UNIVERSITAS COMENIANA
ACTA MATHEMATICA UNIVERSITATIS COMENIANAE
XLVIII—XLIX — 1986

ON GENERALIZED WORDS OF THUE—MORSE

ANTON ČERNÝ, Bratislava

1. Introduction

Axel Thue in his remarkable works [Th 06], [Th 12] on infinite sequences of symbols has shown the existence of infinite words over three-letter alphabet without squares of nonempty words as factors. The construction of such a word in [Th 12] is based on an infinite sequence t over the two-letter alphabet $\{0, 1\}$, not containing a factor of the form $xvxvx$, x being a letter and v a nonempty word. The i-th symbol of t can be described as the parity of occurrences of the symbol 1 in binary notation of the natural number i (however, this is not the way of its description in [Th 12]). The same sequence t appears in the work of Morse [Mo 21] on symbolic dynamics. Therefore we shall call t the sequence of Thue —Morse.

In the sense of Cobham [Co 72], t is a simple example of a uniform tag sequence. In [CKM-FR 80], where some algebraic properties of uniform tag sequences are investigated, generalized sequences of Thue—Morse are introduced. In such a generalized sequence the i-th symbol denotes the parity of occurrences of some fixed factor w over $\{0, 1\}$ in binary notation of i. (In fact, a slightly stronger generalization is given in [CKM-FR 80]). In this paper we show that in such generalized words of Thue—Morse all the factors are of a bounded power. More precisely, there are no factors of the form

$$(xv)^{2^{|w|}}x.$$

2. Notations and definitions

Let A^* be the free monoid generated by a finite alphabet A, with the neutral element ε. Let $A^+ = A^* - \{\varepsilon\}$. Let A^ω be the set of all infinite (to the right) sequences of elements of A. Let $A^\infty = A^* \cup A^\omega$. The elements of A^∞ will be called *words* (finite or infinite). A word $x \in A^*$ is a factor of a word $y \in A^\infty$, iff $y = zxt$ for some $z \in A^*$, $t \in A^\infty$. x is called *initial* (*terminal*) {*proper*} *factor* of y iff $z = \varepsilon$ ($t = \varepsilon$) {$zt \neq \varepsilon$}. The length $|x|$ of a finite word x is the number of its symbols, $|\varepsilon| = 0$.

Let φ: $A^* \to B^*$ be a morphism of monoids. φ can be extended to the mapping φ: $A^\infty \to B^\infty$ satisfying

$$\varphi(xy) = \varphi(x)\varphi(y)$$

for all $x \in A^*$, $y \in B^\infty$. φ is called *prolongable in* $a \in A$ if $\varphi(a) = ax$ for some $x \in A^+$. In this case for each $n \geqslant 0$ $\varphi^n(a)$ is a proper initial factor of $\varphi^{n+1}(a)$. There exists a limit

$$z = \lim_{n \to \infty} \varphi^n(a) \in A^\omega$$

such that each $\varphi^n(a)$ is an initial factor of z. Moreover, z is a fixpoint of φ, i.e. $\varphi(z) = z$. A morphism φ is called *m-uniform* for some $m \geqslant 0$ iff $|\varphi(b)| = m$ for all $b \in A$.

Let μ: $A^i \to A^j$ be a mapping, $i, j \geqslant 1$. μ can be extended to the mapping μ: $A^\omega \to A^\omega$ defined by

$$\mu(x_0x_1x_2\ldots) = y_0y_1y_2\ldots,$$

where

$$y_{k\cdot j}y_{k\cdot j+1}\ldots y_{k\cdot j+j-1} = \mu(x_{k\cdot i}x_{k\cdot i+1}\ldots x_{k\cdot i+i-1})$$

for all $k \geqslant 0$. This extension is called *(i, j)-substitution.*

We shall use two devices for formal description of infinite words — uniform tag systems and sorting automata.

A *tag system* is a quintuple $T = (\Sigma, a, \sigma, \Gamma, \tau)$, where Σ and Γ are alphabets, $\sigma: \Sigma^* \to \Sigma^*$ is a morphism prolongable in $a \in \Sigma$, τ: $\Sigma^* \to \Gamma^*$ is a morphism such that $\tau(\Sigma) \subseteq \Gamma$. The *internal* (*external*) *tag sequence* generated by T is

$$\text{intseq}_T = \lim_{n \to \infty} \sigma^n(a)$$

$$(\text{seq}_T = \lim_{n \to \infty} \tau(\sigma^n(a)) = \tau(\text{intseq}_T)).$$

The tag system and the corresponding sequences are called *m-uniform* iff σ is m-uniform.

