matrix of φ^n is

$$M(\varphi^n) = \begin{bmatrix} F_n & F_{n-1} \\ F_{n-1} & F_{n-2} \end{bmatrix}.$$

The vector of rises with gaps of φ is $RG(\varphi) = \begin{bmatrix} 1 & 0 \end{bmatrix}$. Moreover, from equation (1), we obtain for $n \geq 1$

$$\begin{array}{rcl} |\varphi^{n+1}(a_1)|_{1\#2} & = & m_{n,1,1}.m_{n,2,2} + |\varphi^n(a_1)|_{1\#2} + |\varphi^n(a_2)|_{1\#2} \\ & = & F_n.F_{n-2} + |\varphi^n(a_1)|_{1\#2} + |\varphi^n(a_2)|_{1\#2} \\ & = & F_{n-1}^2 + |\varphi^n(a_1)|_{1\#2} + |\varphi^n(a_2)|_{1\#2} + \begin{cases} 1 & \text{if } n \text{ is even,} \\ -1 & \text{if } n \text{ is odd} \end{cases} \qquad \text{(see Property 2).} \end{array}$$

The vector of descents with gaps of φ is $DG(\varphi) = \begin{bmatrix} 0 & 0 \end{bmatrix}$. Moreover, from equation (3), we obtain for $n \geq 1$

$$\begin{array}{rcl} |\varphi^{n+1}(a_1)|_{2\#1} & = & m_{n,2,1} \cdot m_{n,1,2} + |\varphi^n(a_1)|_{2\#1} + |\varphi^n(a_2)|_{2\#1} \\ & = & F_{n-1}^2 + |\varphi^n(a_1)|_{2\#1} + |\varphi^n(a_2)|_{2\#1}. \end{array}$$

Now, $|\varphi^{n+1}(a_2)|_{1\#2} = |\varphi^n(a_1)|_{1\#2}$ and $|\varphi^{n+1}(a_2)|_{2\#1} = |\varphi^n(a_1)|_{2\#1}$ because $\varphi(a_2) = a_1$.

From this we obtain direct formulas to compute, for every $n \geq 0$, $|\varphi^{n+2}(a_1)|_{1\#2}$ and $|\varphi^{n+2}(a_1)|_{2\#1}$ from the sequence of Fibonacci numbers.

Corollary 2 *For every integer $n \geq 0$,*

$$\begin{array}{rcl} |\varphi^{n+2}(a_1)|_{2\#1} & = & \sum_{p=0}^{n} F_p F_{n-p}^2 \,, \\ |\varphi^{n+2}(a_1)|_{1\#2} & = & |\varphi^{n+2}(a_1)|_{2\#1} + F_n + \left\{ \begin{array}{rl} 1 & \textit{if } n \textit{ is odd,} \\ -1 & \textit{if } n \textit{ is even.} \end{array} \right. \end{array}$$

Proof. Since $F_0 = 1$ and $\varphi^2(a_1) = a_1a_2a_1$, the result is obviously true if $n = 0$.

Also, since $F_0 = 1$, $F_1 = 1$, and $\varphi^3(a_1) = a_1a_2a_1a_1a_2$, the result is true for $n = 1$.

Now suppose the assertions are true for all $m < n$. We prove they are true for n.

- We first compute $|\varphi^{n+2}(a_1)|_{2\#1}$.

$$\begin{array}{rcl} |\varphi^{n+2}(a_1)|_{2\#1} & = & F_n^2 + |\varphi^{n+1}(a_1)|_{2\#1} + |\varphi^{n}(a_1)|_{2\#1} \\ & = & F_n^2 + \sum_{p=0}^{n-1} F_p F_{n-1-p}^2 + \sum_{p=0}^{n-2} F_p F_{n-2-p}^2 . \\ \text{But } \sum_{p=0}^{n-2} F_p F_{n-2-p}^2 & = & \sum_{p=1}^{n-1} F_{p-1} F_{n-2-(p-1)}^2 \\ & = & \sum_{p=1}^{n-1} F_{p-1} F_{n-1-p}^2 . \\ \text{Thus } |\varphi^{n+2}(a_1)|_{2\#1} & = & F_n^2 + F_0 F_{n-1}^2 + \sum_{p=1}^{n-1} (F_p + F_{p-1}) F_{n-1-p}^2 \\ & = & F_n^2 + F_{n-1}^2 + \sum_{p=1}^{n-1} F_{p+1} F_{n-(p+1)}^2 \\ & = & F_n^2 + F_{n-1}^2 + \sum_{p=2}^{n} F_p F_{n-p}^2 \\ & = & \sum_{p=0}^{n} F_p F_{n-p}^2 \text{ (because } F_0 = F_1 = 1). \end{array}$$

- For $|\varphi^{n+2}(a_1)|_{1\#2}$, we remark that if n is even then $n-2$ is even, and $n-1$, $n+1$ are odd. And if n is odd then $n-2$ is odd, and $n-1$, $n+1$ are even. Consequently,

$$\begin{array}{rcl} |\varphi^{n+2}(a_1)|_{1\#2} & = & F_n^2 + |\varphi^{n+1}(a_1)|_{1\#2} + |\varphi^{n}(a_2)|_{1\#2} + \left\{ \begin{array}{rl} 1 & \text{if } n+1 \text{ is even } (n \text{ odd}), \\ -1 & \text{if } n+1 \text{ is odd } (n \text{ even}) \end{array} \right. \\ & = & F_n^2 + \sum_{p=0}^{n-1} F_p F_{n-1-p}^2 + F_{n-1} + 1 + \sum_{p=0}^{n-2} F_p F_{n-2-p}^2 + F_{n-2} - 1 + \left\{ \begin{array}{rl} 1 & \text{if } n \text{ is odd,} \\ -1 & \text{if } n \text{ is even} \end{array} \right. \\ & = & \sum_{p=0}^{n} F_p F_{n-p}^2 + F_{n-1} + F_{n-2} + \left\{ \begin{array}{rl} 1 & \text{if } n \text{ is odd,} \\ -1 & \text{if } n \text{ is even} \end{array} \right. \\ & = & \sum_{p=0}^{n} F_p F_{n-p}^2 + F_n + \left\{ \begin{array}{rl} 1 & \text{if } n \text{ is odd,} \\ -1 & \text{if } n \text{ is even.} \end{array} \right. \blacksquare \end{array}$$

Regarding repetitions of one letter, $R_p(\varphi) = \left[\begin{array}{cc} \binom{1}{p} + \binom{1}{p} & \binom{1}{p} \end{array} \right]$ and, for $n \geq 0$, the vector $R_p(\varphi^{n+2})$ is obtained from equation (5).

