Theoretical Informatics and Applications Will be set by the publisher
Informatique Théorique et Applications

A GENERATOR OF MORPHISMS FOR INFINITE WORDS

Pascal Ochem[1]

Abstract. We present an algorithm which produces, in some cases, infinite words avoiding both large fractional repetitions and a given set of finite words. We use this method to show that all the ternary patterns whose avoidability index was left open in Cassaigne's thesis are 2-avoidable. We also prove that there exist exponentially many $\frac{7}{4}^+$-free ternary words and $\frac{7}{5}^+$-free 4-ary words. Finally we give small morphisms for binary words containing only the squares 0^2, 1^2 and $(01)^2$ and for binary words avoiding large squares and fractional repetitions.

AMS Subject Classification. — Give AMS classification codes —.

1. Introduction

We assume that the reader is familiar with Combinatorics on Words (see for instance [11]). A *pattern* is a finite word of variables. An infinite word w *avoids* pattern P if for any substitution ϕ of the variables of P with non-empty words, $\phi(P)$ is not a factor of w. Let Σ_i denote the i-letter alphabet $\{0, 1, \dots, i-1\}$. The *avoidability index* $\mu(P)$ of P is the smallest integer k such that there exists an infinite word w over Σ_k avoiding P.

In the early 1900's, Thue [16,17] (see also [3]) showed that there exists an infinite word over a three-letter alphabet which contains no square, i.e. two consecutive occurrences of the same factor. He thus proved that $\mu(AA) = 3$. He also obtained that $\mu(AAA) = 2$.

In 1979, Zimin [18] and Bean, Ehrenfeucht, and McNulty [2] indepently considered the avoidability of patterns and gave a caracterization of the patterns P such that $\mu(P) = \infty$ (unavoidable patterns). For more informations on pattern avoidability, we refer to chapter 3 in [11].

[1] LaBRI, Université Bordeaux I
351, cours de la Libération, 33405 Talence Cedex, FRANCE
ochem@labri.fr

Baker, McNulty, and Taylor [1] found a pattern P_4 such that $\mu(P_4) = 4$ and Clark [5] found a pattern P_5 such that $\mu(P_5) = 5$. The question whether there exists a pattern P_k such that $k \le \mu(P_k) < \infty$ for every k remains open.

Let $t(n)$ be the number of words of length n satisfying a property $\mathcal{P}$. We say that there exist polynomially (resp. exponentially) many words satisfying $\mathcal{P}$ if there exists a constant $c > 1$ such that $t(n) \le n^c + c$ (resp. $t(n) \ge c^n$) for every n. A surprising result in [1] states that there exist polynomially many words avoiding P_4 over Σ_4.

Cassaigne [4] obtained the avoidability index of every binary pattern and of some ternary patterns. In Section 3, we end the determination of the avoidability index of ternary patterns.

Let $\alpha > 1$ be a rational number and let $\ell \ge 1$ be an integer. A word w is an (α, ℓ)-repetition if we can write it as $w = x^n x'$ where x' is a prefix of x, $|x| = \ell$, and $|w| = \alpha|x|$. Let β be a real number and let $n \ge 1$ be an integer. A word is (β^+, n)-free if it contains no (α, ℓ)-repetition such that $\alpha > \beta$ and $\ell \ge n$. A word is β^+-free if it is $(\beta^+, 1)$-free. The repetition threshold is the smallest real number R_k such that there exists an infinite R_k^+-free word over Σ_k. The exact value of the repetition threshold is known for $2 \le k \le 11$: $R_2 = 2$ [3, 16, 17], $R_3 = \frac{7}{4}$ [7], $R_4 = \frac{7}{5}$ [13], and $R_k = \frac{k}{k-1}$ for $5 \le k \le 11$ [12]. Dejean [7] conjectured that $R_k = \frac{k}{k-1}$ for every $k \ge 5$. In Section 4, we show that there exist exponentially many $\frac{7}{4}^+$-free ternary words and $\frac{7}{5}^+$-free 4-ary words. We use this result to prove that for every pattern P considered in Section 3, there exist exponentially many words avoiding P over $\Sigma_{\mu(P)}$.

The length of a square u^2 is $|u|$. In Section 5, we give a simple construction for four special types of binary words containing squares of bounded length only.

The main results in Section 3, 4, and 5 were obtained using the method presented in Section 2. Notice that we also used this method in [9] to obtain upper bounds on some generalized repetition thresholds.

2. The method

A morphism is q-uniform if the image of every letter has length q. A uniform morphism $h : \Sigma_s^* \to \Sigma_e^*$ is *synchronizing* if for any $a, b, c \in \Sigma_s$ and $v, w \in \Sigma_e^*$, if $h(ab) = vh(c)w$, then either $v = \varepsilon$ and $a = c$ or $w = \varepsilon$ and $b = c$.