Let $m > 0$. Denote $[m] = \{0, 1, \ldots, m-1\}$. A *sorting automaton* over $[m]$ is a quintuple $A = (S, \delta, s_0, F, G)$, where S is a finite set (of states), $s_0 \in S$ is the initial state, δ: $S \times [m] \to S$ is the transition function satisfying $\delta(s_0, 0) = s_0$, G is an alphabet, and $F = \{F_g\}_{g \in G}$ is a disjoint partition of S. δ can be extended to the domain $S \times [m]^*$ by setting $\delta(s, \varepsilon) = s$, $\delta(s, xd) = \delta(\delta(s, x), d)$ for $s \in S$, $x \in [m]^*$,

$d \in [m]$. The *state* (*sorting*) *sequence* of the automaton A is defined by

$$\text{state}_A = y_0 y_1 \ldots \in S^\omega$$

$$(\text{sort}_A = x_0 x_1 \ldots \in G^\omega),$$

where $y_i = \delta(s_0, i_{[m]})$, $i_{[m]}$ being the m-ary notation of i (since $\delta(s_0, 0) = s_0$ there are no problems with leading zeros)

$$(x_i = g \quad \text{iff} \quad y_i \in F_g)$$

Thus the sorting automaton is a slight generalization of the notion of finite automaton which in fact sorts to two classes of objects (accepted — rejected).

The relation between tag systems and sorting automata can be expressed as in the following proposition.

Proposition 1 [Co 72]. Let $T = (\Sigma, a, \sigma, \Gamma, \tau)$ be an m-uniform tag system and let $A = (\Sigma, \delta, s_0, F, \Gamma)$ be a sorting automaton over $[m]$ such that $\delta(s, i) =$ $= i$-th symbol of $\sigma(s)$, $s_0 = a$, and $s \in F_g$ iff $\tau(s) = g$, where $s \in S$, $i \in [m]$, $g \in \Gamma$.

Then $\text{intseq}_T = \text{state}_A$, $\text{seq}_T = \text{sort}_A$.

Finally, let us define the generalized words of Thue—Morse. Let $w \in$ $\in \{0, 1\}^* - 0^*$. Denote

$$a_w = a(0)a(1)a(2)\ldots$$

the infinite word with i-th symbol

$$a(i) = \#_w(i_{[2]}) \bmod 2$$

where $\#_w(x)$ denotes the number of occurrences of the factor w in the word x and $i_{[2]}$ is the binary notation of i with at least $|w|$ leading zeros. For example, 000010101010101 contains five occurrences of the factor 0101.

From [CKM-FR 80] we know the following important property of the words a_w.

Proposition 2 [CKM-FR 80]. Let $w \in \{0, 1\}^* - 0^*$, let μ be a $(2^{|w|-1}, 2^{|w|})$-substitution on $\{0, 1\}^\omega$ defined by

$$\mu(x_0 x_1 \ldots x_{2^{|w|-1}-1}) = y_0 y_1 \ldots y_{2^{|w|}-1},$$

$$y_i = x_{i/2} + \chi_w(i) \bmod 2 \qquad i \in \{0, 1, \ldots, 2^{|w|} - 1\}$$

where $\chi_w(i) =$ **if** w is a terminal factor of $i_{[2]}$ **then** 1 **else** 0.

Then $\mu(a_w) = a_w$.

As one can easily see, there is exactly one $j \in \{0, 1, \ldots, 2^{|w|} - 1\}$ such that $\chi_w(j) = 1$.

In the case $w = 1$ we obtain

$$a_1 = t = 0110100110010110\ldots$$

— the word of Thue—Morse. It is well known that t does not contain any factor of the form $xvxvx$, $x \in \{0, 1\}$, $v \in \{0, 1\}^*$. In particular, t contains no cubes $x^3 = xxx$.

3. Proof of the result

Our goal is to prove that there are no factors of the form $(xv)^{2|w|}x$ in a_w. It is well known to be true [Fi 80, Pa 81] for $w = 1$, thus in the following we consider $w \in \{0, 1\}^* - 0^*$ to be a fixed word of length at least 2.

The proof is based on the method from [Pa 81]. The proof is divided to a series of lemmas. In the first of them, the minimal sorting automaton for a_w is described. Since the notion of the sorting automaton is derived directly from the notion of the finite automaton, the results from the theory of finite automata concerning the minimality can be applied to sorting automata, too. This fact is used in the proof of the first lemma.

