Available online at www.sciencedirect.com
SCIENCE DIRECT®
Theoretical Computer Science 340 (2005) 280–290
Theoretical Computer Science
www.elsevier.com/locate/tcs

Growth of repetition-free words—a review

Jean Berstel

Institut Gaspard-Monge (IGM), Université de Marne-la-Vallée, 5 Boulevard Descartes, F-77454 Marne-la-Vallée, Cédex 2, France

Abstract

This survey reviews recent results on repetitions in words, with emphasis on the estimations for the number of repetition-free words.

Keywords: Repetitions in words; Square-free words; Overlap-free words; Combinatorics on words

1. Introduction

A *repetition* is any bordered word. Quite recently, several new contributions were made to the field of repetition-free words, and to counting repetition-free words. The aim of this survey is to give a brief account of some of the methods and results.

The terminology deserves some comments. Let $\alpha > 1$ be a rational number. A nonempty word w is an α-*power* if there exist words x, x' with x' a prefix of x and an integer n, such that $w = x^n x'$ and $\alpha = n + |x'|/|x| = |w|/|x|$. For example, the French word entente is a $\frac{7}{3}$-power, and the English word outshout is a $\frac{8}{5}$-power. If $\alpha = 2$ or 3, we speak about a square and a cube, like for murmur or kokoko (the examples are taken from [41]). A word w is an *overlap* if it is a α-power for some $\alpha > 2$. For instance, entente is an overlap.

Let $\beta > 1$ be a real number. A word w is said to *avoid* β-powers or is β-*power-free* if it contains no factor that is an α-power for $\alpha \geqslant \beta$. A word w is β^+-*power-free* if it contains no factor that is an α-power for $\alpha > \beta$. Thus, a word is overlap-free if and only if it is 2^+-power-free.

E-mail address: jean.berstel@univ-mlv.fr.

0304-3975/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.tcs.2005.03.039

This review reports results on the growth of the number of α-free words of length n over an q-letter alphabet. In some cases, growth is bounded by a polynomial in n, in other cases, it is shown to be exponential in n. We consider overlap-free words in the next section, square-free words in Section 3 and some generalizations in the final section. For basics and complements, the reader should consult the book of Allouche and Shallit [3].

2. Counting overlap-free words

We first review estimations for the number of overlap-free words over a binary alphabet. Let V be the set of binary overlap-free words and let $v(n)$ be the number of overlap-free binary words of length n. This sequence starts with 2, 4, 6, 10, 14, 20 (Sloane's sequence A007777, see [42]).

It is clear that V is factorial (factor-closed). It follows that, as for any factorial set, one has $v(n+m) \leqslant v(n)v(m)$. Thus the sequence $(v(n))$ is *submultiplicative* or the sequence $(\log v(n))$ is *subadditive*. This in turn implies, by a well-known argument, that the sequence $\lim_{n\to\infty} v(n)^{1/n}$ has a limit, or equivalently, that the limit

$$\lambda(V) = \lim_{n\to\infty} \frac{1}{n} \log v(n)$$

exists. The number $\lambda(V)$ is called the (topological) *entropy* of the set V. For a general discussion about entropy of square-free words, see [4]. The entropy of the set of square-free words is strictly positive, as we will see later. On the contrary, the entropy of the set of overlap-free words is zero. This is a consequence of the following result of Restivo and Salemi [34,35].

Theorem 1. *The number $v(n)$ of binary overlap-free words of length n is bounded from above by a polynomial in n.*

They proved that $v(n)$ is bounded by n^4. The proof is based on the following structural property of overlap-free words which we state in the more general setting of [22]. Recall first that the Thue–Morse morphism is defined by

$$\mu : \begin{array}{ccc} 0 & \mapsto & 01 \\ 1 & \mapsto & 10 \end{array}$$

Lemma 2. *Let $2 < \alpha < 7/3$, and let x be a word that avoids α-powers. There exist words u, y, v with $u, v \in \{\mathrm{e}, 0, 1, 00, 11\}$ and y avoiding α-powers such that $x = u\mu(y)v$. This factorization is unique if $|x| \geqslant 7$.*

First, observe that the lemma does not hold for $\alpha \geqslant 7/3$ since $x = 0110110$ is a $7/3$-power and has no factorization of the required form. Next, consider as an example the word $x = 011001100$ which is a $9/4$-power, and contains no higher repetition. One gets $x = \mu(0101)0$, and $y = 0101$ itself avoids repetitions of exponent greater than $9/4$.

