INFORMATIQUE THÉORIQUE ET APPLICATIONS

ALDO DE LUCA
STEFANO VARRICCHIO

A combinatorial theorem on p-power-free words and an application to semigroups

Informatique théorique et applications, tome 24, nº 3 (1990), p. 205-228

<http://www.numdam.org/item?id=ITA_1990__24_3_205_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 24, n° 3, 1990, p. 205 à 228)

A COMBINATORIAL THEOREM ON p-POWER-FREE WORDS AND AN APPLICATION TO SEMIGROUPS (*)

by Aldo de LUCA [1] and Stefano VARRICCHIO [1]

Communicated by J.-E. PIN

Abstract. – *Some combinatorial properties of infinite words having a subword complexity which is linearly upper-bounded are considered. The main result is a theorem giving a characterization of infinite words having a linear subword complexity in terms of the number of completions of the factors which do not contain multiple overlaps. An interesting application of this theorem is that the monoid of the factors of an infinite p-overlap-free word is weakly-permutable. This generalizes previous results obtained by Restivo and the authors for the Fibonacci and the Thue-Morse monoids respectively.*

Résumé. – *On considère certaines propriétés combinatoires des mots infinis ayant une croissance linéaire du nombre des facteurs. Le résultat principal est un théorème caractérisant les mots infinis possédant une croissance linéaire en termes de nombres de prolongements des facteurs qui ne contiennent pas de chevauchements multiples. Une application intéressante de ce théorème est que le monoïde des facteurs d'un mot infini sans chevauchement d'ordre $p>0$ possède la propriété de permutation faible. Ceci généralise les résultats précédents obtenus par Restivo et les auteurs pour les mots de Fibonacci et de Thue-Morse respectivement.*

0. INTRODUCTION

The paper is concerned with the study of some combinatorial properties of infinite sequences of letters (or *infinite words*) over a finite alphabet. Infinite words can be described in terms of the factors (or subwords) of finite length occurring in it. A relevant role in our analysis is played by the so-called *special factors*. A factor f of an infinite word $\mathbf{w}$ is called special if there exist at least two distinct letters, x and y, such that fx and fy are still factors.

(*) Received February 1988, revised November 1988.
[1] Dipartimento di Matematica dell'Università di Roma "La Sapienza", Piazzale A. Moro, 2.00185 Roma, Italy.

0988-3754/90/03 205 24/$4.40/© AFCET-Gauthier-Villars

For any infinite word $\mathbf{w}$ one can consider two functions called *subword complexity* and *special-subword complexity* counting for any length n, respectively, the number of factors and the number of special factors of length n occurring in $\mathbf{w}$. These two functions are related as shown in Sec. 2. Cases of particular interest are when (1) the subword complexity of $\mathbf{w}$ is linearly upper-bounded, or (2) the word $\mathbf{w}$ is *p-overlap-free* (or *p-power-free*), *i. e.* $\mathbf{w}$ has no factors of the kind $(uv)^p u$ (resp. u^p) with u different from the empty word and $p>1$.

By using some results of combinatorics on words proved in Sec. 3 it is shown in Sec. 4 (*cf.* Theorem 4.1) that condition (1) is verified if and only if $\mathbf{w}$ satisfies the following property (*k-completion-property*): *for any fixed* $p>1$ *and for any* n *and* k *the number of factors of* $\mathbf{w}$ *of length* $(k+1)n$ *having a common prefix* u *which is a p-overlap-free word of length* n, *is upper-bounded by* Dk, *where* D *is a constant which does not depend on* n.

Moreover in Sec. 5 it is proved that if $\mathbf{w}$ satisfies conditions (1) and (2) then the special-subword complexity is upper-bounded by a constant (*cf.* Theorem 5.2). In Sec. 6 it is shown by examples that the k-completion-property does not hold, in general, if condition (1) is not verified. Also Proposition 5.1 does not hold, in general, if condition (2) is not verified.

In Sec. 7 we consider DOL-infinite words, *i. e.* infinite words such that the set of their finite factors is a DOL-language. To this class of words belong the Fibonacci and the Thue-Morse words in two and three symbols. By using the previous results and a theorem of Ehrenfeucht and Rozenberg [5] on the subword complexity of DOL- languages, one can prove that a p-power-free infinite word having a constant distribution of the letters verifies the k-completion-property.

In Sec. 8 an application to semigroups of the foregoing results is shown. We are able to construct a very large class of infinite monoids which are *weakly permutable* in the sense of Blyth [2] and not *permutable* in the sense of Restivo and Reutenauer [9]. This construction generalizes widely the previous results obtained by Restivo [10] in the case of the Fibonacci monoid and by us (*cf.* [3], [4]) in the case of Thue-Morse monoids in two and three symbols respectively.

1. PRELIMINARIES

Let A be a finite non-empty set or *alphabet* and A^* the *free-monoid* over A. The elements of A are usually called *letters* and those of A^* *words*. The identity element of A^* is called *empty word* and denoted by Λ. In the following we set $A^+ = A^* \setminus \{\Lambda\}$. For any word $w \in A^*$, $\mathrm{alph}(w)$ denotes the set of letters which occur in w and $|w|$ the *length* of w. The length of Λ is taken equal to 0. A *language* L over A is any subset of A^*.

A word f is a *factor* of a word w if there exist h, $h' \in A^*$ such that $w = hfh'$. If $h = \Lambda$ (resp. $h' = \Lambda$) then f is called a *prefix* (resp. *suffix*) of w. If h or h' are different from Λ then the factor f is said to be proper. For any word $w \in A^*$ we denote by $F(w)$ the set of all its factors. A same factor f of a word w can occur many times as a factor in w. Any particular occurrence is then determined by the pair (h, h'), or context, of $A^* \times A^*$ such that $w = hfh'$.

A word w has a *period* q, $0 < q \leqq |w|$, if w can be factorized as $w = (uv)^p u$, with u, $v \in A^*$, $p > 0$ and $q = |uv|$. Let $p > 1$; if $u = \Lambda$ (*resp.* $u \neq \Lambda$) then $w = v^p$ and we say that w is a *p-power* (resp. *p-overlap*). A word w which is a 2-power (resp. 2-overlap) is simply called *square* (resp. *overlap*). It follows that a word w is a square (resp. an overlap) if and only if w has a period q such that $2q = |w|$ (resp. $2q > |w|$). A word w is called *primitive* if it is not a p-power for all $p > 1$.

For any integer $p > 1$ we say that the word w is *p-power-free* (resp. *p-overlap-free*) if $F(w)$ does not contain p-powers (resp. p-overlaps). If w is p-overlap-free then it is obviously $(p+1)$-power-free. A 2-power-free word (resp. 2-overlap-free word) is usually called *square-free* (resp. *overlap-free*).

Let H, K be languages over A; $H^{-1}K$ and KH^{-1} will denote the sets $H^{-1}K = \{w \in A^* \mid Hw \cap K \neq \varnothing\}$, $KH^{-1} = \{w \in A^* \mid wH \cap K \neq \varnothing\}$. For any real number $x \geqq 0$, $[x]$ and $\lceil x \rceil$ denote, respectively, the integer part of x and the smallest integer greater than or equal to x.

An *infinite word* $\mathbf{w}$ over the alphabet A is any map $\mathbf{w} : N \to A$. We denote $\mathbf{w}$ also as:

$$\mathbf{w} = w_0 w_1 \ldots w_n \ldots,$$

where for any $n \in N$, $w_n = \mathbf{w}(n)$.

For any infinite word $\mathbf{w}$, we denote by $F(\mathbf{w})$ the set of all its factors of finite length. Then we say that $\mathbf{w}$ is p-power-free (resp. p-overlap-free) if all the words of $F(\mathbf{w})$ are p-power-free (resp. p-overlap-free). We set $\mathrm{alph}(\mathbf{w}) = \bigcup_{u \in F(\mathbf{w})} \mathrm{alph}(u)$.

