On maximal repetitions of arbitrary exponent

Roman Kolpakova, Gregory Kucherov, Pascal Ochem

Abstract

The first two authors have shown [1,2] (Kolpakov and Kucherov, 1999, 2000) that the sum of the exponents (and thus the number) of maximal repetitions of exponent at least 2 in a word (also called runs) is linear with respect to the length of the word. The exponent 2 in the definition of a run may seem arbitrary. In this paper, we consider maximal repetitions of exponent strictly greater than 1.

1. Introduction

Repetitions (periodicities) are fundamental concepts in word combinatorics [3–5]. A great deal of work in word combinatorics has been devoted to the study of words that do not contain subwords of a given exponent [4]. Another research direction, of a more algorithmic nature, is the efficient identification of all subwords of a given exponent in a word [5], which raises the combinatorial question of the possible number of such subwords.

In [1,2], the first two authors considered the notion of maximal repetitions of a word, which are subword occurrences that cannot be extended outwards without changing their period. They proved that the number of maximal repetitions of exponent at least 2 in a word is linearly bounded with respect to the word length. It has been conjectured that this number is actually smaller than the word length. This conjecture is now commonly known as the “runs conjecture”. They also proved that not only the number of maximal repetitions of exponent 2 or more is linearly bounded, but the sum of exponents of these repetitions is linearly bounded too. The linear bound on the number of repetitions, in turn, allowed them to prove that all such maximal repetitions can be found in linear time. More recently, other researchers attempted to improve these results by finding a simpler proof of the linear bound implying a smaller multiplicative constant. The last current achievement in this direction is presented in [6].

A big question that remained open in this development concerns the lower bound of 2 on the exponent of considered repetitions. While this bound is intuitively natural (as it requires some subword to be consecutively repeated at least twice), it has no formal justification. Moreover, word combinatorics provides many separation results when the “right” bound on the exponent is not an “intuitive” number. For example, the famous Dejean’s result states that the exponents that can be avoided on a ternary alphabet are exponents greater than $7/2$ [7]. As another example, there are exponentially many binary words avoiding exponents greater than $2/4$, while there are only polynomially many of them avoiding smaller exponents [8].

In this paper, we completely lift the lower bound on the exponent and focus on the maximal repetitions of any...
exponent greater than 1. Note that repetitions with exponent between 1 and 2 are subwords of the form \( uvv \) that can be viewed as non-consecutive repetitions. Therefore, in this paper we consider both consecutive (periodicities) and non-consecutive repetitions. To the best of our knowledge, the number of repetitions of exponent smaller than 2 has not been studied.

Instead of directly counting the repetitions or the sum of their exponents, we consider the sum of exponents decremented by 1. The main idea is that repetitions with exponents close to 1 (i.e. subwords \( uvu \) with \( |v| \gg |u| \)) contribute to the sum with an amount close to 0. We prove that this sum is upper-bounded by \( n \ln(n) \) (Theorem 1) which immediately implies that the number of maximal repetitions of any exponent greater than 1 is bounded by \( \frac{n}{\ln(n)} \). On the other hand, the number of all maximal repetitions can be quadratic (Theorem 5). We also show that the lower bound for the sum is \( \frac{1}{2} n \ln(n) \), where \( k \) is the alphabet size, and we characterize the words achieving this lower bound (Theorem 6). Finally, we study this sum for the words containing only repetitions with a period bounded by a constant.

While the “whole picture” of the count of the number of maximal repetitions with exponent smaller than 2 is still incomplete, we believe that our results represent the first step in this direction.

2. Definitions

For a word \( v \), its length will be denoted by \( |v| \). Given a word \( v \), \( p \) is a period of \( v \) if any two letters at distance \( p \) in \( v \) match. Formally, a natural number \( p \), \( 0 < p \leq |v| \), is a period of \( v \) if \( v[i] = v[i+p] \) whenever positions \( i \) and \( i+p \) both exist in \( v \). A period is called primitive if it is not a multiple of any other period. Trivially, any word \( v \) has a period equal to \( |v| \). For example, the word \( ababa \) has periods 2, 4 and 5, where 4 is not primitive and 2 and 5 are primitive. The word \( aabaab \) has primitive periods 3, 4 and 5. We refer to [3,4] for further definitions and properties of periods.

