View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Hope's Institutional Research Archive

Algebraic Properties of Parikh Matrices of Words under an Extension of Thue Morphism

K.G. Subramanian[1,*], Sastha Sriram[2], A.S. Prasanna Venkatesan[3] and Atulya K. Nagar[1]

[1]Department of Mathematics and Computer Science, Faculty of Science, Liverpool Hope University Hope Park, Liverpool, L16 9JD U.K; * Honorary Visiting Professor
[2]Department of Science and Humanities (Mathematics), Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha Nagar, Thandalam, 602 105 Chennai, India
[3]Department of Mathematics, Government Arts College, Dharmapuri, Tamil Nadu, 636 705 India

E-mail: kgsmani1948@gmail.com, sriram.discrete@gmail.com, pacchu79@gmail.com, nagara@hope.ac.uk

Abstract. The Parikh matrix of a word w over an alphabet $\{a_1, \cdots, a_k\}$ with an ordering $a_1 < a_2 < \cdots a_k$, gives the number of occurrences of each factor of the word $a_1 \cdots a_k$ as a (scattered) subword of the word w. Two words u, v are said to be $M-$equivalent, if the Parikh matrices of u and v are the same. On the other hand properties of image words under different morphisms have been studied in the context of subwords and Parikh matrices. Here an extension to three letters, introduced by Séébold (2003), of the well-known Thue morphism on two letters, is considered and properties of Parikh matrices of morphic images of words are investigated. The significance of the contribution is that various classes of binary words are obtained whose images are $M-$equivalent under this extended morphism.

1. Introduction

In the exciting topic of *Combinatorics on words* [7, 9, 10], there has been a great interest and intensive research in the recent past on problems and properties of words although as early as in the beginning years of 20th century, pioneering work on various combinatorial problems related to words has been done by Axel Thue [6, 7]. Parikh matrix of a word, introduced by Mateescu et al. [13] is a comparatively recent research area in this field and is an extension of the classical notion of Parikh vector [14, 15] which tells the number of occurrences of symbols of an alphabet, in a given word. The Parikh matrix of a word w over an ordered alphabet $\{a_1, \cdots, a_k\}$ is an upper triangular matrix, with $1's$ on the main diagonal and $0's$ below it but the entries above the main diagonal give information on the number of certain subwords (or also called scattered subwords) in w. In fact the Parikh matrix gives the number of occurrences of each factor of the word $a_1 \cdots a_k$ as a subword of the word w and has the Parikh vector in the second diagonal above the main diagonal.

There has been a number of studies establishing various properties of words based on the Parikh matrix. In fact $M-$equivalence of words which requires the Parikh matrices of the words to be the same, is an intensively investigated property. Another study is on obtaining properties of Parikh matrices of words under certain mappings, called morphisms on words. Atanasiu [2] considers Istrail morphisms and investigates amiability or ambiguity of image words, based on $M-$equivalence and points out the use of Istrail morphism in the problem of "disambiguation" of binary amiable words. Teh [19] provides a general investigation of this problem and proves that no morphism can completely separate $M-$equivalent words. In [8, 18] certain properties of Parikh matrices of morphic image words on two or three letters, under extensions [16, 11] of the well-known [10] Thue morphism and the Fibonacci morphism, are obtained. Here we consider the extension to three letters, introduced by Séébold [16], of the well-known Thue morphism on two letters. We obtain several classes of binary words whose images are $M-$equivalent under this extended morphism, which we call as the Séébold morphism.

2. Basic Definitions and Results

Basic notions and results [10, 13] needed for the study in the subsequent sections are first recalled.

An alphabet is a finite set of symbols. An ordered alphabet is an alphabet endowed with an ordering, denoted by $<$, on its elements. For instance, the alphabet $\{a, b, c\}$ with an ordering $a < b < c$, is an ordered alphabet, which is written as $\{a < b < c\}$. We are mainly concerned with alphabets with two or three symbols. A word is a finite sequence of symbols belonging to an alphabet. For example the word *aabbabb* is over the binary alphabet $\{a, b\}$ while the word *acbabcab* is a word over the ternary alphabet $\{a, b, c\}$. The set of all words over an alphabet Σ is denoted by Σ^*. A subword (also referred to as scattered subword) v of a given word w is a subsequence of w. The number of subwords v in a given word w is denoted as $|w|_v$. For example, the number of subwords aab in a given word $aabbabb$ over $\{a < b\}$, is $|aabbabb|_{aab} = 8$.