Lemma 1. Let $A_w = (S, \delta, s, (F_0, F_1), \{0, 1\})$ be a sorting automaton over $\{0, 1\}$, where $S = \{\langle\alpha\rangle_0, \langle\alpha\rangle_1 \mid \alpha \text{ is a proper initial factor of } w\}$

$$\delta(\langle\alpha\rangle_i, x) = \begin{cases} \langle\alpha x\rangle_i & \text{if } \langle\alpha x\rangle_i \in S \\ \langle\alpha'\rangle_{1-i} & \text{if } \alpha x = w \\ \langle\alpha'\rangle_i & \text{otherwise,} \end{cases}$$

where $i \in \{0, 1\}$, $x \in \{0, 1\}$, α' is the longest proper terminal factor of αx, being a proper initial factor of w, $s = 0^k$, where 0^k, $k \geqslant 0$ is the longest initial factor of w not containing 1,

$$F_i = \{\langle\alpha\rangle_i\}, \qquad i = 0, 1$$

Then A_w is minimal among the sorting automata with the sorting sequence a_w.

Proof. By induction on $|z|$ one can easily show for $z \in \{0, 1\}^*$, $i = 0, 1$

$$\delta(s_0, z) = \langle\alpha\rangle_i \quad \text{iff} \quad \#_w(0^{|w|}z) \equiv i \pmod 2$$

and α is the longest terminal factor of $0^{|w|}z$, being the proper initial factor of w.

The nonequivalence of each pair of distinct states is evident. (Two states s_1, s_2 are equivalent iff for each $x \in \{0, 1\}^*$ $\delta(s_1, x) \in F_0$ iff $\delta(s_2, x) \in F_0$.)

It is sufficient to show the accessibility of each state (from the initial state). Obviously, each state is accessible from $\langle\varepsilon\rangle_0$ or $\langle\varepsilon\rangle_1$. On the other hand, the states $\langle\alpha_0\rangle_0$, $\langle\alpha_0\rangle_1$, where α_0 is the longest proper initial factor of w are accessible. Thus it is sufficient to show that $\langle\varepsilon\rangle_i$ is accesible from $\langle\alpha_0\rangle_i$, $i = 0, 1$. The proof is based on induction. For each $\alpha \in \{0, 1\}^*$, $\alpha \neq \varepsilon$ a word $\gamma \in \{0, 1\}^*$ is given such that

$$\delta(\langle\alpha\rangle_i, \gamma) = \langle\alpha'\rangle_i \quad \text{and} \quad |\alpha'| < |\alpha|.$$

Let $w = \alpha\beta$. We consider several cases.

1. $w = 1^k$, $k \geqslant 1$. Then $\gamma = 0$.
2. $w = x^k\bar{x}^m$, $k \geqslant 1$, $m \geqslant 1$, x, $\bar{x} \in \{0, 1\}$, $\bar{x} \neq x$. Then $\gamma = \beta w$.
3. $w = x^k\bar{x}^m xs$, k, m, x, $\bar{x}$ like in case 2., $s \in \{0, 1\}^*$
   - 3.1 $|\alpha| \leqslant k$. Then $\gamma = x^{k-|\alpha|}\bar{x}^{m+1}$
   - 3.2 $k + 1 \leqslant |\alpha| \leqslant k + m$. Then $\gamma = \bar{x}^{m+k-|\alpha|+1}$
   - 3.3 $k + m + 1 \leqslant |\alpha|$. Then $\gamma = y$, where y is the inverse of the first letter of β. ∎

Let $T = (S, s, \sigma, \{0, 1\}, \tau)$ be the 2-uniform tag system corresponding to the automaton A from Lemma 1 according to Proposition 1. Hence $\text{seq}_T = \text{sort}_A = a_w$. Denote

$$b_w = b(0)b(1)b(2)\ldots = \text{intseq}_T = \text{state}_A.$$

Lemma 2. σ is an injective mapping.

Proof. Let $s_1, s_2 \in S$, $s_1 \neq s_1$, $\sigma(s_1) = \sigma(s_2)$. From minimality of the automaton A we get $\tau(s_1) \neq \tau(s_2)$. Let now d be the rightmost symbol of w and $\bar{d}$ its inverse. Since no occurrence of w in the word scanned by A can be terminated by $\bar{d}$, we get

$$\begin{aligned} \tau(s_1) &= \tau(\delta(s_1, \bar{d})) \\ &= \tau(\delta(s_2, \bar{d})) \qquad \text{since } \sigma(s_1) = \sigma(s_2) \\ &= \tau(s_2) \qquad \text{— a contradiction.} \end{aligned}$$ ∎

To obtain our main result we will first investigate the structure of the word b_w; the results for a_w will follow directly as can be seen from the following Lemma 4.