In this paper, we call repetition any word having a period smaller than the word length. We will be interested in subwords of a given word \( w \) which are repetitions, i.e. in subwords \( w[i...j] \) having a period \( p < |w[i...j]| = j - i + 1 \). Formally, a repetition \( r \) is a pair \((i,j), p\), where \( p \) is a primitive period of \( w[i...j] \) strictly smaller than \( j - i + 1 \). \( p(r) \) will denote the period \( p \) of \( r \), and \( |r| \) the length \( |w[i...j]| = j - i + 1 \). The exponent of \( r \) is defined as \( e(r) = \frac{|r|}{p(r)} \).

A maximal repetition in \( w \) is a repetition \( r = (i,j), p \) such that (i) if \( i \neq 1 \), then \( p \) is not a period of \( w[i-1...j] \), and (ii) if \( j \neq |w| \), then \( p \) is not a period of \( w[i...j+1] \). The two conditions are equivalent to saying that \( w[i-1] \neq w[i-1+p] \) and \( w[j+1] \neq w[j+1+p] \) respectively. Informally, “maximality” means that the subword is extended outwards as much as possible so long as its period is preserved. The set of all maximal repetitions in \( w \) will be denoted \( \mathcal{M}(w) \).

As an example, the word \( aabaac \) has two maximal repetitions \( aa \) with period 1 (exponent 2), a maximal repetition \( aabaaaabab \) with period 3 (exponent 5\( \frac{1}{3} \)), and a maximal repetition \( aabaab \) with period 4 (exponent 5\( \frac{1}{4} \)). Note in this example that two occurrences of \( aa \) form two distinct maximal repetitions, i.e. we consider repetition occurrences and not mere subwords. Note also that the same subword (\( aabaab \) in the example) can form two distinct maximal repetitions with two different primitive periods.

Maximal repetitions \( r \) with \( e(r) \geq 2 \) (or equivalently \( p(r) \leq \frac{|w|}{2} \)) are called runs [6] and have been extensively studied [5]. By Fine and Wilf’s Theorem, if a word has two periods no larger than a half of the word length, then one of them is not primitive. This means that if we restrict ourselves only to periods of at most half of the word length, a repetition has only one primitive period which is its smallest period and all the other periods are multiples of it.

This situation changes if periods larger than a half of the word length are considered. As an example above illustrates, repetition \( aabaab \) has periods 3 and 4 that are both primitive, which leads to two different repetitions. This is why the repetitions considered in this paper are specified by both the corresponding subword and the period.

Note that we could have considered, similar to runs, only minimal periods, which would eliminate the situation when one subword corresponds to several repetitions. The number of repetitions would then become smaller and therefore this definition is weaker with respect to the counting results that we will present in this paper. All results of this paper stay true under this definition too.

Note that under our definition of repetition, any two occurrences of the same letter at positions \( i < j \) in \( w \) participate in a maximal repetition with period either \( j - i \) or a divisor of \( j - i \). This repetition is uniquely defined: its period is the minimal divisor of \( j - i \), say \( p = (j - i)/k \), such that \( w[i...j-1] = (w[i...i+p-1])^k \), and, on the other hand, two maximal repetitions with the same period \( p \) intersecting on at least \( p \) letters must coincide. Therefore, we will speak about a maximal repetition defined by a letter match \( w[i] = w[j] \).

3. Sum of decremented exponents

For a word \( w \), we will be interested in the sum of exponents of all maximal repetitions, decremented by 1:

\[
\sum_{r \in \mathcal{M}(w)} (e(r) - 1).
\]

This quantity can be viewed as the difference between the sum of exponents of all maximal repetitions and the number of these repetitions.

**Theorem 1.** For every word \( w \) of length \( n \), we have

\[
\sum_{r \in \mathcal{M}(w)} (e(r) - 1) \leq n \ln(n).
\]

**Proof.** For each maximal repetition \( r \) with period \( p \), we distribute the value \( e(r) - 1 = \frac{|r|}{p} - 1 \) over \((|r| - p)\) pairs of matching letters \((w[i], w[i+p])\). Note that \( w[i] = w[i+p] \) within the repetition. Each such pair contributes to the sum with weight \( \frac{1}{2} \). Consider two positions \( i \) and \( j \), \( 1 \leq i < j \leq n \), in \( w \). If \( w[i] = w[j] \), then this match defines a maximal repetition, and it should be counted if and only if
the period of this repetition is \( j - i \) (and not a proper divisor of it), in which case it contributes to the sum with the amount \( \frac{1}{n^\varepsilon} \). If we count all pairs of positions in \( w \), we obtain the upper bound \( \sum_{r \in M(w)} (e(r) - 1) \leq \sum_{1 \leq i < j \leq n} \frac{n^r}{n^\varepsilon} = n \sum_{i=1}^{n-1} \frac{1}{\varepsilon} - (n - 1) \leq n \ln(n) \) for \( n > 2 \). \( \Box \)

If we restrict only to maximal repetitions of period at most \( p \), then the following bound holds.