The Parikh vector [14, 15] of a word w over an alphabet $\Sigma = \{a < b < c\}$ is given by $(n_a(w), n_b(w), n_c(w))$ where $n_x(w)$ is the number of occurrences of the symbol x in the word w. For example, the Parikh vector of the word *acbabcab* over the alphabet Σ is $(3, 3, 2)$. The notion of *Parikh matrix* of a word w over an ordered alphabet Σ, introduced by Mateescu et al. [13], is an extension of the notion of Parikh vector of w. We recall the notion of Parikh matrix of a word restricting ourselves to only a binary or a ternary alphabet. For more details and a formal definition of the Parikh matrix of a word over any ordered alphabet, we refer to [13]. If $\Sigma_2 = \{a < b\}$ and $\Sigma_3 = \{a < b < c\}$ then the Parikh matrix $M(u)$ of a word u over Σ_2 and the Parikh matrix of a word v over Σ_3 are given by

$$M(u) = \begin{pmatrix} 1 & |u|_a & |u|_{ab} \\ 0 & 1 & |u|_b \\ 0 & 0 & 1 \end{pmatrix}, \quad M(v) = \begin{pmatrix} 1 & |v|_a & |v|_{ab} & |v|_{abc} \\ 0 & 1 & |v|_b & |v|_{bc} \\ 0 & 0 & 1 & |v|_c \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

For example, the word $u = aabbabb$ over $\{a < b\}$ has three $a's$, four $b's$ and ten subword ab. The word $v = acbabcab$ over $\{a < b < c\}$ has three $a's$, three $b's$, two $c's$ and six subword ab, two subword bc and three subword abc. Thus the Parikh matrices are

$$M(u) = \begin{pmatrix} 1 & 3 & 10 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}, \quad M(v) = \begin{pmatrix} 1 & 3 & 6 & 3 \\ 0 & 1 & 3 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In fact instead of recalling the formal definition [13], we mention here the ingenious technique of deriving the Parikh matrix described in [13]. For a binary ordered alphabet $\Sigma_2 = \{a < b\}$, with each of a and b, a 3×3 triangular matrix is associated as follows:

$$a \mapsto M(a) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; b \mapsto M(b) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

For a word over Σ_2, for example, $u = aabbabb$, the Parikh matrix $M(u)$ is the matrix product

$$M(u) = M(a)M(a)M(b)M(b)M(a)M(b)M(b) = \begin{pmatrix} 1 & 3 & 10 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}.$$

Like wise, for a ternary ordered alphabet $\Sigma_3 = \{a < b < c\}$, with each of a, b and c, a 4×4 triangular matrix is associated as follows:

$$a \mapsto M(a) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, b \mapsto M(b) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; c \mapsto M(c) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Thus for $v = acbabcab$ over $\Sigma_3 = \{a < b < c\}$,

$$M(v) = M(a)M(c)M(b)M(a)M(b)M(c)M(a)M(b) = \begin{pmatrix} 1 & 3 & 6 & 3 \\ 0 & 1 & 3 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Two words u, v are said to be $M-$equivalent (also called $M-$ambiguous or amiable) [13], denoted $u \equiv_M v$, if they have the same Parikh matrices i.e. $M(u) = M(v)$. We note that the words $u = abaabbb, v = aabbabb$ over $\{a < b\}$ have the same Parikh matrix $M(u) = M(v) = \begin{pmatrix} 1 & 3 & 10 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}$ and hence are $M-$equivalent.

We next recall a weak-ratio property considered in [17]. Two words u, v over $\{a < b < c\}$ satisfy a weak-ratio property, written $u \sim_{wr} v$, if there is a constant $k > 0$, such that $|u|_a = k|v|_a$, $|u|_b = k|v|_b$ and $|u|_c = k|v|_c$. If the alphabet is $\{a, b\}$, $u \sim_{wr} v$, if there is a constant $k > 0$, such that $|u|_a = k|v|_a$ and $|u|_b = k|v|_b$. We next recall the notion of a morphism [10] on words as a mapping $\varphi : \Sigma^* \to \Gamma^*$, where Σ and Γ are two alphabets, such that $\varphi(uv) = \varphi(u)\varphi(v)$, for words u, v over Σ.

