International Journal of Applied Mathematics

Volume 31 No. 2 2018, 181-197

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v31i2.2

SOME COMBINATORIAL PROPERTIES OF THE TERNARY THUE-MORSE WORD

Idrissa Kaboré[1 §], Boucaré Kientéga[2]

[1,2] Department of Mathematics and Informatics
Université Nazi Boni
Bobo-Dioulasso, 01 BP 1091, BURKINA FASO

Abstract: In this paper we study some combinatorial properties of the ternary Thue-Morse word, $\mathbf{t}_3$. More precisely, we focus on squares of letters and the factors of $\mathbf{t}_3$ which separate them. We also establish that the number of return words of a given factor of $\mathbf{t}_3$ is 7, 8 or 9.

AMS Subject Classification: 68R15, 11B85
Key Words: infinite word, square of letters, separator factors, return words

1. Introduction

The study of infinite words goes back at least to 1906 with the works of Thue [22], [23]. The Thue-Morse word, $\mathbf{t}_2$, over the binary alphabet $\{0, 1\}$ is the infinite word generated by the morphism μ_2 defined by $\mu_2(0) = 01$, $\mu_2(1) = 10$. This word was extensively studied [4] and some numerous combinatorial properties are known [21, 16, 2] in the literature.

The Thue-Morse word can be naturally generalized over an alphabet $\mathcal{A}_q$ of size $q \geq 3$. More precisely, on the alphabet $\mathcal{A}_q = \{0, 1, ..., q-1\}$, it is the infinite word $\mathbf{t}_q$ generated by the morphism μ_q defined by: $\mu_q(k) = k(k+1)...(k+q-1)$, where the letters are expressed modulo q. A study of this word has been done in [3].

In an infinite word u, a return word of a factor v appearing at least twice in u is a factor w such that wv is a factor of u beginning with v and containing

Received: December 19, 2017

§Correspondence author

exactly two occurrences of v. This notion was introduced by Durand in [9]. In this paper return words were used as a technical tool to give a characterization of primitive morphic words. Recently, this notion appears regularly in the literature dedicated to infinite recurrent words [11, 15, 10]. In particular, the return words of some classes of words are studied [16, 2, 24, 13]. In this paper, we are interested in the combinatorial study of the Thue-Morse word $\mathbf{t}_3$ over the alphabet $\mathcal{A}_3 = \{0, 1, 2\}$: description of separator of squares of letters, bispecial biprolongable factors and return words.

After introducing the necessary notations and definitions, in Section 2 we state some properties of $\mathbf{t}_3$. Section 3 is devoted to the study of separator factors of the squares of letters in $\mathbf{t}_3$. Mainly, we determine the structure and the lengths of these factors. The number of occurrences of squares of letters in a given factor of $\mathbf{t}_3$ are estimated (Section 4). We end the paper with the study of the cardinality of the set of return words of an arbitrary factor of $\mathbf{t}_3$ (Section 5).

2. Preliminaries

Let $\mathcal{A}$ be a finite alphabet. The set of finite words over $\mathcal{A}$ is noted $\mathcal{A}^*$ and ε represents the empty word. The set of non-empty finite words over $\mathcal{A}$ is denoted by $\mathcal{A}^+$. For all $u \in \mathcal{A}^*$, $|u|$ designates the length of u and the number of occurrences of a letter a in u is denoted $|u|_a$. A word u of length n formed by repeating a single letter x is denoted x^n.

An infinite word is a sequence of letters of $\mathcal{A}$, indexed by $\mathbb{N}$. We denote by $\mathcal{A}^\omega$ the set of infinite words on $\mathcal{A}$. The set of finite or infinite words on $\mathcal{A}$ is denoted $\mathcal{A}^\infty$.

Let $u \in \mathcal{A}^\infty$ and $v \in \mathcal{A}^*$. The word v is called factor of u if there exist $u_1 \in \mathcal{A}^*$ and $u_2 \in \mathcal{A}^\infty$ such that $u = u_1 v u_2$. The factor v is called prefix (resp. suffix) if u_1 (resp. u_2) is empty. The set of the prefixes (resp. the suffixes) of u is denoted $pref(u)$ (resp. $suff(u)$).

Let u and v be two non empty finite words. If v is a prefix (resp. a suffix) of u the $v^{-1}u$ (resp. uv^{-1}) is the word obtained from u by erasing the prefix v (resp. the suffix v).

Let u be an infinite word. The set of factors of length n of u is denoted $\mathcal{F}_n(u)$. The set of all the factors of u is denoted $\mathcal{F}(u)$.

Let u, v_1, v_2 and w be some finite words such that $u = v_1 w v_2$. Then, w is called median factor of u if $|v_1| = |v_2|$.

Let v be a factor of u and a be a letter of $\mathcal{A}$. We say that v is right (resp.

left) prolongable by a, if va (resp. av) is also a factor of u. In this case, the word va (resp. av) is called a right (resp. left) extension of v in u; simply we say that the letter a is a right (resp. left) extension of v. The factor v is said to be right (resp. left) special if it admits at least two right (resp. left) extensions in u. If v is both right special factor and left special factor, it is called bispecial factor of u.

An infinite word u is said to be recurrent if any factor of u appears infinitely many times in u. It is said to be uniformly recurrent if for any natural n, it exists a natural n_0 such that any factor of length n_0 contains all the factors of length n of u.

Let u be an infinite recurrent word and v, w be two factors of u such that wv occurs in u. If v is prefix of wv and wv contains exactly two occurrences of v then w is called return word of v in u. The set of the return words of v is denoted $Ret(v)$.

In order to get the sets of the return words in a recurrent word u, we need to use its bispecial factors [2].