It follows from the lemma that an overlap-free word x has a factorization

$$x = u_1\mu(u_2)\cdots\mu^{h-1}(u_h)\mu^h(x_h)\mu^{h-1}(v_h)\cdots\mu(v_2)v_1,$$

where each u_i and v_i has length at most 2, and x_h has length at most 4. A simple computation shows that $\log|x| - 3 < h \leqslant \log|x|$. Thus, the value of h and each u_i and v_i and x_h may take a finite number of values, from which the total number of overlap-free words results to be bounded by $c \cdot d^{\log n} = c \cdot n^{\log d}$ for some constants c and d.

Another consequence of the lemma is that the Thue–Morse word $t = \mu^\omega(0)$ is not only overlap-free but avoids $7/3$-powers. A clever generalization, by Rampersad [32], of a proof of [39,40] shows that t (and its opposite $\bar{t}$) is the only infinite binary word avoiding $7/3$-powers that is a fixed point of a nontrivial morphism.

Restivo and Salemi's theorem says that $v(n) \leqslant Cn^s$ for some real s. The upper bound $\log 15$ for s given by Restivo and Salemi has been improved by Kfoury [24] to 1.7, by Kobayashi [25] to 1.5866 and by Lepistö in his Master thesis [26] to 1.37; Kobayashi [25] gives also a lower bound. So

Theorem 3. *There are constants C_1 and C_2 such that*

$$C_1 n^r < v(n) < C_2 n^s,$$

where $r = 1.155\ldots$ and $s = 1.37\ldots$.

One might ask what the "real" limit is. In fact, a result by Cassaigne [12] shows that there is no limit. More precisely, he proves

Theorem 4. *Set $r = \liminf \frac{\log v(n)}{\log n}$ and $s = \limsup \frac{\log v(n)}{\log n}$. Then $r < 1.276$ and $1.332 < s$.*

It is quite remarkable that the sequence $v(n)$ is 2-regular. This was shown by Carpi [9] (see [3] for the definition of regular sequences). As we shall see in the next section, the number of square-free ternary words grows exponentially. In fact, Brandenburg [6] proves also that the number of binary cube-free words grows exponentially. The exact frontier between polynomial and exponential growth has been shown to be the exponent $7/3$ by Karhumäki and Shallit [22].

Theorem 5. *There are only polynomially many binary words of length n that avoid $7/3$-powers, but there are exponentially many binary words that avoid $7/3^+$-powers.*

3. Counting square-free words

We report now estimations for the number of square-free words over a ternary alphabet. Let S be the set of ternary square-free words and let $s(n)$ be the number of square-free ternary words of length n.

Since S is factorial (factor-closed), the sequence $(s(n))$ is submultiplicative and the (topological) entropy $\lambda(S)$ exists. We will show that $\lambda(S)$ is not zero, and give bounds for $\lambda(S)$. The

sequence $s(n)$ starts with 3, 6, 12, 18, 30, 42, 60 (Sloane's sequence A006156, see [42]). The sequence $s(n)$ is tabulated for $n \leqslant 90$ in [4] and for $91 \leqslant n \leqslant 110$ in [21].

3.1. Getting upper bounds

There is a simple method to get upper bounds for the number of ternary square-free words, based on using better and better approximations by regular languages.