Lemma 1 *[3] Let Σ, Γ be two finite ordered alphabets. Let α, β be two words over Σ in weak ratio property and $\varphi : \Sigma^* \longrightarrow \Gamma^*$ be a morphism. Then $\varphi(\alpha) \sim_{wr} \varphi(\beta)$.*

Thue morphism [10] (also called Thue-Morse morphism), is a well-known morphism extensively investigated in different studies on combinatorics on words (see, for example, [16]). It is a mapping μ on Σ^*, where $\Sigma = \{a, b\}$ and is given by $\mu(a) = ab, \mu(b) = ba$. Séébold [16] introduced a natural generalization of Thue morphism on n symbols, which we refer to as Séébold morphism. We recall the definition of these Séébold morphism restricting the alphabet to three letters. Unless stated otherwise, we denote the ordered alphabets $\{a < b\}$ and $\{a < b < c\}$ respectively by Σ_2 and Σ_3.

Definition 1 *The Séébold morphism $\sigma : \Sigma_3^* \mapsto \Sigma_3^*$ is defined by $\sigma(a) = ab, \sigma(b) = bc, \sigma(c) = ca$.*

We now recall needed known results.

Lemma 2 *[3] Let* Γ_1, Γ_2 *be two ordered alphabets and* $f : \Gamma_1^* \longrightarrow \Gamma_2^*$ *be a morphism. Then, for all* $w \in \Gamma_1^*$, *and for all* $a \in \Gamma_2$, *we have*

$$|f(w)|_a = \sum_{r \in \Gamma_1} |w|_r \cdot |f(r)|_a.$$

Lemma 3 *[3] Let* $f : \Sigma_3^* \longrightarrow \Sigma_3^*$ *be a morphism. For* $x, y \in \Sigma_3$, $w \in \Sigma_2^+$, *we have*

$$|f(w)|_{xy} = \sum_{r \in \Sigma_2} |w|_r |f(r)|_{xy} + \sum_{r,t \in \Sigma_2} |w|_{rt} |f(r)|_x |f(t)|_y.$$

Lemma 4 *[1] Let* α, β *be any two* $M-$*equivalent words over* Σ_2. *Then*

$$M(\varphi(\alpha)) - M(\varphi(\beta)) = \begin{pmatrix} 0 & 0 & 0 & n \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

where $\varphi : \Sigma_2^* \to \Sigma_3^*$ *is a morphism and* $n = |\varphi(\alpha)|_{abc} - |\varphi(\beta)|_{abc}$ *is an integer.*

Lemma 5 *[12] Let* u, v *be two words over* Σ_3^* *satisfying*

i) the weak-ratio property, namely, $u \sim_{wr} v$ *and*

ii) $|u|_a |u|_{bc} = |u|_{ab} |u|_c$ *and* $|v|_a |v|_{bc} = |v|_{ab} |v|_c$.

Then the words uv, vu *are* $M-$*equivalent i.e.* $M(uv) = M(vu)$.

Atanasiu and Teh [4] considered a restricted shuffle operator $SShuf$ on two binary words over Σ_2 as follows: If $u, v \in \Sigma_2^+$ such that $u = a_1 a_2 \cdots a_n$, $v = b_1 b_2 \cdots b_n$ for $n \geq 1$, then $SShuf(u, v) = a_1 b_1 a_2 b_2 \cdots a_n b_n$ where $a_i, b_i \in \Sigma_2$, $(1 \leq i \leq n)$. Properties on $M-$equivalence in the context of such binary words $SShuf(u, v)$ are obtained in [4]. We recall two of these properties in the following lemmas.

Lemma 6 *[4] If* v_1, v_2, w_1, w_2 *are binary words over* Σ_2, *such that* $|v_1| = |v_2|$ *and* $v_1 \equiv_M w_1$, $v_2 \equiv_M w_2$, *then* $SShuf(v_1, v_2) \equiv_M SShuf(w_1, w_2)$.

