Publ. I.R.M.A. Strasbourg, 1988, 358/S-18
Actes 16^e Séminaire Lotharingien, p. 87-92

SOME RESULTS ON THE COMBINATORICS OF THE THUE-MORSE SEQUENCES

BY

ALDO DE LUCA AND STEFANO VARRICCHIO

1. Introduction

An infinite word w over an alphabet A is a map $w : N \to A$ (cf. [6]). For any infinite word w , F(w) will denote the set of all its finite factors . The structure function of w is defined for any $n > 0$ as

$$F_w(n) = \mathrm{Card}(\ F(w) \cap A^n).$$

Let $A = \{a, b\}$ be a two-symbol alphabet and w an infinite word on A. A finite factor f of w is called special if fa and fb are still factors of w. We denote by S(w) the set of all special factors of w and by $\phi_w : N \to N$ the function which gives for any $n \geq 0$ the number of special factors of w of length n. The importance of the notion of special factor is due to the fact that for any $n > 0$

$$F_w(n+1)=F_w(n)+\phi_w(n),$$

so that if one knows the function ϕ_w then by iteration one can compute the structure-function F_w.

Special factors of the Fibonacci word f have been studied in [1]. In this case $\phi_f(n) = 1$ for any $n > 0$. In [4] we have studied special factors of the Thue-Morse word t which can be introduced as the limit word obtained by iterating on the letter "a" the morphism μ : $A^* \to A^*$ defined as $\mu(a) = ab$, $\mu(b) = ba$. Thus

$$t = abbabaabbaababba......$$

The word **t** has remarkable combinatorial properties (cf. [6]). We recall, in particular, that **t** is overlap-free, i. e. it does not have two overlapping occurrences of the same finite factor.

Let now $B = A \cup \{c\}$ and $i : B^* \to B^*$ the morphism defined as:

$$i(a) = abc, \; i(b) = ac, \quad i(c) = b.$$

The Thue-Morse word **m** on three symbols can be defined as the limit sequence obtained by iterating the morphism i on the letter "a"; thus

$$\mathbf{m} = abcacbabcbacabcacbaca......$$

The word **m** is square-free, i. e. it does not have two consecutive equal blocks of letters (cf. [6]).

The relation between **t** and **m** is the following: if $\delta: B^* \to A^*$ is the morphism defined as $\delta(a) = abb$, $\delta(b) = ab$, $\delta(c) = a$ then $\delta(\mathbf{m}) = \mathbf{t}$ (cf. [6]).

The concept of special factor can be extended to **m** as follows: a factor f of **m** is special if there exist two distinct letters x, y of B such that fx and fy are still factors of **m**.

If w is an infinite word on an alphabet A one can introduce the monoid $M(w) = F(w) \cup \{0\}$ where the product is defined as follows. For any $m_1, m_2 \in M(w)$, $m_1 \circ m_2 = m_1 m_2$ if $m_1 m_2 \in F(w)$ and $m_1 \circ m_2 = 0$, otherwise.

In [4] and [5] we have analyzed some combinatorial properties of the special factors of the words **t** and **m** by means of which we found some results concerning the enumeration of special factors and of the factors of **t** and **m**. Moreover an algoritm which allows us to construct the special factors of **t** and **m** has been given.

The completions of factors of **t** (and of **m**) having a common prefix has also been considered and an application of these results to a problem in semigroup has been given.

2. Results

a) Enumeration results

In the following we denote ϕ_t (resp. ϕ_m) and F_t (resp. F_m) simply by ϕ (resp. ψ) and F (resp. G). For any $x \geq 0$, [x] is the integer part of x.

Proposition.1 One has $\phi(0) = 1$, $\phi(1) = 2$, $\phi(3) = 4$. Moreover

$$\phi(n) = \begin{cases} \phi(n/2) & \text{if } n \text{ is even} \\ \phi(n+1) & \text{if } n \text{ is odd}. \end{cases}$$

From this proposition it follows that for any length n there exist only 2 or 4 special factors of t. More precisely one has:

Proposition 2. For any $n \geq 2$:

$$\phi(n) = \begin{cases} 4, & \text{if } n \leq 3\cdot 2^{[\log_2(n-1)]} - 1 \\ 2, & \text{if } n > 3\cdot 2^{[\log_2(n-1)]} - 1 . \end{cases}$$

By the iteration formula $F(n+1) = F(n) + \phi(n)$ one obtains:

Proposition 3. For any $n \geq 2$:

$$F(n+1) = \begin{cases} 4n - 2^{[\log_2(n)]} , & \text{if } n \leq 3\cdot 2^{[\log_2(n)]} - 1 \\ 2n + 2^{[\log_2(n)]} + 1, & \text{if } n > 3\cdot 2^{[\log_2(n)]} - 1 . \end{cases}$$

A similar formula for the function F has also obtained independently by Brlek [3].

In [4] we have shown the existence for any $n > 0$ of a bijection f_n of the set $M_n = S(t) \cap A^n$ and $(T_{2n+1} \cap T_{2n+2}) \cap aA^*$. The map f_n is defined by setting for any $s \in S(m)$, $f_n(s) = \delta(s)a$ if $s \in B^*a \cup B^*b$, and $f_n(s) = \delta(s)ab$ if $s \in B^*c$. From this result it follows that:

Proposition 4. For any $n > 1$, $\psi(n) = \phi(n+1)$.