Corollary 3 *For any integer $n \geq 0$,*

$$R_p(\varphi^{n+2}) = \left[\begin{array}{cc} \binom{F_{n+2}}{p} + \binom{F_{n+1}}{p} & \binom{F_{n+1}}{p} + \binom{F_n}{p} \end{array} \right].$$

4 A particular family of morphisms

Let k be an integer ($k \geq 2$) and A the k-letter ordered alphabet $A = \{a_1 < a_2 < \cdots < a_k\}$. In this section we are interested in morphisms f having the following particularities:

1. there exists a positive integer m such that $|f(a_1)|_{a_i} = m$, $1 \leq i \leq k$,
2. there exists a positive integer d such that $|f(a_2 \ldots a_k)|_{a_i} = d$, $1 \leq i \leq k$,
3. for every i, j, $1 \leq i, j \leq k$, $|f(a_ia_j)|_{1\#2}^{ext} = |f(a_ja_i)|_{1\#2}^{ext}$.

In this case we are able to give direct formulas to compute $|f^{n+1}(a_1)|_{1\#2}$ and others from m, d, and n.

Proposition 3 *For every positive integer n,*

$$|f^{n+1}(a_1)|_{1\#2} = m(d+m)^{n-1}\sum_{i=1}^{k}|f(a_i)|_{1\#2} + \frac{[m(d+m)^{n-1}-1]m(d+m)^{n-1}}{2}\sum_{j=1}^{k}|f(a_ja_j)|_{1\#2}^{ext} + m^2(d+m)^{2n-2}\sum_{1\le i<j\le k}|f(a_ia_j)|_{1\#2}^{ext},$$

$$|f^{n+1}(a_2\dots a_k)|_{1\#2} = d(d+m)^{n-1}\sum_{i=1}^{k}|f(a_i)|_{1\#2} + \frac{[d(d+m)^{n-1}-1]d(d+m)^{n-1}}{2}\sum_{j=1}^{k}|f(a_ja_j)|_{1\#2}^{ext} + d^2(d+m)^{2n-2}\sum_{1\le i<j\le k}|f(a_ia_j)|_{1\#2}^{ext}.$$

Proof. Let $n\ge 1$. As in Proposition 1, let $f^n(a_1) = a_{1'_1}\dots a_{1'_{q_1}}$. Equation (2) gives

$$\begin{aligned}|f^{n+1}(a_1)|_{1\#2} &= \sum_{1\le i<j\le q_1}\left(\sum_{r=1}^{k-1}(m_{1,r,1'_i}\cdot\sum_{s=r+1}^{k}m_{1,s,1'_j})\right) + \sum_{t=1}^{k}|f(a_t)|_{1\#2}\cdot m_{n,t,1}\\ &= \sum_{1\le i<j\le q_1}|f(a_{1'_i}a_{1'_j})|_{1\#2}^{ext} + \sum_{t=1}^{k}|f(a_t)|_{1\#2}\cdot|f^n(a_1)|_{a_t}.\end{aligned}$$

Now, conditions 1. to 3. above imply that the incidence matrix of f^n is rather special. From 1. and 2., $|f^n(a_1)|_{a_t} = m(d+m)^{n-1}$ for each t, $1\le t\le k$.

This implies that $\sum_{t=1}^{k}|f(a_t)|_{1\#2}\cdot|f^n(a_1)|_{a_t} = m(d+m)^{n-1}\sum_{i=1}^{k}|f(a_i)|_{1\#2}$.

This also implies that $q_1 = km(d+m)^{n-1}$.

But, from 3., the computation of $\sum_{1\le i<j\le km(d+m)^{n-1}}|f(a_{1'_i}a_{1'_j})|_{1\#2}^{ext}$, realized from the word $f^n(a_1)$ which contains $m(d+m)^{n-1}$ occurrences of each letter, can be equivalently realized from the word $a_1^{m(d+m)^{n-1}}a_2^{m(d+m)^{n-1}}\cdots a_k^{m(d+m)^{n-1}}$.

Then the first letter a_1 ($i=1$) gives $[m(d+m)^{n-1}-1]\cdot|f(a_1a_1)|_{1\#2}^{ext}+\sum_{j=2}^{k}m(d+m)^{n-1}|f(a_1a_j)|_{1\#2}^{ext}$.

The second letter a_1 ($i=2$) gives $[m(d+m)^{n-1}-2]\cdot|f(a_1a_1)|_{1\#2}^{ext}+\sum_{j=2}^{k}m(d+m)^{n-1}|f(a_1a_j)|_{1\#2}^{ext}$.

$\vdots$

The last but one letter a_1 ($i=m(d+m)^{n-1}-1$) gives $1\cdot|f(a_1a_1)|_{1\#2}^{ext}+\sum_{j=2}^{k}m(d+m)^{n-1}|f(a_1a_j)|_{1\#2}^{ext}$.

The last letter a_1 ($i=m(d+m)^{n-1}$) gives $0\cdot|f(a_1a_1)|_{1\#2}^{ext}+\sum_{j=2}^{k}m(d+m)^{n-1}|f(a_1a_j)|_{1\#2}^{ext}$.

The first letter a_2 ($i=m(d+m)^{n-1}+1$) gives $[m(d+m)^{n-1}-1]\cdot|f(a_2a_2)|_{1\#2}^{ext}+\sum_{j=3}^{k}m(d+m)^{n-1}|f(a_2a_j)|_{1\#2}^{ext}$.

And so on.