The *subword complexity* (or *structure-function*) of an infinite word $\mathbf{w}$ is the map $f_{\mathbf{w}} : N \to N$ defined for all $n \in N$ as:

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

To give some examples we introduce now the following three important infinite words:

$$\mathbf{f} = abaababaabaab\ldots, \qquad \mathbf{t} = abbabaabbaaba\ldots$$

$$\mathbf{m} = abcacbabcbacab\ldots$$

f is called the Fibonacci word and **t** and **m** the Thue-Morse words in two and three symbols respectively. They can be constructed as the limit-words obtained by iterating on the letter "a" the following three morphisms $\theta, \mu : \{a, b\}^* \to \{a, b\}^*$ and $\psi : \{a, b, c\}^* \to \{a, b, c\}^*$ defined as (*cf.* [1, 8]):

$$\theta(a) = ab, \qquad \theta(b) = a, \qquad \mu(a) = ab, \qquad \mu(b) = ba$$

$$\psi(a) = abc, \qquad \psi(b) = ac, \qquad \psi(c) = b.$$

It is well known that **f** is 4-power-free (*cf.* [7]), **m** is square-free and **t** is overlap-free. Moreover for any $n \in N$ one has that (*cf.* [1, 3, 4]):

$$f_{\mathbf{f}}(n) = n + 1, \qquad 3n \leqq f_{\mathbf{t}}(n+1) \leqq 10n/3$$

and $f_{\mathbf{m}}(n) = f_{\mathbf{t}}(n+1)$.

2. SPECIAL FACTORS OF AN INFINITE WORD

Let $\mathbf{w}$ be an infinite word over the alphabet A of cardinality $q > 1$. A factor s of $F(\mathbf{w})$ is called a *right special factor*, or simply, *special factor* if there exist at least two distinct letters $x, y \in A$ such that $sx, sy \in F(\mathbf{w})$. We denote by $S(\mathbf{w})$ the set of all special factors of $\mathbf{w}$ and by $\varphi_{\mathbf{w}} : N \to N$ the map defined for all $n \in N$ as

$$\varphi_{\mathbf{w}}(n) = \operatorname{Card}(S(\mathbf{w}) \cap A^n).$$

For any $n \in N$, $\varphi_{\mathbf{w}}(n)$ counts the number of special factors of $\mathbf{w}$ of length n; we call $\varphi_{\mathbf{w}}$ the *special-subword complexity* of $\mathbf{w}$. The value $\varphi_{\mathbf{w}}(0)$ can be taken as equal to 1 since Λ has to be considered as a special factor. It is obvious from the definition that *any suffix of a special factor is still a special factor.*

Special factors of the Fibonacci word and of the Thue-Morse words in two and three symbols have been studied in [1] and [3, 4] respectively. It has been shown that in the case of the Fibonacci word for all $n \in N$ there exists one and only one special factor of length n whereas in the case of the words **t** and **m** for all $n > 1$ the number of special factors is either 2 or 4.

The subword complexity of $\mathbf{w}$ and the special-subword complexity of $\mathbf{w}$ are related for all $n \in N$ by the following basic inequality

$$f_{\mathbf{w}}(n)+\varphi_{\mathbf{w}}(n) \leqq f_{\mathbf{w}}(n+1) \leqq f_{\mathbf{w}}(n)+(q-1)\,\varphi_{\mathbf{w}}(n). \tag{2.1}$$

In fact, for any factor u of $\mathbf{w}$ of length n there exists at least one letter x such that $ux \in F(\mathbf{w})$; however when u is a special factor then there exist $r \in [1, q-1]$ more letters $x_i \in A \setminus \{x\}$ $(i=1, \ldots, r)$ each different from the other and such that $ux_i \in F(\mathbf{w})$. By iteration of Eq (2.1) one obtains

$$f_{\mathbf{w}}(1)+\sum_{s=1,\ldots,n} \varphi_{\mathbf{w}}(s) \leqq f_{\mathbf{w}}(n+1) \leqq f_{\mathbf{w}}(1)+(q-1)\sum_{s=1,\ldots,n} \varphi_{\mathbf{w}}(s) \tag{2.2}$$

Hence if one knows the special-subword complexity $\varphi_{\mathbf{w}}$ then one can determine an upper and a lower bound to the subword complexity $f_{\mathbf{w}}$. We remark that when $q=2$, Eq (2.1) becomes simply

$$f_{\mathbf{w}}(n+1)=f_{\mathbf{w}}(n)+\varphi_{\mathbf{w}}(n), \qquad n>0.$$

The same formula holds if $\mathbf{w}$ is a square-free word in a three letter alphabet A. In fact in this case any factor of $\mathbf{w}$ of length n is right-prolongable by the letters of A, at least in one factor and at most in two factors of $\mathbf{w}$.

From Eq. (2.1) and the fact that the suffixes of special factors are still special factors one easily derives that $f_{\mathbf{w}}$ *is upper limited by a constant if and only if there exists an integer* n_0 *such that* $\varphi_{\mathbf{w}}(n_0)=0$. From this and Eq. (2.2) it follows that when $\lim f_{\mathbf{w}}(n)$ is infinite then for all $n \geqq 0$, $f_{\mathbf{w}}(n) \geqq n+1$.

Another consequence of the fact that the set $S(\mathbf{w})$ of the special factors of $\mathbf{w}$ is closed by suffixes is that for all $n \geqq 0$

$$\varphi_{\mathbf{w}}(n) \geqq (1/q)\,\varphi_{\mathbf{w}}(n+1) \tag{2.3}$$

in fact if one drops the first letter to the $\varphi_{\mathbf{w}}(n+1)$ special factors of $\mathbf{w}$ of length $n+1$ one obtains at least $(1/q)\,\varphi_{\mathbf{w}}(n+1)$ special factors of length n.

In a symmetric way one can define also *left-special factors*. A factor s of $\mathbf{w}$ is a left-special factor if there exist at least two distinct letters $x, y \in A$ such that $xs, ys \in F(\mathbf{w})$. We denote by $\psi_{\mathbf{w}}$ the function $\psi_{\mathbf{w}}: N \to N$ which gives for any integer $n \geqq 0$ the number $\psi_{\mathbf{w}}(n)$ of the left special factors of $\mathbf{w}$ of length n.

We observe that for any factor $s \in F(\mathbf{w})$ there exists always at least a letter x for which $xs \in F(\mathbf{w})$ with the only exception when s occurs in $\mathbf{w}$ as a prefix of $\mathbf{w}$ only. Hence for any n the number of factors of $\mathbf{w}$ of length n which cannot be completed on the left by one letter in $F(\mathbf{w})$ is always $\leqq 1$. From this one derives that

$$f_w(n+1) \geqq f_w(n) + \psi_w(n) - 1 \tag{2.4}$$

By comparing Eq. (2.4) and (2.1) it follows that for all $n \geqq 0$

$$\psi_w(n) \leqq 1 + (q-1)\,\varphi_w(n) \tag{2.5}$$

3. MULTIPLE OVERLAPS

A word $w \in A^+$ has an overlap if w has a factor of the kind $(uv)^2 u$, with $u \neq \Lambda$. This is also equivalent to saying that in w there are two distinct and overlapping occurrences of the same non-empty factor of w (*cf.* [8]). We can generalize this result by introducing the concept of *multiple overlap*. Let p be a positive integer; we say that a word w has a multiple overlap of *order p* if there exist $p+1$ distinct occurrences of the same non-empty factor u of w such that any two distinct occurrences of u are overlapping. A multiple overlap of order 1 reduces itself to the usual concept of overlap. It holds the following.

