

## GENERALIZATIONS OF PARIKH MAPPINGS

ANTON ČERNÝ<sup>1</sup>

**Abstract.** Parikh matrices have become a useful tool for investigation of subword structure of words. Several generalizations of this concept have been considered. Based on the concept of formal power series, we describe a general framework covering most of these generalizations. In addition, we provide a new characterization of binary amiable words – words having a common Parikh matrix.

**Mathematics Subject Classification.** 68R15.

### INTRODUCTION

Parikh matrices [15–17] attracted remarkable attention in recent years. They are a generalization of the notorious Parikh mapping [18], which assigns to a given word a vector, expressing the number occurrences of each symbol in the word. Parikh matrices proved to be a powerful tool for investigation of subword occurrences in finite words [14,20–22]. Further matrix generalizations of the Parikh mapping occurred in literature [8,10,11,23], however, some of them lacking the full strength of the classical matrix algebra. Typically, they are able to express the number of occurrences of some specific, very limited, set of subwords in a given word. In this paper we propose a general framework, based on formal power series, capable to express a wide range of Parikh mapping generalizations. It provides an algebraic tool for investigation, capable to express the number of occurrences of subwords from any factorial language.

After providing the basic definitions in Sections 1 and 2, we describe our framework in Section 3 and we document in examples, how several particular generalizations of the Parikh mapping can be expressed within the framework. Theorem 3.5 is a generalization of the construction of the inverse Parikh matrix and inverse

---

*Keywords and phrases.* Parikh mapping, Parikh matrix, formal power series, Prouhet-Tarry-Escott problem, subword, amiable words.

<sup>1</sup> Department of Information Science, Kuwait University, P.O. Box 5969 Safat 13060, Kuwait; [anton.cerny@ku.edu.kw](mailto:anton.cerny@ku.edu.kw)

generalized Parikh matrix, while the simplicity of its proof illustrates the power of our approach.

One of the widely investigated question in the literature is the problem of ambiguity of expressing a word by its Parikh matrix [2,3,12,15]. We generalize this question in Section 4 and provide a new characterization of ambiguity for binary Parikh matrices.

## 1. BASIC NOTIONS

We will use several terms and facts of abstract algebra. We will recall them shortly, a more detailed explanation can be found, *e.g.*, in [6] or at [24].

Throughout the text, if there is no danger of confusion, we will identify the singleton set  $S = \{a\}$  with the element  $a$ . A binary operation  $\circ : A \times A \rightarrow A$  on a set  $A$  is *commutative*, if, for every  $a, b \in A$ ,  $a \circ b = b \circ a$ . It is *associative* if, for every  $a, b, c \in A$ ,  $(a \circ b) \circ c = a \circ (b \circ c)$ . If  $\equiv$  is an equivalence relation on a set  $A$ , the *factor set* of  $A$  with respect to  $\equiv$  is the partition  $A/\equiv$  of  $A$  corresponding to  $\equiv$ . If  $f : A \rightarrow B$  is a mapping, the equivalence relation  $\equiv_f$  on  $A$  defined, for  $a, b \in A$ , as  $a \equiv_f b$  iff  $f(a) = f(b)$  is called the *kernel equivalence relation* of  $f$ .

A (*commutative*, if  $\circ$  is commutative) *semigroup* is a pair  $(A, \circ)$  where  $A$  is a set and  $\circ$  is an associative binary operation on  $A$ . An *identity element* of  $\circ$  is an element  $e \in A$  satisfying  $e \circ a = a \circ e = a$  for each  $a \in A$ . A *monoid* is a semigroup possessing an identity element. In a monoid, an *inverse element* of  $a \in A$  is an element  $a^{-1}$  satisfying  $a^{-1} \circ a = a \circ a^{-1} = e$ . A *group* is a monoid where each element has an inverse element. A *semigroup morphism* is a mapping  $h : A \rightarrow B$ , for semigroups  $(A, \circ_A)$ ,  $(B, \circ_B)$  satisfying  $h(a \circ_A b) = h(a) \circ_B h(b)$  for every  $a, b \in A$ . A *left* (resp. *right*) *congruence* on a semigroup  $(A, \circ)$  is an equivalence relation  $\sim$  on  $A$  such that, for every  $a, b, c \in R$ ,  $a \sim b$  implies  $c \circ a \sim c \circ b$ , (resp.  $a \circ c \sim b \circ c$ ).

