Maximum Complexity of Automatic Sequences on Two Symbols

Théodore Tapsoba
Ecole supérieure d'informatique
Centre universitaire polytechnique de Bobo - Dioulasso
Université de Ouagadougou
01 B.P. 1091 Bobo Dioulasso 01
Burkina Faso

Key-words : constant - length substitution ; automatic sequence ; complexity .

Résumé : Nous déterminons la complexité maximum des suites automatiques (points fixes de substitutions injectives uniformes) sur des alphabets de cardinal deux.

1. Introduction

A finite sequence $m=(m_0, m_1, ..., m_n)$ over a finite set A of letters will often be seen as a (finite) word $m_0m_1...m_n$ over the alphabet A and by extension, a sequence u: $N \to A$ will be seen as an infinite word $u = u_0u_1u_2$..... A simple way to construct such words is to proceed by iterating a substitution; Every letter of a word m is replaced by a word and so on. For example, over the symbols 1, 2, the substitution f given by f(1) = 12, f(2) = 21 brings us to the infinite word 1221211221121 22121..... (if one start by 1) which is the celebrated Morse ([12]) sequence.

The study of factors of infinite sequences goes back at least to Thue([20], [21]) and one of the questions which has been addressed is the problem of computing the complexity function P, where P(n) is the number of distinct factors of length n. It is well known in fact that for an infinite word u, the sequence $n \to P(n)$ gives a good indication on the "complexity" of the language of u and on the associated dynamical system; that's why many papers has been devoted to this sequence (see for instance [16], [13], [9], [4], [14] ...)... The interested reader can find in [1] a recent survey on the complexity of infinite sequences, very rich in bibliography.

If a sequence is not periodic, one can easily prove that the complexity function is strictly increasing. In this case, one has for any $n \in N$, $P(n) \geq n+1$. So, sequences such as $P(n) = n+1$, called Sturmian sequences ([7]), have minimun complexity. The minimum complexity of automatic non-Surmian sequences is given in [19]. In the same way, what can we say about the maximum complexity?

We clearly have $P(n) \leq (Card(A))^n$. For infinite words, fixed points of constant-length substitutions, Cobham ([6]) shows that there exists a constant C depending of the infinite word, such as $P(n) \leq C.n$ and Bleuzen-Guernalec ([2]) gets $\sigma(Card(A))^2$, where σ is the length of the substitution, as an upper bound for C. As far I know, the last majoration we have is due to Bleuzen-Guernalec and Blanc (see [3]). Unfortunately all these majorations are very far from the real values of P(n)...

After the investigation of an example, we determine with precision, the maximum complexity of automatic sequences (fixed points of injective constant-length substitutions) on sets of two elements.

2. Preliminaries

Let A* be the free monoid generated by a non - empty finite set A called the 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 a word of length 0. Let a ∈ A; we simply denote by a^* the set $\{a\}^*$ of the finite words composed of the unique letter a. The infinite word composed of the letter a is denoted by a^∞. A word v is said to be a factor of w if w = xvy for some x, y in A^*. We then write v|w. If x = ε (resp. y = ε), v is called a prefix (resp. suffix) of w and we then write v = PR(w) (resp. SU(w)). A prefix or a suffix of w is said to be strict if it is different from w. A word v is said to be "a power" of w if v=ww...w (n times with n>1). We then write $v=w^n$. v is said to be "primitive" if it is not a power. Two words v and w are said "conjugates" if there exists words B and C such as v = BC and w = CB. (One may refer to [10], [8] or [11] for more details on these three last notions).

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, growing if $|f(i)| \geq 2$. If there exists a letter a ∈ A such as f(a) = am with |m| > 0, then the set of the words with prefix a has a fixed point $u = amf(m)f^2(m)....... f^k(m).......$

When an infinite word u is a fixed point of a constant-length substitution on an alphabet, it is called automatic. In fact it is well known ([6], [5]) that such a sequence is recognizable by a q-automaton. We denote by F the set of the finite factors of u and by F(n) its subset consisting of the factors of length n. It is trivial to verify that every factor of a word v of F is a word of F and that there exists a letter a such as va is in the set F. The factor v of u is said to be special if for any letter i of A, vi is a factor of u. We denote by FS the set of the special factors of u and by FS(n) the set of the special factors of length n.