Lemma 2.1. *Let $\alpha, \beta \in \mathbb{Q}$, $1 < \alpha < \beta < 2$ and $n \in \mathbb{N}^*$. Let $h : \Sigma_s^* \to \Sigma_e^*$ be a synchronizing q-uniform morphism (with $q \ge 1$). If $h(w)$ is (β^+, n)-free for every α^+-free word w such that $|w| < \max\left(\frac{2\beta}{\beta-\alpha}, \frac{2(q-1)(2\beta-1)}{q(\beta-1)}\right)$, then $h(t)$ is (β^+, n)-free for every (finite or infinite) α^+-free word t.*

Proof. Suppose w is an α^+-free word such that $h(w)$ is not (β^+, n)-free and w is of minimum length with this property. Thus $h(w)$ contains a β^+-repetition, that is, a factor uvu such that $\frac{|uvu|}{|uv|} > \beta$. Denote $x = |u|$ and $y = |v|$. Since

$\frac{|uvu|}{|uv|} = \frac{2x+y}{x+y} > \beta$, we have $y < \frac{2-\beta}{\beta-1}x$. If $x \geq 2q-1$, then each occurrence of u contains at least one full h-image of a letter. As h is synchronizing, the two occurrences of u in uvu contain the same h-images and in the same positions. Let U be the factor of w that contains all letters whose h-images are contained in u, and let V be the factor of w that contains all letters whose h-images intersect v. Denoting $X = |U|$ and $Y = |V|$, we have $Yq < y + 2q$ and $Xq > x - 2q$, or equivalently $x < (X+2)q$. Since UVU is a factor of w and w is α^+-free, then $\frac{2X+Y}{X+Y} \leq \alpha$, which gives $X \leq \frac{\alpha-1}{2-\alpha}Y$. Now we have

$$Yq < y + 2q < \frac{2-\beta}{\beta-1}x + 2q < \frac{2-\beta}{\beta-1}(X+2)q + 2q \leq \frac{2-\beta}{\beta-1}\left(\frac{\alpha-1}{2-\alpha}Y + 2\right)q + 2q,$$

implying that $Y < \frac{2(2-\alpha)}{\beta-\alpha}$. By the minimality of w we get

$$|w| \leq 2 + Y + 2X \leq 2 + Y\left(1 + 2\frac{\alpha-1}{2-\alpha}\right) < 2 + \frac{2(2-\alpha)}{\beta-\alpha}\frac{\alpha}{2-\alpha} = \frac{2\beta}{\beta-\alpha}.$$

If $x \leq 2q-2$, then $y < \frac{2-\beta}{\beta-1}(2q-2)$ and thus $2x+y < \frac{2\beta}{\beta-1}(q-1)$. The minimality of w implies that $(|w|-2)q \leq |uvu| - 2 = 2x + y - 2$. By the above we get that $|w| < \frac{2(q-1)(2\beta-1)}{q(\beta-1)}$, which completes the proof. ■

To obtain the results in Section 3, 4, and 5 we need to construct infinite words over Σ_e that satisfies a property $\mathcal{P}$ consisting of some (β^+, n)-freeness properties and the avoidance of a set S of finite words. Such an infinite word is obtained as the h-image of any infinite α^+-free word $t \in \Sigma_s^*$ by a synchronizing morphism $h : \Sigma_s^* \to \Sigma_e^*$ such that for every α^+-free word $t \in \Sigma_s^*$, $h(t)$ satisfies a property $\mathcal{P}$, where $\mathcal{P}$ consists of some (β^+, n)-freeness properties and the avoidance of a set S of finite words. It is easy to check that h is synchronizing and that $h(t)$ avoids the set S. Using Lemma 2.1 with $\alpha \geq R_s$ allows us to check that the h-image of any (infinite) α^+-free word over Σ_s is (β^+, n)-free with a finite amount of computation.

We fix $s, q \in \mathbb{N}$ and $\alpha \in \mathbb{Q}$ such that $s \leq 11$ and $R_s \leq \alpha < 2$, to ensure that there exist an infinite α^+-free word over Σ_s. We use depth-first search to find a word w over Σ_e of size $s \times q$ which defines the q-uniform morphism h by posing $w = h(0)h(1)\ldots h(s-1)$. Obviously, we can restrict the search to words satisfying $h(s-1) \prec \ldots \prec h(1) \prec h(0)$, where $\prec$ is the lexicographic order of Σ_e^q. We prune the search tree by checking property $\mathcal{P}$ on the prefixes of a potential w. If no morphism is found, we increase the value of q and try again [1].

3. Pattern avoidance

We consider here the ternary patterns whose 2-avoidability was left open in Cassaigne's thesis [4]. One of them, namely $ABCBABC$, is shown to be 2-avoidable

[1]The C++ sources of the programs used to find and check the morphisms discussed in this paper are available at: `http://dept-info.labri.fr/~ochem/morphisms/`

in [9]. We add to that list the binary pattern $AABBA$ (resp. $ABAAB$) which was already known to be 2-avoidable, and is here 2-avoided together with its reverse $ABBAA$ (resp. $ABBAB$). In particular, the 2-avoidability of $ABCACB$ was one of Currie's open problems [6], which was mentioned mainly because $ABCACB$ and its reverse are not simultaneously 2-avoidable.