Let $x \in S$, $x = \langle\alpha\rangle_i$. Denote $\bar{x} = \langle\alpha\rangle_{1-i}$. Elements $x, y \in S$ will be called *associated* ($x \sim y$) if $x = y$ or $x = \bar{y}$.

Remark 1. If for some $x, y \in S$ we have $x \sim y$ and $\tau(x) = \tau(y)$, then $x = y$.

Lemma 3. For each $i \geqslant 0$, $b(i) \sim b(i + 2^{|w|-1})$.

Proof. Let $s \in S$. Then

$$\delta(s, i_{[2]}) = \langle\alpha\rangle_j \quad \text{for some} \quad \langle\alpha\rangle_j \in S$$

iff α is a terminal factor of $i_{[2]}$ (as in Lemma 1)
iff α is a terminal factor of $(i + 2^{|w|-1})_{[2]}$
iff $\delta(s, (i + 2^{|w|-1})_{[2]}) = \langle\alpha\rangle_k$ for some $k \in \{0, 1\}$. ∎

Lemma 4. If $a_w = au^{2^{|w|}}\ldots$ for some $a, u \in \{0, 1\}^*$, then $b_w = \alpha'(u')^2\ldots$ for some $\alpha', u' \in S^*$ such that $|\alpha'| = |\alpha|$, $|u'| = |u^{2^{|w|-1}}|$.

Proof. Since $|u^{2^{|w|-1}}|$ is a multiple of $2^{|w|-1}$, the assertion follows from Lemma 3 and Remark 1. ∎

Lemma 5.

(i) Let $i \geqslant 0$. Then

$$\tau(b(2i)) + \tau(b(2i+1)) \equiv \tau(b(2i + 2^{|w|})) + \tau(b(2i + 1 + 2^{|w|})) \pmod 2$$

(ii) There is exactly one $0 \leqslant j \leqslant 2^{|w|-2} - 1$ such that for all $i \geqslant 0$ and all $0 \leqslant k \leqslant 2^{|w|-2} - 1$

$$\tau(b(2^{|w|}\cdot i + 2\cdot k)) + \tau(b(2^{|w|}\cdot i + 2\cdot k + 1)) \not\equiv \tau(b(2^{|w|}\cdot i + 2\cdot k + 2^{|w|-1})) + \tau(b(2^{|w|}\cdot i + 2\cdot k + 1 + 2^{|w|-1})) \pmod 2$$

iff $k = j$.

Proof. The assertions follow from Proposition 2. ∎

A word $x \in S^*$ will be called *m-block* ($m \geqslant 0$) iff $x = \sigma^m(d)$ for some $d \in S$. A word $x \in S^*$ is *m-factorizable* iff it is a (possible empty) concatenation of m-blocks. The set of all m-blocks will be denoted $\mathscr{B}_m$, the set of all m-factorizable words will be denoted $\mathscr{F}_m$.

Remark 2. Each m-block is of length 2^m.

Each initial factor of b_w is of length divisible by 2^m iff it is m-factorizable.

An m-block x will be called *even* (*odd*) iff for some $i \geqslant 0$ $x = \sigma^m(b(2i))$ ($x = \sigma^m(b(2i+1))$).

Remark 3. For $m \geqslant 1$ each m-block is a concatenation of some even $(m-1)$-block with some odd $(m-1)$-block.

Lemma 6. For $m \geqslant 0$ no m-block can be both even and odd.

Proof. Let for some $m, i, k \geqslant 0$ $\sigma^m(b(2i)) = \sigma^m(b(2k+1))$.

1. Let $m = |w| - 1$. Let j be as in (ii) of Lemma 5. Since

$$\tau(\sigma^m(b(2i))) = \tau(\sigma^m(b(2k+1))),$$

using (i) of Lemma 5 we get

$$\tau(b(2^{m+1}k + 2j)) + \tau(b(2^{m+1}k + 2j + 1)) \equiv$$
$$\tau(b(2^{m+1}i + 2j)) + \tau(b(2^{m+1}i + 2j + 1)) \equiv$$
$$\tau(b(2^{m+1}k + 2j + 2^m)) + \tau(b(2^{m+1}k + 2j + 1 + 2^m)) \pmod 2$$

— a contradiction to (ii) of Lemma 5.