A morphism on $\mathcal{A}^*$ is a map $f : \mathcal{A}^* \to \mathcal{A}^*$ such that $f(uv) = f(u)f(v)$, for all $u, v \in \mathcal{A}^*$. A morphism σ is said to be:

• primitive if there exists a positive integer n such that, for any letter a of $\mathcal{A}$, $f^n(a)$ contains all the letters of $\mathcal{A}$.

• k-uniform, if $|\sigma(a)| = k$ for any letter a of $\mathcal{A}$.

• left (resp. right) marked on the alphabet $\mathcal{A} = \{a_1, a_2, ..., a_d\}$, if the first (resp. last) letters of $f(a_i)$ and $f(a_j)$ are different, for all $i \neq j$. If f is both left marked and right marked, it is called marked morphism.

An infinite word u is generated by a morphism f if there exists a letter a such that the words a, $f(a)$, ..., $f^n(a)$, ... are longer and longer prefixes of u. We note $u = f^\omega(a)$. An infinite word generated by a morphism is called purely morphic word.

Let $u = f^\omega(a)$ be a purely morphic word and w be a factor of u verifying

$$|w| \geq \max\{|f(a)| : a \in \mathcal{A}\}.$$

Then, w can be decomposed in the form

$$p_0 f(a_1)f(a_2)...f(a_n)s_{n+1},$$

where

• $n \geq 0$, a_0, a_1, ..., $a_{n+1} \in \mathcal{A}$;

• p_0 is a suffix of $f(a_0)$ and s_{n+1} is a prefix of $f(a_{n+1})$.

The word $v = a_1a_2...a_n$ is called ancester of w.

Consider the alphabet $\mathcal{A} = \{0, 1, 2\}$. The ternary Thue-Morse word denoted $\mathbf{t}_3$ is the infinite word generated by the morphism μ_3 defined by: $\mu_3(0) = 012$, $\mu_3(1) = 120$ and $\mu_3(2) = 201$.

Definition 1. ([4]) Let u be an infinite word on an alphabet $\mathcal{A}$. One says that u admits :

• frequencies of letters if for any letter a and for any sequence (u_n) of prefixes of u such that $\lim_{n \longrightarrow +\infty} |u_n| = +\infty$, then $\lim_{n \longrightarrow +\infty} \frac{|u_n|_a}{|u_n|}$ exists.

• uniform frequencies of letters if for any letter a and for any sequence (v_n) of factors of u such that $\lim_{n \longrightarrow +\infty} |v_n| = +\infty$, then $\lim_{n \longrightarrow +\infty} \frac{|v_n|_a}{|v_n|}$ exists.

Definition 2. Let u be an infinite word on an alphabet $\mathcal{A}$. The word u admits:

• frequencies of factors if for any factor v of u and for any sequence of (u_n) of prefixes of u such that $\lim_{n \longrightarrow +\infty} |u_n| = +\infty$, then $\lim_{n \longrightarrow +\infty} \frac{|u_n|_v}{|u_n|}$ exists.

• uniform frequencies of factors if for any factor v and for any sequence (u_n) of factors of u such that $\lim_{n \longrightarrow +\infty} |u_n| = +\infty$, then $\lim_{n \longrightarrow +\infty} \frac{|u_n|_v}{|v_n|}$ exists.

Notation: The frequency of the non-empty factor v in u is denoted $f_v(u)$, if it exists.

In the following, we consider the alphabet $\mathcal{A}_3 = \{0, 1, 2\}$.

In [2], the authors obtained the frequencies of the factors in the generalized Thue-Morse word. The frequencies of the factors of length 2 of the word $\mathbf{t}_3$ have the form $\frac{f}{3^k}$, $k \in \{0, 1, 2\}$ and $f = f_{01}(\mathbf{t}_3) = \frac{3}{13}$. So, $f_{00}(\mathbf{t}_3) = \frac{1}{39}$ and $f_{10}(\mathbf{t}_3) = \frac{1}{13}$.

Let $u = u_1 u_2 \ldots u_n \in \mathcal{A}^*$. We call mirror of u the word denoted $\overline{u}$ and defined by $u_n \ldots u_2 u_1$.

The $(0, 2)$-complement of u, denoted $\widetilde{u}$ is the word defined by $\widetilde{u} = (2 - u_1)(2 - u_2)\ldots(2 - u_n)$.

For all $u \in \mathcal{A}^*$ we have $\overline{\widetilde{u}} = \widetilde{\overline{u}}$. So, we set $\widehat{u} = \overline{\widetilde{u}}$.

The operation $\widehat{.}$ is involutive: $\widehat{\widehat{u}} = u$ for all $u \in \mathcal{A}^*$.

Proposition 1. *Consider the sequences of words* (u_n), (v_n) *and* (w_n) *defined by* $u_0 = 0$, $v_0 = 1$, $w_0 = 2$ *and for all natural* n, $u_{n+1} = u_n v_n w_n$,

$v_{n+1} = v_n w_n u_n$ and $w_{n+1} = w_n u_n v_n$. Let $w \in \mathcal{F}(\mathbf{t}_3)$. Then

1. *For all $n \geq 0$ we have: $u_n = \mu_3^n(0)$, $v_n = \mu_3^n(1)$, $w_n = \mu_3^n(2)$.*
2. *The sequences (u_n), (v_n) and (w_n) verify:*
   - $\widehat{u}_{3n} = w_{3n}$, $\widehat{v}_{3n} = v_{3n}$ *and* $\widehat{w}_{3n} = u_{3n}$;
   - $\widehat{u}_{3n+1} = u_{3n+1}$, $\widehat{v}_{3n+1} = w_{3n+1}$ *and* $\widehat{w}_{3n+1} = v_{3n+1}$;
   - $\widehat{u}_{3n+2} = v_{3n+2}$, $\widehat{v}_{3n+2} = u_{3n+2}$ *and* $\widehat{w}_{3n+2} = w_{3n+2}$.

Proof. We process by induction on n.