Lemma 3.1. *Any factor uvu of a (β^+, n)-free word w, with $\beta < 2$, is such that*

$$|u| \leq \max\left(n - 1 - |v|, \left\lfloor \frac{\beta - 1}{2 - \beta}|v| \right\rfloor\right).$$

Proof. If $|uv| < n$ then $|u| \leq n - 1 - |v|$ and we are done, so suppose $|uv| \geq n$. Since w is (β^+, n)-free, we have $\frac{|uvu|}{|uv|} \leq \beta \Longrightarrow |u| \leq \frac{\beta-1}{2-\beta}|v|$. ■

Theorem 3.2. *The ternary patterns listed in Table 1 are 2-avoidable.*

Proof. Suppose that we are given a synchronizing morphism $h : \Sigma_s^* \to \Sigma_e^*$ and a pattern P over the alphabet $\{A, B, \ldots\}$. We can try to prove that $h(t)$ avoids P for every α^+-free word $t \in \Sigma_s^*$ in three steps:

(1) Check useful (β^+, n)-freeness properties of $h(t)$ thanks to Lemma 2.1.
(2) Consider a potential occurrence $\phi(P)$ of P, where ϕ is a non-erasing morphism. Then use Lemma 3.1 and the results of step (1) to obtain upper bounds on the quantities $a = |\phi(A)|,\ b = |\phi(B)|,\ \ldots$.
(3) Use the bounds of step (2) to exhaustively check by computer that no occurrence of P appears in $h(t)$.

Let P^R denote the reverse of the pattern P. We have $\mu(P^R) = \mu(P)$ since a word w avoids P if and only the reverse of w avoids P^R. Notice that the bounds obtained in step (2) using Lemma 3.1 that hold for potential occurrences of a pattern P also hold for those of P^R. Thus we try, when possible, to avoid simultaneously P and P^R. The method discussed in Section 2 is thus used so that the set S contains the small occurrences of P (and maybe P^R). Each line of Table 1 contains one of these pattern P and informations about the morphism h we used to show that $\mu(P) = 2$: the q-uniform morphism $h : \Sigma_s^* \to \Sigma_2^*$ is such that for every α^+-free word $t \in \Sigma_s^*$, $h(t)$ avoids P. We also precise whether such a word $h(t)$ also avoids P^R. Now, for each pattern, we give the bounds obtained in step (2) of the proof and the morphism found with the method in Section 2.

The following 8-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $AABAACBAAB$.
The word $h(t)$ is $\left(\frac{24}{13}^+, 3\right)$-free. It contains no square of length at least three, and since the square AA occurs in the pattern, we have that $a \leq 2$. By Lemma 3.1, the factor $BAAB$ implies $b \leq 11a$. For each occurrence of $BAAB$ appearing in $h(t)$, we have checked that the corresponding occurrence of $AABAA$ does not appear. For instance, the occurrence $\phi(BAAB) = 0110$ (where $\phi(A) = 1,\ \phi(B) = 0$)

Pattern	s	α	q	Comments
AABAACBAAB	3	7/4	8	self-reverse
AABACCB	3	7/4	24	avoided with its reverse
AABBA	3	7/4	21	avoided with its reverse
AABBCABBA	3	7/4	102	unavoidable with its reverse
AABBCAC	3	7/4	86	avoided with its reverse
AABBCBC	3	7/4	34	avoided with its reverse
AABBCC	3	7/4	52	self-reverse
AABCBC	3	7/4	46	avoided with its reverse
AABCCAB	3	7/4	34	avoided with its reverse
AABCCBA	3	7/4	56	avoided with its reverse
ABAAB	3	7/4	10	avoided with its reverse
ABAACBC	4	7/5	17	avoided with its reverse
ABAACCB	3	7/4	74	avoided with its reverse
ABACACB	3	7/4	12	avoided with its reverse
ABACBC	4	7/5	29	self-reverse
ABACCAB	3	7/4	19	avoided with its reverse
ABACCBA	3	7/4	14	avoided with its reverse
ABBACCA	3	7/4	12	self-reverse
ABBACCB	3	7/4	42	avoided with its reverse
ABBCACB	4	7/5	16	avoided with its reverse
ABBCBAC	4	7/5	14	avoided with its reverse
ABBCBCA	3	7/4	22	avoided with its reverse
ABBCCAB	3	7/4	20	avoided with its reverse
ABCAACB	3	7/4	24	avoided with its reverse
ABCACAB	3	7/4	10	avoided with its reverse
ABCACB	6	5/4	810	unavoidable with its reverse
ABCBBAC	4	7/5	18	avoided with its reverse

Table 1. Table of 2-avoidable ternary patterns.

appears in $h(t)$, but the factor $\phi(AABAA) = 11011$ does not.

$$\begin{aligned} 0 &\mapsto 01101011 \\ 1 &\mapsto 00111010 \\ 2 &\mapsto 00101110 \end{aligned}$$

The following 24-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $AABACCB$ and its reverse.
The word $h(t)$ is $\left(\frac{31}{16}^+, 3\right)$-free, so $a \leq 2$ and $c \leq 2$. The factor $BACCB$ implies

$b \leq 15(a + 2c)$.

$$
\begin{aligned}
0 &\mapsto 000101001101011001010111 \\
1 &\mapsto 000101000111010111001011 \\
2 &\mapsto 000101000110100111001011
\end{aligned}
$$

The following 21-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $AABBA$ and its reverse.
The word $h(t)$ is $\left(\frac{33}{17}^+, 4\right)$-free, so $a \leq 3$ and $b \leq 3$.

$$
\begin{aligned}
0 &\mapsto 001001011100011101101 \\
1 &\mapsto 001000111000101101101 \\
2 &\mapsto 000111000100101101101
\end{aligned}
$$

The following 102-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $AABBCABBA$.
The word $h(t)$ is $\left(\frac{31}{16}^+, 27\right)$-free, so $a \leq 26$ and $b \leq 26$. For each occurrence of $AABB$ we have checked that the corresponding occurrence of $ABBA$ does not appear. Notice that the k-avoidability of $AABBCABBA$ implies the k-avoidability of $AABBA$. A simple backtracting algorithm shows that $AABBCABBA$ and $ABBAA$ (i.e. the reverse of $AABBA$) are not simultaneously 2-avoidable, so that the two previous results are tight, in a way.