A *ring* is a quintuple  $(R, +, \cdot, 0, 1)$  where  $(R, +)$  is a commutative group with the identity element 0,  $(R, \cdot)$  is a monoid with the identity element 1 and  $\cdot$  distributes over  $+$ , *i.e.*,  $a \cdot (b + c) = a \cdot b + a \cdot c$  and  $(b + c) \cdot a = b \cdot a + c \cdot a$  for every  $a, b, c \in R$ . For short, we will refer to the ring  $(R, +, \cdot, 0, 1)$  as  $R$  if the operations and identity elements are understood from the context<sup>1</sup>. A ring morphism is a mapping between two rings being a semigroup morphism with respect to both the additive and the multiplicative operations. The ring operations can be extended to subsets of  $R$  as follows. For  $B, C \subseteq R$  and  $\circ \in \{+, \cdot\}$  we define  $B \circ C = \{b \circ c \mid b \in B \text{ and } c \in C\}$ . A *congruence* on  $R$  is an equivalence relation  $\sim$  on  $R$  such that, for every  $a, b, c \in R$ ,  $a \sim b$  implies  $a + c \sim b + c$ ,  $c \cdot a \sim c \cdot b$ , and  $a \cdot c \sim b \cdot c$ . An *ideal* of a ring  $R$  is a subset  $I \subseteq R$  satisfying  $I + I \subseteq I$ ,  $R \cdot I \subseteq I$ , and  $I \cdot R \subseteq I$ . If  $I$  is an ideal in  $R$ , then the set of sets  $R/I = \{a + I \mid a \in R\}$  together with  $+$ ,  $\cdot$  and the identity elements  $0 + I = I$  and  $1 + I$  is a ring.  $R/I$  can be equivalently

<sup>1</sup>In accordance with the common practice, we use the symbols  $+$  and  $\cdot$  to denote ring operations of various different rings; the actual meaning is to be recognized from the context. We introduce different notation only in the case when application of the operations in different rings containing some common elements may result in different values.

described as  $R/\equiv$  where  $\equiv$  is the congruence relation defined, for  $a, b \in R$ , as  $a \equiv b$  iff  $a - b \in I$  (here  $a - b$  stands for  $a + (-b)$  where  $-b$  is the additive inverse of  $b$ ).

The terms and results on combinatorics on words and formal power series are taken from [13] (though we sometimes use a slightly different notation), where more details can be found. Assume a fixed alphabet  $\Sigma = \{s_1, s_2, \dots, s_k\}$  of size  $|\Sigma| = k \geq 1$ . Throughout the text, the notation  $\Sigma$ ,  $s_i$ , and  $k$  will be used exclusively with this meaning. The set of all words, the set of all words of length  $p \geq 0$ , and the set of all words of length at most  $p$ , over  $\Sigma$ , will be denoted as  $\Sigma^*$ ,  $\Sigma^p$ , and  $\Sigma_{\leq p}$ , respectively. The empty word will be denoted as  $\lambda$ , the size of a finite language  $L$  as  $|L|$ , and the mirror image of a word  $\alpha$  (resp. of a language  $L$ ) as  $\alpha^R$  (resp. as  $L^R$ ). If not stated otherwise, all words and languages are over  $\Sigma$ . The structure  $(\Sigma, \cdot)$ , where  $\cdot$  denotes the concatenation operation, is a (generally noncommutative) monoid with the identity element  $\lambda$ . If a word  $\alpha$  can be expressed as  $\alpha = tuv$  then  $t, u, v$  are called *prefix*, *factor*, *suffix* of  $\alpha$ , respectively (any of them may be empty).

