Available online at www.sciencedirect.com
ScienceDirect
Discrete Applied Mathematics 155 (2007) 1506 – 1511
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

Note

On the number of words containing the factor $(aba)^k$

Ioan Tomescu

Faculty of Mathematics and Computer Science, University of Bucharest, Str. Academiei, 14, 010014 Bucharest, Romania

Received 10 May 2005; received in revised form 23 December 2006; accepted 1 February 2007
Available online 3 March 2007

Abstract

In this paper a recurrence relation satisfied by the number $L(n)$ of words of length n over an alphabet A of cardinality m ($m \geqslant 2$) not containing the factor $(aba)^k$ ($a \neq b$) is deduced. Let k_n be a sequence of positive integers. From [I. Tomescu, A threshold property concerning words containing all short factors, Bull. EATCS 64 (1998) 166–170] it follows that if $\limsup_{n\to\infty} k(n)/\ln n < 1/(3\ln m)$ then almost all words of length n over A contain the factor $(aba)^{k_n}$ as $n \to \infty$. Using the properties of the roots of the recurrence satisfied by $L(n)$ it is shown that if $\limsup_{n\to\infty} k(n)/\ln n > 1/(3\ln m)$ then this property is false. Moreover, if $\lim_{n\to\infty}(\ln n - 3k\ln m) = \eta \in \mathbb{R}$ then $\lim_{n\to\infty} |\mathscr{W}(n, (aba)^{k_n}, A)|/m^n = 1 - \exp(-(1 - 1/m^3)\exp(\eta))$, where $\mathscr{W}(n, (aba)^{k_n}, A)$ denotes the set of words of length n over A containing the factor $(aba)^{k_n}$.

Keywords: Word; Factor; Recurrence relation; Characteristic equation; Pseudo-Vandermonde determinant

1. Notation and preliminary results

Let A be a finite alphabet of cardinality $|A| = m \geqslant 2$. A factor of a word $\alpha \in A^*$ is a word $\beta \in A^*$ for which there exist $p, q \in A^*$ such that $\alpha = p\beta q$ [3]. In [7] Marcus and Păun proposed a program of studying infinite factors of words over a finite alphabet and bridging the gap between the formal language theory and the theory of infinite words.

In this paper we consider the problem of determining the number of infinite words over a finite alphabet A containing as factor the word $(aba)^{k_n}$ as $k_n \to \infty$ and prove that there exists a threshold for this property depending only on the cardinality of A.

If a and b are two distinct letters of A, let $L(n)$ denote the number of words $\alpha \in A^*$ such that their lengths $|\alpha| = n$ and α does not contain the factor $(aba)^k$ of length $3k$. Guibas and Odlyzko [5] proved that the number of words of length n in A^* that do not contain a single fixed factor w satisfies a linear recurrence equation with constant coefficients. For $w = (aba)^k$ this equation is deduced in the next section.

Let $\mathscr{W}(n, k_n, A)$ denote the set of words w of length n over A having the property that each word of length k_n over A is a factor of w and $P(m, k_n)$ the property that almost all words of length n over A contain as factors all words of length k_n over A as $n \to \infty$, that is, $\lim_{n\to\infty} |\mathscr{W}(n, k_n, A)|/m^n = 1$.

Theorem 1.1 (*Tomescu [11]*). *Let k_n be a sequence of positive integers. If* $\limsup_{n\to\infty} k_n/\ln n < 1/\ln m$ *then* $P(m, k_n)$ *holds, but if* $\limsup_{n\to\infty} k_n/\ln n > 1/\ln m$ *then* $P(m, k_n)$ *does not hold.*

E-mail address: ioan@fmi.unibuc.ro.

0166-218X/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2007.02.009

Notice that for fixed k_n this property is well-known and it follows from the property that almost all random strings (in both Kolmogorov–Chaitin and Chaitin senses) satisfy various conditions of normality (first introduced by Borel) (see [1,2]). Also it may be noted that in the binary case ($m=2$) this property is closely related to a result by Flajolet et al. [4] about the number of occurrences of the factors of length k_n in random binary strings of length n as $n \to \infty$.