2. If $m \neq |w| - 1$, then by several applications of σ or σ^{-1} (Lemma 2) one obtains case 1. ∎

Lemma 7. If $b_w = xB\ldots$, where $B \in \mathscr{B}_m$, then $x \in \mathscr{F}_m$.

Proof. Induction on m. The case $m = 0$ is evident.

Let $m > 0$. Then $B = B_0B_1$, where $B_0, B_1 \in \mathscr{B}_{m-1}$, B_0 is even. By induction hypothesis, $x \in \mathscr{F}_{m-1}$. If $x \notin \mathscr{F}_m$, then B_0 is odd — a contradiction to Lemma 6. ∎

Lemma 8. If $b_w = x_1u\ldots = x_2u\ldots$, where $x_1 \in \mathscr{F}_m - \mathscr{F}_{m+1}$, $x_2 \in \mathscr{F}_{m+1}$, then u is a proper initial factor both of some even and some odd m-block.

Proof. u is an initial factor of an infinite word starting with an odd m-block, and of some other starting with an even m-block. Since no m-block can be both even and odd, $|u| < 2^m$. ∎

Lemma 9. If $b_w = xuBu\ldots$, B being a word of length divisible by 2^m, and $x \in \mathscr{F}_m$, then $u \in \mathscr{F}_m$.

Proof. Let $u \in \mathscr{F}_{m'} - \mathscr{F}_{m'+1}$, $m' \geqslant 0$. If $m' < m$, then $x \in \mathscr{F}_{m'+1}$, $xuB \in \mathscr{F}_{m'+1}$. From Lemma 8 and the fact that $u \in \mathscr{F}_{m'}$ follows $u = \varepsilon$ hence $u \in \mathscr{F}_{m'+1}$ — a contradiction. ∎

Lemma 10. b_w contains no factors of the form $uBuBu$ where $u \in S^*$, $B \in \mathscr{B}_m$, $m \geqslant 0$.

Proof. Let $b_w = xuBuBu\ldots$. We use induction on $|u|$.

1. $|u| = 0$. In this case $b_w = xBB\ldots$, from Lemma 7 we obtain $x \in \mathscr{F}_m$ thus B is both even and odd — a contradiction.

2. $|u| > 0$. Lemma 7 implies $xu \in \mathscr{F}_m$, $xuBu \in \mathscr{F}_m$ hence $u \in \mathscr{F}_m$.

2.1 If B is an even m-block then u can be factorized as $u = Cv$, $C \in \mathscr{B}_m$, $|v| < |u|$ and $b_w = xCvBCvBCv\ldots$, $BC \in \mathscr{B}_{m+1}$ — a contradiction to induction hypothesis.

2.2 If B is an odd m-block then a similar contradiction can be obtained using the factorization $u = vC$, $C \in \mathscr{B}_m$. ∎

Corollary 1. b_w contains no cubes.

Proof. If $b_w = xv^3\ldots$ then (if $v \neq \varepsilon$) $v = uB$ for some $B \in \mathscr{B}_0 = S$ and $b_w = xuBuBuB\ldots$. ∎

Lemma 11. b_w contains no factor of the form $xyBzxyBzx\ldots$ where $x \in S$, $B \in \mathscr{B}_m$, $zxy \in \mathscr{B}_m$, $m = |w| - 2$.

Proof. If b_w contains such a factor then it is of length $2^{|w|} + 1$. Let the first occurrence of x in this factor have in b_w index i. Then for $i \leqslant k \leqslant i + 2^{|w|-1}$ we have $b(k) = b(k + 2^{|w|-1})$, which yields a contradiction to (ii) of Lemma 5. ∎

Lemma 12. b_w contains no factor of the form $xyBzxyBzx$ where $x \in S$, $B \in \mathscr{B}_m$, $zxy \in \mathscr{B}_m$, $m \geqslant 0$.

Proof. If $B \in \mathscr{B}_m$ then $\sigma(B) \in \sigma_{m+1}$. Thus if b_w contains a factor $xyBzxyBzx$ for some m, then it contains a similar factor for $m + 1$. Lemma 1 now direct'y implies that b_w does not contain a factor $xyBzxyBzx$ for $m \leqslant |w| - 2$. For $m > |w| - 2$ we proceed by induction.

Let $b_w = \alpha xyBzxyBzx\ldots$ for some $m > |w| - 2$.