Consequently $\sum_{1\le i<j\le km(d+m)^{n-1}}|f(a_{1'_i}a_{1'_j})|_{1\#2}^{ext}$

$$\begin{aligned}&= \sum_{i=0}^{m(d+m)^{n-1}-1} i\cdot\sum_{j=1}^{k}|f(a_ja_j)|_{1\#2}^{ext} + m(d+m)^{n-1}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}m(d+m)^{n-1}|f(a_ia_j)|_{1\#2}^{ext}\\ &= \frac{[m(d+m)^{n-1}-1]m(d+m)^{n-1}}{2}\sum_{j=1}^{k}|f(a_ja_j)|_{1\#2}^{ext} + m^2(d+m)^{2n-2}\sum_{1\le i<j\le k}|f(a_ia_j)|_{1\#2}^{ext}.\end{aligned}$$

Thus $|f^{n+1}(a_1)|_{1\#2} = m(d+m)^{n-1}\sum_{i=1}^{k}|f(a_i)|_{1\#2} + \frac{[m(d+m)^{n-1}-1]m(d+m)^{n-1}}{2}\sum_{j=1}^{k}|f(a_ja_j)|_{1\#2}^{ext} + m^2(d+m)^{2n-2}\sum_{1\le i<j\le k}|f(a_ia_j)|_{1\#2}^{ext}$.

Now, for $|f^{n+1}(a_2\dots a_k)|_{1\#2}$, equation (2) gives

$$\begin{aligned}|f^{n+1}(a_2\dots a_k)|_{1\#2} &= \sum_{\ell=2}^{k}\left(\sum_{1\le i<j\le q_\ell}|f(a_{\ell'_i}a_{\ell'_j})|_{1\#2}^{ext} + \sum_{t=1}^{k}|f(a_t)|_{1\#2}\cdot|f^n(a_\ell)|_{a_t}\right)\\ &= \sum_{\ell=2}^{k}\sum_{1\le i<j\le q_\ell}|f(a_{\ell'_i}a_{\ell'_j})|_{1\#2}^{ext} + \sum_{t=1}^{k}\sum_{\ell=2}^{k}|f(a_t)|_{1\#2}\cdot|f^n(a_\ell)|_{a_t}.\end{aligned}$$

Again from 1. and 2., $\sum_{\ell=2}^{k}|f^n(a_\ell)|_{a_t} = d(d+m)^{n-1}$ for each t, $1\le t\le k$.

In particular, $\sum_{\ell=2}^{k}\sum_{1\le i<j\le q_\ell}|f(a_{\ell'_i}a_{\ell'_j})|_{1\#2}^{ext} = \sum_{1\le i<j\le kd(d+m)^{n-1}}|f(a_{\ell'_i}a_{\ell'_j})|_{1\#2}^{ext}$.

As above, the computation can be realized from the word $a_1^{d(d+m)^{n-1}} a_2^{d(d+m)^{n-1}} \cdots a_k^{d(d+m)^{n-1}}$. This gives $\sum_{1\le i<j\le kd(d+m)^{n-1}} |f(a_{\ell'_i} a_{\ell'_j})|^{ext}_{1\#2}$

$= \sum_{i=0}^{d(d+m)^{n-1}-1} i \cdot \sum_{j=1}^{k} |f(a_j a_j)|^{ext}_{1\#2} + d(d+m)^{n-1} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} d(d+m)^{n-1} |f(a_i a_j)|^{ext}_{1\#2}$

$= \frac{[d(d+m)^{n-1}-1]d(d+m)^{n-1}}{2} \sum_{j=1}^{k} |f(a_j a_j)|^{ext}_{1\#2} + d^2(d+m)^{2n-2} \sum_{1\le i<j\le k} |f(a_i a_j)|^{ext}_{1\#2}$. ■

Now the same reasoning can be applied for $|f^{n+1}(a_1)|_{2\#1}$ and $|f^{n+1}(a_2 \ldots a_k)|_{2\#1}$, because of the following obvious property.

Property 3 *Let f be a morphism on A. For every non-negative integer n, and for every integers i, j, $1 \le i, j \le k$, $|f^n(a_i a_j)|^{ext}_{1\#2} = |f^n(a_j a_i)|^{ext}_{2\#1}$.*

Thus, using equation (4), we have the following.

Proposition 4 *For every positive integer n,*

$$\begin{aligned} |f^{n+1}(a_1)|_{2\#1} = \;& m(d+m)^{n-1} l \sum_{i=1}^{k} |f(a_i)|_{2\#1} + \frac{[m(d+m)^{n-1}-1]m(d+m)^{n-1}}{2} \sum_{j=1}^{k} |f(a_j a_j)|^{ext}_{2\#1} \\ & + m^2(d+m)^{2n-2} \sum_{1\le i<j\le k} |f(a_i a_j)|^{ext}_{2\#1}, \end{aligned}$$

$$\begin{aligned} |f^{n+1}(a_2 \ldots a_k)|_{2\#1} = \;& d(d+m)^{n-1} l \sum_{i=1}^{k} |f(a_i)|_{2\#1} + \frac{[d(d+m)^{n-1}-1]d(d+m)^{n-1}}{2} \sum_{j=1}^{k} |f(a_j a_j)|^{ext}_{2\#1} \\ & + d^2(d+m)^{2n-2} \sum_{1\le i<j\le k} |f(a_i a_j)|^{ext}_{2\#1}. \end{aligned}$$

The previous reasoning can of course be applied if conditions 1. and 2. are verified for any partition of the alphabet (in Propositions 3 and 4 the partition is in two sets $A = \{a_1\} \cup \{a_2 \ldots a_k\}$). Then we obtain the following general result.