Let S be the shift defined by $S(a_0a_1a_2.....) = a_1a_2........$ and let Ω be the closure of the set $\{S^k(u) ; k \in N\}$ where the distance d is given by $d(v, w) = \exp(-\inf\{n \in N ; v_n \neq w_n\})$. The sequence u is associated to 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.

We give here a simple criterium for minimality when A={1, 2} proved in [17] :

Proposition 1 : *Let* u *be a fixed point of the substitution* f *on the alphabet* A={1, 2} *such as* 1*is a prefix of* u *and* $u \neq 1^\infty$.

(i) *Let us suppose* f *growing, then:* u *minimal* ⇔ $f(2) \notin 2^*$

(ii) *Let us suppose* $|f(1)| \geq 2$ *and* f(2) = 2, *then :* u *minimal* ⇔ $f(1) \in 1A^*1$.

3. Factors and Special Factors

Let u be a minimal sequence, fixed point of an injective constant-length σ substitution f on an alphabet A and let w be a factor of u. It can be decomposed as follows : (1) w = xf(v)y. In equation (1) x is a strict suffix of a word $f(v_1)$, y is a strict prefix of a word $f(v_2)$ and v_1vv_2 is a factor of u.

A factor w of u is said to be "rythmical" if it has a unique decomposition with condition (1). We recall here some useful properties (see[18]) of factors and special factors.

Proposition 2 : *There exists* L_0 (= L_0 (σ, Card (A))) *such as every factor of* u *of length* > Lo *is rythmic.*

Proposition 3 : *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 rythmic.*

Proposition 4 : *Every suffix of a special factor is special.*

Proposition 5 : *If* FS (p) *is empty, then for any* n $\geq$ p, FS (n) *is empty.*

Let us suppose now A={1, 2}.

Let n > L_0 be an integer, where L_0 is the constant of Proposition 2. PR (resp.SU) will denote the greatest common prefix (resp. suffix) of f(1) and f(2). Let us set α = |PR| and μ = |SU|.

The following results describe an inductive method to determine the rythmical special factors and have been proved in ([17]):

Proposition 6 : *Let* k *be the least integer such as* σ k + $\alpha \geq$ n. *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* PR *on the right hand side.*

Proposition 7 : *In the construction above, with the same hypothesis, two distincts special factors of length* k *give the same factor of length* $\in$ [σ(k-1) + α + 1 , σk + α] *if and only if* μ > 0 *and* n $\leq$ σ(k-1) + α + μ.

Remark : Without constraint in the cardinal of the alphabet, Proposition 6 have been proved in ([18]).

The study of an example will be of some help to understand the general case.

4. Example

Let us consider the sequence u $\in$ 1A^*, fixed point of the substitution 1→ 1121, 2→ 2122. A direct counting gives P(1) = 2, P(2) = 4, P(3) = 8, P(4) = 13 and P(5) = 16. Thus, Card (FS(2)) = 4, Card(FS(3)) = 5 and Card (FS(4)) = 3. The 16 words of F(5) are decomposed as follow: 11121 ; 11211 ; 11212 ; 12111 ; 12121 ; 12122 ; 12211 ; 12221 ; 21112 ; 21121 ; 21212 ; 21221 ; 21222; 22112 ; 22122 ; 22212. These decompositions are unique so that every factor of length 5 is rythmic. Hence, Proposition 3 yields that every factor of u of length $\geq$ 5 is rythmic.

As SU = PR = ε , one has μ = α = 0 . It follows from Propositions 6 and 7 that for any q $\geq$ 5, Card (FS(q)) = i (i = 3, 4 or 5) if and only if Card (FS(k)) = i where k is the unique integer such as $\frac{q}{4} \leq k < \frac{q}{4} + 1$. Thus, the integers q such as Card (FS(q)) = 3 are 9, 10, 11, 12, 33, 34 ...., those for which Card (FS(q)) = 4 are 5, 6, 7, 8, 17, 18 .... and those for which Card (FS(q)) = 5 are 13, 14, 15, 16, 49, 50......... .