Clearly, any square-free word over $A = \{0, 1, 2\}$ contains no factor 00, 11 or 22, so $S \subset A^* \setminus A^*\{00, 11, 22\}A^*$. Since the latter is a regular set, its generating function is a rational function. It is easily seen to be $f(t) = (1 + t)/(1 - 2t)$. Indeed, once an initial letter is fixed in a word of this set, there are exactly two choices for the next letter (this remembers Pansiot's encoding [31], see also [28]). So $s(n) \leqslant 2^n + 2^{n-1}$ for $n \geqslant 1$. Moreover, since a word of length at most 3 is square-free if and only if is in $A^* \setminus A^*\{00, 11, 22\}A^*$, the equality $s(n) = 2^n + 2^{n-1}$ holds for $n \leqslant 3$, and thus $s(2) = 6$ and $s(3) = 12$.

One can continue in this way: clearly none of the 6 squares of length 4: 0101, 0202, 1010, 1212, 2020, 2121 is a factor of a word in S, and it suffices to compute the generating function of the set $A^* \setminus A^*XA^*$, where $X = \{00, 11, 22, 0101, 0202, 1010, 1212, 2020, 2121\}$ to get better upper bound for $s(n)$. Some of these generating functions are given explicitly in [36]. For words without squares of length 2 or 4, the series is $(1 + 2t + 2t^2 + 3t^3)/(1 - t - t^2)$ (see [36]). Again, a direct argument gives the reason: a ternary word without squares of length 2 or 4 either ends with aba for $a \neq b$, or with abc where the letters a, b, c are distinct. Denote by u_n (resp. by v_n) the number of words of the first (of the second) type, and by $s^{(2)}(n)$ the total number. Then it is easily seen that, for $n \geqslant 4$, $u_n = v_{n-1}$ and $v_n = s^{(2)}(n - 1)$, and consequently $s^{(2)}(n) = s^{(2)}(n - 1) + s^{(2)}(n - 2)$. This shows of course that $s(n) \leqslant C\tau^n$, for some constant C, with $\tau = (1 + \sqrt{5})/2$ the golden ratio.

More generally, we consider any finite alphabet A, a finite set X and the set $K = A^* \setminus A^*XA^*$. We may assume that X contains no proper factor of one of its elements, so it is a code. Since the set K is a quite particular regular set, we will compute its generating function by using special techniques. There exist at least two (related) ways to compute these generating functions.

First, we consider the *semaphore code* $C = A^*X \setminus A^*XA^+$. Semaphore codes (see e.g. [5]) were introduced by Schützenberger [38] under the name $\mathcal{J}$ codes. The computation below remembers of course also recurrent events in the sense of Feller [18]. The set C is the set of words that have a suffix in X but have no other factor in X. Thus the set K is also the set of proper prefixes of elements in C, and since C is a maximal prefix code, one has

$$C^*K = A^*. \tag{1}$$

Next, one has (see [5] or [27])

$$Kx = \sum_{y \in X} C_y R_{y,x} \quad (x \in X), \tag{2}$$

where $C_y = C \cap A^*y$ and $R_{y,x}$ is the *correlation set* of y and z, given by

$$R_{y,z} = \{z^{-1}x \mid z \in S(y) \cap P(x)\}.$$

Here, $S(y)$ (resp $P(x)$) is the set of proper suffixes of y (proper prefixes of x). Of course,

$$C = \bigcup_{y \in X} C_y. \tag{3}$$

Eqs. (1)–(3) are $\mathrm{Card}\, X + 2$ equations in $\mathrm{Card}\, X + 2$ unknowns and allow to compute the languages or their generating series.

As an example, consider $X = \{00, 11, 22\}$. Denote by f_Z the generating function of the set Z. Then Eqs. (1)–(3) translate into

$$(1-3t) f_K = 1 - f_C,$$
$$f_{Kaa} = t^2 f_K = (1+t) f_{C_{aa}}, \quad f_C = 3 f_{C_{aa}} \quad (a \in A)$$

since $R_{aa,aa} = \{1, a\}$ and $R_{aa,bb} = \emptyset$ for $a \neq b$. Thus $3t^2 f_K = (1+t) f_C$ and $(1-3t) f_K = 1 - f_C = 1 - \frac{3t^2}{1+t} f_K$, whence