Theorem 1 *Let k be an integer ($k \ge 2$), and A the k-letter ordered alphabet $A = \{a_1 < a_2 < \ldots < a_k\}$. Let f be a morphism on A fulfilling the following conditions:*

- *there exist a positive integer p and a set of p positive integers $\{m_1, \ldots, m_p\}$ such that A can be partitioned into p subsets $A_1, \ldots, A_p$ with $\sum_{a\in A_\ell} |f(a)|_{a_i} = m_\ell$, $1 \le i \le k$,*
- *for every i, j, $1 \le i, j \le k$, $|f(a_i a_j)|^{ext}_{1\#2} = |f(a_j a_i)|^{ext}_{1\#2}$.*

Let $M = m_1 + \ldots + m_p$ and let $u = 1\#2$ or $u = 2\#1$. Then, for every positive integer n and for each A_ℓ, $1 \le \ell \le p$,

$$\begin{aligned} \sum_{a\in A_\ell} |f^{n+1}(a)|_u = \;& m_\ell M^{n-1} \sum_{i=1}^{k} |f(a_i)|_u + \frac{(m_\ell M^{n-1}-1)m_\ell M^{n-1}}{2} \sum_{j=1}^{k} |f(a_j a_j)|^{ext}_u \\ & + m_\ell^2 M^{2n-2} \sum_{1\le i<j\le k} |f(a_i a_j)|^{ext}_u. \end{aligned}$$

5 Examples

5.1 The Thue-Morse morphism

The Thue-Morse morphism (see Section 3.3) is the simplest example of a morphism fulfilling conditions 1. to 3. above. Indeed $m = d = 1$, and $|\mu(a_1a_2)|^{ext}_{1\#2} = |a_1a_2a_2a_1|^{ext}_{1\#2} = 1 = |a_2a_1a_1a_2|^{ext}_{1\#2} = |\mu(a_2a_1)|^{ext}_{1\#2}$, $|\mu(a_1a_1)|^{ext}_{1\#2} = |\mu(a_2a_2)|^{ext}_{1\#2} = 1$. Since $|\mu(a_1)|_{1\#2} = |\mu(a_2)|_{2\#1} = 1$, and $|\mu(a_1)|_{2\#1} = |\mu(a_2)|_{1\#2} = 0$, we obtain from Propositions 3 and 4 that $|\mu^{n+1}(a_1)|_{1\#2} = |\mu^{n+1}(a_1)|_{2\#1} = |\mu^{n+1}(a_2)|_{1\#2} = |\mu^{n+1}(a_2)|_{2\#1} = 2^{2n-1}$ (see Corollary 1 above).

5.2 The Istrail morphism

In Istrail, 1977 [14] is given the following well known example of a weakly square-free morphism. The morphism h is defined on the three-letter ordered alphabet $A = \{a_1 < a_2 < a_3\}$ by

$$h(a_1) = a_1a_2a_3, \qquad h(a_2) = a_1a_3, \qquad h(a_3) = a_2$$

(remark that h generates a square-free infinite word, $h^\omega(a_1)$, but is not a square-free morphism: $h(a_1a_2a_1) = a_1a_2a_3a_1a_3a_1a_2a_3$ contains the square $a_3a_1a_3a_1$).

The word $h^\omega(a_1)$ is closely related to the Thue-Morse word $\mathbf{t}$. Indeed, let B be the two-letter alphabet $B = \{a_1, a_2\}$, and consider the morphism

$$\begin{array}{rccl} \delta: & A^* & \to & B^* \\ & a_1 & \mapsto & a_1 \\ & a_2 & \mapsto & a_1a_2 \\ & a_3 & \mapsto & a_1a_2a_2 \end{array}$$

Then $\mathbf{t} = \delta(h^\omega(a_1))$.

Here again the morphism h fulfills conditions 1. to 3. with $m = d = 1$. Moreover $|h(a_1)|_{1\#2} = |h(a_1a_1)|^{ext}_{1\#2} = 3$, $|h(a_2)|_{1\#2} = |h(a_2a_2)|^{ext}_{1\#2} = 1$, $|h(a_3)|_{1\#2} = |h(a_3a_3)|^{ext}_{1\#2} = 0$, and $|h(a_1a_2)|^{ext}_{1\#2} = 2$, $|h(a_1a_3)|^{ext}_{1\#2} = |h(a_2a_3)|^{ext}_{1\#2} = 1$. Then, from Proposition 3, for every integer $n \geq 1$, $|h^{n+1}(a_1)|_{1\#2} = |h^{n+1}(a_2a_3)|_{1\#2} = 3 \cdot 2^{2n-1} + 2^n$.

From Property 3, the values for 2#1 are the same as for 1#2, except for $|h(a_i)|_{2\#1}$, $1 \leq i \leq 3$. Here $|h(a_1)|_{2\#1} = |h(a_2)|_{2\#1} = |h(a_3)|_{2\#1} = 0$. Thus, from Proposition 4, for every integer $n \geq 1$, $|h^{n+1}(a_1)|_{2\#1} = |h^{n+1}(a_2a_3)|_{2\#1} = 3 \cdot 2^{2n-1} - 2^n$.

5.3 The Prouhet morphisms

In 1851, Prouhet ([22]) gave an algorithm to realize an arithmetic construction. This algorithm produces intermediate infinite words that are a generalization of the Thue-Morse word (see above). It was proved in Séébold, 2002 [25] that these words can be generated by morphisms (see also allouche and Shallit, 2000 [1]).

Let $k \geq 2$, and let A be the k-letter ordered alphabet $A = \{a_1 < \cdots < a_k\}$. The *Prouhet morphism* π_k is defined on A by

$$\pi_k(a_i) = a_ia_{i+1} \ldots a_ka_1 \ldots a_{i-1}, \qquad 1 \leq i \leq k.$$

Example. Let $k = 6$. The morphism π_6 is given by

$$\begin{array}{lcl} a_1 & \mapsto & a_1a_2a_3a_4a_5a_6 \\ a_2 & \mapsto & a_2a_3a_4a_5a_6a_1 \\ a_3 & \mapsto & a_3a_4a_5a_6a_1a_2 \\ a_4 & \mapsto & a_4a_5a_6a_1a_2a_3 \\ a_5 & \mapsto & a_5a_6a_1a_2a_3a_4 \\ a_6 & \mapsto & a_6a_1a_2a_3a_4a_5 \end{array}$$

For every k the morphism π_k fulfills the conditions of Theorem 1. Since, for every i, $1 \leq i \leq k$, the word $\pi_k(a_i)$ contains exactly one occurrence of each letter of A, there are a lot of possibilities to choose the partition of A. Here we choose $p = k$ and $A = A_1 \cup \ldots \cup A_k$, $A_i = \{a_i\}$, $1 \leq i \leq k$. This implies that $m_i = 1$, $1 \leq i \leq k$ and, of course, $M = k$.