A straightforward computation gives: Card (FS(q)) = 3 $\Leftrightarrow$ there exists an integer $r \geq 1$ such as $2.4^r < q \leq 3.4^r$, Card (FS(q)) = 4 $\Leftrightarrow$ there exists an integer $r \geq 1$ such as $4^r < q \leq 2.4^r$, Card (FS(q)) = 5 $\Leftrightarrow$ there exists an integer $r \geq 1$ such as $3.4^r < q \leq 4.4^r$.

Since $P(n+1) = P(n) + \text{Card}(FS(n))$ for any integer $n \geq 1$, one has

$$P(n+1) = P(5) + \sum_{q=5}^{n} \text{Card}(FS(q))$$

$$= 16 + (n-5+1) \sum_{i}(i-1)(\text{Card}\ \{q \in [5, n]\ \ \text{Card}(FS(q) = i\})) \quad (i = 3, 4 \text{ or } 5)$$

$$= n+12 + \sum_{i}(i-1)(\text{Card}\ \{q \in [5, n]\ \text{Card}(FS(q) = i\}))$$

Let us recall that the number of integers $q \in\ ]\ i.4^r, (i+1).\ 4^r]$ is equal to 4^r and that for every integer n, there exists a unique integer k such as $4^k < n \leq 4.4^k$.

Thus if $4^k < n \leq 2.\ 4^k$, one has

$$P(n+1) = n + 12 + \left(\sum_{i}(i-1)\right)\left(\sum_{r=1}^{k-1} 4^r\right) + 3(n - 4^k)$$

$$= n+12 + 9.\left(\sum_{r=1}^{k-1} 4^r\right) + 3(n - 4^k)$$

$$= 4n + 12 + 9.4.\frac{4^{k-1}-1}{4-1} - 3.4^k$$

$$= 4n\ ;$$

if $2.\ 4^k < n \leq 3.\ 4^k$, one has

$$P(n+1) = n + 12 + 9.\left(\sum_{r=1}^{k-1} 4^r\right) + 3.4^k + 2(n - 2.4^k)$$

$$= 3n + 12 + 3.4^k - 12 - 4^k$$

$$= 3n + 2.\ 4^k\ ;$$

and if $3.\ 4^k < n \leq 4.\ 4^k$, one has

$$P(n+1) = n + 12 + 9.\left(\sum_{r=1}^{k-1} 4^r\right) + 3.4^k + 2.4^k + 4(n - 3.4^k)$$

$$= 5n + 12 + 3.4^k - 12 - 7.4^k$$

$$= 5n - 4.\ 4^k\ .$$

Let us remark that if $4^k < n \leq 4^{k+1}$, then $k < \log_4(n) \leq k+1$ so that $k = -[-\log_4(n)] - 1$, where $[x]$ denote the largest integer contained in the real x. Therefore we finally have:

$P(n+1) = 4n \quad \text{if } n \in\,]\,4^{-[-\log_4(n)]-1}, 2.\,4^{-[-\log_4(n)]-1}]$
$P(n+1) = 3n + 2.\,4^{-[-\log_4(n)]-1} \quad \text{if } n \in\,]\,2.\,4^{-[-\log_4(n)]-1}, 3.\,4^{-[-\log_4(n)]-1}]$
$P(n+1) = 5n - 4.\,4^{-[-\log_4(n)]-1} \quad \text{if } n \in\,]\,3.\,4^{-[-\log_4(n)]-1}, 4.\,4^{-[-\log_4(n)]-1}]$

Let $n \in\,]\,4^k, 4^{k+1}]$. One has $[-\log_4(n)] = [-\log_4(n-1)]$ if $n \neq 4^k+1$ and $[-\log_4(4^k+1)] = [-\log_4(4^k)] - 1$. Hence, after replacing n's by (n-1)'s in the formula above, we obtain a new formula that allows us to compute P(n) for every $n > 4$:

