ELSEVIER Journal of Pure and Applied Algebra 108 (1996) 301–313

JOURNAL OF
PURE AND
APPLIED ALGEBRA

Special factors of automatic sequences

Théodore Tapsoba

Département de Mathématiques, Faculté des Sciences et Techniques, Université de Ouagadougou, 03 B P 7021 Ouagadougou 03, Burkina Faso

Communicated by M.-F. Roy, received 1 June 1994; revised 15 December 1994

Resumé

Nous décrivons une méthode récurrente de détermination des facteurs spéciaux de suites automatiques.

Abstract

We give an inductive method to determine the special factors of some automatic sequences.

1. Introduction

The study of factors of infinite sequences goes back at least to Thue [10, 11]. Among the questions which have been addressed is the problem of computing the complexity function P, where $P(n)$ is the number of factors of length n.

We quote here some results obtained by the analysis of the special factors of particular sequences on sets of two elements:

- computing the complexity function of the Thue–Morse sequence [3, 6] and the Fibonacci's one [2].
- construction of an automaton for computing the sequence $P(n+1)$-$P(n)$ for some infinite words [9].

The study of the special factors of sequences whose complexity function is $2n+1$ (therefore defined on sets of three elements) and satisfying some technical requirements, show that they can be represented by an exchange of six intervals [1]. (This extends a classical result on representation of Sturmian sequences by rotations [8]). The sequence is represented by a graph which is used as a foundation to build a kind of abstract generalization of continued fraction expansion, on three elements. The proofs use classical results of graph theory and can easily be extended to sequences whose complexity function is $(k+1)n+1$ (therefore defined on sets of k elements) and satisfying the following condition (*):

SSDI 0022-4049(95)00043-7

(∗) For any integer n, there exists a unique special factor of length n. Such sequences can be represented by an exchange of $2k$ intervals.

We think that a generalization of the procedure used in [1] will allow us to get a geometric representation of (at least) sequences with linear increase. Before trying to get an eventual representation, we have to investigate sequences satisfying (∗).

Our aim is to give an inductive method to determine the special factors of automatic sequences (fixed points of injective constant length substitutions) without constraint in the number of the elements of the set. This (certainly) will make us know more about sequences satisfying (∗).

2. Preliminaries

Let A^* be the free monoid generated by a non-empty finite set A called alphabet. The elements of A are called letters and those of A^* words. For any word v in A^*, $|v|$ denotes the length of v, namely the number of its letters. The identity element of A^* denoted by ε is the empty word; it is the word of length 0. A word v is said to be a factor of w if $w = xvy$ for some x, y in A^*. We then write $v \mid w$. If $x = \varepsilon$ (resp. $y = \varepsilon$), v is called a prefix (resp. suffix) of w. A prefix or a suffix of w is said to be a strict one if it is different from w.

We denote by $M(A)$ the set $A^* \cup A^\infty$ where A^∞ is the set of infinite words with letters in A.

We call substitution, a morphism $f: A \to A^*$. It can be naturally extended to a morphism from A^* to A^*. A substitution is said to be a constant length σ substitution if $\sigma = |f(i)|$ for any letter i of A. If there exists a letter $a \in A$ such that $f(a) = am$ with $|m| > 0$, then the set of words with prefix a has a fixed point $u = am\ f(m)\ f^2(m) \dots f^k(m) \dots$

Let $k \geq 2$ be an integer and let $[k]$ denote the set $\{0, 1, \dots, k-1\}$. A k-automaton is given by

(i) to alphabets Σ and E,
(ii) an initial point $x_0 \in \Sigma$,
(iii) an application $\varphi: [k] \times \Sigma \to \Sigma$,
(iv) an application $\tau: \Sigma \to E$.

For any couple $(j, x) \in [k] \times \Sigma$, let $\varphi(j, x) = j(x)$ or more simply jx. The application $\varphi: [k] \times \Sigma \to \Sigma$ can be naturally extended to an application from $[k]^* \times \Sigma$ to Σ in this way: let $e_g, e_{g-1}, \dots, e_0 \in [k]$ and let $x \in \Sigma$. Inductively, set $e_g e_{g-1} \dots e_0(x) = e_g e_{g-1} \dots e_1(e_0(x))$ and $\varphi(\varepsilon, x) = x$.

Let $n \geq 1$ be an integer. We develop n as follows: $n = \sum_{j=0}^{\infty} e_j(n)\ k^j$. Let g be the largest integer contained in $\log n / \log k$. If $j > g$ put $e_j(n) = 0$ and $e_g(n) \neq 0$. Therefore, n can be represented by $e_g(n)\ e_{g-1}(n) \dots e_0(n) \in [k]^*$ and 0 is represented by the empty word. For any $x \in \Sigma$, define $nx = e_g(n)\ e_{g-1}(n) \dots e_0(n)\ x$. Then when n strokes the sequence of all integers, $(nx_0)_{n \in N}$ is an infinite sequence of elements in Σ and

$(\tau(nx_0))_{n\in N}$ is an infinite sequence of elements in E. A sequence t is said to be k-recognizable if there exists a k-automaton $(\Sigma, x_0, \varphi, E, \tau)$ such that $t = (\tau(nx_0))_{n\in N}$.

We have the following result [4].