$$f_K = \frac{1}{1 - 3t + \frac{3t^2}{1+t}} = \frac{1+t}{1-2t}.$$

The second technique is called the "Goulden–Jackson clustering technique" in [29]. The idea is to *mark* occurrences of words in X in a word, and to weight a marked word with an indicator of the number of its marks. If a word w has r marks, then its weight is $(-1)^r t^{|w|}$. As an example, if X is just the singleton $X = \{010\}$, the word $w = 01001010$ exist in eight marked versions, namely 01001010, $\underline{010}01010$, $010\underline{010}10$, $01001\underline{010}$, $\underline{010}\,\underline{010}10$, $\underline{010}01\underline{010}$, $01001\underline{010}$, $\underline{01001010}$. Let us write $\underline{w}$ for a marked version of w, and $p(\underline{w})$ for its weight. The sum of the weights of the marked versions of a word w is 0 if w contains a factor in X, and is 1 otherwise. In other terms, the generating series of the set $K = A^* \setminus A^* X A^*$ is

$$f_K = \sum_{\underline{w} \in \underline{A}^*} p(\underline{w}),$$

where the sum is over all marked versions of all words. Now, it appears that this series is rather easy to compute when one considers clusters: a *cluster* is a marked word $\underline{w}$ where every position is marked, and that is not the product of two other clusters. Thus, for $X = \{010\}$, the word $\underline{01001010}$ is not a cluster since it is the product of the two clusters $\underline{010}$ and $\underline{01010}$. A marked word is a unique product of unmarked letters and of clusters. Thus, a marked word $\underline{w}$ is either the empty word, or its last letter is not marked, or it ends with a cluster. Thus

$$f_K = 1 + f_K(t) k t + f_K(t)\, p(C),$$

where k is the size of the alphabet and $p(\mathcal{C})$ is the generating series of the set $\mathcal{C}$ of clusters. It follows that

$$f_K(t) = \frac{1}{1 - kt - p(\mathcal{C})}. \tag{4}$$

A cluster ends with a word in X. Let $\mathcal{C}_x = \mathcal{C} \cap A^* x$ be the clusters ending in x. Then the generating series $p(C_x)$ are the solutions of the system

$$p(\mathcal{C}_x) = -t^{|x|} - \sum_{y \in X} (y : x)\, p(\mathcal{C}_y), \tag{5}$$

where $y : x$ is the (strict) *correlation polynomial* of y and x defined by

$$y : x = \sum_{z \in R_{y,x} \setminus \{e\}} t^{|z|}.$$

Eq. (5) is a system of linear equations, and the number of equations is the size of X. Solving this system gives the desired expression.

Consider the example $X = \{010\}$ over $A = \{0, 1\}$. Then the generating series of $K = A^* \setminus A^*010A^*$ is

$$f_K(t) = \frac{1}{1 - 2t - p(\mathcal{C}_{010})}$$

and $p(\mathcal{C}_{010}) = -t^3 - t^2 p(\mathcal{C}_{010})$, whence $p(\mathcal{C}_{010}) = \frac{-t^3}{1+t^2}$ and

$$f_K(t) = \frac{1}{1 - 2t + \frac{t^3}{1+t^2}} = \frac{1 + t^2}{1 - 2t + t^2 - t^3}.$$

Both methods are just two equivalent formulations of the same computation, as pointed out to me by Dominique Perrin. When $X = \{x\}$ is a singleton, Eq. (2) indeed becomes

$$Kx = CR$$

with $R = R_{x,x}$, and in noncommuting variables, Eq. (1) is just

$$K(1 - A) = 1 - C$$

so

$$K(1 - A) = 1 - KxR^{-1}$$

whence

$$K(1 - A + xR^{-1}) = 1. \tag{6}$$

Now, the coefficients of the series $-xR^{-1}$ are precisely the weights of the cluster of x. So Eq. (6), converted to a generating series, yields precisely Eq. (4)! In the general case one considers the (row) vectors $\vec{X} = (x)_{x \in X}$ and $\vec{C} = (C_x)x \in X$ and the $X \times X$ matrix $R = (R_{x,y})_{x,y \in X}$. Then Eq. (2) is $K\vec{X} = \vec{C}R$ and the same computation as above gives