Proposition 5. $G(1) = 3$ and for any $n > 1$, $G(n) = F(n+1)$.

b) <u>Construction of special factors.</u>

In [4] we gave an algorithm in order to construct for any $n \geq 1$ the special factors of t of length n. The algoritm is based on same lemmas. For any word $w \in A^*$ we denote by w~ the reversed word of w and by w' the word obtained from w by interchancing the letter a with b.

Lemma 6. If $s \in S(t)$ then $s' \in S(t)$.

Lemma 7. If s is a prefix of t then $s^{\sim} \in S(t)$.

Lemma 8. Let $k \geq 0$ and uv a prefix of t such that $|u| = |v| = 3 \cdot 2^k$ then $u^{\sim}$, $(u^{\sim})'$, v and v' are distinct special factors.

Hence the algorithm for costructing for any n the special factors of t of length n is the following. If $n > 3 \cdot 2^{[\log_2(n-1)]} - 1$ then by proposition 2 there are only two special factors. If u is the prefix of t of length n then the two special factors are $u^{\sim}$ and $(u^{\sim})'$.

If $n \leq 3 \cdot 2^{[\log_2(n-1)]} - 1$ then one considers the prefix uv of t of length $3 \cdot 2^{[\log_2(n-1)]}$, with $|u| = |v| = 3 \cdot 2^{[\log_2(n-1)]} - 1$. In this case one can prove that the special factors of length n are the suffixes of length n of the four special factors $u^{\sim}$, $(u^{\sim})'$, v and v' of length $3 \cdot 2^{[\log_2(n-1)]} - 1$.

An algorithm to construct for each n the special factors of m can be obtained by taking into account the bijection f_n of M_n and $(T_{2n+1} \cap T_{2n+2}) \cap aA^*$. The map f_n can be effectively constructed by using the morphism $\delta: B^* \rightarrow A^*$.

c) <u>Completions.</u>

The problem we have considered is to count for any $k > 1$ the number of special factors having the prefix u whose lengths lye in the interval $(|u|(k-1), |u|k)$. The next proposition shows that in any such interval the number of special factors is at most 4.

Proposition 9. Let $u \in F(t)$, $|u| = n \geq 1$. For any $k > 1$ set $S^{(k)} = \{w \in A^+ / uw \in S(t) \text{ and } n(k-1) < |uw| \leq nk\}$. Then $Card(S^{(k)}) \leq 4$.

A similar proposition holds true also in the case of the factors of m. From the preceding proposition one derives an upper bound to the number of possible completions of a factor of t in factors of t having the same length.

Corollary 10. Let $u \in F(t)$, $|u| = n$, k be a fixed positive integer and $m \leq kn$. If $T = \{w \in A^+ / uw \in F(t)$ and $|w| = m\}$ then $Card(T) \leq 4k+2$.

The following proposition shows that two different special factors of t having a common prefix have to be quite "distant".

Proposition 11. Let $uv_1, uv_2 \in S(t)$, $u \in A^*$, $|v_1| \neq |v_2|$. One has

$$||uv_1| - |uv_2|| > |u|/6.$$

d) Thue-Morse monoids.

A semigroup is called weakly-permutable if there exists $n > 1$ such that for any sequence $s_1, s_2, \ldots s_n$ of elements of S there exist two distinct permutations σ and τ of the set $\{1,\ldots,n\}$ such that

$$s_{\sigma(1)}s_{\sigma(2)}\cdots s_{\sigma(n)} = s_{\tau(1)}s_{\tau(2)}\cdots s_{\tau(n)}.$$

S is called permutable if S is weakly-permutable and moreover one of the two permutations can be always taken equal to identity.

It has been shown by Restivo and Reutenauer [7] that a finitely generated and permutable semigroup is finite if and only if it is permutable.

In the case of groups the permutation and the weakly-permutation properties coincide as it has been shown by Blyth [2]. However this is not the case for semigroups even if one makes the hypothesis that S is periodic and finitely generated.

This latter result has been shown by Restivo for the Fibonacci monoid [8] and by us for the monoids M(t) and M(m) (cf.[4]). The monoids M(t) and M(m) are finitely generated periodic and infinite. Moreover by using the previous propositions one derives also that M(t) and M(m) are weakly-permutable (and not permutable).

References

[1].J.Berstel, Mots de Fibonacci, Seminaire d'Informatique Théorique , L.I.T.P.,Universitè Paris VI et VII, Année 1980/81,pp.57-78.

[2].R.D.Blyth,Rewriting products of groups elements, P.H.D.Thesis, 1987, University of Illinois at Urbana-Champain.

[3].S.Brlek, Enumeration of factors in the Thue-Morse word, Proc.s Colloque Montrealais sur la Combinatoire et l'Informatique,to appear.

[4]. A.de Luca and S.Varricchio, Some combinatorial properties of the Thue-Morse sequence and a problem in semigroups ,Dipartimento di Matematica Università di Roma "La Sapienza",June 1987, preprint.

[5].A.de Luca and S.Varricchio, On the factors of the Thue-Morse word on three symbols, Dipartimento di Matematica dell'Università di Roma "La Sapienza" ,September 1987,preprint.

[6].M.Lothaire,Combinatorics on words ,Addison Wesley, Reading, MA, 1983.

[7].A.Restivo and C.Reutenauer, On the Burnside problem for semigroups, J.of Algebra, 89(1984)102-104.

[8].A.Restivo, Permutation properties and the Fibonacci semigroup , 1987,preprint

Aldo De Luca and Stefano Varricchio,
Dipartimento di Matematica,
Università di Roma La Sapienza,
Piazzale A. Moro 5,
I-00185 Roma 1.