$$
\begin{aligned}
0 &\mapsto 0001000101101110111000101100010001011011101100010110111000101101 \\
&\qquad 11011100010110111011000101101110001011 \\
1 &\mapsto 0001000101101110110001011011101110001011000100010110111000101101 \\
&\qquad 11011100010110111011000101101110001011 \\
2 &\mapsto 0001000101101110001011011101110001011000100010110111011000101101 \\
&\qquad 11011100010110111011000101101110001011
\end{aligned}
$$

The following 86-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $AABBCAC$ and its reverse.
The word $h(t)$ is $\left(\frac{43}{22}^+, 3\right)$-free, so $a \leq 2$ and $b \leq 2$.
The factor CAC implies $c \leq \max(2 - a, 21a) = 21a$.

$$
\begin{aligned}
0 &\mapsto 00010101100101001101010110010101001101011001010011010 \\
&\qquad 100010111010100110101100101011101 \\
1 &\mapsto 00010101100101001101010001011101010011010110010101001 \\
&\qquad 101010110010100110101100101011101 \\
2 &\mapsto 00010101100101001101010001011101010011010110010100110 \\
&\qquad 101011001010100110101100101011101
\end{aligned}
$$

The following 34-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $AABBCBC$ and its reverse.

The word $h(t)$ is $\left(\frac{33}{17}^+, 3\right)$-free, so $a \le 2$ and $b = c = 1$.

$$
\begin{aligned}
0 &\mapsto 111010111000101000111000101010001 1\\
1 &\mapsto 111010111000101010001110001010001 1\\
2 &\mapsto 111010111000111010100011100010100 0
\end{aligned}
$$

The following 52-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $AABBCC$.
The word $h(t)$ is $\left(\frac{59}{30}^+, 3\right)$-free, so $a \le 2$, $b \le 2$, and $c \le 2$.

$$
\begin{aligned}
0 &\mapsto 1101001110001101000101100011010011100101100011100101\\
1 &\mapsto 1101001110001101000101100011100101100011010011100101\\
2 &\mapsto 1101001110001101000101100011100101110100111001011000
\end{aligned}
$$

The following 46-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $AABCBC$ and its reverse.
The word $h(t)$ is $\left(\frac{367}{184}^+, 3\right)$-free, so $a \le 2$ and $b = c = 1$. Notice that the only occurrences of $AABB$ in $h(t)$ are 0011 and 1100.

$$
\begin{aligned}
0 &\mapsto 0011010011100011001011000110100110001110010110\\
1 &\mapsto 0011010011100011001011000110100111001011000111\\
2 &\mapsto 0011010011100011001011000111001101001110010110
\end{aligned}
$$

The following 34-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $AABCCAB$ and its reverse.
The word $h(t)$ is $\left(\frac{257}{136}^+, 4\right)$-free, so $a \le 3$ and $c \le 3$. The factor $ABCCAB$ implies $a + b \le \left\lfloor \frac{242}{15} c \right\rfloor$, thus $b \le \left\lfloor \frac{242}{15} c \right\rfloor - a$.

$$
\begin{aligned}
0 &\mapsto 0000101111110100001111110100010111 1\\
1 &\mapsto 0000101110100001111010001011111100\\
2 &\mapsto 0000101110100000011110100010111111
\end{aligned}
$$

The following 56-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $AABCCBA$ and its reverse.
The word $h(t)$ is $\left(\frac{36}{19}^+, 3\right)$-free, so $a \le 2$ and $c \le 2$. The factor $BCCB$ implies $b \le 17c$.

$$
\begin{aligned}
0 &\mapsto 00010110001110010111010011100011010001011000111001011101\\
1 &\mapsto 00010110001101001110010110001101000101110100111001011101\\
2 &\mapsto 00010110001101000101110100111001011000110100111001011101
\end{aligned}
$$

The following 10-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABAAB$ and its reverse.
The word $h(t)$ is $\left(\frac{39}{20}^+, 3\right)$-free, so $a \leq 2$. The factor $BAAB$ implies $b \leq 38a$.

$$\begin{aligned} 0 &\mapsto 0001110101 \\ 1 &\mapsto 0000111101 \\ 2 &\mapsto 0000101111 \end{aligned}$$

The following 17-uniform morphism h is such that for any $\frac{7}{5}^+$-free word $t \in \Sigma_4^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABAACBC$ and its reverse.
The word $h(t)$ is $\left(\frac{13}{7}^+, 7\right)$-free, so $a \leq 6$. The word $h(t)$ is $\left(\frac{23}{16}^+, 20\right)$-free. Suppose $b + c \geq 20$, then the factor CBC implies $c \leq \left\lfloor \frac{7}{9} b \right\rfloor$ and the factor $BAACB$ implies $b \leq \left\lfloor \frac{7}{9}(2a + c) \right\rfloor$. From these relations we can deduce $b \leq 22$ and $c \leq 17$. Thus we have $b \leq 22$ and $c \leq 18$.

$$\begin{aligned} 0 &\mapsto 01110000110000111 \\ 1 &\mapsto 01011100110011101 \\ 2 &\mapsto 01000110011000101 \\ 3 &\mapsto 00011110011110001 \end{aligned}$$