$$K = \left(1 - A + \sum_{x \in X} (\vec{X}R^{-1})_x\right) = 1.$$

The computation of the generating functions for sets K of the form above, or more generally of the series $\sum_{w \in K} \pi(w) t^{|w|}$, where π is a probability distribution on A^*, is an important issue both in concrete mathematics [20], in the theory of codes [5] and in computational biology (see e.g. chapters 1, 6 and 7 in [27]). Extensions are in [30,33].

In their paper [29], Nanoon and Zeilberger present a package that allows to compute the generating functions and their asymptotic behaviour for the regular sets of words without squares yy of length $|y| = \ell$ for ℓ up to 23. Richard and Grimm [36] go one step further, to $\ell = 24$. The entropy $\lambda(S)$ of the set of square-free ternary words is now known to be at most 1.30194.

3.2. Getting lower bounds

In order to get an exponential lower bound on the number of ternary square-free words, there are two related methods, initiated by Brandenburg [6] and Brinkhuis [7]. The first method is used for instance in [22], the second one, which gives now sharper bounds, was recently used in [2]. Both rely on the notion of a *finite square-free substitution* from A^* into B^*, for some alphabet B. Let us recall that a *substitution* in formal language theory is a morphism f from some free monoid A^* into the monoid of subsets of B^* that is a function satisfying $f(\mathrm{e}) = \{\mathrm{e}\}$ and $f(xy) = f(x)f(y)$, where the product on the right-hand side is the product of the sets $f(x)$ and $f(y)$ in B^*. The substitution is *finite* if $f(a)$ is a finite set for each letter $a \in A$ (and so for each word $w \in A^*$), it is called *square-free* if each word in $f(w)$ is square-free whenever w is a square-free word on A. For an overview of recent results about power-free morphisms in connection with open problems, see [36].

Brandenburg's method goes as follows. Let $A = \{0, 1, 2\}$ and let $B = \{0, 1, 2, \bar{0}, \bar{1}, \bar{2}\}$. Let $g : B^* \to A^*$ be the morphism that erases bars. Define a substitution f by $f(a) = g^{-1}(a)$. Clearly, f is finite and square-free. Also each square-free word w of length n over A is mapped onto 2^n square-free words of length n over B.

The second step consists in finding a square-free morphism h from B^* into A^*. Assume that h is uniform of length r. Then each square-free word w of length n over B is mapped into a square-free word of length rn over A by the morphism h. It follows that there are 2^n square-free words of length rn for each square-free word of length n, that is

$$s(rn) \geqslant 2^n s(n).$$

Since $s(n)$ is submultiplicative, one has $s(rn) \leqslant s(n)^r$. Reporting in the previous equation yields $s(n) \geqslant 2^{n/(r-1)}$ and proves that growth is exponential.

It remains to give a square-free morphism h from B^* into A^*, where $B = \{0, 1, 2, \bar{0}, \bar{1}, \bar{2}\}$. It appears that

$$h : \begin{array}{rcl} 0 & \mapsto & 010201202101210201021 2 \\ 1 & \mapsto & 01020120212012101202 12 \\ 2 & \mapsto & 01020121020102101202 12 \\ \bar{0} & \mapsto & 01020121021202101202 12 \\ \bar{1} & \mapsto & 01020121012021012102 12 \\ \bar{2} & \mapsto & 01020121012021201210 12 \end{array}$$

is a square-free morphism. Here $r = 22$, and consequently $s(n) \geqslant 2^{n/21}$. The following is a slight variation of Brandenburg's result:

Theorem 6. *The number $s(n)$ of square-free ternary words of length n satisfies the inequality $s(n) \geqslant 6 \cdot 1.032^n$.*

A more direct method was initiated by Brinkhuis [7]. He considers a 25-uniform substitution f from A^* into itself defined by

$$f : \begin{array}{rcl} 0 & \mapsto & \{U_0, V_0\} \\ 1 & \mapsto & \{U_1, V_1\} \\ 2 & \mapsto & \{U_2, V_2\} \end{array}$$

where $U_0 = x1\tilde{x}$, $V_0 = y0\tilde{y}$ and $x = 012021020102$ and $y = 012021201021$. The words $U_1, \ldots, V_2$ are obtained by applying the circular permutation $(0, 1, 2)$. He proves that f is square-free, and thus every square-free word w of length n is mapped onto 2^n square-free words of length $25n$. His bound is only $2^{n/24}$.