**Corollary 2.** For every word \( w \) of length \( n \), we have \( \sum_{r \in M(w), p(r) \leq p} (e(r) - 1) \leq n \ln(p + 1) \).

**Proof.** If only repetitions of period at most \( p \) are considered, then the proof of Theorem 1 yields the following bound: \( \sum_{r \in M(w), p(r) \leq p} (e(r) - 1) \leq n \sum_{1 \leq j \leq \min\{i, p, n\}} \frac{1}{j^\varepsilon} \leq n \ln(p + 1) \). \( \Box \)

Similarly, if we count only maximal repetitions of period at least \( p \), then we get

**Corollary 3.** For every word \( w \) of length \( n \), we have \( \sum_{r \in M(w), p(r) \geq p} (e(r) - 1) \leq n \ln(p) \).

**Proof.** Similar to Corollary 2. \( \Box \)

Let us now focus only on maximal repetitions of exponent \( (1 + \varepsilon) \) or more, and let us count their number. Theorem 1 immediately provides a nontrivial upper bound.

**Corollary 4.** For every word \( w \) of length \( n \) and every \( \varepsilon > 0 \), the number of maximal repetitions of exponent at least \( (1 + \varepsilon) \) in \( w \) is at most \( \frac{1}{\varepsilon} \ln(n) \).

**Proof.** Consider the sum in the statement of Theorem 1. Each repetition contributes at least \( \varepsilon \) to it and therefore the number of those is at most \( \frac{n \ln(n)}{\varepsilon} \). \( \Box \)

Similarly, Corollaries 2 and 3 imply respective upper bounds \( \frac{1}{\varepsilon} \ln(p + 1) \) and \( \frac{1}{\varepsilon} \ln(n/p) \) on the number of maximal repetitions of exponent at least \( (1 + \varepsilon) \) and of period respectively at most \( p \) and at least \( p \).

The following theorem shows that the upper bound of Theorem 1 is asymptotically tight within a factor of 8 and that the number of all repetitions of arbitrary exponent can be quadratic (to be compared with Corollary 4).

**Theorem 5.** Let \( w = (0011)^{n/4} \). Then
\[
(i) \sum_{r \in M(w)} (e(r) - 1) \geq \frac{1}{3} n \ln(n).
(ii) The number of all maximal repetitions of \( w \) is \( \Theta(n^2) \).
\]

**Proof.** (i) The entire word \( w \) is an obvious repetition of period 4, its contribution to the sum is \( (n/4 - 1) \). Any other repetition can be specified by a match between two 0’s or two 1’s that occur at a distance other than a multiple of 4.

Consider a repetition \( r \) in which the letter 0 at some position \( m \), \( m \equiv 1 \mod 4 \), matches letter 0 at a position \( \ell > m \), \( \ell \equiv 2 \mod 4 \). This match corresponds to end letters of the repetition, as \( w[m - 1] = 1 \) if \( m \neq 1 \) while \( w[\ell - 1] = 0 \), and \( w[\ell + 1] = 1 \) if \( \ell \neq n \) while \( w[m + 1] = 0 \).

Furthermore, this repetition has a period \( \ell - m = |r - 1| \) and this period is primitive, as the word \( w[m .. \ell - 1] \) contains one more 0 than 1’s and therefore the number of 0’s and the number of 1’s in \( w[m .. \ell - 1] \) are relatively prime, which shows that \( w[m .. \ell - 1] \) is a primitive word (i.e. not an integer power of some shorter word).

Therefore, any two such positions \( m \) and \( \ell \) define a repetition that contributes \( 1/(\ell - m) \) to the sum. In total, all such repetitions contribute \( \sum_{r \in M(w)} (n/4 - i + 1)/(4i - 3) \geq \frac{1}{3} n \ln(n) \).