The following 74-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABAACCB$ and its reverse.
The word $h(t)$ is $\left(\frac{193}{104}^+, 3\right)$-free, so $a \leq 2$ and $c \leq 2$. The factor $BAACCB$ implies $b \leq \left\lfloor \frac{89}{15}(a + c) \right\rfloor$.

$$\begin{aligned} 0 &\mapsto 000001011001110100110101100101110011010000011111010011000 \\ &\qquad 10100011001011111 \\ 1 &\mapsto 000001011001010100110001011001010011010000011111010011000 \\ &\qquad 10100011001011111 \\ 2 &\mapsto 000001011001010011010001100101010011010000011111010011000 \\ &\qquad 10100011001011111 \end{aligned}$$

The following 12-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABACACB$ and its reverse.
The word $h(t)$ is $\left(\frac{15}{8}^+, 3\right)$-free, so $a = c = 1$. The factor $BACACB$ implies $b \leq 14(a + c) = 28$.

$$0 \mapsto 001010011111 \qquad 1 \mapsto 000110100111 \qquad 2 \mapsto 000001101011$$

The following 29-uniform morphism h is such that for any $\frac{7}{5}^+$-free word $t \in \Sigma_4^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABACBC$.

The word $h(t)$ is $\left(\frac{41}{29}^+, 291\right)$-free. Suppose $a+b \geq 291$ and $b+c \geq 291$. The factors ABA, CBC, and $BACB$ respectively imply that $17a \leq 12b$ (i), $17c \leq 12b$ (ii), and $17b \leq 12(a+c)$ (iii). The combination $17 \times (i) + 17 \times (ii) + 24 \times (iii)$ gives $a + c \leq 0$, a contradiction. So we can suppose without loss of generality that $b + c \leq 290$ (iv). If $291 \leq a + b$ (v) then (i) and (iii) still hold and the combination $2324 \times (i) + 1649 \times (iii) + 19788 \times (iv) + 19720 \times (v)$ gives $213b \leq 0$. This contradiction shows that $a + b \leq 290$ and $b + c \leq 290$.

$$
\begin{aligned}
0 &\mapsto 000110101100000111100101001110\\
1 &\mapsto 000110101100000011100101001111\\
2 &\mapsto 000011010110001111011110011110\\
3 &\mapsto 000011010110000011110011101111
\end{aligned}
$$

The following 19-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABACCAB$ and its reverse.
The word $h(t)$ is $\left(\frac{35}{19}^+, 5\right)$-free, so $c \leq 4$. The factor $ACCA$ implies $a \leq \max\left(4 - 2c, \left\lfloor \frac{32}{3}c \right\rfloor\right) = \left\lfloor \frac{32}{3}c \right\rfloor$. The factor $ABACCAB$ implies $a + b \leq \left\lfloor \frac{16}{3}(a + 2c) \right\rfloor$, thus $b \leq \left\lfloor \frac{13a+32c}{3} \right\rfloor$.

$$
\begin{aligned}
0 &\mapsto 0101110010100000111\\
1 &\mapsto 0101100000010011100\\
2 &\mapsto 0100010111010100011
\end{aligned}
$$

The following 14-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABACCBA$ and its reverse.
If w is an occurrence of $ABACCBA$ such that $a > 1$, then the suffix of w of size $|w| - a + 1$ is a smaller occurrence of $ABACCBA$ such that $a = 1$. So we assume without loss of generality that $a = 1$. The word $h(t)$ is $\left(\frac{23}{12}^+, 3\right)$-free, so $c \leq 2$. The factor $BACCBA$ implies $a + b \leq 22c$, thus $b \leq 22c - 1$.

$$
\begin{aligned}
0 &\mapsto 10101100001110\\
1 &\mapsto 01010111100011\\
2 &\mapsto 01010000111110
\end{aligned}
$$

The following 12-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABBACCA$.
The word $h(t)$ is $\left(\frac{31}{16}^+, 4\right)$-free, so $b \leq 3$ and $c \leq 3$. The factor $ABBA$ implies $a \leq \max(3 - 2b, 30b) = 30b$.

$$
\begin{aligned}
0 &\mapsto 000111001011\\
1 &\mapsto 000101111010\\
2 &\mapsto 000100111011
\end{aligned}
$$

The following 42-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABBACCB$ and its reverse.
The word $h(t)$ is $\left(\frac{53}{28}^+, 3\right)$-free, so $b \le 2$ and $c \le 2$. The factor $ABBA$ implies $a \le \left\lfloor \frac{50}{3} b \right\rfloor$.

$$
\begin{aligned}
0 &\mapsto 000010111110001011101000111101000011101011\\
1 &\mapsto 000010111110001010110010100110101100101011\\
2 &\mapsto 000010111110001010001111010000111010001011
\end{aligned}
$$

The following 16-uniform morphism h is such that for any $\frac{7}{5}^+$-free word $t \in \Sigma_4^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABBCACB$ and its reverse.
The word $h(t)$ is $\left(\frac{9}{5}^+, 4\right)$-free, so $b \le 3$. The word $h(t)$ is $\left(\frac{233}{160}^+, 49\right)$-free. Suppose $a + c \ge 49$. The factors $ABBCA$ and CAC respectively imply that $a \le \left\lfloor \frac{73}{87}(2b + c) \right\rfloor$ and $c \le \left\lfloor \frac{73}{87} a \right\rfloor$. From these relations we can deduce $a \le 15$ and $c \le 12$. This contradiction shows that $a + c \le 48$.