Lemma 7 *[4] Let* v, w *be two binary words over* Σ_2, *such that* $|v| = |w|$. *Then* $SShuf(v, w) \equiv_M SShuf(w, v)$ *if and only if* $\psi(v) = \psi(w)$ *where* $\psi(x)$ *is the Parikh vector of the word* x.

Remark 1 *Results corresponding to Lemma 6 and Lemma 7 are not known. In fact it is pointed out in [4] in Page 7, that these results cannot be immediately extended to the ternary alphabet.*

An extension of the operator $SShuff$ [4], denoted by $S_{m,n}$, is introduced in [5]. Informally expressed, this shuffle operator $S_{m,n}$, $m, n \geq 1$, in operating on a pair of binary words (u, v) forms a binary word $S_{m,n}(u, v)$ obtained by concatenating alternately consecutive factors of u and v, with each factor in u of length m and in v of length n respectively. Here we recall this extension operator but with $m = n = 2$.

Definition 2 *Let* u, v *be two binary words over* Σ_2 *such that* $u = u_1 u_2 \cdots u_l$ *and* $v = v_1 v_2 \cdots v_l$, *where* $u_i, v_i \in \Sigma_2^*$, $|u_i| = |v_i| = 2$, *for* $1 \leq i \leq l$. *The shuffle operator, denoted as* $S_{2,2}$ *is defined on the pair* (u, v) *as follows:* $S_{2,2}(u, v) = u_1 v_1 u_2 v_2 \cdots u_l v_l$. *In particular, when* $m = n = 1$, *we have* $S_{1,1}(u, v) = SShuf(u, v)$.

3. Main Results

Using the results in Lemma 2 and Lemma 3, we first obtain formulae which give the counts of certain subwords in the image under the Séébold morphism σ of a word over Σ_2. These formulae are obtained in [18] but the proof given here is more precise and is given for completeness.

Theorem 1 *[18] For a word $w \in \Sigma_2^*$,*

(i) $|\sigma(w)|_a = |w|_a$, $|\sigma(w)|_b = |w|_a + |w|_b = |w|$, $|\sigma(w)|_c = |w|_b$

(ii) $|\sigma(w)|_{ab} = \frac{1}{2}|w|_a\,(|w|_a + 1) + |w|_{ab}$

(iii) $|\sigma(w)|_{bc} = \frac{1}{2}|w|_b\,(|w|_b + 1) + |w|_{ab}$

where $\sigma : \Sigma_3^* \mapsto \Sigma_3^*$ is the Séébold morphism.

Proof. Let w be a nonempty word over Σ_2. By Lemma 3, we have $|\sigma(w)|_a = \sum_{r\in\Sigma_2} |w|_r \cdot |\sigma(r)|_a$ $= |w|_a$ as $|\sigma(a)|_a = 1$ and $|\sigma(b)|_a = 0$. Likewise, $|\sigma(w)|_b = |w|_a + |w|_b$ as $|\sigma(a)|_b = |\sigma(b)|_b = 1$ while $|\sigma(w)|_c = |w|_b$ as $|\sigma(a)|_c = 0$ and $|\sigma(b)|_c = 1$. This proves (i).
By Lemma 3,

$$|\sigma(w)|_{ab} = \sum_{r\in\Sigma_2} |w|_r|\sigma(r)|_{ab} + \sum_{r,t\in\Sigma_2} |w|_{rt}|\sigma(r)|_a|\sigma(t)|_b$$

$$= |w|_a + |w|_{aa} + |w|_{ab} = |w|_a + \frac{1}{2}|w|_a(|w|_a - 1) + |w|_{ab} = \frac{1}{2}|w|_a\,(|w|_a + 1) + |w|_{ab}$$

since $|\sigma(a)|_a = |\sigma(a)|_b = 1$ and $|\sigma(b)|_a = 0, |\sigma(b)|_b = 1$. This proves (ii). The proof of (iii) is similar.

We now consider certain special binary words over Σ_2 and show that the Séébold morphism retains $M-$equivalence only when the binary word has an equal number of $a's$ and $b's$. We make use of Lemma 4.