There are three other symmetric cases: one corresponds to another way of matching two 0’s and the other two correspond to matching two 1’s. The four cases together yield \( \sum_{r \in M(w)} (e(r) - 1) \geq \left( \frac{7}{3} - 1 \right) + 4 \frac{1}{\varepsilon} \ln(n) \geq \frac{1}{\varepsilon} \ln(n) \).

(ii) Is obvious from the above considerations, as the number of pairs of 0’s and pairs of 1’s defining repetitions is quadratic. \( \Box \)

We now focus lower on the bound for sum (1). In the rest of the paper, we assume that we have a \( k \)-letter alphabet \( A_k = \{a_1, a_2, \ldots, a_k\} \).

**Theorem 6.** For all \( w \in (A_k)^k \), \( \sum_{r \in M(w)} (e(r) - 1) \geq \frac{n}{k} - 1 \) and the equality holds if and only if \( w = (a_1 a_2 \ldots a_k)^k \) modulo a permutation of alphabet letters.

**Proof.** Given a word \( w \in (A_k)^k \), consider all occurrences of a letter \( a_i \in A_k \) in \( w \), and let \( d_1(i), d_2(i), \ldots, d_j(i) \) be the distances between all consecutive occurrences of \( a_i \) in \( w \). Consider the sum
\[
\sum_{a_i \in A_k} \sum_{j=1}^{\ell_i} \frac{1}{d_j(i)}.
\]

Observe that \( \sum_{r \in M(w)} (e(r) - 1) \geq \sum_{a_i \in A_k} \sum_{j=1}^{\ell_i} \frac{1}{d_j(i)} \) since two consecutive occurrences of \( a_i \) necessarily participate in a repetition with period equal to the distance between these occurrences, and then contribute to sum (1) (see proof of Theorem 1).

Therefore, if we construct a word that minimizes sum (2) and for which \( \sum_{r \in M(w)} (e(r) - 1) = \sum_{a_i \in A_k} \sum_{j=1}^{\ell_i} \frac{1}{d_j(i)} \), this will prove that this word also minimizes sum (1). Our goal is to prove that this minimum is reached if and only if for any letter \( a_i \), \( d_j(i) = k \) for all \( j \), i.e. on words of the form \( w = (a_1 a_2 \ldots a_k)^k \) (modulo a permutation of alphabet letters). Clearly, for such words, sum (1) and sum (2) are both equal to \( \frac{n}{k} - 1 \).

By contradiction, consider a word \( w \) that does not have the form \( (a_1 a_2 \ldots a_k)^k \) and assume that it minimizes sum (2). Then there exists a pair of positions \( m_1 \neq m_2 \), such that \( w[m_1] = w[m_2] \) and \( m_2 - m_1 < k \). Among all such pairs, consider the one with minimal \( m_2 \).

If \( m_2 \leq k \), then there is an alphabet letter, say \( a_j \), that does not occur in \( w[1 .. m_1] \). Let \( w[m_1] = a_j \). We then replace, for all positions \( m_2 \geq m \), each occurrence of \( a_j \) by
that the only distance between consecutive occurrences of some letter that gets modified is \((m_r - m_l)\) which is replaced by a larger distance (or by no distance at all if \(a_j\) does not occur in \(w\)). This makes sum (2) smaller, which is a contradiction.

If \(m_r > k\), then show that for any position \(m, k < m < m_r\), we must have \(w[m] = w[m - k]\). This is because the letter \(w[m]\) cannot repeat on the left at a distance smaller than \(k\), as this would contradict the definition of \(m_r\). On the other hand, the closest occurrence of \(w[m]\) to the left cannot be at a distance larger than \(k\) either. Indeed, if \(w[m] = w[m']\) for some \(m' < m\) and \(m - m' > k\) and there is no occurrence of \(w[m]\) in \(w[m' + 1..m - 1]\), then the subword \(w[m' + 1..m + k]\) is composed of \(k - 1\) letters and has length \(k\), and therefore contains a letter repeated at a distance at most \(k - 1\). This contradicts again the definition of \(m_r\).

By the above, we can assume that \(w[1..m_r - 1] = (a_1..a_q)\) \(\forall a_i, a_j\) (up to a permutation of alphabet letters), \(q \geq 1\), and \(w[m_i] = a_j\) for some \(j \neq i'\) where \(i' = i + 1\) if \(i < k\) and \(i' = 1\) if \(i = k\). Consider the closest position of \(a_r\) to the right of \(m_r\), which we denote by \(m'\). (If such a position does not exist, the proof below will trivially apply.)