PROPOSITION 3.1: *Let $w \in A^+$ and p a positive integer. The following conditions are equivalent*

(i) *w has a multiple overlap of order p.*

(ii) *There exist $p+1$ factors $u, f_1, \ldots, f_p$ of w such that for all $i (i=1, \ldots, p-1)$ f_i is a proper prefix of f_{i+1} and, moreover, for all $i (i=1, \ldots, p)$, f_i can be factorized as $f_i = \lambda_i s_i \lambda'_i$, $\lambda_i, s_i, \lambda'_i \in A^+$, having*

$$u = \lambda_i s_i = s_i \lambda'_i.$$

(iii) *w has a $2p$-overlap, i.e. w has a factor $(\alpha\beta)^h \alpha$, with $\alpha \neq \Lambda$ and $h \geqq 2p$.*

Proof: (i) $\Rightarrow$ (ii). By hypothesis in w there are $p+1$ distinct occurrences of a non- empty factor u of w. We can then factorize w as $w = k_i u k'_i$, $k_i, k'_i \in A^*$ $(i=0, \ldots, p)$. Since the occurrences are distinct we can denumerate them in such a way that $|k_0| < |k_1| < \ldots < |k_p|$. Thus (k_0, k'_0) determines the left-most occurrence of u in w. By hypothesis all the occurrences determined by the contexts (k_i, k'_i) $(i=1, \ldots, p)$ have to overlap with the left-most. Hence one has for any fixed $i (i=1, \ldots, p)$:

$$w = k_0 u k'_0 = k_i u k'_i \tag{3.1}$$

Since $|k_i| > |k_0| (i=1, \ldots, p)$ one can write

$$k_i = k_0 \lambda_i, \qquad \lambda_i \in A^+, \tag{3.2}$$

with $|\lambda_i| < |u|$. By replacing in Eq. (3.1) k_i by $k_0 \lambda_i$ one obtains

$$u k'_0 = \lambda_i u k'_i \tag{3.3}$$

From this one derives

$$u = \lambda_i s_i, \qquad s_i k'_0 = u k'_i,\ s_i \in A^+.$$

From the second equation it follows $u = s_i \lambda'_i$, $\lambda'_i \in A^+$ so that one has

$$u = \lambda_i s_i = s_i \lambda'_i \qquad (i=1, \ldots, p) \tag{3.4}$$

From Eq. (3.3) one has

$$u k'_0 = \lambda_i u k'_i = \lambda_i s_i \lambda'_i k'_i = f_i k'_i$$

having set $f_i = \lambda_i s_i \lambda'_i$. One has then $f_i k'_i = f_{i+1} k'_{i+1}$ $(i=1, \ldots, p)$. By Eq. (3.2) and (3.4) since $|k_i| < |k_{i+1}|$, it follows $|\lambda_i| = |\lambda'_i| < |\lambda_{i+1}| = |\lambda'_{i+1}|$. Thus $|f_i| < |f_{i+1}|$ and then f_i is a prefix of f_{i+1}.

(ii) $\Rightarrow$ (iii). Under the condition (ii) one has that there exists a non-empty factor u of w such that

$$u = \lambda_i s_i = s_i \lambda'_i \qquad (i=1, \ldots, p) \tag{3.5}$$

From a classical result of Lyndon and Schützenberger on the equations in free monoids one has that for any fixed $i (i=1, \ldots, p)$ the solution of Eq. (3.5) can be expressed as:

$$\lambda_i = (\alpha_i \beta_i)^{q_i}, \qquad \lambda'_i = (\beta_i \alpha_i)^{q_i}$$

$$s_i = (\alpha_i \beta_i)^{r_i} \alpha_i, \qquad u = (\alpha_i \beta_i)^{q_i + r_i} \alpha_i,$$

where $\alpha_i \beta_i$, $\beta_i \alpha_i$ are primitive words, $q_i > 0$, $r_i \geqq 0$, $\alpha_i \neq \Lambda$ $(i=1, \ldots, p)$. Moreover one has that

$$f_i = \lambda_i s_i \lambda'_i = (\alpha_i \beta_i)^{2 q_i + r_i} \alpha_i.$$

We set in the next $\alpha_1 = \alpha$ and $\beta_1 = \beta$. We prove now that for any fixed $i (i=1, \ldots, p)$ one has that $\alpha_1 \beta_1 = \alpha_i \beta_i$.

Since $f_i = f_1 \xi$, $\xi \in A^+$ the word f_1 has periods $m = |\alpha\beta|$ and $n = |\alpha_i \beta_i|$. Moreover $|f_1| = |\lambda_1 s_i \lambda'_1| > m+n$ as $|\lambda_1| \geqq |\alpha\beta|$ and $|u| \geqq |\alpha_i \beta_i|$. Hence from the theorem of Fine and Wilf (*cf.* [8]) f_1 has also a period equal to the greatest common divisor d of m and n. We can then set $m=rd$, $n=sd$ and $\alpha\beta=z^r$, $\alpha_i\beta_i=z^s$, $z\in A^+$. Since $\alpha\beta$ and $\alpha_i\beta_i$ are primitive it follows that $\alpha\beta=\alpha_i\beta_i$.

Let us now prove that for all i, $(i=1, \ldots, p)$, $\alpha=\alpha_i$ (and then $\beta=\beta_i$). One has, in fact, setting $k=q_1+r_1$ and $k_i=q_i+r_i$

$$u=(\alpha\beta)^k\alpha=(\alpha_i\beta_i)^{k_i}\alpha_i=(\alpha\beta)^{k_i}\alpha_i.$$

If $k=k_i$ then $\alpha=\alpha_i$. Suppose now $k>k_i$. One has

$$(\alpha\beta)^{k-k_i}\alpha=(\alpha_i\beta_i)^{k-k_i}\alpha=\alpha_i.$$

This equality implies $k=k_i$ and $\alpha=\alpha_i$, otherwise one would reach a contradiction. In the case $k<k_i$ by a similar argument one derives that $\alpha=\alpha_i$. Hence one has that for all $i\,(i=1, \ldots, p)\,\alpha=\alpha_i$ and $\beta=\beta_i$. From this result one has that

$$f_i=(\alpha\beta)^{2q_i+r_i}\alpha \qquad (i=1, \ldots, p).$$

Since $|\lambda_1|<|\lambda_2|<\ldots<|\lambda_p|$ and $\lambda_i=(\alpha\beta)^{q_i}\,(i=1, \ldots, p)$ one has that $q_1<q_2<\ldots<q_p$. From this it follows that

$$q_p\geqq q_1+p-1\geqq p.$$

Thus $2q_p\geqq 2p$ and $f_p=(\alpha\beta)^h\alpha$, with $h\geqq 2p$.

(iii) $\Rightarrow$ (i). Suppose that w has a factor $(\alpha\beta)^h\alpha$ with $\alpha\neq\Lambda$ and $h\geqq 2p$. We can write w as $w=k(\alpha\beta)^{2p}\alpha k'$, $k, k'\in A^*$. The word w has then $p+1$ distinct occurrences of the factor $u=(\alpha\beta)^p\alpha$ determined by the contexts $(k, (\beta\alpha)^p k')$, $(k\alpha\beta, (\beta\alpha)^{p-1}k')$, $\ldots$, $(k(\alpha\beta)^p, k')$. One easily verifies that the left-most occurrence of u overlaps with all the others. Thus w has a multiple overlap of order p.

Q.E.D.

As an application of the preceding proposition we give the following:

COROLLARY 3.2: *Let u be a non-empty word over A and denote by $f_u(k)$, $k>0$, the number of factors of u of length k. If u is q-overlap-free and $p=\lceil q/2\rceil$ then for any k, $1\leqq k\leqq\lceil |u|/2\rceil$,*

$$f_u(k)\geqq k/p.$$

Proof: Let k be a positive integer such that $1\leqq k\leqq\lceil |u|/2\rceil$. We can factorize u as $u=h_i v_i h_i'$ with $|h_i|=i$, $|v_i|=k$, $0\leqq i\leqq k$. We have then the $k+1$

factors $v_0, \ldots, v_k$ of length k. If $k > f_u(k)p$ then, by the "pigeon-hole" principle, $p+1$ of the above factors will be equal to a same factor v of u. Since any two of these occurrences of v overlap then u has a multiple overlap of order p. From Proposition 3.1, u has a factor which is a q-overlap and this is a contradiction. Hence $f_u(k) \geqq k/p$.

Q.E.D.

A consequence of Corollary 3.2 is that if $\mathbf{w}$ is an infinite q-overlap-free word then for all $n \geqq 0$, $f_{\mathbf{w}}(n) \geqq n$. Indeed by taking any factor u of $\mathbf{w}$ of length $2n\,(n>0)$ one has that $f_u(n) \geqq n/p$ and then $f_{\mathbf{w}}(n) \geqq f_u(n) \geqq n/p$; thus $\lim f_{\mathbf{w}}(n) = \infty$.

4. A COMBINATORIAL THEOREM ON p-POWER-FREE WORDS

Let $\mathbf{w}$ be a given infinite word. We say that $f_{\mathbf{w}}$ is linearly upper bounded if there exists a constant $c>0$ such that $f_{\mathbf{w}}(n) \leqq cn$, for all $n>0$. For any $u \in A^*$ and $r>0$ we denote by $F_{u,r}$ the set:

$$F_{u,r} = \{ uv \in F(\mathbf{w}) \mid |v| = r|u| \},$$

i. e. $F_{u,r}$ is the set of factors of $\mathbf{w}$ of length $(r+1)|u|$ having a common prefix u. In this section we prove the following.