1. The formulae are verified for $n = 0$ since, by definition, we have $u_0 = 0 = \mu_3^0(0)$, $v_0 = 1 = \mu_3^0(1)$ and $w_0 = 2 = \mu_3^0(2)$.

   Assume that for all $n \geq 1$, $u_n = \mu_3^n(0)$, $v_n = \mu_3^n(1)$ and $w_n = \mu_3^n(2)$. Then, we have:

   $u_{n+1} = u_n v_n w_n = \mu_3^n(0)\mu_3^n(1)\mu_3^n(2) = \mu_3^n(012) = \mu_3^{n+1}(0)$;
   $v_{n+1} = v_n w_n u_n = \mu_3^n(1)\mu_3^n(2)\mu_3^n(0) = \mu_3^n(120) = \mu_3^{n+1}(1)$;
   $w_{n+1} = w_n u_n v_n = \mu_3^n(2)\mu_3^n(0)\mu_3^n(1) = \mu_3^n(201) = \mu_3^{n+1}(2)$.

2. The relations are verified for $n = 0$. Assume that for $n \geq 1$ we have

$$\begin{cases} \widehat{u}_{3n} = w_{3n}, \ \widehat{v}_{3n} = v_{3n} \\ \widehat{u}_{3n+1} = u_{3n+1}, \ \widehat{v}_{3n+1} = w_{3n+1} \\ \widehat{u}_{3n+2} = v_{3n+2}, \ \widehat{w}_{3n+2} = w_{3n+2} \end{cases}.$$

   Then

   - $\begin{cases} \widehat{u}_{3(n+1)} = \widehat{w}_{3n+2}\widehat{v}_{3n+2}\widehat{u}_{3n+2} = w_{3n+2}u_{3n+2}v_{3n+2} = w_{3(n+1)} \\ \widehat{v}_{3(n+1)} = \widehat{u}_{3n+2}\widehat{w}_{3n+2}\widehat{v}_{3n+2} = v_{3n+2}w_{3n+2}u_{3n+2} = v_{3(n+1)} \end{cases}.$

   Similarly we check that:

   - $\widehat{u}_{3(n+1)+1} = u_{3(n+1)+1}$, $\widehat{v}_{3(n+1)+1} = w_{3(n+1)+1}$;
   - $\widehat{u}_{3(n+1)+2} = u_{3(n+1)+2}$, $\widehat{v}_{3(n+1)+2} = w_{3(n+1)+2}$.

□

The exchange morphism on $\mathcal{A}_3$ is the morphism E defined by: $E(0) = 1$, $E(1) = 2$, $E(2) = 0$.

Proposition 2. *1. For all n we have: $E(\mathcal{F}_n(\mathbf{t}_3)) = \mathcal{F}_n(\mathbf{t}_3)$.*

2. *For all $a \in \mathcal{A}_3$, the word aaa does not occur in* $\mathbf{t}_3$.

3. *The words* 0120120, 1201201 *and* 2012012 *do not occur in* $\mathbf{t}_3$.

Proof. 1. Let $w \in \mathcal{F}_n(\mathbf{t}_3)$. Then, $|E(w)|$ and w have the same length because E just exchange the letter in w. Furthermore, there exists n_0 such that w occurs in $u_{n_0} = \mu_3^{n_0}(0)$. So, $E(w)$ occurs in $E(\mu_3^{n_0}(0))$. Moreover, we have $\mu_3 E = E\mu_3$. Thus, $E(w)$ occurs in $\mu_3^{n_0}(1)) = v_{n_0}$. Therefore, $E(w) \in \mathcal{F}_n(\mathbf{t}_3)$ since $v_{n_0} \in \mathcal{F}(\mathbf{t}_3)$. So, $E\left(\mathcal{F}_n(\mathbf{t}_3)\right) \subset \mathcal{F}_n(\mathbf{t}_3)$. As E is bijective it results $E\left(\mathcal{F}_n(\mathbf{t}_3)\right) = \mathcal{F}_n(\mathbf{t}_3)$.

2. It suffices to observe that any square of letters begins (resp. ends) with the last (resp. first) letter of the image of some letter. As no image of letter begins or ends by a square of letters, then $\mathbf{t}_3$ can not contain a cube.

3. This assertion follows from 2.

□

3. Squares of Letters and Separator Factors

In this section the structure of the factors which separate the squares of letters in $\mathbf{t}_3$ are studied.

We call separator factor, a factor which separates two consecutive squares of letters in $\mathbf{t}_3$.

Proposition 3. *Let w be a separator factor. Then $|w| \in \{7, 16\}$.*

Proof. From a given square of letters we construct w by using μ_3 until the next square of letters. As we have $E(F(\mathbf{t}_3)) = F(\mathbf{t}_3)$ it is enough to handle the case of one of the squares of letters. So, consider the square 11 to continue.

Let w_1, w_2 and u be some factors of $\mathbf{t}_3$ such that $v = w_1 11 w_2$. Observe that 11 is due to $\mu_3(21)$. We can write $v = w'_1\mu_3(21)w'_2$, where w'_1 (resp. w'_2) is a prefix (resp. suffix) of w_1 (resp. w_2). As the factor 21 begins (resp. ends) by the last (resp. the first) letter of the image of 0 (resp. 1), then it admits a unique left (resp. right) extension. So, $v = w''_1\mu_3(012120)w''_2$, where w''_1 (resp. w''_2) is a prefix (resp. suffix) of w'_1 (resp. w'_2). For further, take $v_1 = w''_1\mu_3(012)$, $v_2 = \mu_3(120)w''_2$ and use v_2.