The *symbol decomposition* of a word  $\alpha$  is the concatenation  $\alpha = a_1a_2 \dots a_n$ , where  $a_i \in \Sigma$ ,  $i = 1, 2, \dots, n$ . Then  $[n] = \{1, 2, \dots, n\}$  is the set of *positions* in  $\alpha$  (positions are numbered starting from 1). The ordered set  $\iota = \{i_1 < \dots < i_m\} \subseteq [n]$ ,  $m \geq 0$ , is called *subword occurrence* in  $\alpha$ . The (*scattered*) *subword* occurring at  $\iota$  in  $\alpha$  is the word  $a_{i_1}a_{i_2} \dots a_{i_m}$ . We denote by  $|\alpha|_u$  the number of occurrences of the subword  $u$  in  $\alpha$ . For example, the subword occurring at  $\{2, 3, 6, 7\}$  in *babbaba* is *abba*, the set of all occurrences of *abba* in *babbaba* is  $\{\{2, 3, 4, 5\}, \{2, 3, 4, 7\}, \{2, 3, 6, 7\}, \{2, 4, 6, 7\}\}$ , and  $|\textit{babbaba}|_{\textit{abba}} = 4$ . The only occurrence of  $\lambda$  in a word  $\alpha$  is  $\emptyset$ , thus  $|\alpha|_{\lambda} = 1$  for every  $\alpha$ . An alternative notation may be used (but will not be used here) – see, *e.g.*, Section 6.3 in [13] – denoting  $|\alpha|_u$  as the generalized combinatorial number  $\binom{\alpha}{u}$ . This notation is based on the fact that, in the case of a unary alphabet  $\Sigma = \{a\}$ ,  $|a^r|_{a^s} = \binom{r}{s}$ , for  $r \geq s \geq 0$ .

Since, for  $a, b \in \Sigma$ ,  $a \neq b$ ,

$$|\alpha|_{ab} + |\alpha|_{ba} = |\alpha|_a |\alpha|_b \tag{1.1}$$

we have

**Proposition 1.1.** *For  $a, b \in \Sigma$ ,  $\alpha_1, \alpha_2 \in \Sigma^*$ , if  $|\alpha_1|_a = |\alpha_2|_a$ ,  $|\alpha_1|_b = |\alpha_2|_b$  and  $|\alpha_1|_{ab} = |\alpha_2|_{ab}$  then  $|\alpha_1|_{ba} = |\alpha_2|_{ba}$ .*

A language  $L$  is *factorial* if it contains all factors of its words. The class of all factorial languages is closed under union and intersection. For a language  $L$ , we denote as  $\mathcal{F}(L)$  the *factorial closure* of  $L$ , *i.e.*, the smallest factorial language containing  $L$ .

A *formal power series* (with integer coefficients) over  $\Sigma$  is a mapping  $\mathbf{x} : \Sigma^* \rightarrow \mathbb{Z}$ , where  $\mathbb{Z}$  is the ring of integers. Following the usual conventions, we denote the value  $\mathbf{x}(\alpha)$  as  $\langle \mathbf{x}, \alpha \rangle$  and express  $\mathbf{x}$  as  $\sum_{v \in \Sigma^*} \langle \mathbf{x}, v \rangle v$ . The *support* of  $\mathbf{x}$  is the set  $\{v \in \Sigma^* \mid \langle \mathbf{x}, v \rangle \neq 0\}$ . The set of all power series over  $\Sigma$ , together with the