$P(n) = 4(n-1) \quad \text{if } n \in [4^{-[-\log_4(n)]-1}+1, 2.\,4^{-[-\log_4(n)]-1}+1]$
$P(n) = 3(n-1) + 2.\,4^{-[-\log_4(n)]-1} \quad \text{if } n \in\,]\,2.4^{-[-\log_4(n)]-1}+1, 3.4^{-[-\log_4(n)]-1}+1]$
$P(n) = 5(n-1) - 4.\,4^{-[-\log_4(n)]-1} \quad \text{if } n \in\,]\,3.4^{-[-\log_4(n)]-1}+1, 4.4^{-[-\log_4(n)]-1}+1[$

We can now tackle the general case.

5. The Maximum Complexity

Let us consider a non periodic minimal sequence $u \in 1A^*$, fixed point of an injective constant-length σ substitution f on the set A={1, 2} and show the following.

Proposition 8 : *If* f(1) *and* f(2) *have a common prefix then they cannot be both special factors.*

Proof : Let us set f(1) = 1u and f(2) = 1v where (u, v)∈ F(σ-1). Let us suppose f(1) and f(2) be special factors; then 1u1, 1u2, 1v1 and 1v2 are factors. As 2 is not a prefix of f(i) (i = 1, 2), we have $u = u_1u_2$ and $v = v_1v_2$ with u_2 and v_2 different from the empty word. Hence $1u_1 = SU(f(i))$, $u_2 2 = PR(f(j))$, $1v_1 = SU(f(k))$, $v_2 2 = PR(f(p))$ where i, j, k, p ∈ {1, 2}.Thus there exists words B, C, D and E such as $f(1) = 1u_1u_2$
$= B1u_1$ or $C1v_1$
$= u_2 2D$ or $v_2 2E$

and $f(2) = 1v_1v_2$
$= C1v_1$ or $B1u_1$
$= v_2 2E$ or $u_2 2D$

Four cases can occur:
1) $f(1) = B1u_1 = u_2 2D$ and $f(2) = C1v_1 = v_2 2E$
2) $f(1) = B1u_1 = v_2 2E$ and $f(2) = C1v_1 = u_2 2D$
3) $f(1) = C1v_1 = u_2 2D$ and $f(2) = B1u_1 = v_2 2E$
4) $f(1) = C1v_1 = v_2 2E$ and $f(2) = B1u_1 = u_2 2D$

Trivially, cases 1) and 4) are the same and so it is for cases 2) and 3). So, we have to examine only cases 1) and 2).

Case 1 : $f(1) = 1u_1u_2 = B1u_1 = u_2 2D$ and $f(2) = 1v_1v_2 = C1v_1 = v_2 2E$ with $|u_1| + |u_2| = |v_1| + |v_2| = \sigma - 1$, $|B| = \sigma - 1 - |u_1| = |u_2|$, $|C| = \sigma - 1 - |v_1| = |v_2|$, $|D| = \sigma - 1 - |u_2| = |u_1|$, $|E| = \sigma - 1 - |v_2| = |v_1|$. Since $|D| = |u_1|$, we have $f(1) = 1u_1u_2 = u_2 2u_1$. Likewise since $|E| = |v_1|$, $f(2) = 1v_1v_2 = v_2 2v_1$.

If $|u_1| = |u_2|$ then $u_1 = u_2$ and $f(1) = 1u_1u_1 = u_12u_1$.
Hence $1u_1 = u_12 \Rightarrow 1u_1 = 1u'_12 \Rightarrow 11u'_1 = 1u'_12 \Rightarrow u'_1=1u''_1$.
Then $111u'_1 = 11u''_12$ so that $u''_1 = 1u'''_1$. Inductively, we get $u_1 = 11....1$ which yield the contradiction $11....1 = 11....2$.
If $|u_1| <> |u_2|$, let us suppose $|u_1| > |u_2|$ (For $|u_1| < |u_2|$, the arguing is almost identical) . Then $f(1) = 1u_1u_2 = u_22\,Xu_2$. Hence $1Xu_2u_2 = u_22\,Xu_2$.
If $|u_2| <> |X|$, I repeat the process above and by induction we get that f(1) have two different values; same conclusion if $|u_2| = |X|$.