Proposition 1. *Let E be a not empty finite set, $t = (t_n) \in E^N$ and k a prime integer. Then the following are equivalent:*

(i) *The sequence t is recognizable by a k-automaton.*

(ii) *t is generated by a substitution of constant length k.*

(iii) *There exists a finite field K of characteristic k and an injective application $\alpha : E \to K$ such that $\alpha(t)$ is algebraic over $K[X]$.*

Remark 1. Without supposing k prime, the equivalence between (i) and (ii) has been proved in [5].

Example 1. The sequence $1 < 2 < 4 < 7 < 8 < \ldots$ of the integers which sum of the digits in their base 2 expansion is odd is recognizable by the 2-automaton $(\Sigma, x_0, \varphi, E, \tau)$ where $\Sigma = \{i, s\}$, $E = \{0, 1\}$, $x_0 = i$, $\varphi(0, i) = i$, $\varphi(0, s) = s$, $\varphi(1, i) = s$, $\varphi(1, s) = i$, $\tau(i) = 0$, $\tau(s) = 1$. Note that it is the fixed point in $0A^*$ of the substitution f_0 on the alphabet $A = \{0, 1\}$ defined by $f_0(0) = 01$ and $f_0(1) = 10$.

Let u be a finite or infinite word. We denote by $F(u)$ the set of finite factors of u and by $F_n(u)$ its subset consisting of the factors of length n. If u is an infinite word, it is trivial to verify that every factor of a word v of $F(u)$ is a word of $F(u)$ and that there exists a letter a such that va is in the set $F(u)$. The factor v of u is said to be special if for any letter i of A, vi is a factor of u. Denote by $FS(u)$ the set of the special factors of u and by $FS_n(u)$ the set of the special factors of length n.

Let S be the shift defined by $S(a_0\, a_1\, a_2 \ldots) = a_1\, a_2 \ldots$ and let Ω be the closure of the set $\{S^k(u); k \in N\}$ where the distance d is given by $d(v, w) = \exp(\text{-}\inf\{n \in N; v_n \neq w_n\})$. The sequence u is associated with the dynamical system (Ω, T) (where T is the restriction of S to Ω) and it is said to be minimal when the empty set and Ω are the only closed subsets of Ω invariant under T.

Example 2. The Morse sequence, fixed point in $1A^*$ of the substitution f_1 on the alphabet $A = \{1, 2\}$ defined by $f_1(1) = 12, f_1(2) = 21$ is minimal.

Example 3. The fixed point in $1A^*$ of the substitution f_2 on the alphabet $A = \{1, 2, 3\}$ defined by $f_2(1) = 121, f_2(2) = 232$ and $f_2(3) = 323$ is not minimal.

The following classical characterization has been proved in [7].

Proposition 2. *The word u is minimal if and only if for any factor m of u, there exists an integer j depending on m such that*

$$\text{for any } k \in N, \quad m \mid u_k\, u_{k+1} \ldots u_{k+j}. \tag{1}$$

Remark 2. The Condition (1) means that every factor m of u appears in u with bounded lacunas.

We give here a simple criterium for minimality.

Proposition 3. *Let u be a fixed point of the substitution f on the alphabet A. If a is a prefix of u with $|f(a)| \geq 2$ and if every letter of A is a factor of u, then the following are equivalent:*

(i) *u is minimal and* $\lim |f^k(b)| = +\infty$ *for every letter $b \in A$.*
(ii) *There exists $L \leq \mathrm{Card}(A)$ such that for any $b \in A, a \mid f^L(b)$.*
(iii) *For any $b \in A$, there exists $k(b) \in N$ such that $a \mid f^{k(b)}(b)$.*

Proof. (i) $\Rightarrow$ (iii) is a consequence of Proposition 2 and of the fact that every letter of A is a factor of u. Let us suppose (iii) and let us prove (ii). We set $L = \mathrm{Max}\{k(b); b \in A\}$ where $k(b) = \mathrm{Min}\{n; a \mid f^n(b)\}$. As $a \mid f(a)$, one has $a \mid f^s(b)$ for every $s \geq k(b)$. Hence, $a \mid f^L(b)$ for every $b \in A$. Furthermore, for any letter $b \neq a$, there is a letter c such that $k(c) = k(b) - 1$; hence $L \leq \mathrm{Card}(A)$. Let us suppose (ii) and let us prove (i). In particular, $|f(b)| > 0$ for every $b \in A$. As $f(a) \in aA^+$, $\lim |f^k(a)| = +\infty$. For every $b, a \mid f^L(b)$; hence, $\lim |f^k(b)| = +\infty$. As a is prefix of u note that $u = f^{\omega}(a)$. Let m be a factor of u: There is j such that $m \mid f^j(a)$. Hence, $m \mid f^{j+L}(b)$ for every $b \in A$.

So m appears in u with lacunas bounded by $2(\mathrm{Max}\{|f(b)|; b \in A\})^{j+L} - 2$. By Proposition 2, u is minimal. □

From now, we will consider minimal sequence u, fixed points of injective constant length σ substitutions which are not periodic.

3. Some properties of factors and special factors

Let w be a factor of u. It can be decomposed as follows:

$$w = xf(v)y. \tag{2}$$

In (2) x is a strict suffix of a word $f(v_1)$, y is a strict prefix of a word $f(v_2)$ and $v_1, v_2 \in A$.

A factor w of u is said to be rythmical if it has a unique decomposition with condition (2).

Example 4. In the Morse sequence, 122 and 221 are rythmical while 121 and 212 are not rythmical.