We modify \(w\) by replacing, for all positions \(\geq m_r\), each occurrence of \(a_j\) by \(a_i\) and, vice versa, each occurrence of \(a_i\) by \(a_j\). We show that this modification makes sum (2) smaller.

The only distances between consecutive occurrences of letters that will be affected by the modification of \(w\) are the distance \(m_r - m_l\) between the corresponding occurrences of \(a_j\) and the distance \(m' - (m_r - k)\) between the occurrences of \(a_i\). The new distances become respectively \(k\) (between occurrences \(m_r\) and \(m_r - k\) of \(a_r\)) and \(m' - m_l\) (between corresponding occurrences of \(a_j\)). We show that

\[
\frac{1}{m_r - m_l} + \frac{1}{m' - (m_r - k)} > \frac{1}{k} + \frac{1}{m' - m_l}.
\]

This will show that sum (2) becomes smaller after the modification. For this, we show that

\[
\frac{1}{m_r - m_l} - \frac{1}{k} > \frac{1}{m' - m_l} - \frac{1}{m' - (m_r - k)}
\]

or

\[
\frac{k - m_l + m_r}{(m_r - m_l)k} > \frac{k - m_l + m_r}{(m' - m_l)(m' - (m_r - k))}.
\]

The numerators of both sides are equal. In the denominator, we have \(m' - m_l > m_r - m_l\) and \(m' - (m_r - k) > k\), which proves the inequality.

We obtained a contradiction with the assumption that \(w\) minimizes sum (2). This shows that a word that minimizes sum (2) must have the form \(w = (a_1 a_2 \ldots a_k)^z\) (modulo a permutation of alphabet letters). On this word, sum (1) and sum (2) are both equal \(\frac{z}{k} - 1\). This proves that \(w\) also minimizes sum (1).

4. Words with repetitions of bounded period

In this section, we study sum (1) in the case when all repetitions in \(w\) have period at most \(p\). Recall that \(k\) is the alphabet size.

**Theorem 7.** For a word \(w\), \(|w| = n\), assume that for all \(r \in \mathcal{M}(w)\), \(p(r) \leq p\) for some constant \(p\). Then \(\sum_{r \in \mathcal{M}(w)} \ell(r) - 1 \leq n + 3kp(\ln(p) + 1)\).

The proof will use Fine and Wilf’s Theorem (see e.g. [3]) asserting that if \(w\) has periods \(p_1\) and \(p_2\) and \(|w| \geq p_1 + p_2 - \gcd(p_1, p_2)\), then \(w\) has also the period \(\gcd(p_1, p_2)\). This implies, in particular, that two different repetitions with primitive periods \(p_1\) and \(p_2\) cannot intersect on \((p_1 + p_2)\) letters or more.

For a word \(w\), we call a root of \(w\) any subword of \(w\) of length \(p\). The prefix (resp. suffix) root of \(w\) is the prefix (resp. suffix) of \(w\) of length \(p\).\(w\).

**Proof.** Consider a word \(w\) over \(A_k\) such that the period of any repetition in \(w\) is bounded by \(p\).

Assume that for some letter \(a\), two occurrences of \(a\) are located at a distance \(3p\) or more from each other. Consider the repetition \(r\) defined by the match of these two occurrences of \(a\). We will show that \(r\) has a very particular form, namely

(a) all letters within a root of \(r\) are different,
(b) any letter of \(r\) does not occur outside \(r\).

First observe that since the period of \(r\) cannot exceed \(p\), then the two occurrences of \(a\) are separated by at least three periods \(p(r)\). To prove (a), assume that there is another occurrence of \(a\) in the suffix root of \(r\) (cf. Fig. 1). Then, there is a repetition \(r'\) formed by matching this occurrence of \(a\) with the left occurrence of \(a\). These two occurrences are separated by \(3p - p(r) > 2p\) letters. Since \(p(r') < p\), there are at least \(2p(r')\) letters between these two occurrences of \(a\). This means that repetitions \(r\) and \(r'\) intersect on length at least \(2 \cdot \max(p(r), p(r'))\) and by Fine and Wilf’s Theorem, \(r\) and \(r'\) must coincide. This contradiction proves that \(a\) cannot have another occurrence within a root of \(r\). The above argument shows that any letter occurs in a root only once.