case 2 : With almost similar transformations than those used in case 1, the same arguments bring us to the same contradictions . ◊

It follows from Proposition 5 that if a factor has a suffix which is not a special factor, it cannot be a special factor. Hence, according to Proposition 8, to have a sequence with maximum complexity, we may have $f(1) \in 1A^*$ and $f(2) \in 2A^*$ so that $\alpha = 0$. Moreover, one derives from Propositions 6 and 7 that f(1) and f(2) may not have a common suffix so that $\mu = 0$.

Proposition 9 : *If* f(1) *and* f(2) *are conjugates , then the complexity of the sequence* u *cannot be maximal.*

Proof : As stated above, if $\alpha <> 0$ or if $\mu <> 0$, the Proposition is clearly true. Let us suppose $\alpha = \mu = 0$. Thus, two cases can occur: f(1) = 1x22y1 and f(2) = 2y11x2 or f(1) = 1x12y2 and f(2) = 2y21x1, where x and y are words eventually empty. Since 1 and 2 are special factors, we have f(1) and f(2) which are both special factors. Unfortunately it does not exist a third special factor of length σ, so that the complexity cannot be maximal. ◊

We now prove a Proposition about rythmical factors of sequences which are able to have a maximum complexity.

Proposition 10 : *If* $\alpha = \mu = 0$, *and if* f(1) *and* f(2) *are not conjugates, then every factor of* u *of length* $(\sigma+1)$ *is rythmic.*

Proof : By contraposition. Let us suppose that there exists a non rythmical factor $w = a_1a_2..........a_\sigma a_{\sigma+1}$. Then there exists integers r and p such as
$w = a_1a_2.......... a_ra_{r+1}..........a_{\sigma+1} = a_1a_2.......... a_pa_{p+1}.........a_{\sigma+1}$ where $a_1.......... a_r$ and $a_1..........a_p$ (resp. $a_{r+1}.......... a_{\sigma+1}$ and $a_{p+1}.......... a_{\sigma+1}$) are suffixes (resp. prefixes) of f(1) and/or f(2). Without restriction in the generality we can suppose p>r. Hence, if ____________ represente an unknown factor, four cases can occur:

1°) $f(1) = \_\_\_\_\_\_\_\_\_\_\_\_\_\_ a_1 a_2 \dots\dots a_r$

$f(1) = a_{r+1} \dots\dots a_p a_{p+1} \dots\dots a_{\sigma+1} \_\_\_\_\_\_\_\_$

$f(1) = \_\_\_\_\_\_\_\_ a_1 a_2 \dots\dots a_r a_{r+1} \dots\dots a_p$

$f(2) = a_{p+1} \dots\dots a_{\sigma+1} \_\_\_\_\_\_\_\_\_\_\_\_$

2°) $f(1) = \_\_\_\_\_\_\_\_\_\_\_\_\_\_ a_1 a_2 \dots\dots a_r$

$f(1) = a_{r+1} \dots\dots a_p a_{p+1} \dots\dots a_{\sigma+1} \_\_\_\_\_\_\_\_$

$f(2) = \_\_\_\_\_\_\_\_ a_1 a_2 \dots\dots a_r a_{r+1} \dots\dots a_p$

$f(1) = a_{p+1} \dots\dots a_{\sigma+1} \_\_\_\_\_\_\_\_\_\_\_\_$

3°) $f(1) = \_\_\_\_\_\_\_\_\_\_\_\_\_\_ a_1 a_2 \dots\dots a_r$

$f(1) = a_{r+1} \dots\dots a_p a_{p+1} \dots\dots a_{\sigma+1} \_\_\_\_\_\_\_\_$

$f(2) = \_\_\_\_\_\_\_\_ a_1 a_2 \dots\dots a_r a_{r+1} \dots\dots a_p$

$f(2) = a_{p+1} \dots\dots a_{\sigma+1} \_\_\_\_\_\_\_\_\_\_\_\_$

4°) $f(1) = \_\_\_\_\_\_\_\_\_\_\_\_\_\_ a_1 a_2 \dots\dots a_r$

$f(2) = a_{r+1} \dots\dots a_p a_{p+1} \dots\dots a_{\sigma+1} \_\_\_\_\_\_\_\_$

$f(2) = \_\_\_\_\_\_\_\_ a_1 a_2 \dots\dots a_r a_{r+1} \dots\dots a_p$

$f(1) = a_{p+1} \dots\dots a_{\sigma+1} \_\_\_\_\_\_\_\_\_\_\_\_$

Let us set $X = a_1 a_2 \dots\dots a_r$, $Y = a_{r+1} \dots\dots a_p$ and $Z = a_{p+1} \dots\dots a_{\sigma+1}$. We then have $|X| = r$, $|Y| = p-r$ and $|Z| = \sigma-p+1$.

case 1 : $f(1) = \_\_\_\_\_\_\_\_\_\_\_\_ a_1 a_2 \dots\dots a_r = \_\_\_\_\_\_ X$

$f(1) = \_\_\_\_\_\_ a_1 a_2 \dots\dots a_r a_{r+1} \dots\dots a_p = \_\_\_\_\_\_ XY$

If $|X| \geq |Y|$ then $X = WY$ so that $f(1) = \_\_\_\_\_\_ WY$ and $f(1) = \_\_\_\_ WYY$. Arguing with the suffixes give inductively $f(1) = SU(Y)YY....Y$. Hence f(1) is not primitive and the infinite word is periodic; contradiction. For $|X| < |Y|$, we are led to the same conclusion.

case 2 : $f(1) = YZ\_\_\_\_\_\_$

$f(1) = Z\_\_\_\_\_\_$

Arguing as in the preceding case, not with the suffixes but with the prefixes gives again the contradiction that the infinite word is periodic.

case 3 : $f(1) = \_\_\_\_\_\_ X$

$f(1) = YZ\_\_\_\_\_\_$

and $f(2) = \_\_\_\_\_\_ X\,Y$

$f(2) = Z\_\_\_\_\_\_$

Looking the length of the words bring us to $f(1) = YZa_2 \dots\dots a_r = Ya_{p+1} \dots\dots a_\sigma X$ and $f(2) = Za_2 \dots\dots a_r Y = a_{p+1} \dots\dots a_\sigma XY$.

Hence $f(1) = a_{r+1} \dots\dots a_p a_{p+1} \dots\dots a_\sigma\, a_1 \dots\dots a_r = BC$

$f(2) = a_{p+1} \dots\dots a_\sigma a_1 \dots\dots a_r a_{r+1} \dots\dots a_p = CB$

which yield the contradiction f(1) and f(2) are conjugates.

case 4 :   f(1) = ________X
f(1) = Z________

and f(2) = YZ________
f(2) = ________XY

This case is almost the same than the preceding and does not bring any new difficult. ◊

Let us summarize: To have a sequence with maximum complexity we may have $\alpha = \mu = 0$, f(1) and f(2) not conjugates and clearly the numbers of integers q such as Card(FS(q)) = k must be upper with k's biggers than it is possible. We know that for such a sequence, every factor of length > σ is rythmic.

For $i \in [1, \sigma[$, let us set a_i = Card(FS(i+1)). Thus, every infinite word u is associated to a unique sequence $(a_1, a_2, \ldots\ldots, a_{\sigma-1})$ of integers ≥ 2. Furthermore for any integer $q \geq \sigma+1$, there exists $i \in [1, \sigma[$ such as Card(FS(q)) = a_i. The multitude of examples studied, and many oder considerations bring us to assume that $\sum_{i=1}^{\sigma-1} a_i = 4(\sigma-1)$.