Also, for every i, j, $1 \leq i, j \leq k$, $|\pi_k(a_ia_j)|^{ext}_{1\#2} = |\pi_k(a_ia_j)|^{ext}_{2\#1} = \frac{k(k-1)}{2}$.

Now it is easy to verify that, due to the particular form of the images of the letters by π_k, one has for every ℓ, $1 \leq \ell \leq k$, $|\pi_k(a_\ell)|_{1\#2} = \frac{[k-(\ell-1)](k-\ell)}{2} + \frac{(\ell-1)(\ell-2)}{2}$ and $|\pi_k(a_\ell)|_{2\#1} = (\ell-1)[k-(\ell-1)]$.

Thus we obtain the following corollary of Theorem 1.

Corollary 4 *For every i, $1 \leq i \leq k$, and for every positive integer n,*

$$|\pi_k^{n+1}(a_i)|_{1\#2} = \frac{k^{n-1}}{4}\left(k^{n+2} \cdot (k-1) + \sum_{\ell=1}^{k}(\ell-1)(\ell-2)\right),$$

$$|\pi_k^{n+1}(a_i)|_{2\#1} = \frac{k^{n-1}}{4}\left(k^{n+2}\cdot(k-1) - \sum_{\ell=1}^{k}(\ell-1)(\ell-2)\right).$$

Proof. From Theorem 1 and from what precedes,

$$\begin{aligned}
|\pi_k^{n+1}(a_i)|_{1\#2} &= k^{n-1}\cdot\sum_{\ell=1}^k\left(\tfrac{[k-(\ell-1)](k-\ell)}{2}+\tfrac{(\ell-1)(\ell-2)}{2}\right)+\tfrac{(k^{n-1}-1)k^{n-1}}{2}\sum_{j=1}^k\tfrac{k(k-1)}{2}\\
&\quad+k^{2n-2}\cdot\sum_{1\le j<\ell\le k}\tfrac{k(k-1)}{2}\\
&= k^{n-1}\cdot\left[\sum_{\ell=1}^k\left(\tfrac{[k-(\ell-1)](k-\ell)}{2}+\tfrac{(\ell-1)(\ell-2)}{2}\right)+\tfrac{(k^{n-1}-1)}{2}\cdot\tfrac{k^2(k-1)}{2}+k^{n-1}\cdot\left[\tfrac{k(k-1)}{2}\right]^2\right]\\
&= k^{n-1}\cdot\left[\sum_{\ell=1}^k\left(\tfrac{[k-(\ell-1)](k-\ell)}{2}+\tfrac{(\ell-1)(\ell-2)}{2}\right)-\tfrac{k^2(k-1)}{4}+k^{n-1}\cdot\left(\tfrac{k^2(k-1)}{4}+\tfrac{k^2(k-1)^2}{4}\right)\right].
\end{aligned}$$

Since $\sum_{\ell=1}^k\left(\frac{[k-(\ell-1)](k-\ell)}{2}+\frac{(\ell-1)(\ell-2)}{2}\right)-\frac{k^2(k-1)}{4}=\frac{1}{2}\sum_{\ell=1}^k\frac{(\ell-1)(\ell-2)}{2}$, we obtain

$$\begin{aligned}
|\pi_k^{n+1}(a_i)|_{1\#2} &= k^{n-1}\cdot\left[k^{n-1}\cdot\left(\tfrac{k^2(k-1)+k^2(k-1)^2}{4}\right)+\tfrac{1}{4}\sum_{\ell=1}^k(\ell-1)(\ell-2)\right]\\
&= \tfrac{k^{n-1}}{4}\cdot\left[k^{n+2}\cdot(k-1)+\sum_{\ell=1}^k(\ell-1)(\ell-2)\right]
\end{aligned}$$

The proof is the same for $|\pi_k^{n+1}(a_i)|_{2\#1}$, using $\sum_{\ell=1}^k(\ell-1)[k-(\ell-1)]-\frac{k^2(k-1)}{4}=-\frac{1}{2}\sum_{\ell=1}^k\frac{(\ell-1)(\ell-2)}{2}$.
■

Example (continued).

$$\begin{array}{lcrclcl}
|\pi_6(a_1)|_{1\#2} & = & 15, & \quad & |\pi_6(a_1)|_{2\#1} & = & 0,\\
|\pi_6(a_2)|_{1\#2} & = & 10, & & |\pi_6(a_2)|_{2\#1} & = & 5,\\
|\pi_6(a_3)|_{1\#2} & = & 7, & & |\pi_6(a_3)|_{2\#1} & = & 8,\\
|\pi_6(a_4)|_{1\#2} & = & 6, & & |\pi_6(a_4)|_{2\#1} & = & 9,\\
|\pi_6(a_5)|_{1\#2} & = & 7, & & |\pi_6(a_5)|_{2\#1} & = & 8,\\
|\pi_6(a_6)|_{1\#2} & = & 10, & & |\pi_6(a_6)|_{2\#1} & = & 5.
\end{array}$$

For every i, $1\le i\le k$, and for every $n\ge 1$,

$$\begin{aligned}
|\pi_6^{n+1}(a_i)|_{1\#2} &= \tfrac{6^{n-1}}{4}\cdot\left(6^{n+2}\cdot 5+\sum_{\ell=1}^6(\ell-1)(\ell-2)\right)\\
&= 6^{n-1}\cdot(45\cdot 6^n+10),\\
|\pi_6^{n+1}(a_i)|_{2\#1} &= 6^{n-1}\cdot(45\cdot 6^n-10).
\end{aligned}$$

5.4 The Arshon morphisms

In a paper written in 1937 [3], Arshon gives an algorithm to construct for each integer n, $n\ge 3$, an infinite square-free word over an n-letter alphabet, and in the case of two letters a cube-free word. It appears now that this construction is closely connected to the use of Prouhet morphisms. In the case of two letters the Arshon word is the Thue-Morse word and Arshon's algorithm gives exactly the Thue-Morse morphism which is a particular case of Prouhet morphism.

The Arshon words were proved to be, in the odd case, an exemple of infinite words that can be generated by a tag-system but not by a morphism (Berstel, 1980 [5], Currie, 2002 [13], Kitaev, 2003 [15]). In Séébold, 2003 [26] is given a family of morphisms which generates the even case Arshon words (see also Currie, 2002 [13], Kitaev, 2003 [15]). These morphisms are the following.