sum and product operations defined as  $\langle \mathbf{x} + \mathbf{y}, \alpha \rangle = \langle \mathbf{x}, \alpha \rangle + \langle \mathbf{y}, \alpha \rangle$ ,  $\langle \mathbf{x} \cdot \mathbf{y}, \alpha \rangle = \sum_{uv=\alpha} \langle \mathbf{x}, u \rangle \langle \mathbf{y}, v \rangle$ , respectively, forms a ring denoted as  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$ . The elements of  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$  with finite support are called *polynomials* over  $\Sigma$ . They form the subring  $\mathbb{Z} \langle \Sigma \rangle$  of  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$ . The elements with a singleton support are called *monomials*. We identify a word  $v$  with the monomial  $1v$  and, for  $m \in \mathbb{Z}$ , the monomial  $m\lambda$  with the number  $m$  (thus  $\langle m \cdot \mathbf{x}, v \rangle = m \langle \mathbf{x}, v \rangle$ ). The monomial  $0\lambda$ , being the additive identity of  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$  is thus identified with the number 0, and the monomial  $1\lambda$ , being the multiplicative identity of  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$ , can be written both as the word  $\lambda$  or as the number 1.

## 2. THE RING $\mathbb{Z}_L$

Parikh matrices allow to apply a strong mathematical tool of matrix algebra to investigate the counts of subword occurrences in words. We will consider here, for a language  $L \subseteq \Sigma^*$ , the set  $\mathbb{Z}_L^{(1)}$  consisting of all formal power series from  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$  with support being a subset of  $L$  and with the coefficient of  $\lambda$  being equal to 1. We will show that, in the case when  $L$  is a factorial language,  $\mathbb{Z}_L^{(1)}$  possesses several nice algebraic properties induced by the structure of  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$ . Consequently, in Section 3, we will show that the algebra  $\mathbb{Z}_L^{(1)}$  can be used to deal with subword counts in a similar way as the algebra matrices and we will provide a generalization of the Parikh matrix mapping.

For a language  $L \subseteq \Sigma^*$ , we will consider the set  $\mathbb{Z}_L$  consisting of all series from  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$  with support being a subset of  $L$ . We first show that, in the case when  $L$  is a factorial language, one can equip this set with a ring structure induced by the structure of  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$ .

**Definition 2.1.** Let  $L$  be a language. The *L-projection* of a series  $\mathbf{x} \in \mathbb{Z} \langle \langle \Sigma \rangle \rangle$  is the series  $\pi_L(\mathbf{x}) = \sum_{v \in L} \langle \mathbf{x}, v \rangle v$ . We denote  $X_{\overline{L}} = \{\mathbf{x} \in \mathbb{Z} \langle \langle \Sigma \rangle \rangle \mid \pi_L(\mathbf{x}) = 0\}$ .

**Lemma 2.2.**  $X_{\overline{L}}$  is an ideal in  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$  iff  $L$  is a factorial language.

*Proof.* The assertion is true if  $L = \emptyset$ . Let  $L \neq \emptyset$ . Assume first that  $L$  is a factorial language. Take  $\mathbf{x} \in X_{\overline{L}}$  and  $\mathbf{y} \in \mathbb{Z} \langle \langle \Sigma \rangle \rangle$ . Let  $z \in L$ . Then  $\langle \mathbf{x} \cdot \mathbf{y}, z \rangle = \sum_{u \in \Sigma^* - L, v \in \Sigma^*, uv=z} \langle \mathbf{x}, u \rangle \langle \mathbf{y}, v \rangle = 0$  since no factor of  $z$  is in  $\Sigma^* - L$ . Hence  $\mathbf{x} \cdot \mathbf{y} \in X_{\overline{L}}$ . A similar argument shows that  $\mathbf{y} \cdot \mathbf{x} \in X_{\overline{L}}$ . Assume now that  $X_{\overline{L}}$  is an ideal and there is a word  $z \in L$  containing a factor not being from  $L$ , i.e., there are words  $t, u, v$ , such that  $z = tuv$ , and  $u \notin L$ . Then, clearly,  $u \in X_{\overline{L}}$  (as a monomial), but  $t \cdot u \cdot v \notin X_{\overline{L}}$ , yielding a contradiction.  $\square$