$$\text{Hence, } P(\sigma+1) = P(1) + \sum_{i=1}^{\sigma} \text{Card(FS(i))} = 2 + \text{Card(FS(1))} + \sum_{i=2}^{\sigma} \text{Card(FS(i))}$$

$$= 4 + \sum_{i=1}^{\sigma-1} a_i = 4 + 4(\sigma-1) = 4\sigma.$$

Since $P(n+1) = P(\sigma+1) + \sum_{q=\sigma+1}^{n} \text{Card(FS(q))}$, counting the numbers of integers $q \in]\, i.\,\sigma^r, (i+1).\,\sigma^r]$ $(1 \leq i < \sigma)$ and arguing as in Paragraph 4 gives:

$$\text{if } n \in]\, i.\,\sigma^k, (i+1).\,\sigma^k] ,$$

$$P(n+1) = P(\sigma+1) + (n-(\sigma+1)+1) + \sum_{i=1}^{\sigma-1} \left((a_i-1) \sum_{r=1}^{k-1} \sigma^r\right) + \sum_{j=1}^{i-1} (a_j-1)\,\sigma^k + (a_i-1)(n-i.\,\sigma^k)$$

$$= n+3\sigma + \sum_{i=1}^{\sigma-1} (a_i-1).\sigma.\,\frac{\sigma^{k-1}-1}{\sigma-1} + \sum_{j=1}^{i-1} (a_j-1)\,\sigma^k + (a_i-1)(n-i.\,\sigma^k)$$

$$= n+3\sigma + \sigma.\,\frac{\sigma^{k-1}-1}{\sigma-1}\,((4(\sigma-1) - (\sigma-1)) + (a_i-1)n - (a_i-1)\,i.\,\sigma^k + \sum_{j=1}^{i-1} (a_j-1)$$

$$= n+3\sigma + 3.\,\sigma^k - 3\sigma + (a_i-1)n - i\,(a_i-1)\,\sigma^k + \sigma^k\left(\sum_{j=1}^{i-1} (a_j-1)\right)$$

$$= a_i\, n + \sigma^k (3 + i - i\, a_i + \sum_{j=1}^{i-1} a_j - i + 1)$$

$$= a_i\, n + \sigma^k (4 + \sum_{j=1}^{i-1} a_j - i\, a_i) .$$

Moreover, as $k = -[- \log_\sigma(n)] - 1$, one finally has:

$$P(n+1) = a_i\, n + (4 - i.\, a_i + \sum_{j=1}^{i-1} a_j)\, \sigma^{- [- \log_\sigma(n)] - 1}$$

if $n \in]\, i.\, \sigma^{- [- \log_\sigma(n)] -}\, , (i+1)\, \sigma^{- [- \log_\sigma(n)] - 1}]$.

As for the example, one can easily see that for $n \in]\, \sigma^k, \sigma^{k+1}]$ if $[- \log_\sigma(n)] \neq [- \log_\sigma(n-1)]$ then $n = \sigma^k + 1$ and $[- \log_\sigma(n)] = [- \log_\sigma(n-1)] - 1$. Hence after replacing n's by (n-1)'s in the formula above, and remark that $a_1 = 4$, we have thus proved the following theorem on maximum complexity of automatic sequences on sets of two elements:

Theorem : *Let us set* a_i = Card(FS(i+1)) $(1 \leq i < \sigma)$. *For* $n > \sigma$, *the complexity function* P *is given by*

$P(n) = 4(n-1)$ *if* $n \in [\, \sigma^{- [- \log_\sigma(n)] -1}\, , 1 + 2.\, \sigma^{- [- \log_\sigma(n)] - 1}]$.

$$P(n) = a_i\, (n-1) + (4 - i.\, a_i + \sum_{j=1}^{i-1} a_j)\, \sigma^{- [- \log_\sigma(n)] - 1}$$

if $n \in]\, 1 + i.\, \sigma^{- [- \log_\sigma(n)] -1}\, , 1 + (i+1)\, \sigma^{- [- \log_\sigma(n)] - 1}]$ $(1 < i < \sigma-1)$.

$$P(n) = a_{\sigma-1}\, (n-1) + (4 - (\sigma-1).\, a_{\sigma-1} + \sum_{j=1}^{\sigma-2} a_j\ \sigma^{- [- \log_\sigma(n)] - 1}$$

if $n \in]\, 1 + (\sigma-1).\, \sigma^{- [- \log_\sigma(n)] -1}\, , 1 + \sigma.\, \sigma^{- [- \log_\sigma(n)] - 1} [$