1. Let $|\alpha|$ be even, i.e. x is an even 0-block. Then $y \neq \varepsilon$ otherwise the m-block zxy would be terminated by an even 0-block. Thus $y = dv$, $d \in S$, $v \in S^*$, and

$$b_w = \alpha xdvBzxdvBzx\ldots$$

From Lemma 3 we obtain

$$b_w = \alpha xdvBzxdvBzxd'\ldots$$

where $d \sim d'$. From (i) of Lemma 5 and from Remark 1 we get

$$\tau(x) + \tau(d) = \tau(x) + \tau(d')$$
$$\tau(d) = \tau(d')$$
$$d = d'.$$

Since σ is injective, b_w can be factorized as follows:

$$b_w = \alpha' x' y' B' z' x' y' B' z' x' \ldots$$

where $\sigma(\alpha') = \alpha$, $\sigma(x') = xd$, $\sigma(y') = v$, $\sigma(B') = B$, $\sigma(z') = z$ — a contradiction to induction hypothesis.

2. If $|\alpha|$ is odd then by factorization $z = vd$ one can obtain a contradiction analogically to case 1. ∎

Lemma 13. b_w contains no factor of the form $xyBzxyBzx$ where $x \in S$, $B \in \mathscr{B}_m$, $z, y \in S^*$.

Proof. Induction on $|zxy|$.

1. $|zxy| = 1$ i.e. $z = y = \varepsilon$. The assertion follows from Lemma 10.

2. $|zxy| > 1$. Let $b_w = \alpha xyBzxyBzx\ldots$. From Lemma 7 we obtain $zxy \in \mathscr{F}_m$. Moreover, $|zxy|$ is an odd multiple of 2^m, otherwise B would be simultaneously odd and even.

2.1 $y = vA$, A being an even m-block. Then $b_w = \alpha xvABzxvABzx\ldots$, $AB \in \mathscr{B}_{m+1}$, $|zxv| < |zxy|$ a contradiction to induction hypothesis.

2.2 $z = Av$, $A \in \mathscr{B}_{m+1}$. Then $b_w = \alpha xyBAvxyBavx\ldots$, $|vxyB| < |Avxy| = |zxy|$ — a contradiction to i.h.

2.3 $z = Av$, A being an odd m-block. Then $b_w = \alpha xyBAvxyBAvx\ldots$, $BA \in \mathscr{B}_{m+1}$, $|vxy| < |zxy|$ — a contradiction to i.h.

2.5 $zxy \in \mathscr{B}_m$ — a contratiction to Lemma 12. ∎

We have now proved the following properties of b_w and a_w:

Theorem 1. b_w does not contain a factor of the form $xvxvx$, $x \in S$, $v \in S^*$.

Proof. The case $v = \varepsilon$ follows from Corollary 1.

In the case $v \neq \varepsilon$ using the factorization $v = dz$, $d \in \mathscr{B}_0 = S$, $z \in S^*$, one obtains $xvxvx = xdzxdzx$ and Lemma 13 implies that such a factor cannot be contained in b_w. ∎

Theorem 2. a_w does not contain a factor of the form $(xv)^{2^{|w|}}x$, $x \in \{0, 1\}$, $v \in \{0, 1\}^*$.

Proof. Applying Lemma 4 for $x = xv$ and $u = vx$, one obtains that b_w contains a factor of the form $x'v'x'v'x'$ — a contradiction to Theorem 1. ∎

Theorem 2 does not exclude the possibility that a_w contains a factor of the form $u^{2^{|w|}}$. Our next goal is to find some necessary conditions for appearing of such a factor in a_w. First we shall describe how the squares in b_w look like.

Lemma 14. Let $b_w = \alpha uBu\ldots$, $B \in \mathscr{B}_m$, $u \neq \varepsilon$.

Then $\alpha, u \in \mathscr{F}_m$.

Proof. Let $\alpha \in \mathscr{F}_{m'} - \mathscr{F}_{m'+1}$. Let $m' < m$. Lemma 7 implies $u \in \mathscr{F}_{m'}$. Since $\alpha \notin \mathscr{F}_{m'+1}$, the first m'-block of u is odd. Since $B \in \mathscr{F}_{m'+1}$, the same block is even — a contradiction.

Thus $m' \geqslant m$, $\alpha \in \mathscr{F}_m$. Lemma 7 implies that $u \in \mathscr{F}_m$. ∎

Lemma 15. Let $b_w = \alpha u B u \ldots$, $B \in \mathscr{B}_m$. Then $|u| = 2^q - 2^m$ for some $q \geqslant m$.