THEOREM 4.1: *The subword complexity $f_{\mathbf{w}}$ of an infinite word $\mathbf{w}$ is linearly upper-bounded if and only if the word $\mathbf{w}$ satisfies the following property* (r-completion-property)*: For any fixed $p>1$ and for any p-overlap-free word $u \in A^*$ and $r>0$ one has that* $\mathrm{Card}(F_{u,r}) \leqq Dr$, *where D is a constant which does not depend on the length of u.*

To prove the theorem we need some preliminary lemmas:

LEMMA 4.2: *Let $\mathbf{w}$ be an infinite word such that $f_{\mathbf{w}}$ is linearly upper-bounded. For all positive integers n, r and h such that $n \equiv 0$ (mod. h) consider the partition of the interval $[0, rn]$ in the sub-intervals $[sn/h, (s+1)n/h)$. Let us denote by μ_s the minimal value of $\varphi_{\mathbf{w}}$ in the interval $[sn/h, (s+1)n/h)$. One has that $\sum_{s \in [0, rh)} \mu_s$ is upper-bounded by Cr, where C is a constant which does not depend on n.*

Proof: By hypothesis $f_{\mathbf{w}}$ is linearly upper bounded, *i. e.* $f_{\mathbf{w}}(n) \leqq cn$, for all $n>0$, where c is a suitable positive integer. From Eq. (2.2) one has that

$$f_{\mathbf{w}}(rn) \geqq f_{\mathbf{w}}(0) + \sum_{x \in [0, rn)} \varphi_{\mathbf{w}}(x) \geqq \sum_{s \in [0, rh)} \sum_{j \in [sn/h, (s+1)n/h)} \varphi_{\mathbf{w}}(j) \geqq (n/h) \sum_{s \in [0, rh)} \mu_s.$$

Since $f_w(rn) \leqq crn$ one derives that

$$\sum_{s \in [0, rh)} \mu_s \leqq hcr. \tag{4.1}$$

Q.E.D.

LEMMA 4.3: *Let H be a finite subset of A^+ whose elements have a common prefix (or suffix) $u \in A^+$ and such that the lengths of the words of $u^{-1} H$ (resp. $H u^{-1}$) lie in the interval $[l, L]$ with $l < L$ and $L - l < \lceil |u|/2 \rceil$. If u is p-overlap-free then*

$$\mathrm{Card}(H) \leqq \lceil p/2 \rceil m,$$

where m is the minimal value of cardinality of the set $(A^*)^{-1} H \cap A^k$ [resp. $H(A^*)^{-1} \cap A^k$] for k ranging in the interval $([|u|/2] + L, |u| + l]$.

Proof: Let $t = \lceil p/2 \rceil$. We prove that if $\mathrm{Card}(H) > tm$ then u has a multiple overlap of order t, so that by Proposition 3.1 the result will follow. We consider here only the case in which the elements of H have a common prefix u which is a p-overlap-free word. A perfect symmetric proof can be done in the case in which the elements of H have a common suffix u which is a p-overlap-free word.

We denote by k_0 an integer in the interval $([|u|/2] + L, |u| + l]$ such that $\mathrm{Card}((A^*)^{-1} H \cap A^{k_0}) = m$. Since there exist only m distinct suffixes of length k_0 of the words of H, from the "pigeon-hole" principle there must exist $t+1$ words $h_0, h_1, \ldots, h_t \in H$ having the same suffix s of length k_0. Moreover the words of H have the same prefix u, so that we can write

$$h_i = u w_i \qquad (i = 0, ..., t).$$

Since $k_0 > L$ then for any pair $i, j \in \{0, ..., t\}$, $i \neq j$, one has $|h_i| \neq |h_j|$. In fact, otherwise, $h_i = h_j$ which is a contradiction. One can then always suppose that

$$|h_0| < |h_1| < \ldots < |h_t|. \tag{4.2}$$

Since $k_0 < |u| + l$ one can write

$$h_i = p_i s, \qquad p_i \in A^* \qquad (i = 0, \ldots, t).$$

Moreover $k_0 = |s| > [|u|/2] + L \geqq [|u|/2] + |w_i|$ $(i = 0, \ldots, t)$ so that one has

$$s = v_i \lambda w_i \text{ and } u = p_i v_i \lambda \qquad (i = 0, \ldots, t) \tag{4.3}$$

with $v_i \in A^*$ and $|\lambda| = [|u|/2]+1$.

From Eqs. (4.2) and $(4.3)_1$ one derives

$$|v_0| > |v_1| > \ldots > |v_t|;$$

hence from Eq. $(4.3)_1$ one has that for all $i (i=1, \ldots, t)$ $v_i \lambda$ is a prefix of $v_0 \lambda$, *i. e.*

$$v_0 \lambda = v_i \lambda \zeta_i \, (i=1, \ldots, t), \qquad \zeta_i \in A^+.$$

From Eq. $(4.3)_2$ one has $|v_i| \leqq |u| - |\lambda| < [|u|/2]+1 = |\lambda| \, (i=0, \ldots, t)$. Thus it follows that in $v_0 \lambda$ there are $t+1$ distinct occurrences of the factor λ such that any two distinct occurrences of λ are overlapping. Since $v_0 \lambda$ is a factor of u, one has from Proposition 3.1 that u contains a p-overlap that is a contradiction.

Q.E.D.

LEMMA 4.4: *Let* $\mathbf{w}$ *be an infinite word and denote for any* $u \in A^+$ *and* $r > 0$ *by* $S_{u,r}$ *the set:*

$$S_{u,r} = \{ uv \in S(\mathbf{w}) \mid 0 \leqq |v| < r |u| \}.$$

One has that $\mathrm{Card}(F_{u,r}) \leqq 1 + (q-1)\,\mathrm{Card}(S_{u,r})$, *where* $q = \mathrm{Card}(A)$.

Proof: Suppose that r is a fixed positive integer and denote by $P_i (i=0, 1, \ldots, r|u|)$ the set of prefixes of length $|u|+i$ of the words of $F_{u,r}$. One has that $P_0 = \{u\}$ and $P_{r|u|} = F_{u,r}$. We set, moreover, $T_i = P_i \cap S(\mathbf{w})$ $(i=0, 1, \ldots, r|u|)$. One has for $i \in [0, r|u|)$

$$\mathrm{Card}(P_{i+1}) \leqq \mathrm{Card}(P_i) + (q-1)\ \mathrm{Card}(T_i).$$

In fact for any $f \in P_i$ there exists at least one letter x such that $fx \in P_{i+1}$; moreover when f is special (*i.e.* $f \in T_i$) there exist $r \in [1, q-1]$ more letters $x_j \in A \setminus \{x\}$ $(j=1, \ldots, r)$ different each other and such that $fx_j \in P_{i+1}$. By iterating the above formula one has:

$$\mathrm{Card}\ (F_{u,r}) \leqq \mathrm{Card}\ (P_0) + (q-1) \sum_{i \in [0, r|u|)} \mathrm{Card}\ (T_i).$$

Since Card $(P_0)=1$ and $S_{u,r}=\bigcup_{i\in[0,r|u|)} T_i$ the result follows.

Q.E.D.

Proof of Theorem 4.1

By hypothesis $f_{\mathbf{w}}$ is linearly upper-bounded, *i.e.* $f_{\mathbf{w}}(n)\leqq cn$ for all $n>0$, where c is a suitable positive constant. If u is not a factor of $\mathbf{w}$ then $F_{u,r}=\emptyset$ and the result is trivially true. Let us then suppose that $u\in F(\mathbf{w})$ and denote by n the length of u. We consider first the case that $n\equiv 0$ (mod. 4).