Theorem 2 *Let $\delta \in \Sigma_2^*$ with an equal number of $a's$ and $b's$. Then the word $\sigma(ab\delta ba)$ is $M-$equivalent to the word $\sigma(ba\delta ab)$ where σ is the Séébold morphism.*

Proof. The words $\alpha = ab\delta ba$ and $\beta = ba\delta ab$ are $M-$equivalent with Parikh matrix

$$\begin{pmatrix} 1 & 2+|\delta|_a & 2+|\delta|+|\delta|_{ab} \\ 0 & 1 & 2+|\delta|_b \\ 0 & 0 & 1 \end{pmatrix}.$$

Then by definition, $\sigma(\alpha)$ and $\sigma(\beta)$ are $M-$equivalent if $M(\sigma(\alpha)) = M(\sigma(\beta))$. Using Lemma 4, $\sigma(\alpha)$ and $\sigma(\beta)$ are $M-$equivalent if $|\sigma(\alpha)|_{abc} - |\sigma(\beta)|_{abc} = 0$. Now $\sigma(\alpha) = abbc\sigma(\delta)bcab$ and $\sigma(\beta) = bcab\sigma(\delta)abbc$. By Lemma 1, $|\sigma(\delta)|_a = |\delta|_a$, $|\sigma(\delta)|_c = |\delta|_b$, and by hypothesis $|\delta|_a = |\delta|_b$ so that

$$|\sigma(\alpha)|_{abc} = 5 + |\sigma(\delta)|_{abc} + 2|\sigma(\delta)|_c + |\sigma(\delta)|_{bc} + |\sigma(\delta)|_a + |\sigma(\delta)|_{ab} + |\sigma(\delta)|_b$$

$$= 5 + |\sigma(\delta)|_{abc} + 2|\sigma(\delta)|_a + |\sigma(\delta)|_{bc} + |\sigma(\delta)|_c + |\sigma(\delta)|_{ab} + |\sigma(\delta)|_b = |\sigma(\beta)|_{abc}.$$

This proves the result.

Remark 2 *We note that the Theorem 2 does not hold for any binary word δ. For example, if $\delta = aba$, then $|\delta|_a \neq |\delta|_b$ and $|\sigma(ab\delta ba)|_{abc} = 20$ while $|\sigma(ba\delta ab)|_{abc} = 21$.*

A more general situation on the $M-$equivalence of $\sigma(\alpha)$ and $\sigma(\beta)$ for binary words α, β is given in the following result.

Theorem 3 *Let α, β be two words over Σ_2 satisfying the following condition*

$$(A): |\alpha|_{aab} + |\alpha|_{abb} = |\beta|_{aab} + |\beta|_{abb}.$$

Then $\sigma(\alpha)$ and $\sigma(\beta)$ are $M-$equivalent if and only if α, β are $M-$equivalent, where σ is the Séébold morphism.

Proof. Under an application of σ to the word α, the subwords ab, aab, abb yield the subword abc in $\sigma(\alpha)$. Thus $|\sigma(\alpha)|_{abc} = 2|\alpha|_{ab} + |\alpha|_{aab} + |\alpha|_{abb}$, $|\sigma(\beta)|_{abc} = 2|\beta|_{ab} + |\beta|_{aab} + |\beta|_{abb}$. Assume that the binary words α, β are $M-$equivalent. Then we have $|\alpha|_{ab} = |\beta|_{ab}$. Also, by Lemma 4, it is enough to show that $|\sigma(\alpha)|_{abc} = |\sigma(\beta)|_{abc}$, and this is true by hypothesis. Conversely, suppose $\sigma(\alpha)$ and $\sigma(\beta)$ are $M-$equivalent. Then by Theorem 1, $|\alpha|_a = |\sigma(\alpha)|_a = |\sigma(\beta)|_a = |\beta|_a$. Again by Theorem 1, $|\alpha|_b = |\sigma(\alpha)|_b - |\alpha|_a = |\sigma(\beta)|_b - |\beta|_a = |\beta|_b$. Also, by hypothesis and the equality $|\sigma(\alpha)|_{abc} = |\sigma(\beta)|_{abc}$ we have $|\alpha|_{ab} = |\beta|_{ab}$, thus proving the assertion.