Proposition 4. *If there exists a rythmical factor R of u with $|R| \geq \sigma$, then every factor of u which has R as a factor is rythmical.*

Proof. Let m be a factor of u such that R is a factor of m. The decomposition of u by blocks of σ letters, namely $u = f(u_0)f(u_1)\ldots$ gives a decomposition of R read in m.

Since there is only one decomposition, one has a prefix B of a word in $f(A)$ of length $|B| < \sigma$ such that if $R = u_k u_{k+1} \dots u_{k+r}$, then $k + |B|$ is a multiple of σ. The position of $u_{k+|B|}$ in R, and so in m, gives a decomposition of m induced by that of u, but which does not depend on the reading of m in u. Let $m = \alpha(m_1)(m_2)\dots(m_t)\beta$ be this decomposition, where the words m_i belong to $f(A)$, and, α like β satisfy $|\alpha|, |\beta| < \sigma$. Since f is injective, m is a rythmical factor. □

Let m and u be in A^*. u is said to be a "bifix" of m if it is both a prefix and a suffix of m. Hence, there exists v and v' in A^* such that $m = uv = v'u$.

The following two lemmas describe the structure of factors which have the same "bifixes".

Lemma 1. *Let m, u and v be factors such that $m = uv = vu$, u and v non-empty, and let us set $d = \gcd(|u|, |v|)$. Then $m = p^r$ where p is the prefix of m of length d and $r = |m|/d$.*

Proof. Let us first prove that for $k = E(|m|/|v|)$ (where $E(x)$ denotes the largest integer contained in the real x) one has $m = p(v)v^k = vp(v)v^{k-1}$ where $p(v)$ is the prefix of v of length $|m| - k|v|$. This decomposition is trivial if $|u| \le |v|$ since in this case one has $u = p(v)$ and $k = 1$. Let us suppose $|u| > |v|$; from $uv = vu$ one derives $u = vu'$ so that $vu'v = vvu'$ and $u'v = vu'$. Inductively, $u = v^k u''$ with $|u''| \le |v|$ and $u''v = vu''$ so that $u'' = p(v)$ and m has the required form.

The lemma is clearly true if $|u| = |v|$ or if one of the factors u or v is a letter. Let us suppose the lemma true for factors m such that $|m| < K$. We can suppose $|u| > |v|$, and the preceding argument gives $m = p(v)v^k = vp(v)v^{k-1}$. Then $m' = p(v)v(=vp(v))$ and $|m'| < K$. Furthermore, $\gcd(|u|, |v|) = \gcd(|v|, |p(v)|) = d$, so that by the induction hypothesis $m' = \pi^{r'}$ with $r' = |m'|/d$ and π a prefix of v.

Hence, there exists two integers s and t such that $p(v) = \pi^{sd}$ and $v = \pi^{td}$. Finally, $m = \pi^{sd+ktd}$. □

Lemma 2. *Let m, u, v and v' be factors such that $m = uv = v'u$ and $uv' = vu$. Then $v = v'$ and $m \in p^*$ where p is the prefix of m of length $\gcd(|u|, |v|)$.*

Proof. If $|u| \ge |v|(=|v'|)$, v and v' are both prefixes of u and have the same length. Then $v = v'$ and Lemma 1 holds. Let us suppose $|u| < |v|$ and let us set $v = v_1 u$ and $v' = v'_1 u$. Hence, $uv_1u = v'_1uu$ and $uv'_1u = v_1uu$ so that $uv_1 = v'_1$ and $uv'_1 = v_1u$. We are then led to the preceding case with $|v_1| = |v| - |u|$. Inductively, one obtains for $k = E(|v|/|u|)$: $v = v_k u^k$ and $v' = v'_k u^k$ with $uv_k = v'_k u$ and $uv'_k = v_k u$. As $|u| \ge |v_k|$, one has $v_k = v'_k$ and Lemma 2 derives directly from Lemma 1. □

Let u be a fixed point in $1A^*$ of a substitution of constant length σ on the set A, and let us suppose that all the letters of A are factors of u. It is clear that if u is periodic, with the period of length σ^v, then all the factors $f^v(a)$ are the same for any $a \in A$. Reciprocally, if there exists v such that all the factors $f^v(a)$, $a \in A$, are the same, then

$f^v(1)$ is a periodic word of u of length σ^v. If any period of u is not a power of σ, the preceding property is not valid. If u is not periodic on an alphabet of two letters, the property is true, i.e. f is injective. In other words:

Lemma 3. *Let f be a constant length substitution on the set $A = \{1, 2\}$, with the fixed point u in $1A^*$ and not periodic. Then $f^k(1) \neq f^k(2)$ for any integer k and for any factors m and m' of A^*, one has $f(m) = f(m') \Rightarrow m = m'$.*

Let m be a factor of u and set $L_u(m) = \operatorname{Sup}\{k; m^k \mid u\}$. Let us estimate $L_u(m)$ when m is a letter or a word of two distinct letters.

Lemma 4. *Let u be a non-periodic minimal sequence, fixed point of a constant length σ substitution f on the set $A = \{1, 2\}$, and let us suppose u in $1A^\infty$. Then $\sigma^2 + \sigma$ (resp. $\sigma^3 + \sigma^2 + \sigma$) is an upper bound for $L_u(1)$ and $L_u(2)$ (resp. $L_u(12)$ and $L_u(21)$).*

Proof. Since u is not periodic, one has $f(1) \notin 1^*$. Furthermore, $f(2) \notin 2^*$ because u is minimal. Two cases can occur.