If $n=0$ then Card $(F_{u,r})=1$ and also in this case the result is trivial. Let us then suppose that $n\geqq 4$. We consider the set $S_{u,r}=\{uv\in S(\mathbf{w})\mid 0\leqq |v|<rn\}$ and we show first that Card $(S_{u,r})$ is less then or equal to dr, where d is a constant which does not depend on n. For any fixed u we can decompose $S_{u,r}$ as follows:

$$S_{u,r}=\bigcup_{j\in[1,r]}\bigcup_{i\in[0,3]} Q_{i,j}, \tag{4.4}$$

where

$$Q_{i,j}=\{uv\in S(\mathbf{w})\mid (j-1)n+in/4\leqq |v|<(j-1)n+(i+1)n/4\}.$$

We observe that the lengths of the words of $u^{-1}Q_{i,j}$ range in the interval $[(4j+i-4)n/4, (4j+i-3)n/4-1]$. Thus from Lemma 4.3 one has:

$$\text{Card }(Q_{i,j})\leqq\lceil p/2\rceil\, m_{i,j}, \tag{4.5}$$

where $m_{i,j}$ is the minimal value of the cardinality of the set $(A^*)^{-1}Q_{i,j}\cap A^k$ of the suffixes of $Q_{i,j}$ of length k, for k ranging in the interval $[(4j-1+i)n/4, (4j+i)n/4]$.

Since $(A^*)^{-1}Q_{i,j}\cap A^k$ is included in $S(\mathbf{w})$ one has that $m_{i,j}\leqq\mu_{i,j}$ where $\mu_{i,j}$ is the minimal value of $\varphi_{\mathbf{w}}$ in the interval $[(4j-1+i)n/4, (4j+i)n/4)$. Hence from Eq. (4.4) it follows:

$$\text{Card }(S_{u,r})\leqq\lceil p/2\rceil\sum_{j\in[1,r]}\sum_{i\in[0,3]}\mu_{i,j}.$$

For all $s\in[0, 4(r+1))$ denote by μ_s the minimal value of $\varphi_{\mathbf{w}}$ in the interval $[sn/4, (s+1)n/4)$. By Lemma 4.2 and Eq. (4.1) one has

$$\sum_{j\in[1,r]}\sum_{i\in[0,3]}\mu_{i,j}\leqq\sum_{s\in[0,4(r+1))}\mu_s\leqq 4c(r+1).$$

Hence

$$\mathrm{Card}\,(S_{u,\,r}) \leqq 4c \lceil p/2 \rceil (r+1).$$

By using Lemma 4.4 one obtains

$$\mathrm{Card}\,(F_{u,\,r}) \leqq 1+(q-1)\,\mathrm{Card}\,(S_{u,\,r}) \leqq 1+4c(q-1)\lceil p/2 \rceil (r+1) \leqq dr,$$

where d is a constant which does not depend on n. (The constant d can be taken equal to $1+8c(q-1)\lceil p/2 \rceil$).

Let us now suppose that n is not congruent to 0 (mod. 4). In this case let n' be the smallest integer greater than n such that $n' \equiv 0$ (mod. 4). We denote by $u_1, \ldots, u_t$ the completions of u in factors of $\mathbf{w}$ of length n'. Since $n'-n \leqq 3$ one has obviously that $t \leqq (\mathrm{Card}\,A)^3 = q^3$. Moreover any element of $F_{u,\,r}$ is a prefix of some element of $F_{u_j,\,r}$ for some $j \in [1, t]$. Hence one has:

$$\mathrm{Card}\,(F_{u,\,r}) \leqq \mathrm{Card}\,\Big(\bigcup_{j \in [1,\,t]} F_{u_j,\,r}\Big) \leqq \sum_{j \in [1,\,t]} \mathrm{Card}\,(F_{u_j,\,r}) \leqq q^3\,dr.$$

Thus in any case $\mathrm{Card}\,(F_{u,\,r}) \leqq Dr$, where D is a constant depending only on p, q and c.

Conversely let us suppose that $\mathrm{Card}\,(F_{u,\,r}) \leqq Dr$. Any word of length 1 can be completed on the right in a word of length n in at most $D(n-1)$ ways. Hence for any $n>1$ one has

$$f_{\mathbf{w}}(n) \leqq f_{\mathbf{w}}(1)\,D(n-1) < q\,Dn.$$

Q.E.D.

Let us now denote for any $r>0$ by $G_{u,\,r}$ the set

$$G_{u,\,r} = \{ vu \in F(\mathbf{w}) \mid |v| = r\,|u| \},$$

i.e. the set of all the factors of w of length $(r+1)|u|$ having a common suffix u. It holds the following theorem symmetric to Theorem 4.1:

THEOREM 4.5: *The subword complexity $f_{\mathbf{w}}$ of an infinite word is linearly upper bounded if and only if* $\mathbf{w}$ *satisfies the following property: For any fixed $p>1$ and for any p-overlap-free word $u \in A^*$ and $r>0$ one has that* $\mathrm{Card}\,(G_{u,\,r}) \leqq Kr$, *where K is a constant which does not depend on the length of u.*

Proof: The proof is symmetric and similar to the proof of Theorem 4.1. For this reason we shall not give it in the details. One has take into account the set $L(\mathbf{w})$ of left special factors and for any $r>0$ the set

$L_{u,\,r}=\{vu\in L(\mathbf{w})\mid 0\leqq|v|<r|u|\}$. One first prove that Card $(L_{u,\,r})\leqq Cr$ where C is a constant which does not depend on $|u|$. To this end we observe that from Eq. (2.5), $\psi_{\mathbf{w}}(n)\leqq 1+(q-1)\,\varphi_{\mathbf{w}}(n)$ so that Lemma 4.2 holds also for the function $\psi_{\mathbf{w}}$ enumerating the left special factors of $\mathbf{w}$. In fact denoting, for any s in the interval $[0, rh)$, by ω_s the minimal value of $\psi_{\mathbf{w}}$ in the interval $[sn/h, (s+1)\,n/h)$, one has

$$\omega_s\leqq 1+(q-1)\,\mu_s \qquad \text{and} \qquad \sum_{s\in[0,\, rh)} \omega_s\leqq rh+(q-1)\,hcr=(h+(q-1)\,hc)\,r.$$

One has then to use the version of Lemma 4.3 in which H is a set of words having a common p-overlap-free suffix u. From this one derives that Card $(L_{u,\,r})\leqq Cr$. Finally in a way symmetric to that of Lemma 4.4, one proves that Card $(G_{u,\,r})\leqq 1+(q-1)$ Card $(L_{u,\,r})$. From this the result follows.

Q.E.D.

5. THE CASE OF INFINITE p-OVERLAP-FREE WORDS

In this section we shall consider some consequences of Theorems 4.1 and 4.5 in the case of infinite p-overlap-free words. From Theorem 4.1 one obviously derives as a corollary the following:

PROPOSITION 5.1: *Let* $\mathbf{w}$ *be an infinite p-overlap-free word. The subword complexity of* $\mathbf{w}$ *is linearly upper-bounded if and only if for any* $u\in A^*$ *and* $r>0$, Card $(F_{u,\,r})\leqq Dr$, *where* D *is a constant which does not depend on the length of* u.

The next theorem gives some insight on the special factors of a p-overlap-free word:

THEOREM 5.2: *Let* $\mathbf{w}$ *be an infinite p-overlap-free word. The subword complexity* $f_{\mathbf{w}}$ *is linearly upper bounded if and only the special-subword complexity* $\varphi_{\mathbf{w}}$ *is upper-limited by a constant.*

Proof: Let $u\in A^*$ and h be an integer such that $0\leqq h\leqq |u|$. We denote by $Q_{u,\,h}$ the set $Q_{u,\,h}=\{vu\in F(\mathbf{w})\mid 0\leqq|v|=h\leqq|u|\}$. We first prove that Card $(Q_{u,\,h})\leqq g$, where g is a constant which does not depend on h and on $|u|$. To this end we observe that for $h=0$, Card $(Q_{u,\,0})=1$. For $h>0$ one has

$$\text{Card}\,(Q_{u,\,h+1})\leqq\text{Card}\,(Q_{u,\,h})+(q-1)\,\text{Card}\,(Q_{u,\,h}\cap L(\mathbf{w})).$$

One has only to note that all the elements of $Q_{u,\,h}$ with the possible exception of one can be completed on the left in elements of $F(\mathbf{w})$. By iteration of this formula one derives that

$$\operatorname{Card}(Q_{u,\,h+1}) \leqq 1+(q-1) \sum_{j \in [0,\,h]} \operatorname{Card}(Q_{u,\,j} \cap L(\mathbf{w})) \leqq 1+(q-1) \operatorname{Card}(L_{u,\,1}).$$

As we have seen in the proof of Theorem 4.5, Card $(L_{u,\,1}) \leqq C$, where C does not depend on $|u|$. From this it follows that Card $(Q_{u,\,h}) \leqq g$, where the constant g does not depend on h and on $|u|$.