$$
\begin{aligned}
0 &\mapsto 0010000001101111\\
1 &\mapsto 0000111010001111\\
2 &\mapsto 0000100111111011\\
3 &\mapsto 0000001001101011
\end{aligned}
$$

The following 14-uniform morphism h is such that for any $\frac{7}{5}^+$-free word $t \in \Sigma_4^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABBCBAC$ and its reverse.
The word $h(t)$ is $\left(\frac{18}{11}^+, 7\right)$-free, so $b \le 6$. The word $h(t)$ is $\left(\frac{29}{20}^+, 43\right)$-free. Suppose $a + b + c \ge 43$. The factors $ABBCBA$ and $CBAC$ respectively imply that $a \le \left\lfloor \frac{9}{11}(3b + c) \right\rfloor$ and $c \le \left\lfloor \frac{9}{11}(a + b) \right\rfloor$. From these relations we can deduce $a \le 54$ and $c \le 49$.

$$
\begin{aligned}
0 &\mapsto 00101010101011\\
1 &\mapsto 00010001110111\\
2 &\mapsto 00000101011111\\
3 &\mapsto 00000010111111
\end{aligned}
$$

The following 22-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABBCBCA$ and its reverse.
The word $h(t)$ is $\left(\frac{173}{88}^+, 3\right)$-free, so $b = c = 1$. The factor $ABBCBCA$ implies $a \le \left\lfloor \frac{85}{3}(3b + 2c) \right\rfloor = 141$.

$$
\begin{aligned}
0 &\mapsto 0001101011001010100111\\
1 &\mapsto 0001101010110010100111\\
2 &\mapsto 0001101010011100101011
\end{aligned}
$$

The following 20-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABBCCAB$ and its reverse.

The word $h(t)$ is $\left(\frac{15}{8}^+, 4\right)$-free, so $b \leq 3$ and $c \leq 3$. The factor $ABBCCAB$ implies $a + b \leq 7(b + 2c)$, thus $a \leq 6b + 14c$.

$$\begin{aligned} 0 &\mapsto 00010100100101011111 \\ 1 &\mapsto 00010010001110110111 \\ 2 &\mapsto 00000101011011010111 \end{aligned}$$

The following 24-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABCAACB$ and its reverse.
The word $h(t)$ is $\left(\frac{187}{96}^+, 4\right)$-free, so $a \leq 3$. The word $h(t)$ is $\left(\frac{355}{192}^+, 97\right)$-free. The factor $CAAC$ implies $c \leq \max\left(96 - 2a, \left\lfloor \frac{326a}{29} \right\rfloor\right) = 96 - 2a$. The factor $BCAACB$ implies $b \leq \max\left(96 - 2a - 2c, \left\lfloor \frac{326(a+c)}{29} \right\rfloor\right)$.

$$\begin{aligned} 0 &\mapsto 000001011111001000110111 \\ 1 &\mapsto 000001011111000100111011 \\ 2 &\mapsto 000001010111110010011011 \end{aligned}$$

The following 10-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABCACAB$ and its reverse.
The word $h(t)$ is $\left(\frac{79}{40}^+, 3\right)$-free, so $a = c = 1$. The word $h(t)$ is $\left(\frac{149}{80}^+, 41\right)$-free. The factor $ABCACAB$ implies $a + b \leq \max\left(40 - a - 2c, \left\lfloor \frac{69}{11}(a + 2c) \right\rfloor\right)$, thus $b \leq 40 - 2a - 2c = 36$.

$$0 \mapsto 0001110101 \qquad 1 \mapsto 0001011101 \qquad 2 \mapsto 0001010111$$

The 810-uniform morphism $h = m_{4,2} \circ m_{6,4}$ is such that for any $\frac{5}{4}^+$-free word $t \in \Sigma_6^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABCACB$.
The word $h(t)$ is $\left(\frac{1073}{810}^+, 3241\right)$-free. Suppose $a + c \geq 3241$. The factors $ABCA$, $BCACB$, and CAC respectively imply that $547a \leq 263(b + c)$ (i), $547b \leq 263(a + 2c)$ (ii), and $547c \leq 263a$ (iii). The combination $2 \times (i) + (ii) + 2 \times (iii)$ gives $305a + 568b + 42c \leq 0$. This contradiction shows that $a + c \leq 3240$ (iv). The word $h(t)$ is $\left(\frac{29}{14}^+, 4\right)$-free, so the factor CAC implies $c \leq 14a$ (v). Suppose now $3238 \leq b$ (vi), so that $a + b + 2c \geq 3241$ and (ii) still holds. The combination $15 \times (ii) + 7627 \times (iv) + 263 \times (v) + 7632 \times (vi)$ gives $573b + 936 \leq 0$. This contradiction shows that $b \leq 3237$.