Case 1: *Let us first suppose that* $f(2) \notin 1^*$. Then for any $i \in A$, one has $L_u(i) < 2\sigma (\leq \sigma^2 + \sigma)$. In fact, if there exists $j \in A$ such that $L_u(j) \geq 2\sigma$ so $j^{2\sigma}$ is a factor of u. As f is a constant length σ substitution, one of the words $f(1)$ or $f(2)$ is a factor of $j^{2\sigma}$. As $f(1) \notin 1^*$, from $f(1) \in 1A^*$ and $f(2) \notin 2^*$ one derives $f(2) = 1^\sigma$; a contradiction.

Let us prove now that for any word $ij \in A^2$ such that $i \neq j$, one has $L_u(ij) < 2\sigma^2 + \sigma (\leq \sigma^3 + \sigma^2 + \sigma)$. Let us suppose that for a choice of ij one has $L_u(ij) \geq 2\sigma^2 + \sigma$. Thus, the word $(ij)^{2\sigma^2 + \sigma}$ is a factor of u and can be decomposed as follows: $s(f(a))f(b_1) \ldots f(b_{2\sigma})p(f(c))$, where $b_k \in A$ for $k \in \{1, \ldots, 2\sigma\}$ and $a, c \in A \cup \{\varepsilon\}$; $s(f(a))$ (resp. $p(f(c))$) denote a suffix of $f(a)$ (resp. a prefix of $f(c)$). If σ is not odd, then for any k, $f(b_k) = (ij)^{\sigma/2}$ or $f(b_k) = (ji)^{\sigma/2}$ and by Lemma 3, one has $b_1 = b_2 = \ldots = b_{2\sigma}$; this is a contradiction because $L_u(b) < 2\sigma$ for every $b \in A$. If σ is odd, for every k such that $1 \leq k \leq \sigma$ one of these cases occurs: $f(b_{2k-1}) = (ij)^{(\sigma-1)/2}(i)$ and $f(b_{2k}) = (ji)^{(\sigma-1)/2}(j)$ or $f(b_{2k-1}) = (ji)^{(\sigma-1)/2}(j)$ and $f(b_{2k}) = (ij)^{(\sigma-1)/2}(i)$. In any case, one has $f(1) = (12)^{(\sigma-1)/2}(1)$ and $f(2) = (21)^{(\sigma-1)/2}(2)$ so that u is periodic; a contradiction.

Case 2: *Let us now suppose that* $f(2) \in 1^*$. Thus, the letter 2 appears only in $f(1)$. As $f(1) \in 1A^*$, $L_u(2) < \sigma$. Let us prove that $L_u(1) < \sigma^2 + \sigma$. If $L_u(1) \geq \sigma^2 + \sigma$, then $1^{(\sigma^2+\sigma)}$ is a factor of u and its following decomposition $s(f(a))f(b_1) \ldots f(b_\sigma)p(f(c))$ implies $f(b_k) = 1^\sigma$ for any $k \in \{1, \ldots, \sigma\}$. Hence, $b_1 = b_2 = \ldots = b_\sigma = 2$ so that $L_u(2) \geq \sigma$; a contradiction.

Let us now estimate $L_u(ij)$, $i \neq j$. If $L_u(ij) \geq \sigma^3 + \sigma^2 + \sigma$, then as we have just shown there exists a factor m of u of length $\sigma^2 + \sigma$ such that $f(m) \in (ij)^*$ or $f(m) \in (ji)^*$. As 1 is a prefix of $f(1)$ and $f(2)$, one has $f(m) \in (12)^*$. Hence σ is not odd, $f(1) = (12)^{\sigma/2}$ and $m = 1^{(\sigma^2+\sigma)}$; a contradiction because $L_u(1) < \sigma^2 + \sigma$. $\square$

Proposition 5. *Let u be a fixed point of an injective substitution of constant length σ on the set $A = \{1, 2, \dots, q\}$. Let us suppose u in $1A^*$, u minimal and not periodic. Then there exists L_0 depending on σ and q such that every factor of u of length $> L_0$ is a rythmical one.*

Let us prove first Proposition 5 when A is a set of two elements:

Proof. Let m be a factor of u such that $|m| > L_0$ and let us suppose that m is not rythmical. Then there exists two different decompositions of m, $Bf(D)C$ and $B'f(D')C'$ with condition (2). We can suppose $|C| \neq |C'|$. In fact if $|C| = |C'|$, then $C = C'$ and $|B| = |B'|$. Thus, $B = B'$ so that $f(D) = f(D')$ and by Lemma 3, $D = D'$. Hence, we can choose $|C| < |C'|$ and set $C' = X_0 C$, $|X_0| = r, 0 < r < \sigma - 1$.