Let us now prove that for all $n>0$, $\varphi_{\mathbf{w}}(n)$ is upper limited by a constant. We shall do this by considering first the even values of the arguments of $\varphi_{\mathbf{w}}$. Let j be any integer in the interval $[n, 2n]$. From the above result a special factor of length j can be completed at the left in a factor of length $2n$ in at most g ways, where g does not depend on n and on j. Hence one can write that:

$$\varphi_{\mathbf{w}}(2n) \leqq g\, \varphi_{\mathbf{w}}(j), \quad \text{for all } j \in [n, 2n].$$

Since $f_{\mathbf{w}}$ is linearly upper-bounded, *i.e.* $f_{\mathbf{w}}(n) \leqq cn$, $n>0$, one has

$$2nc \geqq f_{\mathbf{w}}(2n) \geqq \sum_{j \in [n,\,2n]} \varphi_{\mathbf{w}}(j) \geqq (n/g)\, \varphi_{\mathbf{w}}(2n)$$

and then $\varphi_{\mathbf{w}}(2n) \leqq 2cg$.

Let us now consider arguments of $\varphi_{\mathbf{w}}$ which are odd integers. From Eq. (2.3) one has that

$$\varphi_{\mathbf{w}}(2n+1) \leqq q\, \varphi_{\mathbf{w}}(2n) \leqq 2cgq.$$

Thus in any case $\varphi_{\mathbf{w}}$ is upper limited by the constant $2cgq$.

Let us now suppose that $\varphi_{\mathbf{w}}$ is upper limited by a constant. From Eq. (2.2) it follows that $f_{\mathbf{w}}$ is linearly upper-bounded.

Q.E.D.

6. REMARKS AND EXAMPLES

A first remark is that since a p-power-free word is p-overlap-free and, conversely, a p-overlap-free word is $(p+1)$-power-free then one can replace in Theorems 4.1, 4.5, 5.2 and Proposition 5.1 the term p-overlap-free by p-power-free obtaining new propositions substantially equivalent to the previous ones.

As we said in Section 2 the Fibonacci word and the Thue-Morse word in two or three symbols are such that the subword complexity is linearly upper-bounded. Hence Proposition 5.1 and Theorem 5.2 hold true for these infinite words.

Example 1 shows that Proposition 5.1, as well as Theorem 4.1, do not in general hold if the subword complexity of a p-overlap-free infinite word is not linearly but only quadratically upper bounded.

Example 1: Let $A=\{a, b\}$, $B=A\cup\{c\}$ and for any $i>0$ denote by t_i the prefix of length i of the Thue-Morse word $\mathbf{t}$ on A. We introduce the infinite word τ on the alphabet B defined as

$$\tau=t_1\, c\, t_2\, cc\, t_3\, c\, t_4\, cc \ldots t_{2n-1}\, c\, t_{2n}\, cc\, t_{2n+1}\, c\, t_{2n+2}\, cc \ldots \tag{6.1}$$

The word τ verifies the following properties the proof of which is reported in the Appendix:

PROPERTY 6.1: The word τ is overlap-free.

PROPERTY 6.2. The subword complexity of τ is of quadratic order, *i.e.* there exist positive integers c and C such that for all $n>0$

$$cn^2 \leqq f_\tau(n) \leqq C n^2.$$

We show now that τ does not verify the 1-completion property. In fact for any $i>0$ the factor t_i of τ has at least $i+1$ completions of length $2i$ in factors of τ. Indeed suppose that i is odd (the case i even is dealt in a similar way). The words $t_{i+2r}\, c\, t_{i+2r+1}$ and $t_{i+2r+1}\, cc\, t_{i+2r+2}$ ($r=0, \ldots, [i/2]$) are $i+1$ factors of τ having $i+1$ distinct prefixes of length $2i$ and the same prefix t_i of length i.

The following example shows that Proposition 5.1 does not hold if we miss the hypothesis that the infinite word is p-overlap-free.

Example 2: Let $\mathbf{u}$ the infinite word over $A=\{a, b\}$:

$$\mathbf{u}=aba^2ba^4b \ldots ba^n b \ldots, \qquad \text{with} \quad n=2^i, \qquad i\geqq 0.$$

The word $\mathbf{u}$ has factors which can be arbitrarily large powers. Moreover one has (*cf.* Appendix) that

PROPERTY 6.3: The subword complexity of $\mathbf{u}$ is linearly upper bounded.

Proposition 5.1 does not hold true for **u**. In fact consider for any $n>0$ the factor $f=a^n$ with $n=2^i$; f has $n+1$ completions of length $2n$ in factors of **u**. They are fv_i $(i=0, \ldots, n)$ with $v_0=a^n$ and $v_i=a^{n-i}ba^{i-1}$ $(i=1, \ldots, n)$.

7. INFINITE DOL-WORDS

Let L be a language over the alphabet A. We denote by $F(L)$ the set of the factors of the words of L and by $F_L: N \to N$ the map defined as

$$F_L(n)=\text{Card } (F(L) \cap A^n).$$

F_L is also called the *subword complexity* of the language L (*cf.* [5, 6]). L is said to have a *constant distribution* if there exists an integer c such that any word $u \in F(L) \cap A^{c+1} A^*$ is such that $\text{alph}(u)=\Delta$, where Δ is a subset of A. In other terms if L has a constant distribution then in all sufficiently long factors of the words of L occur all letters belonging to a same subset of A. Let $p>1$. A language L is called *p-overlap-free* (resp. *p-power-free*) if $F(L)$ does not contain p-overlaps (resp. p-powers).

A language L over A is a DOL-*language* (*cf.* [5]) if there exists a word $\omega \in A^*$ and a morphism $h: A^* \to A^*$ such that

$$L=\{ h^n(\omega) \mid n \geqq 0 \},$$

where $h^0(\omega)=\omega$. The triplet $G=(A, h, \omega)$ is called DOL-*system* and the language L is denoted by $L(G)$.

We say that an infinite word **w** is an infinite DOL-*word* if the set $F(\mathbf{w})$ of its factors is equal to the set $F(L)$ of the factors of a DOL-language L. Moreover we say that **w** has a constant distribution if $F(\mathbf{w})$ (or L) has a constant distribution. One easily recognizes that the Fibonacci word **f** and the Thue-Morse words **t** and **m** are infinite DOL-words since they are defined by iterated morphisms starting on one letter.

The subword complexity of square-free DOL-languages and of DOL-languages with a constant distribution has been studied quite extensively by Ehrenfeucht and Rozenberg in [5] and [6]. We recall here the following important

THEOREM 7.1: *Let L be a DOL-language that thas a constant distribution. Then the subword complexity is linearly upper bounded.*

From this result and Proposition 5.1 one derives in the case of infinite DOL-words with a constant distribution, the following:

PROPOSITION 7.2: *Let* $\mathbf{w}$ *be an infinite p-overlap-free (resp. p-power-free) infinite DOL-word with a constant distribution. Then for any* $r>0$ *and* $u \in A^*$ *one has that* Card $(F_{u,\, r}) \leqq Dr$ *where D is a constant which does not depend on the length of u.*

Proof: By hypothesis $F(\mathbf{w})=F(L)$ where L is a DOL-language with a constant distribution. Since $\mathbf{w}$ is p-overlap-free then the language L will be $(p+1)$-power-free. From Theorem 7.1 it follows that the subword complexity F_L is linearly upper-bounded. Since for any $n \geqq 0$, $f_{\mathbf{w}}(n)=F_L(n)$ then from Proposition 5.1 the result follows.

Q.E.D.

It is noteworthy that in some cases of interest one can eliminate in the preceding proposition the hypothesis that $\mathbf{w}$ has a constant distribution. This occurs, for instance, when Card (alph $(\mathbf{w})) = 2$ and $\mathbf{w}$ is p-overlap-free (or p-power-free) or when Card (alph $(\mathbf{w})) = 3$ and $\mathbf{w}$ is square-free. Indeed in the first case one has that any finite factor v of $\mathbf{w}$ of length $>p$ is such that Card (alph $(v)) = 2$ so that $\mathbf{w}$ has a constant distribution. In the second case $\mathbf{w}$ has a constant distribution since any square-free word in a three letter alphabet of length >3 has to contain three letters. In particular one has that the Fibonacci word and the Thue-Morse words in two and three symbols are DOL-words with a constant distribution.