Moreover, 120 is right triprolongable. Thus, 120012, 120120 and 120201 are factors of $\mathbf{t}_3$. As a consequence, v_2 takes one of the following forms:

$\mu_3(120012)u_1$, $\mu_3(120120)u_2$ or $\mu_3(120201)u_3$, where u_i is a suffix of w'', for all $i \in \{1, 2, 3\}$. In others terms, v_2 is equal $120201012012120210u_1$, 120201012120 $201012u_2$ or 1202010122010– $12120u_3$.

For $v_2 = 1\underline{2020101}2201012120u_3$, it appears that the square which follows 11 is 22 and the factor which separates them is of length 7.

For $v_2 = 120201012012120210u_1$, we have $v_2 = \mu_3(\mu_3(10))u_1$. Moreover, 1 is the unique right extension of 10. So, 1012 is in $\mathbf{t}_3$. Thus, $v_2 = \mu_3(\mu_3(1012))u'_1$, where u'_1 is a suffix of u_1. We get:

$$v_2 = 1\underline{2020101201212020}11202010122010121 20u'_1.$$

For $v_2 = 120201012012120210u_1$, we proceed in a similar way as to the previous case and we get $v_2 = \mu_3(\mu_3(1120))u'_2$, where u'_2 is a suffix of u_2. Thus, we have:

$$v_2 = 1\underline{2020101212020101}2201012120012120201.$$

After all, the length of a separator of squares of letters is either 7 or 16.

□

Observe that the squares which follow the 11 are 11 or 22.

Corollary 4. *Let w be a separator factor. We have the following properties.*

1. *If w separates ii and jj then $i = j$ or $j = E(i)$.*
2. *The set of separators of length 7 is:*

$$\mathcal{S}(7) = \{i^{-1}\mu_3^2(i)j^{-1} : i \in \mathcal{A}_3, j = E(i)\}.$$

3. *If $|w| = 16$ then we can have $iiwii$ or $iiwjj$ in $\mathbf{t}_3$, $i \neq j$.*

   (a) *The separators which separate identical squares of letters are given by the set:*

   $$\mathcal{S}_{i-i}(16) = \{i^{-1}\mu_3^2(ij)i^{-1} : i \in \mathcal{A}_3, j = E^2(i)\}.$$

   Every separator in this set admits a median factor which is image of square of some letters.

   (b) *The separators which separate non identical squares are given by the set:*

   $$\mathcal{S}_{i-j}(16) = \{i^{-1}\mu_3^2(ii)j^{-1} : i \in \mathcal{A}_3, j = E(i)\}.$$

Proof. 1. Recall that in $\mathbf{t}_3$, after 11 the next square of letters is 11 or 22. Let w_1 (resp. w_2) be a separator factor of 11 with 11 (resp. 22). Then, consider the factors $u = 11w_111$ and $v = 11w_222$ of $\mathbf{t}_3$. From Lemma 2, it follows that $E(u) = 22E(w_1)22$ (resp. $E(v) = 22E(w_2)00$) and $E^2(u) = 00E^2(w_1)\ 00$ (resp. $E^2(v) = 00E^2(w_2)11$) are in $\mathbf{t}_3$. Thus, we get the separator factors of 22 (resp. 00).

2. The separator factor of length 7 preceded by 11 is $w = 2020101$ (see the proof of Proposition 3). If we consider the image of w by E (resp. E^2) we get the separator factor of length 7 preceded by 22 (resp. 00). As w separates the squares 11 and 22, then $E(w)$ (resp. $E^2(u)$) separates the squares 22 and 00 (resp. 00 and 11).

3. Let w be a separator factor of length 16. Consider the case where w is preceded by 11. The two other cases coincide with $E(w)$ and $E^2(w)$. We have that follows.

   - $w = 20201\underline{012012}12020 = 1^{-1}\mu_3^2(10)1^{-1}$ if w is followed by 11. Moreover, the image of 00 is a median factor of w.
   - $w = 2020101212020101 = 1^{-1}\mu_3^2(11)2^{-1}$ if w is followed by 22.

□

Observe that no separator factor of length 7 separates two identical squares of letters.

Corollary 5. *Every factor of* $\mathbf{t}_3$ *of length* 19 *contains at least one square of letters and at most two. This length is minimal.*

Proof. Let w be a factor of $\mathbf{t}_3$ of length 19. Then w synchronizes in a unique way of the form $i\mu_3(u)$, $\mu_3(u)j$ or $ij\mu_3(u)kl$ with $i, j, k, l \in \mathcal{A}_3$. We have two cases to discuss.

Case 1: The factor w is of the form $i\mu_3(u)$. Then, the factor u verifies $|u| = 6$. So, w can be written in the form $\mu_3(ij)$, $jk\mu_3(i)l$ or $j\mu_3(i)kl$, where $i, j, k, l \in \mathcal{A}_3$. Consider the different forms of u.

• If $u = \mu_3(ij)$, then ij belongs to $\{00, 01, 02, 10, 11, 12, 20, 21, 22\}$. Without lost of generality, let us see the case where $i = 0$. Observe that 00 (resp. 02) begins with the last letter of the image of some letter. So, 2 is the unique left extension of 00 and 02. Moreover, 01 is left triprolongable. After taking all the left extensions of 00, 02 and 01 in $\mathbf{t}_3$, w takes one of the following values: 0012120201012120201, 0012120201201012120 or i01212020112020−1012, $i \in \mathcal{A}_3$.

• The factor u is of the form $jk\mu_3(i)l$. Consider the case where $i = 0$. Then, we have $u = jk012l$. Moreover, 012 is both left triprolongable and right triprolongable. As a consequence, jk browses all the values of the set $\{01, 12, 20, \}$. If $jk = 12$ (resp. 20), then $l = 1$, since neither 000 nor 002 (resp. neither 102 nor 100) is in $\mathbf{t}_3$. Thus, the factor u takes one of the following forms $010120, 010121, 010122, 120121$, 200121. By applying μ_3 to each of these factors after extension on the left, we get the following values for w: 2120201012120201120, 0201012012120201120, 10121200121− 20201$\mu_3(l)$, $l \in \mathcal{A}_3$.