The substitution f can be viewed as the composition of an inverse morphism and a morphism, when $U_0, \ldots, V_2$ are considered as letters and then each of these letters is mapped to the corresponding word. However, the second mapping is certainly not square-free since the image of U_0V_0 contains the square 00. Thus, the construction of Brinkhuis is stronger. Indeed, Ekhad and Zeilberger [17] found 18-uniform square-free substitution of the same form than Brinkhuis' and thus reduced the bound from $2^{n/24}$ to $2^{n/17}$. A relaxed version of Brinkhuis' construction is used by Grimm [21] to derive the better bound $65^{n/40}$, and by Sun [43] to improve this bound to $110^{n/42}$.

4. Other bounds

We review briefly other bounds on the number of repetition-free words. Concerning cube-free binary words, already Brandenburg [6] gave the following bounds.

Theorem 7. *The number $c(n)$ of binary cube-free words of length n satisfies* $2 \cdot 1.080^n < 2 \cdot 2^{n/9} \leqslant c(n) \leqslant 2 \cdot 1251^{(n-1)/17} < 1.315 \cdot 1.522^n$.

The upper bound was improved by Edlin [16] to $B \cdot 1.4576^n$ for some constant B by using the "cluster" method.

Next, we consider Abelian repetitions. An *Abelian square* is a nonempty word uu', where u and u' are commutatively equivalent, that is u' is a permutation of u. For instance, 012102 is an Abelian square. It is easy to see that there is no infinite Abelian square-free word over three letters. The existence of an infinite word over four letters without Abelian squares was demonstrated by Keränen [23]. Also, the question of the existence of exponentially many quaternary infinite words without Abelian squares was settled by Carpi [10] positively. He uses an argument similar to Brinkhuis' but much more involved. Square-free morphisms from alphabets with more than four letters into alphabets with four letters seem not to exist [8]. He shows

Theorem 8. *The number $d(n)$ of quaternary words avoiding Abelian squares satisfies $d(n) \geqslant C \cdot 2^{19n/(85^3-85)}$ for some constant C.*

This result should be compared to the following, concerning ternary words without Abelian cubes [2].

Theorem 9. *The number $r(n)$ of ternary words avoiding Abelian cubes grows faster than $2^{n/24}$.*

The number of ternary words avoiding Abelian cubes is 1, 3, 9, 24, 66, 180, ... It is the sequence A096168 in [42]. The authors consider the 6-uniform substitution

$$h: \begin{array}{rcl} 0 & \mapsto & 001002 \\ 1 & \mapsto & 110112 \\ 2 & \mapsto & 002212, 122002 \end{array}$$

This does *not* preserve Abelian cube-free words since the word

0010|02110|11200|10021|10112

which contains an Abelian cube is in $h(0101)$. However, the set $\{h^n(0) : n \geqslant 0\}$ is shown to avoid Abelian cubes.

There is an interesting intermediate situation between the commutative and the noncommutative case which is the case where, for the definition of squares, only some of the letters are allowed to commute. To be precise, consider a set Θ of commutation relations of the form $ab = ba$ for a, b letters, and define the relation $u \sim v \bmod \Theta$ as the transitive closure of the relation $uabv \sim ubav$ for all words u, v and $ab = ba$ in Θ. A *Θ-square* is a word uu' such that $u \sim u' \bmod \Theta$. If Θ is empty, a *Θ-square* is just a square, and if Θ is the set of all $ab = ba$ for $a \neq b$, a *Θ-square* is an Abelian square. Since there is an infinite quaternary word that avoids Abelian squares, the same holds for Θ-squares. For 3 letters, the situation is on the edge since ther exist infinite square-free words, but no infinite Abelian square-free word. The result proved by Cori and Formisano [13] is:

Theorem 10. *If the set Θ of commutation relations contains at most one relation, then the set of ternary words avoiding Θ-squares is infinite, otherwise it is finite.*

It has been proved by the same authors [14] that the number of words grows only polynomially with the length.