The 135-uniform morphism $m_{4,2}$ is given by:

$0 \mapsto$ 00000100111101100010111100000101101000011111011000010011110000010
11010000111110100101111100001001111101100000100111010000111110100101111
$1 \mapsto$ 00000100111101100010111100000101101000011111011000010011110000010
11010000111011000000100111110110001011110000010110100001111101001011111
$2 \mapsto$ 00000100111101100001111101001011110000100111110110000010011101000
01111101001011110000010110100001111101100001001110100001111101001011111
$3 \mapsto$ 00000100111101100001111101001011110000010110100001111011000001001
11110110001011110000010110100001111101100001001110100001111101001011111

The 6-uniform morphism $m_{6,4}$ is given by:

$0 \mapsto 032131$	$1 \mapsto 031232$	$2 \mapsto 023121$
$3 \mapsto 021323$	$4 \mapsto 013212$	$5 \mapsto 012313$

The following 18-uniform morphism h is such that for any $\frac{7}{5}^+$-free word $t \in \Sigma_4^*$, $h(t) \in \Sigma_2^*$ avoids the pattern $ABCBBAC$ and its reverse.
The word $h(t)$ is $\left(\frac{8}{5}^+, 7\right)$-free, so $b \leq 6$. The word $h(t)$ is $\left(\frac{527}{378}^+, 181\right)$-free. Suppose $a + 2b + c \geq 181$. The factors $CBBAC$ and $ABCBBA$ respectively imply that $c \leq \left\lfloor \frac{149}{229}(a + 2b) \right\rfloor$ and $a \leq \left\lfloor \frac{149}{229}(3b + c) \right\rfloor$. From these relations we can deduce $a \leq 28$ and $c \leq 26$. This contradiction shows that $a + 2b + c \leq 180$.

$$
\begin{aligned}
0 &\mapsto 010000011011011011 \\
1 &\mapsto 001001001001111101 \\
2 &\mapsto 001000000111111011 \\
3 &\mapsto 000000101010111111
\end{aligned}
$$

■

4. Application to repetition-free words

The repetition threshold for binary words is 2, and this result is tight in the following senses:

(1) There exist polynomially many 2^+-free binary words.
(2) There exist arbitrarily large squares in any infinite 2^+-free binary word.

In this section we show that no similar situation occurs for ternary and 4-ary words. We use the following easy lemma, which is already implicitly used in [10].

Lemma 4.1. *There are at least $2^{\lceil \frac{n}{k} \rceil}$ R_k^+-free words of length n over Σ_{k+1}.*

Proof. Consider an R_k^+-free word w in Σ_k^n. At least one letter in Σ_k, say 0, occurs at least $\lceil \frac{n}{k} \rceil$ times in w. The letter k belongs to Σ_{k+1} but does not belong to Σ_k. Notice that replacing zero or more occurrences of 0 by an occurrence of k in w produces an R_k^+-free word of length n over Σ_{k+1}, and that we can obtain at least $2^{\lceil \frac{n}{k} \rceil}$ such words. ■

Theorem 4.2.

(1) *There exist exponentially many $\frac{7}{4}^+$-free ternary words with no large repetition of exponent $\frac{7}{4}$.*

(2) *There exist exponentially many $\frac{7}{5}^+$-free 4-ary words with no large repetition of exponent $\frac{7}{5}$.*

Proof. By Lemma 4.1, there exist exponentially many $\frac{7}{5}^+$-free words over Σ_5 and exponentially many $\frac{5}{4}^+$-free words over Σ_6

The following 59-uniform morphism h is such that for any $\frac{7}{5}^+$-free word $t \in \Sigma_5^*$, $h(t) \in \Sigma_3^*$ is $\frac{7}{4}^+$-free and $\left(\frac{3}{2}^+, 10\right)$-free.

$0 \mapsto 01020121021201210120102101202102012102120210120102101210212$
$1 \mapsto 01020120210201210212012101201021202102012021201210201021012$
$2 \mapsto 01020120210201210120102101202120121021202102012021201021012$
$3 \mapsto 01020120210201021201210201202101201021012102120210201021012$
$4 \mapsto 01020120210121020102120121012021020120212012102010212021012$

The following 132-uniform morphism h is such that for any $\frac{5}{4}^+$-free word $t \in \Sigma_6^*$, $h(t) \in \Sigma_4^*$ is $\frac{7}{5}^+$-free and $\left(\frac{61}{44}^+, 11\right)$-free.

$0 \mapsto 012031021320123021032013023120321301231021320312301321031201302$1
03201231021301203210231201321031230132031021301231032013023120321023
$1 \mapsto 012031021320123021032013023103213012310213203123013210231201302$1
03201230213203102301203210231201321031230213201231032130231201321023
$2 \mapsto 012031021320123021032013023103213012031023013210312013023103201$2
30213203102301203210231201321031230132031021301231032130231201321023
$3 \mapsto 012031021320123021031201321023120321302310320123021320310230132$1
03120130210320123102132031230132102312032130231032013021031201321023
$4 \mapsto 012031021320123021031201302310321301231021320312301321031201302$1
03201230213203102301203213023120132103123021320123103213023120321023
$5 \mapsto 012031021320123021031201302310321301203102301321031201302103201$2
30213203102301203210231201302103123013203102130123103213023120321023

■

Corollary 4.3. *For each pattern P listed in Table 1, there exist exponentially many words avoiding P over Σ_2.*

Proof. Binary words avoiding $ABCACB$ are constructed from $\frac{5}{4}^+$-free words over Σ_6, and there are exponentially many such words by Lemma 4.1. For each other pattern P listed in Table 1, binary words avoiding P are constructed from either $\frac{7}{4}^+$-free words over Σ_3, or $\frac{7}{5}^+$-free words over Σ_4. In both cases, there are exponentially many such words by Theorem 4.2. ■

We have not been able to extend Theorem 4.2 to Σ_5. However, we believe that the following strong form of Dejean's conjecture holds.