We proceed in a similar way as in the case where $u = j\mu_3(i)kl$.

Case 2: The factor w is of the form $ij\mu_3(u)kl$. Then, $|u| = 5$ and u can be synchronized in the form $i'j'\mu_3(k')$, $\mu_3(k')i'j'$ or $i'\mu_3(k')j'$. Without lost of generality, let us take $k' = 0$.

– Suppose that $u = i'j'\mu_3(0)$. As 0 is left triprolongable, then $i'j'$ browses all the set $\{01, 12, 20\}$.

- If $i'j' = 01$ then $2u = 201012 = \mu_3(20)$. As the factor 20 is left triprolongable, then $w = i201012120012120201$, $i \in \{0, 1, 2\}$.
- If $i'j' = 12$, then $0u = 012012 = \mu_3(00)$. Moreover, the letter 2 is the unique left extension of 00 in $\mathbf{t}_3$. So, $w = 001212020101212020 1$.
- If $i'j' = 20$, then $1u = 120012 = \mu_3(10)$. Moreover, in $\mathbf{t}_3$, 0 is the unique left extension of 10. So, $w = 112020101201212020 1$.

– Suppose that $u = i'\mu_3(0)j'$. In a similar way as in the case 1, we verify that w takes one of the following forms: 010120121202011- 2020, 12120012120201120 20, 20201012120201120 20, 121200121202- 0120101, 1212001212020101212.

Observe to finish that $120201012012120201 = \mu_3^2(10)$ is a factor of length 18 of $\mathbf{t}_3$ without square of letters.

□

Proposition 6. *Let w be a separator factor of length* 16*. Then, w is preceded (resp. followed) by a separator factor of length* 7*.*

Proof. Let w be a separator factor of length 16 preceded by 11. Then, w takes one of the following forms:

$$1^{-1}\mu_3^2(10)1^{-1},\ 1^{-1}\mu_3^2(11)2^{-1}.$$

Note that 10 (resp. 11) admits a unique left extension and a unique right extension. Thus, 0101 and 0112 occur in $\mathbf{t}_3$. By applying μ_3^2 to these two factors, we verify that w is preceded (resp. followed) by $0^{-1}\mu_3^2(0)1^{-1}$ (resp. $2^{-1}\mu_3^2(2)0^{-1}$ or $1^{-1}\mu_3^2(0)2^{-1}$). From Corollary 4 these factors are separator factors of length 7.

One proceeds similarly if we take w preceded by 00 or 22. □

Proposition 7. *Let w_1 and w_2 be two separator factors of length* 16. *Consider u a factor separating w_1 and w_2 such that u does not contain a separator factor of length* 16. *Then, $|u| \in \{11, 38\}$.*

Proof. Let u be a factor of $\mathbf{t}_3$ separating two separator factors w_1 and w_2 of length 16. By Proposition 6, the minimal length of u is 11.

First, show that the length of u is not between 11 and 38.

Suppose that u contains two separator factors of length 7. Then, $|u| = 20$ and u has the form iiu_1jju_2kk, where $j = E(i)$, $k = E^2(i)$. Even if it means changing the roles of letters, we take $i = 0$. As a consequence, $u = 00u_111u_222 = 0\mu_3^2(01)2$. Thus, one of the factors $00w_1uw_200$, $00w_1uw_222$, $22w_1uw_200$ and $22w_1uw_222$ is in $\mathbf{t}_3$ (Corollary 5). These factors can be written respectively as follows:

$$0\mu_3^2(020122)0,\ 0\mu_3^2(020121)2,\ 2\mu_3^2(220122)0, 2\mu_3^2(220121).$$

But, the words 020122, 020121, 220122 and 2202121 do not occur in $\mathbf{t}_3$, since 02 and 22 (resp. 21 and 22) are not suffixes (resp. prefixes) of images of letters. So, u cannot have the form iiu_1jju_2kk.

Suppose that u contains three separator factors of length 7. Then, $|u| = 29$ and u has the form $iiu_1jju_2kku_3ii$, where $j = E(i)$ and $k = E^2(i)$. As previously, we end in a contradiction.

Now, show that the length of u cannot be superior to 38.

Suppose that the length of u is 47, the next possible value after the length 38. Then, u is of the form $iiu_1jju_2kku_3iiu_4jju_5kk$. Even if it means changing the roles of the letters, take $i = 0$. We have $u = 00u_111u_222u_300u_411u_522$ or $u = 0\mu_3^2(01201)2$. If w_1 is preceded by 00 (resp. 22), then $0\mu_3^2(0201201)$ (resp. $2\mu_3^2(2201201)$) occurs in $\mathbf{t}_3$. This is impossible, because the word 0201201 (resp. 2201201) does not occur in $\mathbf{t}_3$. Similarly, we show that u cannot take the following possible length, 56.

Suppose that $|u| > 56$. Then, one verifies that u contains the factor $\mu_3(012)\mu_3(012)\mu_3(012) = \mu_3(012012012)$. This is impossible by Proposition 6.

If the length of u is 38 and u begins by the square 00, then we have

$$u = 00121202011202010122010121200121202011 = 0\mu_3^2(0120)1.$$

Moreover, 0120 is left (resp. right) prolongable by 2 (resp. 1). Consequently, $0\mu_3^2(0120)1$ occurs in $\mathbf{t}_3$. Similarly, we show that u occurs in $\mathbf{t}_3$ if u begins with the square 11 or 22.

After all, $|u| \in \{11, 38\}$. □

4. Estimation of the Numbers of Squares of Letters in the Factors of $\mathbf{t}_3$

In this section, we estimate the number of squares of letters in a given factor of $\mathbf{t}_3$.