Let k be any even positive integer. The morphism β_k is defined, for every r, $1\le r\le k/2$, by

$$\begin{array}{lcl}
a_{2r-1} & \mapsto & a_{2r-1}a_{2r}\ldots a_{k-1}a_k a_1 a_2\ldots a_{2r-3}a_{2r-2},\\
a_{2r} & \mapsto & a_{2r-1}a_{2r-2}\ldots a_2 a_1 a_k a_{k-1}\ldots a_{2r+1}a_{2r}.
\end{array}$$

(Remark that, agai, though they generate square-free infinite words the morphisms β_k are not square-free morphisms.)

Example. Let $k = 6$. The morphism β_6 is given by

$$\begin{array}{lcl} a_1 & \mapsto & a_1a_2a_3a_4a_5a_6 \\ a_2 & \mapsto & a_1a_6a_5a_4a_3a_2 \\ a_3 & \mapsto & a_3a_4a_5a_6a_1a_2 \\ a_4 & \mapsto & a_3a_2a_1a_6a_5a_4 \\ a_5 & \mapsto & a_5a_6a_1a_2a_3a_4 \\ a_6 & \mapsto & a_5a_4a_3a_2a_1a_6 \end{array}$$

Of course, since it is obtained from π_k in an obvious manner, the morphism β_k fulfills the conditions of Theorem 1 for every even k. Since, for every i, $1 \leq i \leq k$, the word $\beta_k(a_i)$ contains exactly one occurrence of each letter of A, there are again a lot of possibilities to choose the partition of A. Here we choose also $p = k$ and $A = A_1 \cup \ldots \cup A_k$, $A_i = \{a_i\}$, $1 \leq i \leq k$. This implies that $m_i = 1$, $1 \leq i \leq k$ and, of course, $M = k$.

Also, for every i, j, $1 \leq i, j \leq k$, $|\beta_k(a_ia_j)|^{ext}_{1\#2} = |\beta_k(a_ia_j)|^{ext}_{2\#1} = \frac{k(k-1)}{2}$.

Now, again because β_k is directly obtained from π_k, one has for every r, $1 \leq r \leq \frac{k}{2}$,

$$\begin{array}{rcl} |\beta_k(a_{2r-1})|_{1\#2} & = & \frac{[k-(2r-2)][k-(2r-1)]}{2} + \frac{(2r-2)(2r-3)}{2}, \\ |\beta_k(a_{2r})|_{1\#2} & = & (2r-1)[k-(2r-1)], \\ |\beta_k(a_{2r-1})|_{2\#1} & = & (2r-2)[k-(2r-2)], \\ |\beta_k(a_{2r})|_{2\#1} & = & \frac{[k-(2r-1)](k-2r)}{2} + \frac{(2r-1)(2r-2)}{2}. \end{array}$$

Thus we obtain another corollary of Theorem 1.

Corollary 5 *Let k be any even positive integer. For every i, $1 \leq i \leq k$, and for every positive integer n,*

$$|\beta_k^{n+1}(a_i)|_{1\#2} = \frac{k^{n-1}}{4}\left[k^{n+2}\cdot(k-1)+2k\right],$$

$$|\beta_k^{n+1}(a_i)|_{2\#1} = \frac{k^{n-1}}{4}\left[k^{n+2}\cdot(k-1)-2k\right].$$

Proof. As for the proof of Corollary 4, we obtain from Theorem 1 and from what precedes,

$$|\beta_k^{n+1}(a_i)|_{1\#2} \quad = \quad k^{n-1}\cdot\left[\sum_{\ell=1}^{k}|\beta_k(a_\ell)|_{1\#2} - \frac{k^2(k-1)}{4}\right] + k^{n-1}\cdot\left[\frac{k^{n+2}\cdot(k-1)}{4}\right].$$

But $\sum_{\ell=1}^{k}|\beta_k(a_\ell)|_{1\#2} = \sum_{r=1}^{k/2}[|\beta_k(a_{2r-1})|_{1\#2} + |\beta_k(a_{2r})|_{1\#2}] = \frac{k^2(k-1)}{4} + \frac{k}{2}$, and the result follows.

The proof is the same for $|\beta_k^{n+1}(a_i)|_{2\#1}$, using $\sum_{\ell=1}^{k}|\beta_k(a_\ell)|_{2\#1} = \frac{k^2(k-1)}{4} - \frac{k}{2}$. ■

Example (continued).

$$\begin{array}{rcrcrcr} |\beta_6(a_1)|_{1\#2} & = & 15, & \quad & |\beta_6(a_1)|_{2\#1} & = & 0, \\ |\beta_6(a_2)|_{1\#2} & = & 5, & & |\beta_6(a_2)|_{2\#1} & = & 10, \\ |\beta_6(a_3)|_{1\#2} & = & 7, & & |\beta_6(a_3)|_{2\#1} & = & 8, \\ |\beta_6(a_4)|_{1\#2} & = & 9, & & |\beta_6(a_4)|_{2\#1} & = & 6, \\ |\beta_6(a_5)|_{1\#2} & = & 7, & & |\beta_6(a_5)|_{2\#1} & = & 8, \\ |\beta_6(a_6)|_{1\#2} & = & 5, & & |\beta_6(a_6)|_{2\#1} & = & 10. \end{array}$$

For every i, $1 \leq i \leq k$, and for every $n \geq 1$,

$$\begin{array}{rcl} |\beta_6^{n+1}(a_i)|_{1\#2} & = & \frac{6^{n-1}}{4}\cdot\left(6^{n+2}\cdot 5 + 2\cdot 6\right) \\ & = & 6^{n-1}\cdot(45\cdot 6^n + 3), \\ |\beta_6^{n+1}(a_i)|_{2\#1} & = & 6^{n-1}\cdot(45\cdot 6^n - 3). \end{array}$$

5.5 Three other examples

To end this list of examples, we give three morphisms that fulfill the conditions of Theorem 1, but are not linked with the Thue-Morse morphism. Moreover they are interesting because the first one is an erasing morphism, the second gives a non trivial partition of the alphabet when applying Theorem 1, and the third is an example of a ternary square-free morphism fulfilling the conditions..