8. AN APPLICATION TO SEMIGROUPS

Let L be a language over A. One can define the *monoid* $M(L)$ *of the factors* of L as $M(L)=F(L) \cup \{0\}$ where the product $(\circ)$ in $M(L)$ is defined as: for any $m_1, m_2 \in M(L)$

$$m_1 \circ m_2 \begin{cases} = m_1 m_2 & \text{if} \quad m_1, m_2 \text{ and } m_1 m_2 \in F(L) \\ = 0, & \text{otherwise.} \end{cases}$$

We observe that $J_L = A^* \setminus F(L)$ is a two-sided ideal of A^* and that $M(L)$ is isomorphic to the Rees-quotient monoid A^*/J_L. It is clear that, by the finiteness of A, that $M(L)$ is finitely generated. Moreover if L is p-power-free (resp. p-overlap-free) then $M(L)$ is torsion. Indeed if L is p-power-free then for any $m \in F(L)$, $m \neq \Lambda$, the power m^p does not belong to $F(L)$ so that the product $m \circ m \circ \ldots \circ m$, p times, is 0. Thus for any $m \in M(L)$ one has $m^p = m^{p+1}$.

We shall refer in the following to the case in which $L = F(\mathbf{w})$ where $\mathbf{w}$ is an infinite word. In this case the monoid $M(F(\mathbf{w}))$ will be simply denoted by $M(\mathbf{w})$ and called the *monoid of the factors* of $\mathbf{w}$. The monoids $M(\mathbf{f})$, $M(\mathbf{t})$ and $M(\mathbf{m})$ are called respectively the *Fibonacci monoid* and the *Thue-Morse monoids* in two and three generators.

We recall now the following property of semigroups called *permutation property* (*cf.* [9]) :

Let S be a semigroup and $n > 1$. S is called *n-permutable* if for any sequence $s_1, \ldots, s_n$ of n elements of S there exists a non-trivial permutation σ of the symmetric group $\mathscr{S}_n$, such that

$$s_1 \ldots s_n = s_{\sigma(1)} \ldots s_{\sigma(n)}.$$

S is called *permutable* if there exists an integer $n > 1$ for which S is n-permutable.

The importance of this permutation property, which is a generalization of commutativity, is due to the following theorem of Restivo and Reutenauer [9]:

THEOREM 8.1: *A finitely generated and torsion semigroup S is finite if and only if S is permutable.*

A weaker permutation property may be introduced by the following definition (*cf.* [2, 3, 4, 10]):

A semigroup S is *n-weakly permutable*, $n > 1$, if for any subset $\{s_1, \ldots, s_n\}$ of n elements of S there exist two permutation $\sigma, \tau \in \mathscr{S}_n$, $\sigma \neq \tau$, for which

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

S is weakly permutable if there exists $n > 1$ such that S is n-weakly permutable.

In the case of groups the two above concepts coincide as it has been shown by Blyth [2]:

THEOREM 8.2: *A weakly permutable group is permutable.*

In the case of semigroups the permutation property and the weak-permutation property do not coincide, in general, even if one makes the hypothesis that the semigroups are finitely generated and torsion. This latter fact has been recently shown by Restivo [10] for the Fibonacci monoid $M(\mathbf{f})$ and de Luca and Varricchio [3, 4] for the Thue-Morse monoids $M(\mathbf{t})$ and $M(\mathbf{m})$. The proof given in [3] makes use of a method which has been widely generalized in this paper so that we are now able to obtain a more general result including all previous cases.

PROPOSITION 8.3: *Let* $\mathbf{w}$ *be an infinite word such that the subword complexity is linearly upper-bounded and p be a fixed integer* >1. *Then there exists a constant D with the porperty that for any integer k such that* $(k-1)!>D$ *the following property holds: For any set* $\{u, u_1, \ldots, u_k\}$ *of* $k+1$ *words of* A^+ *such that u is p-overlap-free and* $|u| \geqq |u_i|$ $(i=1, \ldots, k)$ *one of the two following conditions is verified:*

(i) *there exist two permutations* $\sigma, \tau \in \mathscr{S}_k$, $\sigma \neq \tau$, *for which*

$$u_{\sigma(1)} \ldots u_{\sigma(k)} = u_{\tau(1)} \ldots u_{\tau(k)}.$$

(ii) *there exist at least* $k! - \mathrm{Card}\,(F_{u,k})\,(\geqq k! - Dk)$ *permutations* $\sigma \in \mathscr{S}_k$ *such that the words* $u u_{\sigma(1)} \ldots u_{\sigma(k)}$ *are different each other and such that*

$$u u_{\sigma(1)} \ldots u_{\sigma(k)} \in A^* \setminus F(\mathbf{w}).$$

Proof: Let us consider the set $V = \{u_{\sigma(1)} \ldots u_{\sigma(k)} \in A^* \mid \sigma \in \mathscr{S}_k\}$. If $\mathrm{Card}\,(V) < k!$ then obviously there exist two permutations $\sigma, \tau \in \mathscr{S}_k$, $\sigma \neq \tau$, for which condition (i) is verified. Let us then suppose that $\mathrm{Card}\,(V) = k!$. We set $m = \sum_{i \in [1, k]} |u_i| \leqq k|u|$ and consider the set $T_{u,m} = \{uv \in F(\mathbf{w}) \mid |v| = m\}$. Since any element of $T_{u,m}$ is a prefix of one element of $F_{u,k}$ then from Theorem 4.1 one has that

$$\mathrm{Card}\,(T_{u,m}) \leqq \mathrm{Card}\,(F_{u,k}) \leqq Dk.$$

Let us set

$$V_{u,k} = \{u u_{\sigma(1)} \ldots u_{\sigma(k)} \in F(\mathbf{w}) \mid \sigma \in \mathscr{S}_k\}.$$

One has that $V_{u,k}$ is included in $T_{u,m}$ so that

$$\mathrm{Card}\,(V_{u,k}) \leqq \mathrm{Card}\,(T_{u,m}) \leqq \mathrm{Card}\,(F_{u,k}).$$

Hence there exist at least $k! - \mathrm{Card}\,(F_{u,k})$ permutations $\sigma \in \mathscr{S}_k$ such that condition (ii) is verified.

Q.E.D

From this proposition we derive the following:

COROLLARY 8.4: *Let* $\mathbf{w}$ *be an infinite p-overlap-free (resp. p-power-free) word. If the subword complexity is linearly upper-bounded then* $M(\mathbf{w})$ *is weakly permutable.*

Proof: By Proposition 5.1 for any $k>0$, Card $(F_{u,\,k}) \leqq Dk$. Let now k be the minimal integer such that $k!>Dk+1$. We shall prove that $M(w)$ is $(k+1)$-weakly permutable. In fact let $\{m, m_1, \ldots, m_k\}$ be any subset of $k+1$ elements of $M(\mathbf{w})$. If at least one of the elements is 0 then the result is trivial. Let us then suppose that m and $m_i (i=1, \ldots, k)$ belong to $F(\mathbf{w})$ and that $|m| \geqq |m_i| (i=1, \ldots, k)$. From the preceding proposition if condition (i) is verified then obviously the result follows. If condition (ii) is verified then there exist at least 2 permutations $\sigma, \tau \in \mathscr{S}_k$, $\sigma \neq \tau$, for which

$$m m_{\sigma(1)} \ldots m_{\sigma(k)},\ m m_{\tau(1)} \ldots m_{\tau(k)} \in A^* \setminus F(\mathbf{w}).$$

Thus in all cases one derives that

$$m \circ m_{\sigma(1)} \circ \ldots \circ m_{\sigma(k)} = m \circ m_{\tau(1)} \circ \ldots \circ m_{\tau(k)},$$

i.e. $M(\mathbf{w})$ is $(k+1)$-weakly permutable.

Q.E.D.