Consider the set $G = \{i\mu_3^3(i)\mu_3^2(i)j : i \in \mathcal{A}_3, j = E(i)\}$. Observe that any element of G is in $\mathbf{t}_3$. Moreover, from Proposition 7 these elements does not overlap.

Notation: Let $a, b \in \mathcal{A}_3$, $a \neq b$. We denote R_{ab} the application defined from $F(\mathbf{t}_3)$ to $\mathcal{A}_3^*$ which consisting to replace in each occurrence of element of G the 3^{rd} square of letters by ab.

Proposition 8. *Let w be a factor of $\mathbf{t}_3$ of length n beginning with a square of letters. Let α be the number of occurrences of the elements of G in w. If n is multiple of 27, then*

$$|w|_{00} + |w|_{11} + |w|_{22} = \frac{2n}{27} + \alpha.$$

Proof. Let w be a factor of length n of $\mathbf{t}_3$, beginning with a square of letters. Recall that the elements of G are the longest factors of $\mathbf{t}_3$, which contain only separator factors of length 7. As a consequence, if $|w|$ is multiple of 27, then there exists a natural k such that:

$R_{ab}(w) = (iiXjjY)^k$, where X and Y are separator factors. Thus,

$$|R_{ab}(w)|_{00} + |R_{ab}(w)|_{11} + |R_{ab}(w)|_{22} = \frac{2n}{27}.$$

The result is obtained by adding α to the previous equality.

□

Theorem 9. *Let w be a factor of $\mathbf{t}_3$ of length n. Then, there exist some naturals N and β verifying $N \leq n$ and $\beta \in \{0, 1, 2, 3\}$ such that*

$$|w|_{00} + |w|_{11} + |w|_{22} = \frac{2N}{27} + \alpha + \beta,$$

where $\alpha = \#\{|w|_z : z \in G\}$.

Proof. For n large enough, decompose w in the form $w = w_1 u w_2$, where w_1 is the prefix of w preceding the first square of letters, u the factor of w beginning with the first square of letters such that $|u| \equiv 0[27]$ and w_2 the suffix of w verifying $|w_2| < 27$. Let us set $N = |u| = n - (|w_1| + |w_2|)$. From Proposition 8 we have $|u|_{00} + |u|_{11} + |u|_{22} = \frac{2N}{27} + \alpha$.

Determine now the number of possible squares of letters in w_2. We have two cases to discuss: $w_2 \in G$ or $w_2 \notin G$.

If $w_2 \in G$ then we have:

$$|w_2|_{00} + |w_2|_{11} + |w_2|_{22} = \begin{cases} 1 & \text{if } 2 \leq |w_2| \leq 10 \\ 2 & \text{if } 11 \leq |w_2| \leq 19 \\ 3 & \text{if } 20 \leq |w_2| \end{cases}.$$

If $w_2 \notin G$, then w_2 contains at most two squares of letters. Let us set β the number of squares of letters in w_2. So, it follows:

$$|w|_{00} + |w|_{11} + |w|_{22} = \frac{2N}{27} + \alpha + \beta.$$

□

Proposition 10. *Each element of* G *comes from the image of a square of letters by* μ_3^3.

Proof. Let $w \in G$. Suppose that w begins with the square 11. Then, w is of the form $1\mu_3^3(1)\mu_3^2(1)2 = 1\mu_3^2(1201)2$. Moreover, the factor 1201 is bispecial biprolongable. As a consequence, the factors 012012, 012010 and 212012 are in $\mathbf{t}_3$. Observe that among these factors, only the image of 012012 by μ_3^2 contains w and we have:

$$\mu_3^2(012012) = \mu_3^3(00) = 01212020w01012120.$$

We proceed similarly when we suppose that w begins with the square 00 or 22. □

Consequently, for all $w \in G$, no element of G occurs in $\mu_3(w)$ or $\mu_3^2(w)$ since the squares of letters are the only squares of factors of $\mathbf{t}_3$ occurring in w.

Proposition 11. *The longest factors of* $\mathbf{t}_3$ *which contain squares of letters and do not contain images by* μ_3 *of squares of letters are factors of the form* $u_1 w u_2$, *where* $w \in G$, *and* $|u_1| = |u_2| = 10$.

Proof. By Corollary 4, note that every separator factor of a same square of letters is centered on the image by μ_3 of a square of letters. Now, among the separator factors of squares of letters only those of length 7 do not contain some square of factors. Furthermore, the elements of G are the longest factors of $\mathbf{t}_3$ which do not contain separator factors of length 16. Thus, we just have to determine for all w in G, the longest left (resp. right) extension which does not contain image of square of letters by μ_3. Suppose that w comes from the image of 00 and put $u = \mu_3^3(00)$. We have

$$u = \mu_3^3(00) = 01212020w01012120.$$

Moreover, the letter 2 (resp. 1) is the unique left (resp. right) extension of 00. From the image of 2001 by μ_3^3 one observes that the first square 11 (resp. the last square 22) of w is preceded by 11 (resp. 22). As a result, the factor 01212020 (resp. 01012120) is left (resp. right) prolongable by 012 (resp. 120). Therefore, u_1 (resp. u_2) contains the image of a square of letters if $|u_1| \geq 11$ (resp. $|u_2| \geq 11$).