Conjecture 4.4. for every $k \geq 5$, there exist exponentially many $\frac{k}{k-1}^+$-free words over Σ_k.

5. Binary words avoiding large squares

Fraenkel and Simpson constructed in [8] an infinite binary word containing only three squares. Another construction using uniform morphisms is given in [14]. Shallit [15] also gives uniform morphisms for binary words avoiding:

- squares of length at least 3 and 3^+-repetitions (10-uniform),
- squares of length at least 4 and $\frac{5}{2}^+$-repetitions (1560-uniform),
- squares of length at least 7 and $\frac{7}{3}^+$-repetitions (252-uniform).

In this section we give small $\Sigma_3^* \to \Sigma_2^*$ uniform morphisms producing words having these properties.

The following 50-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ contains only the squares in $\{0^2, 1^2, (01)^2\}$ and is $\left(\frac{37}{19}^+, 3\right)$-free.

$$\begin{aligned}
0 &\mapsto 00011001011000111001011001110001011100101100010111\\
1 &\mapsto 00011001011000101110010110011100010110001110010111\\
2 &\mapsto 00011001011000101110010110001110010111000101100111
\end{aligned}$$

The following 8-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ is 3^+-free, $\left(\frac{5}{2}^+, 2\right)$-free, and $\left(\frac{59}{32}^+, 3\right)$-free.

$$0 \mapsto 01101011 \qquad 1 \mapsto 00111010 \qquad 2 \mapsto 00101110$$

The following 103-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ is $\frac{5}{2}^+$-free, $\left(\frac{7}{3}^+, 3\right)$-free, and $\left(\frac{823}{412}^+, 4\right)$-free.

$$
\begin{aligned}
0 \mapsto\ & 0010011010010110010011011001010011010110010011011001011010011011 \\
& 001001101011001010011011001011010011011 \\
1 \mapsto\ & 0010011010010110010011011001010011010110010011011001011010011011 \\
& 001001101001011001001101011001010011011 \\
2 \mapsto\ & 0010011010010110010011011001010011010110010011010010110010011011 \\
& 001011010011011001001101011001010011011
\end{aligned}
$$

The following 30-uniform morphism h is such that for any $\frac{7}{4}^+$-free word $t \in \Sigma_3^*$, $h(t) \in \Sigma_2^*$ is $\frac{7}{3}^+$-free and $\left(\frac{79}{40}^+, 7\right)$-free.

$$
\begin{aligned}
0 &\mapsto 001011001011010011011001001101 \\
1 &\mapsto 001011001011010011001011001101 \\
2 &\mapsto 001011001001101100101101001101
\end{aligned}
$$

References

[1] K.A. Baker, G.F. McNulty, and W. Taylor, Growth Problems for Avoidable Words, *Theoret. Comput. Sci.* **69** (1989) 319–345.

[2] D.R. Bean, A. Ehrenfeucht, and G.F. McNulty, Avoidable Patterns in Strings of Symbols, *Pacific J. of Math.* **85** (1979 261–294.

[3] J. Berstel, *Axel Thue's Papers on Repetitions in Words: a Translation.* Number 20 in Publications du Laboratoire de Combinatoire et d'Informatique Mathématique. Université du Québec à Montréal, February 1995.

[4] J. Cassaigne, Motifs évitables et régularité dans les mots, Thèse de Doctorat, Université Paris VI, Juillet 1994.

[5] R.J. Clark. Avoidable formulas in combinatorics on words, Ph.D. Thesis, University of California, Los Angeles (2001).

[6] J. D. Currie, Open problems in pattern avoidance, *Amer. Math. Monthly* **100** (1993) 790–793.

[7] F. Dejean. Sur un théorème de Thue, *J. Combin. Theory. Ser. A* **13** (1972) 90–99.

[8] A. S. Fraenkel and R. J. Simpson. How many squares must a binary sequence contain? *Electronic J. Combinatorics* **2** (1995), #R2.

[9] L. Ilie, P. Ochem, and J.O. Shallit. A generalization of Repetition Threshold, *Theoret. Comput. Sci.* To appear.

[10] J. Karhumäki and J. O. Shallit. Polynomial versus exponential growth in repetition-free binary words, *J. Combinatorial Theory Ser. A* **105** (2004) 335–347.

[11] M. Lothaire. *Algebraic Combinatorics on Words.* Cambridge Univ. Press, 2002.

[12] J. Moulin-Ollagnier. Proof of Dejean's conjecture for alphabets with 5,6,7,8,9,10 and 11 letters, *Theoret. Comput. Sci.* **95** (1992) 187–205.

[13] J.-J. Pansiot. A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, *Disc. Appl. Math.* **7** (1984) 297–311.

[14] N. Rampersad, J.O. Shallit, and M.-w. Wang. Avoiding large squares in infinite binary words. *Theoret. Comput. Sci.* **339** (2005), 19–34.

[15] J.O. Shallit. Simultaneous avoidance of large squares and fractional powers in infinite binary words, *Internat. J. Found. Comput. Sci.* **15** (2004) 317–327.

[16] A. Thue. Über unendliche Zeichenreihen, *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **7** (1906), 1–22. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 139–158.

[17] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **1** (1912), 1–67. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 413–478.

[18] A.I. Zimin. Blocking sets of terms, *Math. USSR Sbornik* **47(2)** (1984) 353–364. English translation.

Communicated by (The editor will be set by the publisher).
(The dates will be set by the publisher).