In the case of infinite DOL-words one derives from Theorem 7.1 and Corollary 8.4 the following:

COROLLARY 8.5: *Let* $\mathbf{w}$ *be an infinite p-overlap-free (p-power free) DOL-word with a constant distribution. Then the monoid* $M(\mathbf{w})$ *is weakly permutable.*

It should be remarked that the monoid $M(\mathbf{w})$ in the above two corollaries is finitely generated, torsion and infinite so that from Theorem 8.1 it cannot be permutable. As a consequence of the previous corollaries one derives that $M(\mathbf{f})$, $M(\mathbf{t})$ and $M(\mathbf{m})$ are weakly permutable and not permutable monoids.

ACKNOWLEDGMENTS

We acknowledge the anonymous referees for their useful comments.

REFERENCES

1. J. BERSTEL, *Mots de Fibonacci*, Seminaire d'Informatique Théorique, L.I.T.P. Université Paris-VI et VII, Année 1980/81, pp. 57-78.
2. R. D. BLYTH, *Rewriting Products of Groups Elements*, Ph. D. Thesis, 1987, University of Illinois at Urbana-Champain.
3. A. DE LUCA and S. VARRICCHIO, *Some Combinatorial Properties of the Thue-Morse Sequence and a Problem in Semigroups*, Theoretical Computer Science, Vol. 63, 1989, pp. 333-348.
4. A. DE LUCA and S. VARRICCHIO, *On the Factors of the Thue-Morse Word on Three Symbols*, Information Processing Letters, Vol. 27, 1988, pp. 281-285.

5. A. EHRENFEUCHT and G. ROZENBERG, *On the Subword Complexity of DOL Languages with a Constant Distribution*, Information Processing Letters, Vol. 13, 1981, pp. 108-113.

6. A. EHRENFEUCHT and G. ROZENBERG, *On the Subword Complexity of Square-Free DOL Languages*, Theoretical Computer Science, Vol. 16, 1981, pp. 25-32.

7. J. KARHUMAKI, *On Cube-Free ω-Words Generated by Binary Morphisms*, Discrete Applied Mathematics, Vol. 5, 1983, pp. 279-297.

8. M. LOTHAIRE, *Combinatorics on Words*, Addition Wesley, Reading, MA, 1983.

9. A. RESTIVO and C. REUTENAUER, *On the Burnside Problem for Semigroups*, J. of Algebra, Vol. 89, 1984, pp. 102-104.

10. A. RESTIVO, *Permutation Properties and the Fibonacci Semigroup*, Semigroup Forum, Vol. 38, 1989, pp. 337-345.

APPENDIX

Proof of property 6.1

By construction the word τ has no factors such as (i) $cucvc$ and $cuc\,\xi\,cvc$ with $u, v \in A^+$, $\xi \in B^*$ and $|u|=|v|$, (ii) $ccvcc$, $v \in A^*$ and (iii) $ucvcw$ with $u, v, w \in A^+$. Let us now prove that τ is overlap-free. In fact suppose that τ has an overlap, *i.e.* a factor such as $f=xvxvx$ with $x \in B$ and $v \in B^*$. It follows that $v \in B^* \setminus A^*$. Indeed suppose that $v \in A^*$. If $x \in A$ one reaches a contradiction since $\mathbf{t}$ is overlap-free. If $x=c$ then f will have a factor such as $cvcvc$ which is again a contradiction in view of property (i). We can then factorize v as $v=v_1cv_2$, $v_1 \in A^*$, $v_2 \in B^*$ and write $f=xv_1\,cv_2xv_1cv_2x$. If $v_1 \neq \Lambda$ then from properties (iii) and (ii) one easily derives that v_2 has to contain an occurrence of the letter c. The same occurs if $v_1=\Lambda$. In fact in this case if $v_2 \in A^*$ then $f=xcv_2xcv_2x$ so that from property (ii) x has to be different from the letter c and then from property (iii) one reaches a contradiction. We can then write $v_2=v_3cv_4$, $v_3 \in A^*$, $v_4 \in B^*$ and $f=xv_1\,(cv_3c)\,v_4xv_1\,(cv_3c)\,v_4x$. From property (i) it follows that $v_3=\Lambda$ so that $f=xv_1ccv_4xv_1ccv_4x$. From property (ii) one derives that v_4 has to contain an occurrence of the letter c so that $v_4=v_5cv_6$, $v_5 \in A^*$, $v_6 \in B^*$ and

$$f=xv_1c\,(cv_5c)\,v_6xv_1c\,(cv_5c)\,v_6x.$$

One has that $v_5 \neq \Lambda$ and from property (i) one reaches again a contradiction.

Q.E.D.

Proof of property 6.2

Let us first prove that for any $n>0$, $f_\tau(n) \geqq n(n+1)$. In fact let $n>1$, $i \in \{1, \ldots, n-1\}$ and w any factor of $\mathbf{t}$ of length i. From the combinatorial properties of the Thue-Morse sequence $\mathbf{t}$ (*cf.* [8]) it follows that there exist infinitely many words $\lambda \in A^*$ such that λw is a prefix of $\mathbf{t}$. Hence there exists always a $\lambda \in A^*$ such that $\lambda w = t_k$ and $k > n-i-1$. Thus if k is odd then $wct_{k+1} \in F(\tau)$ and $wct_{n-i-1} \in F(\tau) \cap B^n$ and if k is even then $wcct_{n-i-2} \in F(\tau) \cap B^n$. Since there are $f_{\mathbf{t}}(i)$ factors of $\mathbf{t}$ of length i and for any $i>0$, $f_{\mathbf{t}}(i) \geqq 2i$ (*cf.* [3]) one derives

$$f_\tau(n) \geqq \sum_{i=1, \ldots, n-1} f_{\mathbf{t}}(i) \geqq \sum_{i=1, \ldots, n-1} 2i = n(n+1).$$

We show now that f_τ is quadratically upper bounded. For any $n>0$ let $w \in F(\tau) \cap B^n$. We suppose n to be odd (a similar argument can be followed if n is even). One of the two following possibilities can occur

(i) $w \in F(v)$ with $v = t_1 c t_2 cc t_3 c \ldots c t_{n+1}$

(ii) $w \in F(\mathbf{r})$ with $\mathbf{r} = t_{n+1} cc t_{n+2} c \ldots$

In the case (i) the number of factors of length n is upper limited by the length $|v|$ of v, so that one has

$$\text{Card}\,(F(v) \cap B^n) \leqq |v| = \sum_{i=1, \ldots, n+1} |t_i| + 3(n+1)/2 - 2 < (n+1)(n+5)/2$$

In the case (ii) we have to distinguish three subcases (a) w is a factor of $\mathbf{t}$. In this case the number of factors of length n is upper-limited by $4n$ (*cf.* [3]). (b) w has only one occurrence of the letter c, *i.e.* $w = uct_{n-i-1}$ with $u \in F(\mathbf{t}) \cap A^i$ ($i = 0, \ldots, n-1$). The number of factors of length n is then upper limited by

$$\sum_{i=1, \ldots, n-1} f_{\mathbf{t}}(i) \leqq 2n(n-1).$$

(c) w has two occurrences of the letter c, *i.e.* $w = ucct_{n-i-2}$ with $u \in F(\mathbf{t}) \cap A^i$ ($i = 0, \ldots, n-2$). In a similar way one obtains an upper bound to the number of factors of length n given by $1 + 2(n-2)(n-1)$.

Thus in any case an upper bound to $f_{\tau}(n)$ is given by $5n^2+9$ and the result follows.

Q.E.D.

Proof of property 6.3

We have to show that the subword complexity of the word $\mathbf{u}$ is linearly upper bounded. In fact let $n>0$, k the minimal integer such that $n \leqq 2^k$ and $w \in F(\mathbf{u}) \cap A^n$. We have to consider two cases:

(i) $w \in F(v)$ where $v = aba^2b^2 \ldots a^m b$ with $m = 2^k$.

(ii) $w \in F(\mathbf{r})$ with $\mathbf{r} = a^m ba^{2m} b \ldots$

In the case (i) an upper bound to the number of factors of length n is given by the length of the word v. Since $|v| = k + 2^{k+1}$ and $2^{k-1} < n \leqq 2^k$ one has $|v| \leqq k + 4n < 1 + \log_2 n + 4n \leqq 5n + 1$. In the case (ii) since the number of the occurrences of the letter b in w is at most 1, the number of factors of length n is upper bounded by $n+1$.

Thus in any case one has that $f_{\mathbf{u}}(n)$ is upper limited by $6n+2$.

Q.E.D.