We proceed in a same way when w comes from the image of 11 (resp. 22). □

Proposition 12. *Let w be a factor of $\mathbf{t}_3$ of length n and α be the number of occurrences of elements of G in w. Then $\frac{\alpha}{n}$ tends to $\frac{1}{351}$ when n tends to ∞.*

Proof. By Theorem 9, we have

$$|w|_{00} + |w|_{11} + |w|_{22} = \frac{2N}{27} + \alpha + \beta.$$

Since $\mathbf{t}_3$ admits frequencies of factors [3], then it follows

$$f_{00}(\mathbf{t}_3) + f_{11}(\mathbf{t}_3) + f_{22}(\mathbf{t}_3) = \frac{2}{27} + \delta,$$

where $\delta = \lim_{n \longrightarrow \infty} \frac{\alpha}{n}$. Moreover, $f_{00}(\mathbf{t}_3) = f_{11}(\mathbf{t}_3) = f_{22}(\mathbf{t}_3) = \frac{1}{39}$. Thus, $\frac{1}{13} = \frac{2}{27} + \delta.$

□

5. Bispecial Biprolongable Factors and Return Words in $\mathbf{t}_3$

The bispecial triprolongable factors of $\mathbf{t}_3$ have been studied in [14]. In this section, we establish some properties of the bispecial biprolongable factors of $\mathbf{t}_3$. Then, we determine the cardinality of the set of return words of a given factor in $\mathbf{t}_3$.

For the further, $BSB(\mathbf{t}_3)$ (resp. $BST(\mathbf{t}_3)$) will denote the set of the bispecial biprolongable (resp. triprolongable) factors of $\mathbf{t}_3$.

The following theorem and proposition are obtained in [14]:

Theorem 13. *The set $BST(\mathbf{t}_3)$ is given by:*

$$BST(\mathbf{t}_3) = \bigcup_{n\geq 0}\{\mu_3^n(0),\, \mu_3^n(1),\, \mu_3^n(2),\, \mu_3^n(01),\, \mu_3^n(12),\, \mu_3^n(20)\} \cup \{\varepsilon\}.$$

Proposition 14. *Let u be an element of $BST(\mathbf{t}_3)$ such that $|u| \geq 3$. Then, there exists u' in $BST(\mathbf{t}_3)$ such that $u = \mu_3(u')$.*

Proposition 15. *Let w be a factor of $\mathbf{t}_3$. Then, w is right (resp. left) biprolongable if and only if $\mu_3(w)$ is right (resp. left) biprolongable.*

Proof. Let w be a right biprolongable factor of $\mathbf{t}_3$. Without lost of generality, suppose that $w0$ and $w1$ are in $\mathbf{t}_3$. Then, $\mu_3(w)0$ and $\mu_3(w)1$ are in $\mathbf{t}_3$, since $\mu_3(i)$ begins with i for all $i \in \mathcal{A}_3$. Conversely, let w be a factor of $\mathbf{t}_3$ such that $\mu_3(w)$ is right biprolongable. Then, we have $|w| \geq 2$. Thus, $|\mu_3(w)0|$, $|\mu_3(w)1| \geq 7$ and each of them admits a unique synchronization. Even if it means exchanging the roles of the letters, suppose that $\mu_3(w)0$ and $\mu_3(w)1$ are in $\mathbf{t}_3$. As 0 and 1 are the first letters of their images respectively then the words $\mu_3(w)012$ and $\mu_3(w)120$ are in $\mathbf{t}_3$. These two factors can be written respectively $\mu_3(w0)$ and $\mu_3(w1)$ in a unique way. So, w is right biprolongable in $\mathbf{t}_3$. We proceed in a same way for the left biprolongable factors. □

Thus, every factor w of $\mathbf{t}_3$ is bispecial biprolongable if $\mu_3(w)$ is bispecial biprolongable and conversely.

Proposition 16. *Let $w \in BSB(\mathbf{t}_3)$ such that $|w| \geq 7$. Then, there exists a unique word v in $BSB(\mathbf{t}_3)$ such that $w = \mu_3(v)$.*

Proof. Let $w \in BSB(\mathbf{t}_3)$ verifying $|w| \geq 7$. Then, w admits a unique decomposition in the form $\delta_1\mu_3(v)\delta_2$. As w is bispecial biprolongable, then w

begins (rep. ends) with the image of a letter. So, $\delta_1, \delta_2 = \varepsilon$. Therefore, by Proposition 16, we have $w = \mu_3(v)$ and $v \in BSB(\mathbf{t}_3)$. □

Theorem 17. *The set $BSB(\mathbf{t}_3)$ is given by:*

$$BSB(\mathbf{t}_3) = \bigcup_{n \geq 0} \{\mu_3^{n+1}(i)\mu_3^n(i) : i \in \mathcal{A}_3\}.$$

Proof. By Proposition 16, we have just to find the elements of $BSB(\mathbf{t}_3)$ with length at most 7, since the others are obtained by applying successively μ_3 on these elements. These factors are $i(i+1)(i+2)i = \mu_3(i)i$, $i \in \mathcal{A}_3$ □

Lemma 18. ([2]) *Let u be an infinite word generated by a morphism ϕ and w be a factor of u. If w is the only ancester of $\phi(w)$ then $Ret(\phi(w)) = \phi(Ret(w))$.*

Theorem 19. *Every factor of $\mathbf{t}_3$ admits 7, 8 or 9 return words.*

Proof. As $\mathbf{t}_3$ is uniformly recurrent and non eventually periodic, to describe the cardinality and the structure of $Ret(w)$ for arbitrary factor w it suffices to consider bispecial factors w (see Section 3.1 in [2]). So, by Proposition 14, Proposition 16 and Lemma 18, it is sufficient to consider the initial elements i, ij of $BST(\mathbf{t}_3)$ and $\mu_3(i)i$ of $BSB(\mathbf{t}_3)$, where $i \in \mathcal{A}_3$ and $j = E(i)$. By Proposition 2 we can restrict to the cases 0, 01 in $BST(\mathbf{t}_3)$ and $\mu_3(0)0$ in $BSB(\mathbf{t}_3)$.