1. Let A be the four-letter ordered alphabet $A = \{a_1 < a_2 < a_3 < a_4\}$. Define the morphism f by

$$\begin{array}{rcl} f : A^* & \rightarrow & A^* \\ a_1 & \mapsto & a_1a_3a_2a_4 \\ a_2 & \mapsto & \varepsilon \\ a_3 & \mapsto & a_1a_4 \\ a_4 & \mapsto & a_2a_3 \end{array}$$

The morphism f fulfills the conditions of Theorem 1. Here we choose $p = 3$, $A = A_1 \cup A_2 \cup A_3$ with $A_1 = \{a_1\}$, $A_2 = \{a_2\}$, $A_3 = \{a_3, a_4\}$, and $m_1 = m_3 = 1$, $m_2 = 0$, thus $M = 2$.

One has $|f(a_1)|_{1\#2} = 5$, $|f(a_3)|_{1\#2} = |f(a_4)|_{1\#2} = 1$, $|f(a_1)|_{2\#1} = 1$, $|f(a_3)|_{2\#1} = |f(a_4)|_{2\#1} = 0$, $|f(a_1a_1)|^{ext}_{1\#2} = |f(a_1a_1)|^{ext}_{2\#1} = 6$, $|f(a_3a_3)|^{ext}_{1\#2} = |f(a_3a_3)|^{ext}_{2\#1} = |f(a_4a_4)|^{ext}_{1\#2} = |f(a_4a_4)|^{ext}_{2\#1} = 1$, $|f(a_1a_3)|^{ext}_{1\#2} = |f(a_1a_3)|^{ext}_{2\#1} = |f(a_1a_4)|^{ext}_{1\#2} = |f(a_1a_4)|^{ext}_{2\#1} = 3$, $|f(a_3a_4)|^{ext}_{1\#2} = |f(a_3a_4)|^{ext}_{2\#1} = 2$. All the values with a_2 are of course 0.

Then we have the following corollary of Theorem 1.

Corollary 6 *For every positive integer n,*

$$\begin{array}{rclcl} |f^{n+1}(a_1)|_{1\#2} & = & |f^{n+1}(a_3a_4)|_{1\#2} & = & 3 \cdot 2^{n-1} \cdot (2^{n+1} + 1), \\ |f^{n+1}(a_1)|_{2\#1} & = & |f^{n+1}(a_3a_4)|_{2\#1} & = & 3 \cdot 2^{n-1} \cdot (2^{n+1} - 1), \\ |f^{n+1}(a_2)|_{1\#2} & = & |f^{n+1}(a_2)|_{2\#1} & = & 0. \end{array}$$

2. Let A be the five-letter ordered alphabet $A = \{a_1 < a_2 < a_3 < a_4 < a_5\}$. Define the morphism g by

$$\begin{array}{rcl} g : A^* & \rightarrow & A^* \\ a_1 & \mapsto & a_1a_3a_5a_4a_2 \\ a_2 & \mapsto & a_4a_2a_3 \\ a_3 & \mapsto & a_5a_1 \\ a_4 & \mapsto & a_1a_5 \\ a_5 & \mapsto & a_2a_3a_4 \end{array}$$

The morphism g fulfills the conditions of Theorem 1. Here we choose $p = 3$, $A = A_1 \cup A_2 \cup A_3$ with $A_1 = \{a_1\}$, $A_2 = \{a_2, a_4\}$, $A_3 = \{a_3, a_5\}$, and $m_1 = m_2 = m_3 = 1$, thus $M = 3$.

One has $|g(a_1)|_{1\#2} = 6$, $|g(a_2)|_{1\#2} = |g(a_4)|_{1\#2} = 1$, $|g(a_3)|_{1\#2} = 0$, $|g(a_5)|_{1\#2} = 3$,

$|g(a_1)|_{2\#1} = 4$, $|g(a_2)|_{2\#1} = 2$, $|g(a_3)|_{2\#1} = 1$, $|g(a_4)|_{2\#1} = |g(a_5)|_{2\#1} = 0$,

$|g(a_1a_1)|^{ext}_{1\#2} = 10$, $|g(a_2a_2)|^{ext}_{1\#2} = |g(a_5a_5)|^{ext}_{1\#2} = 3$, $|g(a_3a_3)|^{ext}_{1\#2} = |g(a_4a_4)|^{ext}_{1\#2} = 1$,

$|g(a_1a_2)|^{ext}_{1\#2} = |g(a_1a_5)|^{ext}_{1\#2} = 6$, $|g(a_1a_3)|^{ext}_{1\#2} = |g(a_1a_4)|^{ext}_{1\#2} = 4$, $|g(a_2a_3)|^{ext}_{1\#2} = |g(a_2a_4)|^{ext}_{1\#2} = |g(a_2a_5)|^{ext}_{1\#2} = |g(a_3a_5)|^{ext}_{1\#2} = |g(a_4a_5)|^{ext}_{1\#2} = 3$, $|g(a_3a_4)|^{ext}_{1\#2} = 1$.

To end we recall that, with the second condition of Theorem 1 and property 3, one has for every integers i, j, $1 \leq i, j \leq k$, $|g(a_ia_j)|^{ext}_{1\#2} = |g(a_ia_j)|^{ext}_{2\#1}$.

Then we have the following corollary of Theorem 1.

Corollary 7 *For every positive integer n,*

$$|g^{n+1}(a_1)|_{1\#2} = |g^{n+1}(a_2a_4)|_{1\#2} = |g^{n+1}(a_3a_5)|_{1\#2} = 3^{n-1} \cdot (5 \cdot 3^{n+1} + 2),$$

$$|g^{n+1}(a_1)|_{2\#1} = |g^{n+1}(a_2a_4)|_{2\#1} = |g^{n+1}(a_3a_5)|_{2\#1} = 3^{n-1} \cdot (5 \cdot 3^{n+1} - 2).$$

3. Let A be the three-letter ordered alphabet $A = \{a < b < c\}$. Define the morphism h by

$$\begin{array}{rcl} h : A^* & \rightarrow & A^* \\ a & \mapsto & aba\ cab\ cac\ bab\ cba\ cbc \\ b & \mapsto & aba\ cab\ cac\ bca\ bcb\ abc \\ c & \mapsto & aba\ cab\ cba\ cbc\ acb\ abc \end{array}$$

This morphism was given to be square-free by Brandenburg in [6]. It fulfills the conditions of Theorem 1 with $p = 3$, $A = A_1 \cup A_2 \cup A_3$ with $A_1 = \{a\}$, $A_2 = \{b\}$, $A_3 = \{c\}$, and $m_1 = m_2 = m_3 = 6$, thus $M = 18$.