6. Concluding Remarks

It follows from the egality $\sum_{i=1}^{\sigma-1} a_i = 4(\sigma-1)$ that there is a kind of compensation between integers n such as Card(FS(n)) < 4 and those for which Card(FS(n)) > 4. Hence, it appears like a search of a balance between integers such as $P(n) > 4(n-1)$ and those such as $P(n) < 4(n-1)$. We then come to P(n) almost equal to 4(n-1) with as many slight ups and downs. This result is somewhat surprising since the figure remains unchanged no matter the length of the substitution, but allows us to reaffirm that the upper bounds we had until now were not enough accurate.

References

1 - J. - P. Allouche : *Sur la complexité des suites infinies*; Bull. Belg. Math. Soc. (1994), 133 - 143.

2 - N. Bleuzen-Guernalec : *Suites points fixes de transductions uniformes*; CRAS, Paris, T. 300, Série I, n°3 (1985) , 85-88.

3 - N. Bleuzen-Guernalec et G.Blanc: *Production en temps réel et complexité de structure de suites infinies*; RAIRO-Informatique Theorique et Applications, Vol 23, N°2 (1987), 195-216.

4 - J. Cassaigne : *Special factors of sequences with linear subword complexity*; to appear in Developments in Language Theory.

5 - G. Christol, T. Kamae, M. Mendes-France et G. Rauzy : *Suites algébriques, automates et substitutions* ; Bull. Soc. Math. France 108 (1980), 401-419.

6 - A. Cobham : *Uniform Tag Sequences* ; Math. Syst. Theo. 6 (1972), 164-192.

7 - E. M. Coven and G. A. Hedlung: *Sequences with minimal block growth*, Math. Syst. Theo. 7 (1973), 138-153.

8 - J. P. Duval : *Contribution à la combinatoire du monoïde libre*; Thèse, Université de Rouen (1980).

9 - S. Ferenczy : *Rank and symbolic complexity* ; Ergodic Theory and Dynamical Systems 16, 1 (1996), 1-20.

10 - J. I. Khmelevski : *Equations in free semigroups*; Trudy Mat. Inst. Steklov (1971).

11 - M. Maksimenko : *Algorithme quadratique de calcul de la solution générale d'équations en mots à une variable*; Theorical Informatics and Applications, Vol 29, N°4 (1995), 277-284.

12 - M. Morse : *Recurrent geodesic on a surface of negative curvature*; Trans. Amer. Math. Soc. 22 (1921) , 84-100.

13 - B. Mossé: *Puissances des mots et reconnaissabilité des points fixes de substitutions* ; Theor. Comp. Sci. 99 (1992), 327-334.

14 - B. Mossé : *Reconnaissabilité des substitutions et complexité des suites automatiques*; to appear in Bulletin de la S.M.F.

15 - J.-J. Pansiot : *Complexité des facteurs des mots infinis engendrés par morphismes itérés*; Lecture notes in Computer Science 172 (1984), 380-389.

16 - T. Tapsoba : *Complexité de suites automatiques*.Thèse de troisième cycle, Université d'Aix-Marseille II, (1987).

17 - T. Tapsoba : *Automates calculant la complexité de suites automatiques*. Journal de Théorie des nombres de Bordeaux 6 (1994),127-134.

18 - T. Tapsoba : *Special factors of automatic sequences*. J. Pure Appl. Algebra 108 (1996), 301-313.

19 - T. Tapsoba : *Minimum complexity of automatic non Sturmian sequences* ; Theorical Informatics and Applications, Vol 29, N°4 (1995), 285-291.

20 - A. Thue : *Uber unendliche zeichenreihen* ; Norske Vid. Selsk. Skr., I.Mat. Nat. K1, Christiana 7 (1906), 1-22.

21 - A. Thue : *Uber die gegenseitige Lage gleicher Teile genvisser zeichenreihen*, Norske Vid. Selsk. Skr., I. Math. Nat. K1, Christiana 1(1912), 1-67.