Let us set now $D = a_k \dots a_1$, $D' = a'_g \dots a'_1$. One has $Bf(a_k) \dots f(a_1) = B'f(a'_g) \dots f(a'_1)X_0, k - 1 \leq g \leq k$, where B and B' are, respectively, suffixes of $f(b)$ and $f(b')$ and X_0 prefix of $f(a'_0)$. Moreover, $ba_k \dots a_1$ and $b'a'_g \dots a'_1 a'_0$ are factors of u. Let us set $a_{k+1} = b, a'_{g+1} = b'$ and let us write $f(a_1) = X_1 X_0, f(a'_1) = X_2 X_1$ with $|X_1| = \sigma - r$ and $|X_2| = r$. We then get $Bf(a_k) \dots f(a_2) = B'f(a'_g) \dots f(a'_2)X_2$ and recursively for $i = 1, \dots, k$ we can write

$$f(a_i) = U_i V_i, f(a'_i) = V'_i U_i, f(a_{i+1}) = U_{i+1} V'_i$$

$$\text{with } |U_i| = |U'_i| = r \text{ and } |V_i| = |V'_i| = \sigma - r. \tag{3}$$

Choosing L_0, one has $k \geq \sigma^3 + \sigma^2 + \sigma$. Hence, D is not a factor in a word in $1^* \cup 2^* \cup (12)^*$. Let us write $D = WW'$ with $|W| = E(k/2)$. As $\sigma \geq 2$, $\sigma^2 + \sigma$ is a lower bound for the lengths of W and W' and by Lemma 4, 1 and 2 are both factors of W and W'. If 21 is not a factor of W then W is a factor of a word in $1^* \cup 2^*$ so that 21 is a factor of W'. In the same way if 12 is not a factor of W, then it is a factor of W'. Moreover, if 11 and 22 are not factors of D, then D is a factor of a word in $(12)^*$ which is a contradiction to Lemma 4.

Finally, we have $\{12, 21, 11\} \subset F_2(D)$ or $\{12, 21, 22\} \subset F_2(D)$.

Let us set $f(1) = UV$ and $f(2) = U'V'$ with $|U| = |U'| = r$. By (3), $E = \{VU', V'U, VU\}$ or $E' = \{VU', V'U, V'U'\}$ is included in $\{UV, U'V'\}$. As the hypothesis $f(1) \in 1A^*$ which gives the difference between the letter 1 and the letter 2 will not be used later, we can suppose $E \subset \{UV, U'V'\}$.

If E is a set of one element, then $V = V'$ and $U = U'$ so that $f(1) = f(2)$ which contradicts the fact that u is not periodic. Hence E is a set of two elements. If $U = U'$, then $E = \{VU, V'U\}$ so that $UV = VU$ and $UV' = V'U$ or $UV = V'U$ and $UV' = VU$. Lemma 1 in the first case and Lemma 2 in the second case yields the contradiction $f(1) = f(2)$. If $V = V'$, the same argument gives the same contradiction. We can now suppose $U \neq U'$ and $V \neq V'$. Then E is a set of three elements and this is out of question. □

The preceding proof shows that if a factor D of u is such that the words 12, 21 and one of the words 11 or 22 are both factors of D, then $f(D)$ is rythmical. Let us summarize:

Propositon 6. *Let f be a substitution of constant length σ on the set $A = \{1, 2\}$ and let u be a fixed point of f, minimal and not periodic. Let L_1 be the least integer g such that 12, 21 and one of the words 11 or 22 are both factors of every factor of u of length $\geq g$. Then every factor of u of length $\geq \sigma(L_1 + 1)$ is rythmical.*

Let us prove now Proposition 5 in the general case.

Proof. We shall proceed as follows.

Let u be a fixed point of an injective substitution f of constant length σ on the set $A = \{1, 2, \ldots, q\}$ with $q \geq 3$. Let us suppose that u is minimal and not periodic. We shall show that there exists a constant H depending on σ and q such that if there exists a factor m of u of length $\geq H$ which is not rythmical, then there exists an integer d divisor of σ and a substitution F of constant length σ on an alphabet P such that $P \subset A^d$, $\mathrm{Card}(P) \leq q - 1$ and $F(u) = u$, looking u as an infinite word on the alphabet P. Moreover, m can be read like an element of P^* and one can prove that it is not a rythmical factor on the alphabet P. As the length of m in P is $|m|/d$, we get a contradiction if $|m| \geq \frac{\sigma}{2} L_0(\sigma, q - 1)$.

In the case $A = \{1, 2\}$ the proof was to show that if there does not exists a rythmical factor of any length, then $f(1) = f(2)$. In other words, f could be seen as defined on an alphabet of one letter. We shall follow this idea in three steps, beginning by the generalization of Lemma 2.

Step 1.

Lemma 5. *Let M be a set of words of length σ on an alphabet A. Let $U = P_r(M)$ (resp. $V = S_{\sigma-r}(M)$) be the set of the prefixes (resp. suffixes) of length r (resp. $\sigma - r$) of the words of M. Moreover, let us suppose*

$$V = P_{\sigma-r}(M) \quad \textit{and} \quad U = S_r(M). \tag{4}$$

Let $d = \gcd(\sigma, r)$ and P be the set of the prefixes of length d of the words of M. Then $M \subset P^*$.

Proof. Even if we have to permute U and V, we can suppose $r \leq \sigma - r$. Let v be a word of V. There exists $m \in M$ and $u \in U$ such that $m = uv$. As v is prefix of a word m_1 of M, there exists $u' \in U$ such that $m_1 = vu'$ and by (4), there exists $u_1 \in U$ and $v_1 \in V$ such that $m_1 = vu' = u_1 v_1$. We can then write $v = u_1 w_1$ so that $m = uu_1 w_1$. In other words, $M \subset UUW_1$ $(= U^2 W_1)$ where W_1 is the set of the suffixes of length $\sigma - 2r$ of the words of M. Let W_{k-1} be the set of the suffixes of length $\sigma - kr$ (≥ 0) of the words of M.