One has $|h(a)|_{1\#2} = 70$, $|h(b)|_{1\#2} = |h(c)|_{1\#2} = 66$, $|h(a)|_{2\#1} = 38$, $|h(b)|_{2\#1} = |h(c)|_{2\#1} = 42$.

Moreover, due to the particular form of the morphism h (it is uniform, i.e., $|h(a)| = |h(b| = |h(c)|$, and for every $x, y \in A$, $|h(x)|_y = 6$), one has $|h(xy)|^{ext}_{1\#2} = |h(xy)|^{ext}_{2\#1} = 108$ for every $x, y \in A$.

Then we have the following corollary of Theorem 1.

Corollary 8 *For every $x \in A$ and for every positive integer n,*

$$|h^{n+1}(x)|_{1\#2} = 6 \cdot 18^{n-1} \cdot (9 \cdot 18^{n+1} + 40),$$

$$|h^{n+1}(x)|_{2\#1} = 6 \cdot 18^{n-1} \cdot (9 \cdot 18^{n+1} - 40).$$

6 Further investigations

In this paper we studied in details the counting of rises and descents with gaps in words generated by morphisms. The same study can be done for rises and descents without gaps. However, if similar results to those of Proposition 1 can be easily obtained in the case of nonerasing morphisms, the situation is rather more difficult with erasing morphisms. We hope we will be able to realize soon a complete study of this notion in a next paper.

References

[1] J.-P. Allouche, J. Shallit, Sums of digits, overlaps, and palindromes, *Discrete Math. & Theoret. Comput. Sci.* **4** (2000), 1–10.

[2] J.-P. Allouche, J. Shallit, *Automatic sequences: theory, applications, generalizations*, Cambridge University Press, Cambridge, UK, 2003.

[3] S. Arshon, Démonstration de l'existence de suites asymétriques infinies, *Mat. Sb.* **44** (1937), 769–777 (in Russian), 777–779 (French summary).

[4] E. Babson, E. Steingrímsson, Generalized permutation patterns and a classification of the Mahonian statistics, Sém.Lothar. Comb. **44** (2000), Art. B44b, 18 pp.

[5] J. Berstel, Mots sans carré et morphismes itérés, *Discr. Math.* **29** (1980), 235–244.

[6] F.-J. Brandenburg, Uniformly growing k-th power-free homomorphisms, *Theoret. Comput. Sci.* **23** (1983), 69–82.

[7] A. Burstein, Enumeration of words with forbidden patterns, Ph.D. thesis, University of Pennsylvania, 1998.

[8] A. Burstein, T. Mansour, Words restricted bypatterns with at most 2 distinct letters, *Electronic J. of Combinatorics* **9:2** (2002), #R3.

[9] A. Burstein, T. Mansour, Words restricted by 3-letter generalized multipermutation patterns, *Annals of Combinatorics* **7** (2003), 1–14.

[10] A. Burstein, T. Mansour, Counting occurrences of some subword patterns, *Discrete Math. & Theoret. Comput. Sci.* **6:1** (2003), 1–12.

[11] A. Claesson, Generalised Pattern Avoidance, *European J. Combin.* **22** no. 7 (2001), 961–971.

[12] A. Cobham, Uniform tag sequences, *Math. Systems Th.* **6** (1972), 164–192.

[13] J. Currie, No iterated morphism generates any Arshon sequence of odd order, *Discr. Math.* (2002), 277–283.

[14] S. Istrail, On irreducible languages and nonrational numbers, *Bull. Math. Soc. Sci. Math. R. S. Roumanie* **21** (1977), 301–308.

[15] S. Kitaev, There are no iterated morphisms that define the Arshon sequence and the sigma-sequence, *Automata, Languages and Combinatorics* **8** (2003), 43–50.

[16] S. Kitaev, Multi-avoidance of generalised patterns, *Discr. Math.* **260** (2003), 89–100.

[17] S. Kitaev, T. Mansour, P. Séébold, The Peano curve and counting occurrences of some patterns, *Journal of Automata, Languages and Combinatorics*, to appear.

[18] D. E. Knuth, *The Art of Computer Programming*, 2nd ed., Addison-Wesley, Reading, MA, 1973.

[19] M. Lothaire, *Combinatorics on Words*, vol. 17 of *Encyclopedia of Mathematics and Applications*, Addison-Wesley, Reading, Mass., 1983.
Reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, Cambridge, UK, 1997.

[20] M. Lothaire, *Algebraic Combinatorics on Words*, vol.90 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge, UK, 2002.

[21] M. Morse, Recurrent geodesics on a surface of negative curvature, *Trans. Amer. Math. Soc.* **22** (1921), 84–100.

[22] M. E. Prouhet, Mémoire sur quelques relations entre les puissances des nombres, *Comptes Rendus Acad. Sci. Paris* **33** (1851), 225.

[23] G. Rozenberg, A. Salomaa (Eds), *Handbook of Formal Languages* Vol. 1, Springer, 1997.

[24] A. Salomaa, *Jewels of Formal Language Theory*, Computer Science Press, 1981.

[25] P. Séébold, About some overlap-free morphisms on a n-letter alphabet, *Journal of Automata, Languages and Combinatorics* **7** (2002), 579–597.

[26] P. Séébold, On some generalizations of the Thue-Morse morphism, *Theoret. Comput. Sci.* **292-1** (2003), 283–298.

[27] R. Simion, F. Schmidt, Restricted permutations, *European J. Combin.* **6** no. 4 (1985), 383–406.

[28] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Vidensk.-Selsk. Skrifter. I. Mat. Nat. Kl.* **1** Kristiania (1912), 1–67.
Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, 413-478.