This result is different from [11] where square-free words in partially commutative monoids are investigated.

Another variation concerns circular words. A circular word avoids α-powers if all its conjugates avoid α-powers. For instance, 001101 is a circular 2^+-power free word because each word in the set

{001101, 011010, 110100, 101001, 010011, 100110}

is a 2^+-power free word. On the contrary, the word 0101101 is cube-free but its conjugate 1010101 is not cube-free and not even 3^+-power free; so, viewed as a circular word, 0101101 is not 3^+-power free. It is proved in [1] that there exist infinitely many $5/2^+$-power free binary circular words, whereas every circular word of length 5 either contains a cube or a $5/2$-power. This improves a previous result [15] showing that there are infinitely many cube-free circular binary words, see also [19]. No informations is available about the growth of the number of these words.

Acknowledgements

Many thanks to the anonymous referee who contributed additional references and corrected several misprints.

References

[1] A. Aberkane, J. Currie, There exist binary circular $5/2^+$ power free words of every length, Electron. J. Combin. 11 (2004) R10.
[2] J. Aberkane, J. Currie, N. Rampersad, The number of ternary words avoiding Abelian cubes grows exponentially, in: Workshop on Word Avoidability, Complexity and Morphisms, LaRIA Techn. Report 2004-07, 2004, pp. 21–24.
[3] J.-P. Allouche, J. Shallit, Automatic Sequences: Theory, Applications, Generalizations, Cambridge University Press, Cambridge, 2003.
[4] M. Baake, V. Elser, U. Grimm, The entropy of square-free words, Math. Comput. Modelling 26 (1997) 13–26.
[5] J. Berstel, D. Perrin, Theory of Codes, Academic Press, New York, 1985.
[6] F.-J. Brandenburg, Uniformly growing *k*-th power-free homomorphisms, Theoret. Comput. Sci. 23 (1983) 69–82.
[7] J. Brinkhuis, Nonrepetitive sequences on three symbols, Quart. J. Math. Oxford 34 (1983) 145–149.
[8] A. Carpi, On Abelian power-free morphisms, Internat. J. Algebra Comput. 3 (1993) 151–167.
[9] A. Carpi, Overlap-free words and finite automata, Theoret. Comput. Sci. 115 (2) (1993) 243–260.
[10] A. Carpi, On the number of Abelian square-free words on four letters, Discrete Appl. Math. 81 (1998) 155–167.
[11] A. Carpi, A. De Luca, Square-free words on partially commutative monoids, Inform. Proc. Lett. 22 (1986) 125–131.
[12] J. Cassaigne, Counting overlap-free words, in: P. Enjalbert, A. Finkel, K. Wagner (Eds.), STACS '93, Lecture Notes in Computer Science, Vol. 665, Springer, Berlin, 1993, pp. 216–225.
[13] R. Cori, M. Formisano, Partially Abelian squarefree words, RAIRO Inform. Théoret. Appl. 24 (6) (1990) 509–520.
[14] R. Cori, M. Formisano, On the number of partially Abelian squarefree words on a three-letter alphabet, Theoret. Comput. Sci. 81 (1) (1991) 147–153.
[15] J. Currie, D. Fitzpatrick, Circular words avoiding patterns, in: M. Ito, M. Toyama (Eds.), Developments in Language Theory, DLT 2002, Lecture Notes in Computer Science, Springer, Berlin, 2004, pp. 319–325.
[16] A. Edlin, The number of binary cube-free words of length up to 47 and their numerical analysis, J. Differential Equations Appl. 5 (1999) 153–154.
[17] S.B. Ekhad, D. Zeilberger, There are more than $2^{n/17}$ *n*-letter ternary square-free words, J. Integer Seq. (1998) (Article 98.1.9).
[18] W. Feller, An Introduction to Probability Theory and its Applications, Wiley, New York, 1966.
[19] D. Fitzpatrick, There are binary cube-free circular words of length *n* contained within the Thue–Morse word for all positive integers *n*, Electron. J. Combin. 11 (2004) R14.
[20] R. Graham, D. Knuth, O. Patashnik, Concrete Mathematics, Addison-Wesley, Reading, MA, 1989.
[21] U. Grimm, Improved bounds on the number of ternary square-free words, J. Integer Seq. (2001) (Article 01.2.7).
[22] J. Karhumäki, J. Shallit, Polynomial versus exponential growth in repetition-free binary words, J. Combin. Theory Ser. A 105 (2004) 335–347.
[23] V. Keränen, Abelian squares are avoidable on 4 letters, in: ICALP '92, Lecture Notes in Computer Science, Vol. 623, Springer, Berlin, 1992, pp. 41–52.
[24] R. Kfoury, A linear time algorithm to decide whether a binary word contains an overlap, RAIRO Inform. Théoret. Appl. 22 (1988) 135–145.
[25] Y. Kobayashi, Enumeration of irreductible binary words, Discrete Appl. Math. 20 (1988) 221–232.