Remark 3 *The hypothesis in Theorem 3 is not vacuous in the sense that there are binary words satisfying the condition (A). The words $\alpha = abaab$ and $\beta = aabba$ are $M-$equivalent words over Σ_2. Also $|\alpha|_{aab} + |\alpha|_{abb} = |\beta|_{aab} + |\beta|_{abb} = 4$. In fact $\sigma(\alpha) = abbcababbc$, and $\sigma(\beta) = ababbcbcab$ have the same Parikh matrix, namely,*

$$\begin{pmatrix} 1 & 3 & 10 & 12 \\ 0 & 1 & 5 & 7 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We now derive a sufficient condition on two binary words u, v so that the words $\sigma(uv)$ and $\sigma(vu)$ are $M-$equivalent. We use Lemma 5 to obtain this condition.

Theorem 4 *Let the words $u, v \in \Sigma_2^*$ be such that $|u|_a = |u|_b$ and $|v|_a = |v|_b$ Then $M(\sigma(uv)) = M(\sigma(vu))$ so that $\sigma(uv), \sigma(vu)$ are $M-$equivalent, where σ is the Séébold morphism.*

Proof. Since $|u|_a = |u|_b$ and $|v|_a = |v|_b$, the words u, v satisfy the weak-ratio property, which implies, by Lemma 1, that $\sigma(u)$ and $\sigma(v)$ also satisfy the weak-ratio property. Also, using Theorem 1 and the hypothesis, we have

$$|\sigma(u)|_a |\sigma(u)|_{bc} = |u|_a \left[\frac{1}{2}|u|_b \left(|u|_b + 1\right) + |u|_{ab}\right]$$

$$= \left[\frac{1}{2}|u|_a \left(|u|_a + 1\right) + |u|_{ab}\right] |u|_b = |\sigma(u)|_{ab} |\sigma(u)|_c.$$

Likewise $|\sigma(v)|_a |\sigma(v)|_{bc} = \sigma(v)|_{ab} |\sigma(v)|_c$. Hence by Lemma 5, $M(\sigma(uv)) = M(\sigma(vu))$.

Remark 4 *The condition in the hypothesis of Theorem 4 is only sufficient and is not necessary. For example, consider $u_1 = babab$, $v_1 = bbaaaabbbba$. Although $|u_1|_a \neq |u_1|_b$ and $|v_1|_a \neq |v_1|_b$, the words $\sigma(u_1 v_1)$ and $\sigma(v_1 u_1)$ have the same Parikh matrix*

$$\begin{pmatrix} 1 & 7 & 59 & 198 \\ 0 & 1 & 16 & 76 \\ 0 & 0 & 1 & 9 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In the special case when two binary words u, v over Σ_2 have the same Parikh vector, then the conditions $|u|_a = |u|_b$ and $|v|_a = |v|_b$ are necessary and sufficient for $M-$equivalence of $\sigma(uv)$ and $\sigma(vu)$ as shown in the following Theorem 5.

Theorem 5 *Let the words $u, v \in \Sigma_2^*$ be such that u and v have the same Parikh vector but not $M-$equivalent so that $|u|_a = |v|_a$, $|u|_b = |v|_b$ and $|u|_{ab} \neq |v|_{ab}$. Then $\sigma(uv), \sigma(vu)$ are $M-$equivalent if and only if $|u|_a = |u|_b$ and $|v|_a = |v|_b$, where σ is the Séébold morphism.*

Proof. In view of Theorm 4, it is enough to prove necessity. Assume that $\sigma(uv), \sigma(vu)$ are $M-$equivalent. Then $M(\sigma(uv)) = M(\sigma(vu))$ so that $|\sigma(uv)|_{abc} = |\sigma(vu)|_{abc}$. This implies that

$$|\sigma(u)|_a|\sigma(v)|_{bc} + |\sigma(u)|_{ab}|\sigma(v)|_c = |\sigma(v)|_a|\sigma(u)|_{bc} + |\sigma(v)|_{ab}|\sigma(u)|_c.$$