- Determine $Ret(0)$. As 0 occurs in $\mu_3(i)$, $i \in \mathcal{A}_3$, the return words of 0 are due to $\mu_3(ij)$ for all $i, j \in \mathcal{A}_3$. Thus, we have

$$Ret(0) = \{0, 01, 02, 012, 0112, 0122, 01212\}.$$

- Determine $Ret(01)$. Observe that 01 occurs in $\mu_3(0)$, $\mu_3(2)$ and $\mu_3(11)$. So, the return words of 01 are due to $\mu_3(00)$, $\mu_3(02)$, $\mu_3(22)$, $\mu_3(20)$, $\mu_3(010)$, $\mu_3(011)$, $\mu_3(012)$, $\mu_3(112)$ and $\mu_3(212)$. Thus, the return words of 01 are 01, 012, 0122, 01212, 01202, 011202, 012120 and 0121202.

- Determine $Ret(0120)$. Observe that 0120 occurs in $\mu_3(ii)$, $i \in \mathcal{A}_3$. So, the return words of 0120 are due to $\mu_3(iiwjj)$, where $i, j \in \mathcal{A}_3$ and w is separator. Thus, by Corollary 4, the set of return words of 0120 is:

$$\{\mu_3\left(0\mu_3^2\left(0\right)\right)0^{-1},\ (12)^{-1}\mu_3\left(1\mu_3^2\left(1\right)\right)(01)^{-1},\\ 2^{-1}\mu_3\left(2\mu_3^2\left(2\right)\right)(012)^{-1},\ \mu_3\left(0\mu_3^2\left(02\right)\right)(012)^{-1},\\ (12)^{-1}\mu_3\left(1\mu_3^2\left(10\right)\right)0^{-1},\ 2^{-1}\mu_3\left(2\mu_3^2\left(21\right)\right)(01)^{-1},\\ \mu_3\left(0\mu_3^2\left(00\right)\right)0^{-1},\ (12)^{-1}\mu_3\left(1\mu_3^2\left(11\right)\right)(01)^{-1},\\ \text{and } 2^{-1}\mu_3\left(2\mu_3^2\left(22\right)\right)(012)^{-1}\}.$$

□

References

[1] J.P. Allouche, J. Shallit, *Automatique Sequences: Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, 2003.

[2] Ľ. Balkovǎ, E. Pelantovǎ, W. Steiner, Return words in fixed points of substitutions, *Monatsh. Math.*, **155**, No 3-4 (2008), 251-263.

[3] Ľ. Balkovǎ, Factor frequencies in generalized Thue-Morse words, *Kybernetika*, **48**, No 3 (2012), 371-385.

[4] V. Berthé, M. Rigo, Eds, *Combinatorics, Automata and Number Theory*, Ser. Encyclopedia of Mathematics and its Applications 135, Cambridge University Press, 2010.

[5] P. Borwein, C. Ingalls, The Prouhet-Tarry-Escott problem revisited, *Ensign. Math.*, **40**, No 994 (1994), 3-27.

[6] J. Cassaigne, I. Kaboré, Abelian complexity and frequencies of letters in infinite words, *Int. J. Found. Comput. Sci.*, **27** (2016), 631-649.

[7] J. Cassaigne, I. Kaboré, A word without uniforme frequency, *Preprint.*

[8] J. Cassaigne, G. Richomme, K. Saari, L.Q. Zamboni, Avoiding Abelian powers in binary words with bounded Abelian Complexity, *Int. J. Found. Comput. Sci.*, **22**, No 4 (2011), 905-920.

[9] F. Durand, A characterisation of substitutive sequences using return words, *Discrete Math.*, **179** (1998), 89-101.

[10] S. Ferenczi, Ch. Holton, L.Q. Zamboni, Structure of three interval exchange transformations, II. A combinatorial description of the trajectories, *J. Anal. Math.*, **89** (2003), 239-276.

[11] S. Ferenczi, C. Maudit, A. Nogueira, Substitution dynamical systems: algebraic characterization of eigenvalues, *Ann. Sci. Norm. Sup.*, **29** (1996), 519-533.

[12] A. Frid, On the frequency of factors in a DOL word, *J. Autom. Lang. Comb.*, **33**, No 1 (1998), 29-41.

[13] J. Justin, L. Vuillon, Return words in Sturmian and episturmian words, *Theor. Inform. Appl.*, **34** (2000), 343-356.

[14] I. Kaboré, B. Kientéga, Abelian complexity of the Thue-Morse word over a ternary alphabet, In: *Combinatorics on Words (WORDS 2017)*, S. Brlek et al. (Eds), Ser. *LNCS*, **10432** (2017), 132-143.

[15] M. Morse, D. A. Hedlund, Symbolic dynamics II. Sturmian trajectories, *Amer. J. Math.*, **62** (1940), 1-42.

[16] J. Peltomäki, Privileged factors in the Thue-Morse Word- A comparaison of privileged words and Palindromes, *Discrete Applied Mathematics*, **193** (2015), 187-199.

[17] M. Peter, The asymptotic distribution of elements in automatic sequences, *Theoret. Comput. Sci.*, **301** (2003), 285-312.

[18] E. Prouhet, Mémoire sur quelques relations entre les puissances des nombres, *C. R. Acad. Sci. Paris*, **33** (1851), 225.

[19] M. Queffelec, *Substitution dynamical systems- Spectral analysis*, In: Lecture Notes in Math., **1294**, 1987.

[20] G. Richomme, K. Saari, L.Q. Zamboni, Abelian Complexity of Minimal Subshifts. *J. Lond. Math. Soc.*, **83**, No 1 (2011), 79-94.

[21] K. Saari, On the frequency of letters in morphic sequences, *LNCS*, **3967** (2006), 334-345.

[22] A. Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. Mat. Nat. Kl.*, **7** (1906), 1-22.

[23] A. Thue, Über die gegenseilige lage gleicher Teile gewisser Zeichenreihen, *Norske vid. selsk. Skr. Mat. Nat. Kl.*, **1** (1912), 139-158.

[24] L. Vuillon, A characterization of sturmian words by return words, *Eur. J. Comb.*, **22** (2001), 263-275.