Let us now suppose that $M \subset U^k W_{k-1}$ for an integer k such that $\sigma \geq (k + 1)r$. As $V = P_{\sigma-r}(M)$ one has, as above, $V \subset U^k W_k$. $M \subset V'$ implies $M \subset U^{k+1} W_k$ so that $V \subset U^{k+1} W_{k+1}$.

Let us repeat this process until $t = E(\sigma/r)$. As every word m of M can be written as $m = uv$ and $m = v'u'$, respectively, in UV and VU with $V = U^{t-1}W_{t-1}$, one has

$$m = u_1 \dots u_t w = u'_1 \dots u'_{t-1} w' u'_t, \quad u_t, u'_t \in U, \; w, w' \in W_{t-1}. \tag{5}$$

In particular, $u_t w = w' u'_t$.

If $\sigma = t.r$ then W_{t-1} is empty and the lemma is proved. If $\sigma \neq t.r$, let us set $M' = S_\rho(M)$ for $\rho = \sigma - (t-1)r$, $U' = U$ and $V' = W_{t-1}$.

From the hypothesis on U and V one has $S_{\rho-r}(M') = S_{\rho-r}(M) = S_{\rho-r}(V) = V'$ and by (5), $P_{\rho-r}(M') = V'$. Furthermore, $U = S_r(M) = S_r(M') = S_r(V)$ and by (5), one has $S_r(V) = P_r(M')$. Finally, $U' = P_r(M') = S_r(M')$ and $V' = S_{\rho-r}(M') = P_{\rho-r}(M')$.

Hence, we are led back to the preceding case with ρ and $\rho - r$ replacing, respectively, σ and r, and the same gcd. Thus, if Lemma 5 is proved for M', then U' and V' are in P^* so that M is in P^*. As Lemma 5 is evident for $\sigma = 2d$ and if it is true for the multiples $2d, 3d, \dots, kd$, then the preceding argument shows that it is true for $\sigma = (k+1)d$. Hence, Lemma 5 is proved recursively on the multiples of σ. $\square$

Step 2. Let L_2 depending on σ and q be a constant such that for every factor m of u of length $\geq L_2$ one has $F_2(m) = F_2(u)$.

We may choose L_2 such that it depends on f, but in any case $2\sigma^{2q} - 1$ holds always. In fact if $m \mid u$ and $|m| \geq 2\sigma^{2q} - 1$, then there is a factor of m which can be written as $f^{2q}(a)$. Let $E_k(a)$ be the set of the factors of length 2 which appears in one of the words $f(a), \dots, f^k(a)$. If $E_{k+1}(a) = E_k(a)$, this means that for every $uv \in E_k(a)$, one has $F_2(uv) \subset E_k(a)$. So $E_{k+n}(a) = E_k(a)$ for any $n \geq 0$, and this equality holds as soon as $k \geq \mathrm{Card}(A^2) = q^2$.

Let us choose now $H = \sigma(L_2 + 1)$. Let m be a factor of u such that $|m| \geq H$, and let us suppose m not rythmical. Then there exists two different decompositions of m:

$$Bf(D)C = B'f(D')C' \quad \text{with condition (2).} \tag{6}$$

As for the proof for two letters, we can write $D = a_k \dots a_1$, $D' = a'_g \dots a'_1$ with $k - 1 \leq g \leq k$. Thus, $Bf(a_k) \dots f(a_1) = B'f(a'_g) \dots f(a'_1)X_0$ where B and B' are, respectively, suffixes of $f(b)$ and $f(b')$, X_0 prefix of length r of $f(a'_0)$ and $ba_k \dots a_1$ and $b'a'_g \dots a'_1 a'_0$ factors of u. Let us set $b = a_{k+1}$ and $b' = a'_{g+1}$. We have again the relation (3).

Let $M = f(A)$, $U = P_r(M)$ and $V = S_{\sigma-r}(M)$. For a choice of H, one has $k \geq L_2$. Every letter of A appears in $a_k \dots a_1$ and $a'_g \dots a'_1$ so that by (3), $S_r(M) = U$. Furthermore, every letter of A appears also in $a_{k+1} \dots a_2$ and by (3), $P_{\sigma-r}(M) = V$. By Lemma 5, $M \subset P^*$ for $P = P_d(M) = S_d(M)$ with $d = \gcd(\sigma, r)$.

Let us suppose $\mathrm{Card}(P) = \mathrm{Card}(A)$. Then the images $f(a)$ are determinated by their prefixes (or suffixes) of length d and there exists bijections $\beta: A \to U, \tau: A \to V, \beta': A \to U$ and $\tau': A \to V$ such that for every $a \in A, f(a) = \beta(a)\tau(a) = \tau'(a)\beta'(a)$. If the word of two letters $a'a$ is a factor of D, then by (2), $\beta'(a') = \beta(a)$ so that $a = \beta^{-1}\beta'(a')$. Let us set $\delta = \beta^{-1}\beta'$. Then δ is a bijection in A and $bD = b\delta(b) \dots \delta^k(b)$.

As every letter is in D, δ is a bijection order q. Furthermore, as every factor of u of length 2 is in D, they have necessarily the form $a\delta(a)$ so that $u = 1\delta(1)\delta^2(1) \ldots$. This means $u_k = \delta^k(u_0)$ which yields the contradiction that u is periodic.