For a factorial language  $L \subseteq \Sigma^*$ , we will consider the quotient ring  $\mathbb{Z}_L = \mathbb{Z} \langle \langle \Sigma \rangle \rangle / X_{\overline{L}}$ . We will denote the operations of addition and multiplication in this ring as  $+_L$  and  $\cdot_L$  and identify the set  $\mathbf{x} + X_{\overline{L}}$  (being element of  $\mathbb{Z}_L$ ) with the element  $\pi_L(\mathbf{x})$  of  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$  (with support being a subset of  $L$ ). In particular, the ring  $\mathbb{Z}_{\Sigma^*}$  can be identified with  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle$ . Let us note that the set  $\mathbb{Z}_L$  can be equivalently characterized as  $\mathbb{Z}_L = \mathbb{Z} \langle \langle \Sigma \rangle \rangle / \equiv_L$  where  $\equiv_L$  is the equivalence kernel of the mapping  $\pi_L$ . The set  $\mathbb{Z} \langle \langle \Sigma \rangle \rangle / \equiv_L$  can be constructed even if  $L$  is

not a factorial language, however, in that case, the set  $\mathbb{Z}_L$  does not possess a ring structure.

**Proposition 2.3.**  $\pi_L : \mathbb{Z}\langle\langle\Sigma\rangle\rangle \rightarrow \mathbb{Z}_L$  is a ring morphism.

**Definition 2.4.** Let  $L_1, L_2$  be two factorial languages. The series  $\mathbf{x}_1 \in \mathbb{Z}_{L_1}, \mathbf{x}_2 \in \mathbb{Z}_{L_2}$  are *compatible* if  $\pi_{L_1 \cap L_2}(\mathbf{x}_1) = \pi_{L_1 \cap L_2}(\mathbf{x}_2)$ . We then denote the series  $\sum_{v \in L_1} \langle \mathbf{x}_1, v \rangle v + \sum_{v \in L_2 - L_1} \langle \mathbf{x}_2, v \rangle v$  from  $\mathbb{Z}_{L_1 \cup L_2}$  as  $\mathbf{x}_1 \sqcup \mathbf{x}_2$ .

An easy observation yields the following propositions.

**Proposition 2.5.** Let  $L_1, L_2, L_3$  be factorial languages and let  $\mathbf{x}_1 \in \mathbb{Z}_{L_1}, \mathbf{x}_2 \in \mathbb{Z}_{L_2}, \mathbf{x}_3 \in \mathbb{Z}_{L_3}$ .

- (1) If  $\mathbf{x}_1, \mathbf{x}_2$  are compatible then  $\mathbf{x}_1 \sqcup \mathbf{x}_2 = \mathbf{x}_2 \sqcup \mathbf{x}_1$ .
- (2) Let  $\mathbf{x}_1, \mathbf{x}_2$  be compatible. Then  $\mathbf{x}_1 \sqcup \mathbf{x}_2$  and  $\mathbf{x}_3$  are compatible iff both  $\mathbf{x}_1, \mathbf{x}_2$  and  $\mathbf{x}_2, \mathbf{x}_3$  are compatible. If it is the case then  $(\mathbf{x}_1 \sqcup \mathbf{x}_2) \sqcup \mathbf{x}_3 = \mathbf{x}_1 \sqcup (\mathbf{x}_2 \sqcup \mathbf{x}_3)$ .
- (3) If  $\mathbf{x}_1 \in \mathbb{Z}_{L_1}, \mathbf{x}_2 \in \mathbb{Z}_{L_2}$  are compatible and  $\mathbf{y} \in \mathbb{Z}_{L_1 \cup L_2}$  then  $\mathbf{x}_1 \cdot_{L_1} \pi_{L_1}(\mathbf{y})$  and  $\mathbf{x}_2 \cdot_{L_2} \pi_{L_2}(\mathbf{y})$  are compatible and  $(\mathbf{x}_1 \cdot_{L_1} \pi_{L_1}(\mathbf{y})) \sqcup (\mathbf{x}_2 \cdot_{L_2} \pi_{L_2}(\mathbf{y})) = (\mathbf{x}_1 \sqcup \mathbf{x}_2) \cdot_{L_1 \cup L_2} \mathbf{y}$ .