In the third section we shall prove that a similar property holds for a single word, namely $(aba)^{k_n}$ of length $3k_n$ and for every sequence k_n (bounded or not).

2. An explicit expression for $L(n)$

Lemma 2.1. *Let k be an integer, $k \geqslant 2$. We have*

$$L(n+1) = mL(n) - L(n-3k+2) + mL(n-3k+1) - mL(n-3k) + L(n-3k-1) \\ - (m-1)L(n-3k-2) \tag{1}$$

for every $n \geqslant 3$. Moreover, $L(s) = m^s$ for each $0 \leqslant s \leqslant 3k-1$ and $L(3k+r) = m^{3k+r} - (r+1)m^r$ for $0 \leqslant r \leqslant 2$.

Proof. It is clear that $L(s) = m^s$ for every $1 \leqslant s \leqslant 3k-1$ and $L(3k) = m^{3k} - 1$. If $1 \leqslant r \leqslant 2$ then the words of length $3k+r$ containing the factor $(aba)^k$ have the form $\alpha(aba)^k\beta$, where $|\alpha| + |\beta| = r$ and $|\alpha|, |\beta| \geqslant 0$. This equation has $r+1$ natural solutions, hence $L(3k+r) = m^{3k+r} - (r+1)m^r$. For $r=3$ two choices lead to identical words, namely $aba(aba)^k$ and $(aba)^k aba$. Hence, $L(3k+3) = m^{3k+3} - 4m^3 + 1$ and (1) yields $L(0) = 1$ for $n = 3k+2$. In order to prove (1) let us denote by $W(n)$, $V(n)$ and $U(n)$, respectively, the number of words of length n over A not containing the factor $(aba)^k$, such that they have a suffix equal to $(aba)^{k-1}ab$, $(aba)^{k-1}$ and aba, respectively. It is easy to see that these numbers are related by the following equations:

$$L(n) = m(L(n-1) - W(n-1)) + (m-1)W(n-1)$$

or

$$L(n) = mL(n-1) - W(n-1), \tag{2}$$

$$W(n) = L(n-3k+1) - (U(n-3k+1) + W(n-3k+1)), \tag{3}$$

$$V(n) = L(n-3k+3) - (U(n-3k+3) + W(n-3k+3)) \tag{4}$$

and

$$U(n) = L(n-3) - (W(n-3) + V(n-3)). \tag{5}$$

(2) and (3) yield $W(n) = V(n-2)$ and from (5) we get $U(n) = L(n-3) - W(n-1) - W(n-3)$. By substituting this value in (3) we deduce

$$W(n) = L(n-3k+1) - L(n-3k-2) - W(n-3k+1) + W(n-3k) - W(n-3k-2). \tag{6}$$

From (2), $W(n) = mL(n) - L(n+1)$ and by expressing in (6) W as a function of L we get (1). □

Note that for $k=1$ we obtain the recurrence $L(n+1) = mL(n) - L(n-1) + (m-1)L(n-2)$ for every $n \geqslant 2$ and $L(r) = m^r$ for $0 \leqslant r \leqslant 2$.

To prove that all roots of the characteristic equation of recurrence (1) are simple for k large enough, we will use the following result known as "the continuous dependence of the zeros of a polynomial on its coefficients". The proof can be found in [6,8]. In the following theorem σ_n is the symmetric group of degree n.

Theorem 2.2 (*Ostrowski [8]*). *Let*

$$p(z) = z^n + p_{n-1}z^{n-1} + \cdots + p_1 z + p_0$$

be a monic polynomial with complex coefficients. Then, for every $\varepsilon > 0$, there is $\eta > 0$ such that for any polynomial

$$q(z) = z^n + q_{n-1}z^{n-1} + \cdots + q_1 z + q_0$$