Using Theorem 1, we obtain $|u|_a \left[\frac{1}{2}|v|_b\left(|v|_b + 1\right) + |v|_{ab}\right] + |v|_b \left[\frac{1}{2}|u|_a\left(|u|_a + 1\right) + |u|_{ab}\right] = |v|_a \left[\frac{1}{2}|u|_b\left(|u|_b + 1\right) + |u|_{ab}\right] + |u|_b \left[\frac{1}{2}|v|_a\left(|v|_a + 1\right) + |v|_{ab}\right]$. Using the hypothesis that $|u|_a = |v|_a$, $|u|_b = |v|_b$, we obtain

$$|u|_a|v|_{ab} + |u|_b|u|_{ab} = |u|_a|u|_{ab} + |u|_b|v|_{ab}$$

which implies that

$$|u|_a\left(|v|_{ab} - |u|_{ab}\right) = |u|_b\left(|v|_{ab} - |u|_{ab}\right).$$

By hypothesis, $|u|_{ab} \neq |v|_{ab}$ and so $|u|_a = |u|_b$. Likewise $|v|_a = |v|_b$.

We now consider the shuffle operator $S_{2,2}$ which is a special case of the extended shuffle operator $S_{m,n}$ introduced in [5].

Theorem 6 *If v_1, v_2, w_1, w_2 are binary words over Σ_2, such that $|v_1| = |v_2|$ and $v_1 \equiv_M w_1$, $v_2 \equiv_M w_2$, then $S_{2,2}(\sigma(v_1), \sigma(v_2)) \equiv_M S_{2,2}(\sigma(w_1), \sigma(w_2))$, if*

$$\Sigma_{x\in\{aab,abb\}}|SShuf(v_1, v_2)|_x = \Sigma_{x\in\{aab,abb\}}|SShuf(w_1, w_2)|_x.$$

Proof. It can be seen that

$$S_{2,2}(\sigma(v_1), \sigma(v_2)) = \sigma(SShuf(v_1, v_2)),\ S_{2,2}(\sigma(w_1), \sigma(w_2)) = \sigma(SShuf(w_1, w_2)).$$

By Lemma 6, we have $SShuf(v_1, v_2) \equiv_M SShuf(w_1, w_2)$. Since by hypothesis

$$|SShuf(v_1, v_2)|_{aab} + |SShuf(v_1, v_2)|_{abb} = |SShuf(w_1, w_2)|_{aab} + |SShuf(w_1, w_2)|_{abb},$$

it follows from Theorem 3 that
$\sigma(SShuf(v_1, v_2)) \equiv_M \sigma(SShuf(w_1, w_2))$ which proves $S_{2,2}(\sigma(v_1), \sigma(v_2)) \equiv_M S_{2,2}(\sigma(w_1), \sigma(w_2))$.

We illustrate Theorem 6 by considering $v_1 = abaabb$, $w_1 = aabbab$, $v_2 = bbabba$, and $w_2 = bbbaab$ so that $\sigma(v_1) = abbcababbcbc$, $\sigma(v_2) = bcbcabbcbcab$ and so $S_{2,2}(\sigma(v_1), \sigma(v_2))$ $= abbcbcbcabababbcbcbcbcab$. Likewise, $S_{2,2}(\sigma(w_1), \sigma(w_2))$
$= abbcabbcbcbcbcabababbcbc$. The words $S_{2,2}(\sigma(v_1), \sigma(v_2))$ and $S_{2,2}(\sigma(w_1), \sigma(w_2))$ have the same Parikh matrix

$$\begin{pmatrix} 1 & 5 & 34 & 101 \\ 0 & 1 & 12 & 47 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Theorem 7 *Let v, w be two binary words over Σ_2, such that $|v| = |w|$. Then $S_{2,2}(v, w) \equiv_M S_{2,2}(w, v)$ if $\psi(v) = \psi(w)$ where $\psi(x)$ is the Parikh vector of the word x and if $\Sigma_{x\in\{aab,abb\}}|SShuf(v, w)|_x = \Sigma_{x\in\{aab,abb\}}|SShuf(wv,)|_x$.*

Proof. As in the proof of Theorem 6, we have

$$S_{2,2}(\sigma(v), \sigma(w)) = \sigma(SShuf(v, w)),\ S_{2,2}(\sigma(w), \sigma(v)) = \sigma(SShuf(w, v)).$$

By Lemma 7,$SShuf(v, w) \equiv_M SShuf(w, v)$. Also, by hypothesis

$$|SShuf(v, w)|_{aab} + |SShuf(v, w)|_{abb} = |SShuf(w, v)|_{aab} + |SShuf(w, v)|_{abb}.$$

It follows from Theorem 3 that $\sigma(SShuf(v, w)) \equiv_M \sigma(SShuf(w, v))$ which proves the result.