Proof. Induction on $|u|$.

1. If $|u| = 0$, then $|u| = 2^m - 2^m$.

2. Let $|u| > 0$. Lemma 14 implies that $|u|$ is divisible by 2^m.

2.1 If $|u| = 2^m$, then $|u| = 2^{m+1} - 2^m$.

2.2 Let $|u| > 2^m$. Then $u = AvC$, $A, C \in \mathscr{B}_m$, $v \in S^*$, and $b_w = \alpha AvCBAvC \ldots$. Either CB or BA is an $(m+1)$-block. By induction hypothesis, for some $q \geqslant m+1$, $|Av| = 2^q - 2^{m+1}$ or $|vC| = 2^q - 2^{m+1}$. In both cases

$$|u| = 2^q - 2^{m+1} + 2^m = 2^q - 2^m.$$ ∎

Lemma 16. Let $b_w = \alpha u u \ldots$, $u \neq \varepsilon$, $\alpha \in \mathscr{F}_m$.

Then $|u| = 2^q$ for some $q \geqslant m + |w| - 1$.

Proof. $u = vB$ for some $v \in S^*$, $B \in S = \mathscr{B}_0$. Lemma 15 implies that $|v| = 2^q - 1$ for some $q \geqslant 0$, hence $|u| = 2^q$. Let $q < |w| - 1 + m$. Since σ is injective (Lemma 2) and $b_w = \sigma(b_w)$, for $k = |w| - 1 - q$ (satisfying $-m < k \leqslant$ $\leqslant |w| - 1$), one obtains $|\sigma^k(u)| = 2^{|w|-1}$, $\sigma^k(\alpha) \in \mathscr{F}_1$, and $b_w = \sigma^k(\alpha)\sigma^k(u)\sigma^k(u)\ldots$ — a contradiction to (ii) of Lemma 5. ∎

Lemma 17. Let $b_w = \alpha u u \ldots$, $u \neq \varepsilon$, $\alpha \in \mathscr{F}_m - \mathscr{F}_{m+1}$.

Then $|u| = 2^{m+|w|-1}$ and $u \in \mathscr{F}_m - \mathscr{F}_{m+1}$.

Proof. According to Lemma 16 $|u|$ is divisible by 2^m. $\alpha \in \mathscr{F}_m$ implies that $u \in \mathscr{F}_m$. Since $\alpha \notin \mathscr{F}_{m+1}$ the initial m-block of u is odd, thus $u \notin \mathscr{F}_{m+1}$.

Considering Lemma 16 it is enough to prove $|u| \leqslant 2^{m+|w|-1}$. Let $u = 2^{r+m+|w|-1}$, $r \geqslant 0$. Then $b_w = \sigma^{-m}(b_w) = \alpha' u' u' \ldots$, where $|u'| = 2^{r+|w|-1}$, $\alpha' \in \mathscr{F}_0 - \mathscr{F}_1$. Let $u' = xv$, $x \in S$, $v \in S^*$. If $r \geqslant 1$ then from Lemma 3 and (i) of Lemma 5 it follows (since the rightmost letter of $u'u'$ has in b_w an even index) that $b_w = \alpha' xvxv \ldots = \alpha' xvxvx \ldots$ — a contradiction to Theorem 1. ∎

Corollary 2. If $b_w = \alpha u u \ldots$, $u \neq \varepsilon$, $\alpha \in \mathscr{F}_m - \mathscr{F}_{m+1}$ for some $m \geqslant 0$ then the same is true for $m = 0$.

Lemma 18. Let $b_w = \alpha u u \ldots$, $u \neq \varepsilon$, $\alpha \in \mathscr{F}_0 - \mathscr{F}_1$.

Then either for $y = \alpha$ or for $y = \alpha u$

$$(*) \qquad |y| = \mathrm{v}(w) + \bar{d} \pmod{2^{|w|}},$$

where $\bar{d}$ is the inverse of the rightmost digit of w and $\mathrm{v}(w)$ is the integer whose binary notation is w.

Proof. Lemma 17 implies that $|u| = 2^{|w|-1}$. It is easy to see that (ii) of Lemma 5 is satisfied only if $(*)$ is valid. ∎

Our knowledge of the powers in b_w and a_w can now be summarized in the following theorems:

Theorem 3. If $b_w = \alpha uu\ldots$, $u \neq \varepsilon$, $\alpha \in \mathscr{F}_m - \mathscr{F}_{m+1}$, then

(i) $|u| = 2^{m+|w|+1}$ and $u \in \mathscr{F}_m - \mathscr{F}_{m+1}$,

(ii) $b_w = \alpha' u' u' \ldots$ for some $u' \neq \varepsilon$, $\alpha' \in \mathscr{F}_0 - \mathscr{F}_1$ and either for $y = \alpha'$ or for $y = \alpha' u'$, $|y| \equiv \nu(w) + \bar{d} \pmod{2^{|w|}}$.