[26] A. Lepistö, A characterization of 2^+-free words over a binary alphabet, Master's Thesis, University of Turku, Finland, 1995.
[27] M. Lothaire, Applied Combinatorics on Words, Cambridge University Press, Cambridge, 2005.
[28] 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.
[29] J. Nanoon, D. Zeilberger, The Goulden–Jackson cluster method: extensions, applications and implementations, J. Differential Equations Appl. 5 (1999) 355–377.
[30] P. Nicodème, B. Salvy, P. Flajolet, Motif statistics, Theoret. Comput. Sci. 287 (2002) 593–617.
[31] J.-J. Pansiot, A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, Discrete Appl. Math. 7 (1984) 297–311.
[32] N. Rampersad, Words avoiding 7/3-powers and the Thue–Morse morphism, 2003, Preprint available at http://www.arxiv.org/abs/math.CO/0307401.
[33] M. Régnier, A unified approach to word occurrence probabilities, Discrete Appl. Math. 104 (2000) 259–280.
[34] A. Restivo, S. Salemi, On weakly square free words, Bull. EATCS 21 (1983) 49–56.
[35] A. Restivo, S. Salemi, Overlap-free words on two symbols, in: M. Nivat, D. Perrin (Eds.), Automata on Infinite Words, Lecture Notes in Computer Science, Vol. 192, Springer, Berlin, 1985, pp. 198–206.
[36] C. Richard, U. Grimm, On the entropy and letter frequencies of ternary square-free words, http://arXiv:math.CO/0302302, July 2004.
[37] G. Richomme, P. Séébold, Conjectures and results on morphisms generating *k*-power-free words, Internat. J. Found. Comput. Sci. 15 (2) (2004) 307–316.
[38] M.-P. Schützenberger, On the synchronizing properties of certain prefix codes, Inform. Control 7 (1964) 23–36.
[39] P. Séébold, Morphismes itérés, mot de Morse et mot de Fibonacci, C. R. Acad. Sci. Paris 295 (1982) 439–441.
[40] P. Séébold, Sequences generated by infinitely iterated morphisms, Discrete Appl. Math. 11 (1985) 255–264.
[41] J. Shallit, Avoidability in words: recent results and open problems, in: Workshop on Word Avoidability, Complexity and Morphisms, LaRIA Techn. Report 2004-07, 2004, pp. 1–4.
[42] N.J.A. Sloane, The on-line encyclopedia of integer sequences, http://www.research.att.com/~njas/sequences/.
[43] X. Sun, New lower bound on the number of ternary square-free words, J. Integer Seq. (2003) (Article 03.2.2).