4. Conclusion

We have obtained properties related to subwords and Parikh matrices of the image words under an extension of Thue morphism considered by Séébold [16]. In particular several classes of binary words are obtained, whose images are $M-$equivalent under this extended morphism, which we call as Séébold morphism. Investigating properties of Parikh matrices of special words under specific morphisms as considered here, is of significance in bringing out the capability of a specific morphism in preserving or not preserving $M-$equivalence of words. So it will be of interest to construct other kinds of $M-$ equivalent binary words whose images under special morphisms on three or more letters are $M-$equivalent.

Acknowledgments

The authors thank the reviewers for their useful comments. The author K.G. Subramanian is grateful to Liverpool Hope University and in particular to Prof. Atulya K. Nagar for the support extended to him.

References

[1] Atanasiu, A. : Morphisms on Amiable Words. Annals of Bucharest University. LIX, 99–111 (2010).

[2] Atanasiu, A, : Parikh Matrices, Amiability and Istrail Morphism. Int. J. Found. Comput. Sci. 21,1021–1033 (2010).

[3] Atanasiu, A., Mahalingam, K., Subramanian, K.G. : Morphisms and Weak-Ratio Property. In : Eds. Paun, Gh. et al. Discrete Mathematics and Computer Science, The Publishing House of the Romanian Academy, Romania, pp 13–21 (2014).

[4] Atanasiu, A., Teh, W.C : A New Operator over Parikh Languages. Int. J. Found. Comput. Sci. 27(6): 757–770 (2016).

[5] Bera, S., Mahalingam, K., Subramanian, K.G. : Properties of Parikh Matrices of Binary Words Obtained by an Extension of a Restricted Shuffle Operator. (To appear in Int. J. Found. Comput. Sci.).

[6] Berstel, J. : Axel Thue's papers on repetitions in words: a translation. Publications du LaCIM 20, Université du Québec à Montréal, 1995.

[7] Berstel, J. and Perrin, D.: The origins of combinatorics on words, European J. Combinatorics. 28(3), 996–1022 (2007).

[8] Isawasan, P., Venkat, I., Subramanian, K.G. and Sarmin, N.H. : Morphic Images of Binary Words and Parikh Matrices. In: AIP Conference Proceedings 1605, 708–713 (2014).

[9] Karhumaki, J. : Combinatorics on words: A new challenging topic. TUCS Report No. 645 (2004).

[10] Lothaire, M. : Combinatorics on Words. Cambridge Mathematical Library, Cambridge university Press, 1997.

[11] Lothaire, M. : Applied Combinatorics on Words. Cambridge university Press, 2005. 1997.

[12] Mahalingam, K. and Subramanian, K.G. : Product of Parikh Matrices and Commutativity, Int. J. Found. Comput. Sci. 23, 207–223 (2012).

[13] Mateescu, A., Salomaa, A., Salomaa, K. and Yu, S. : A Sharpening of the Parikh Mapping. RAIRO-Theor. Inform. and Appl. 35(6),551–564 (2001).

[14] Rozenberg, G. and Salomaa, A. : Handbook of Formal Languages, Springer-Verlag New York, Inc., Vols. 1–3, 1997.

[15] Salomaa, A. : Formal Languages, Academic Press, Reading, MA, 1973.

[16] Séébold, P. : On some generalizations of the Thue-Morse morphism. Theor. Comput. Sci., 292(1), 283–298 (2003).

[17] Subramanian, K.G., Huey, A.M. and Nagar, A.K. : On Parikh Matrices. Int. J. Found. Comput. Sci. 20(2), 211–219 (2009).

[18] Subramanian, K.G., Isawasan, P., Venkat, I. : Parikh matrices of words under a generalized Thue morphism. In: AIP Conference Proceedings 1558, 2516–2519 (2013).

[19] Teh, W.C. : Separability of $M-$Equivalent Words by Morphisms. Int. J. Found. Comput. Sci. 27(1), 39–52 (2016).