Let us suppose that $\mathrm{Card}(P) < \mathrm{Card}(A)$ and let us call ψ the canonical imbedding of P^* into A^* which will be extended to $M(P)$. As $f(A) = M \subset P^*$, we can see the words of $M(M)$ like words of $M(P)$. Let $\phi : M(M) \to M(P)$ denote this identification and let $h : M(A) \to M(M)$ be the application induced by the substitution f. Let $F = \phi \circ h \circ \psi$ and $\hat{u} = \phi(h(u))$. By construction, $\hat{u}$ is the sequence u after grouping in successive words of d letters.

This means $\hat{u}_n = u_{nd} \ldots u_{(n+1)d-1}$, and F is a substitution of constant length σ on the alphabet P. The infinite word $\hat{u}$ of P^∞ is a fixed point of F which is minimal by Proposition 2.

Step 3. Let us prove that m (satisfying (6)) can be seen like a factor of $\hat{u}$. In fact, B and B' are suffixes of words of M of length multiples of d and C and C' are prefixes of words of M that are also multiples of d. Let us remark that the decompositions of m come from the decomposition of u in successive blocks of σ letters given by $u = f(u)$.

Considering $\hat{u}$ and F, the decomposition (s) of m is (are) a consequence of the decomposition of $\hat{u}$ in blocks of σd letters:

$$\hat{u} = [(u_0 \ldots u_{d-1}) \ldots (u_{d(\sigma-1)} \ldots u_{d\sigma-1})][(u_{d\sigma} \ldots u_{d(\sigma+1)-1}) \ldots (u_{d(2\sigma-1)} \ldots u_{2d\sigma-1})] \ldots$$

Every block can be seen as σ factors of words in P (of d letters) or as d factors of words in M (of σ letters).

One derives that the decomposition $Bf(D)C$ implies the decomposition $BB_1F(D_1)C_1C$ of m with $B_1 = f(\lambda)$ (resp. $C_1 = f(\mu)$) where λ (resp. μ) is a prefix (resp. a suffix) of D. In the same way the decomposition $B'f(D')C'$ gives $B'f(\lambda')F(D'_1)f(\mu')C'$. In particular, if m is rythmical on the alphabet P, then $|Bf(\lambda)| = |B'f(\lambda')|$ so that $|B| - |B'|$ is a multiple of σ and this is quite of the question. Hence, m is not rythmical on P and the length of m, seen like a word in P, is at least $2|m|/\sigma \geq 2H/\sigma$.

Let us choose now $L_0(\sigma, q) = \sigma^{2(q+1)}$ for $q \geq 3$. (The value for $q = 2$ is bigger than the value we have already obtained.) As $H(\sigma, q) \leq 2\sigma^{2q+1}$, if $|m| \geq L_0(\sigma, q)$, one has $|m| \geq H$ and $2|m|/\sigma \geq 2\sigma^{2q+1} \geq L_0(\sigma, q')$ for any $q' = 2, \ldots, q-1$. If the proposition is proved for alphabets with number of letters $\leq q-1$, then every factor m of u of length $\geq L_0(\sigma, q)$ is rythmical on the alphabet A. $\square$

Remark 3. The constant $L_0 = \sigma^{2(q+1)}$ gives the minimal length of the word m to be rythmical but it is not the "best possible". In fact, for example, $L_0 = 64$ for the Morse sequence but one can easily prove that every factor of length ≥ 4 is rythmical, and 4 is the optimal value.

Remark 4. As shown in the following example, the hypothesis that u is not periodic in Proposition 5 is necessary. Let $A = \{a, b\}, f_3(a) = aba$ and $f_3(b) = bab$. So f_3 is injective and of length 3. Furthermore, $f_3^{\omega}(a) = (ab)^{\omega}$ is a fixed point of f_3. As $a \mid f_3(a)$ and $a \mid f_3(b)$ and from Proposition 3(iii), $f_3^{\omega}(a)$ is minimal. For every $n \geq 1$, $(ab)^{6n}$ is not a rythmical factor of $f^{\omega}(a)$. In fact, $(ab)^{6n} = \varepsilon.f((ab)^{2n}).\varepsilon = a.f((ba)^{2n-1}.b).ab$.

Propositon 7. *Every suffix of a special factor is special.*

Proof. Let $w = xv$ be a special factor. For every letter $i \in A, xvi = wi$ is a factor of u so that vi is a factor of u. Then v is special. □

The following corollary is immediate.

Corollary. *If $FS(p)$ is empty, then for any $n \geq p$, $FS(n)$ is empty.*

4. Inductive construction of special factors

Let $n \geq L_0$ be an integer, where L_0 is the constant of Proposition 5. For two different letters i and j, $P_{i,j}$ (resp. P) will denote the greatest common prefix of $f(i)$ and $f(j)$ (resp. the greatest common prefix of all the $f(i)$). Let us set $\alpha_{i,j} = |P_{i,j}|$ and $\alpha = |P|$.

Theorem 1. *If there exist two different letters i and j such that $P_{i,j} \neq P$, then there is no special factor of length $n > L_0$.*

Proof. By contraposition: Let us suppose that $FS_n(u)$ is non-empty and let $w = xf(v)y$ be a decomposition of a special factor of length $n > L_0$. By Proposition 5 this decomposition is unique. Then for every letter $i \in A$, wi is a factor of u. Any one of the Card(A) different factors yi is a prefix of exactly one $f(k)$.