Condition (b) is proved by a similar argument. Assume that some letter \(b\) of \(r\) occurs outside \(r\), for instance to the right of \(r\). Then consider the match of this occurrence of
\( b \) with the leftmost occurrence of \( b \) inside \( r \). This match defines a repetition \( r' \). Similar to part (a), \( r \) and \( r' \) intersect on length at least \( 2 \cdot \max \{ p(r), p(r') \} \) and therefore must coincide by Fine and Wilf’s Theorem. This contradicts the assumption of an occurrence of \( b \) outside \( r \) and proves (b).

Now, we split all repetitions into two disjoint classes: repetitions verifying conditions (a) and (b) and the others, called respectively repetitions of type 1 and repetitions of type 2. By condition (b), for any word \( w \), repetitions of type 1 and type 2 in \( w \) are non-intersecting. Furthermore, conditions (a) and (b) insure that two distinct repetitions of type 1 cannot intersect. Therefore, all repetitions of type 1 together cannot contribute more than \( n \) to the sum.

On the other hand, repetitions of type 2 cannot take more than \( 3kp \) letters altogether in \( w \), as each letter cannot occur more than \( 3p \) times (otherwise this would lead to a repetition of type 1 by the above reasoning). Therefore, by Corollary 2, sum (1) for repetitions of type 2 is bounded by \( 3kp(\ln(p) + 1) \). This gives the final bound \( n + 3kp(\ln(p) + 1) \).

Notice that the bound in Theorem 7 is optimal in some sense, since sum (1) is \( n - 1 \) for the word \( a^n \) and \( \Theta(kp \ln(p)) \) for the word \( (a_1a_2a_3)^{p/4}(a_2a_3a_4)^{p/4} \ldots \), according to Theorem 5.

5. Concluding remarks

Many questions related to the combinatorics of repetitions of arbitrary exponent remain unanswered. A major such problem is to determine a more precise bound on the number of such repetitions. Corollary 4 provides an \( O(n\log n) \) bound for the exponents at least \((1 + \varepsilon)\), for any fixed \( \varepsilon > 0 \). It would be of great interest to refine this bound, possibly depending on \( \varepsilon \). It is not excluded that, possibly starting from some \( \varepsilon > 0 \), or even for any fixed \( \varepsilon > 0 \), the number of all repetitions of exponent at least \((1 + \varepsilon)\) is \( O(n) \). This is a challenging question, which seems, however, difficult to solve, as it would generalize the result of [1,2] on the linear number of runs.

References

[1] R. Kolpakov, G. Kucherov, On maximal repetitions in words, in: Proceedings of the 12th International Symposium on Fundamentals of Computation Theory, Iasi, Romania, 1999, in: Lecture Notes in Computer Science, Springer-Verlag, 1999, pp. 374–385.
[2] R. Kolpakov, G. Kucherov, On maximal repetitions in words, Journal of Discrete Algorithms 1 (1) (2000) 159–186.
[3] M. Lothaire, Combinatorics on Words, Encyclopedia of Mathematics and its Applications, vol. 17, Addison–Wesley, 1983.
[4] C. Choffrut, J. Karhumäi, Combinatorics of words, in: G. Rozenberg, A. Salomaa (Eds.), Handbook on Formal Languages, vol. I, Springer-Verlag, Berlin, 1997, pp. 329–438.
[5] R. Kolpakov, G. Kucherov, Identification of periodic structures in words, in: J. Berstel, D. Perrin (Eds.), Applied Combinatorics on Words, in: Encyclopedia of Mathematics and its Applications, vol. 104, Cambridge University Press, 2005, pp. 430–477, Lothaire books, Chap. 8.
[6] M. Crochemore, L. Ilie, L. Tinta, Towards a solution to the “runs” conjecture, in: P. Ferragina, G.M. Landau (Eds.), Combinatorial Pattern Matching, 19th Annual Symposium, CPM 2008, Pisa, Italy, June 18–20, 2008, in: Lecture Notes in Computer Science, vol. 5029, Springer, ISBN 978-3-540-69066-5, 2008, pp. 290–302.
[7] F. Dejean, Sur un Théorème de Thue, Journal of Combinatorial Theory, Series A 13 (1972) 90–99.
[8] J. Karhumäki, J. Shallit, Polynomial versus exponential growth in repetition-free binary words, Journal of Combinatorial Theory, Series A 105 (2004) 335–347.