Theorem 4. If $a_w = \alpha u^{2^{|w|}}\ldots$, $u \neq \varepsilon$, $|\alpha|$ divisible by 2^m and not divisible by 2^{m+1}, then

(i) $|u| = 2^m$,

(ii) either for $z = |\alpha|/2^m$ or for $z = |\alpha|/2^m + 2^{|w|-1}$, $z \equiv \nu(w) + \bar{d} \pmod{2^{|w|}}$.

Using Corollary 2 one can show that b_{1101} does not contain squares, and consequently that a_{1101} does not contain a factor of the form u^{16}. On the other hand for each w of the form 1^k, $k > 1$, a_w contains the subword 0^{2k} beginning in a_w at the place with the index (in binary notation) $1w = 1^{k+1}$.

Each a_w contains the factor $0^{2^{|w|}-1}$, as shown in the following table where n is a binary notation of such an index in a_w that

$$a(n+1)a(n+2)\ldots a(n+2^{|w|}-1) = 0^{2^{|w|}-1}$$

and $x \in \{0, 1\}^*$, $k \geq 1$.

w	n	remark
$00x$	$1^{|x|}01w$	$w \notin 0^*$
$10x$	$1^{|w|}w$	$x \notin 0^*$
$11x$	$1w$	
$01x$	$01^{|w|}w$	$x \notin 0^* \cup 1^*$
10^k	$ww1^{k-1}w$	
010^k	$011011w$	
011^k	$w01ww$	
01	010000	

Acknowledgement: The contribution contains the results obtained during the author's stay at Institut de Programation, Université P. et M. Curie Paris VI, France.

The author is deeply indebted to Prof. Jean Berstel for introduction to the problem and for many stimulating discussions.

REFERENCES

CKM-FR 80 Christol, G.—Kamae, T., Mendes-France, M.—Rauzy, G.: Suites algébriques, automates et substitutions, Bull. Soc. math. France 108 (1980), 401—419.

Co 72 Cobham, A.: Uniform tag sequences, Math. Syst. Theory 6 (1972), 164—192.

Fi 80 Fife, E. D.: Binary sequences which contain no BBb, Transaction of the Am. Math. Soc. 261 (1980), № 1, 115—136.

Mo 21 Morse, H. M.: Recurrent geodesics on a surface of negative curvature, Transactions of the Am. Math. Soc., 22 (1921), 84—100.

Pa 81 Pansiot, J. J.: The Morse sequence and iterated morphisms, Inf. Proc. Letters, 12 (1981), № 2, 68—70.

Th 06 Thue, A.: Über unendliche Zeichenreihen, Videnskabs-Selskabets Skifter, Math. Naturv. Klasse, Kristiania 1906, № 7, 1—22.

Th 12 Thue, A.: Über die gegenseitige Lage gleichen Teile gewisser Zeichenreihen, Videnskapsselskapets Skifter, I. Mat. — naturv. Klasse, Kristiania 1912, № 1, 1—67.

Author's address:
Anton Černý,
Katedra teoretickej kybernetiky MFF UK,
Mlynská dolina,
842 15 Bratislava,
Czechoslovakia

Received: 8. 2. 1984

SÚHRN

ZOVŠEOBECNENÉ SLOVÁ THUEHO—MORSEHO

Anton Černý, Bratislava

V práci sa vyšetruje trieda nekonečných slov nad abecedou {0, 1}, ktoré sú zovšeobecnením nekonečného slova vznikajúceho iteráciou morfizmu 0 → 01, 1 → 10 známeho z prác Thueho a Morseho. Ukazuje sa, že takéto nekonečné slová obsahujú ako podslová len ohraničené mocniny slov.

РЕЗЮМЕ

ОБОБЩЕННЫЕ СЛОВА ТЮ—МОРСА

Антон Черны, Братислава

В работе рассуждается один класс бесконечных последовательностей в алфавите {0, 1}, являющийся обобщением последовательности порождаемой итерацией морфизма 0 → 01, 1 → 10 известной из работ Тю и Морса. Показано, что такие слова содержат в виду подслов только слова ограниченной степени.