Since this connexion is bijective, one has $P_{i,j} = y$ for every couple (i, j) of two different letters. □

Theorem 2. *Let k be the least integer such that $\sigma k + \alpha \geq n$ and let us suppose that for any two different letters i and j, $P_{i,j} = P$. Then the special factors of length $n > L_0$ are suffixes of length (n-α) of the images of the special factors of length k to which one has concatenated at right P.*

Proof. Let us prove first that the construction above gives special factors.

Let v be a special factor of length k. Then for every letter $i \in A$, vi is a factor of u. It is the same for its image whose length is $\sigma k + \sigma$. The prefix $f(v).P$ of length $\sigma k + \alpha$ of the image $f(vi)$ is special because $P = P_{i,j}$. From Proposition 7, it is the same for all its suffixes.

Let us prove now that every special factor of length $> L_0$ can be obtained by the construction of Theorem 2. Let w be a special factor of length $n > L_0$. By Proposition 5, w is rythmical: $w = xf(v)y$, where x is a strict suffix of exactly one $f(j)$ and y is a strict prefix of exactly one $f(k)$. So w is a factor of $f(jvk)$ with jvk factor of u. Since w is special, wi is a factor of u for every letter $i \in A$. Any one of these different $\mathrm{Card}(A)$ factors yi is a prefix of exactly one $f(k)$, and this connexion is bijective. Hence $y = P$ and jv special. Finally, $w = xf(v)P$ where $xf(v)$ is a suffix of $f(jv)$ with jv special. $\square$

Example 5. In the Morse sequence, every factor of length ≥ 4 is rythmical; furthermore, $P(5) = 12$ and $P(6) = 16$. As $\alpha = 0$, the 4 special factors of length 5 are the suffixes of length 5 of the images of the special factors of length 3 which are $112, 121, 212$ and 221. One has $f_1(112) = 121221, f_1(121) = 122112, f_1(212) = 211221$ and $f_1(221) = 212112$ so that the special factors of length 5 are $21221, 22112, 11221$ and 12112.

Example 6. Let u be the fixed point in $1A^*$ of the substitution f_4 on the alphabet $A = \{1,2\}$ defined by $f_4(1) = 112$ and $f_4(2) = 111$. Every factor of u of length ≥ 5 is rythmical. Furthermore $P(5) = 8$ and $P(6) = 9$. As $\alpha = 2$, the special factor of length 5 is the suffix of length 3 of the image of the special factor of length 1 to which one has concatenated at right $P = 11$. 1 is the special factor of length 1 and we have $f_4(1) = 112$ so that 11211 is the special factor of length 5.

5. Conclusion

The inductive method described above shows that if a sequence satisfies $(*)$, it must be such that for any two different letters i and j, $P_{i,j} = P$. Other considerations bring us to assume that these sequences have a non-linear complexity between $(k-1)n+1$ and $kn+1$.

Therefore, $P(n+1) - P(n)$ is not constant; so, a geometric representation will be (probably) more difficult to get.

Nevertheless, since this difference takes only a finite number of values, we hope to succeed .... One could start with sequences such that $\mathrm{Card}\{P(n+1) - P(n);\ n \in N^*\} = 2$ before trying to get an inductive procedure.

Acknowledgements

The example of the Remark 4 is given by the referee. The author would like to thank warmly M.-F. Roy and the referee for useful comments and suggestions concerning the first version of the paper.

References

[1] P. Arnoux and G. Rauzy, Représentation géométrique de suites de complexité $2n + 1$, Bull. Soc Math. France 119 (1991) 199–215.

[2] J. Berstel, Mots de Fibonacci: Séminaire d'Informatique Théorique (LITP Université Paris VI et VII, Année 1980:81) 57–78.

[3] S. Brlek, Enumeration of factors in Thue–Morse word, Discrete Appl. Math. 24 (1989) 83–96.

[4] G. Christol, T. Kamae, M. Mendes-France and G. Rauzy, Suites algébriques, automates et substitutions, Bull. Soc. Math. France 108 (1980) 401–419.

[5] A. Cobham, Uniform tag sequences, Math. Systems Theory 6 (1972) 164–192.

[6] A De Lucas and S. Varricchio, Some combinatorial properties of the Thue–Morse sequence and a problem of semigroups, Theoret. Comput. Sci. 63 (1989) 333–348.

[7] W H. Gottschalk and G.A. Hedlung, Topological Dynamics, Amer Math. Soc. Colloq. Publ., Vol. 36 (AMSE, Providence, RI, 1968).

[8] G. Rauzy, Suites à termes dans un alphabet fini, Sém. Théorie Nombres Bodeaux Exp. 35 (1982–1983) 1 16.

[9] T Tapsoba, Automates calculant la complexité de suites automatiques, J. Théorie Nombres Bordeaux 6 (1994) 127–134.

[10] A. Thue, Uber unendliche zeichenreihen, Norske Vid. Selsk. Skr I Math. Nat. Kl. Christiana 7 (1906) 1–22.

[11] A. Thue, Uber die gegenseitige lage gleicher teile genvisser zeichenreihen, Norske Vid. Selsk Skr., I Math. Nat. Kl. Christiana 1 (1912) 1–67.