**Proposition 2.6.** If  $\mathbf{x} \in \mathbb{Z}_{L_1 \cup L_2}$  for factorial languages  $L_1, L_2$ , then  $\pi_{L_1}(\mathbf{x})$  and  $\pi_{L_2}(\mathbf{x})$  are compatible and  $\pi_{L_1}(\mathbf{x}) \sqcup \pi_{L_2}(\mathbf{x}) = \mathbf{x}$ .

We will now restrict our focus to the set  $\mathbb{Z}_L^{(1)} = \{\mathbf{x} \in \mathbb{Z}_L \mid \langle \mathbf{x}, \lambda \rangle = 1\}$ . This set is closed under multiplication. The semigroup  $(\mathbb{Z}_L^{(1)}, \cdot)$  is a group. This follows from the fact, that each element  $\mathbf{x} \in \mathbb{Z}\langle\langle\Sigma\rangle\rangle$  with  $\langle \mathbf{x}, \lambda \rangle = 1$  has a multiplicative inverse. Indeed, if we assume  $\mathbf{x}$  with these properties, then  $\langle 1 - \mathbf{x}, \lambda \rangle = 0$  and the family of series  $\{(1 - \mathbf{x})^r\}_{r=0}^\infty$  is locally finite, *i.e.*, for each  $w \in \Sigma^*$ , just finitely many of the values  $\langle (1 - \mathbf{x})^r, w \rangle, r = 1, 2, \dots$  are non-zero. Thus the sum  $(1 - \mathbf{x})^* = \sum_{r=0}^\infty (1 - \mathbf{x})^r$  is a well-defined series. The proof of the following proposition can be found in [13], Section 1.4:

**Proposition 2.7.**  $(1 - \mathbf{x})^* \mathbf{x} = \mathbf{x}(1 - \mathbf{x})^* = 1$ .

We will denote, for  $\mathbf{x} \in \mathbb{Z}\langle\langle\Sigma\rangle\rangle$  satisfying  $\langle \mathbf{x}, \lambda \rangle = 1$ , the inverse series of  $\mathbf{x}$  as  $\mathbf{x}^{-1}$ , thus  $\mathbf{x}^{-1} = (1 - \mathbf{x})^*$ . Some basic properties of  $\mathbf{x}^{-1}$  may be summarized as follows.

**Proposition 2.8.** Let  $\mathbf{x} \in \mathbb{Z}\langle\langle\Sigma\rangle\rangle$  such that  $\langle \mathbf{x}, \lambda \rangle = 1$ .

- (1) For each  $w \in \Sigma^*$ ,

$$\begin{aligned} \langle \mathbf{x}^{-1}, w \rangle &= - \sum_{uv=w, v \neq \lambda} \langle \mathbf{x}^{-1}, u \rangle \langle \mathbf{x}, v \rangle \quad \text{for } w \neq \lambda \\ \langle \mathbf{x}^{-1}, \lambda \rangle &= 1. \end{aligned}$$

- (2) For each  $w \in \Sigma^*$ ,  $\langle \mathbf{x}^{-1}, w \rangle = \left\langle \sum_{r=0}^{|w|} (1 - \mathbf{x})^r, w \right\rangle$ .

- (3) If  $L$  is a finite factorial language,  $s = \max\{|u| \mid u \in L\}$ ,  $\mathbf{x} \in \mathbb{Z}_L^{(1)}$ , then  $\pi_L(\mathbf{x}^{-1}) = \pi_L(\sum_{r=0}^s (1 - \mathbf{x})^r)$  is the multiplicative inverse of  $\mathbf{x}$  in  $\mathbb{Z}_L^{(1)}$ .

**Lemma 2.9.** *Let  $L$  be a factorial language,  $w \in L, w \neq \lambda$  and  $\mathbf{x} \in \mathbb{Z}_L^{(1)}$ . Then*

$$\langle \mathbf{x}^{-1}, w \rangle = \sum_{u_1 u_2 \dots u_r = w, u_i \neq \lambda, r \geq 1} (-1)^r \langle \mathbf{x}, u_1 \rangle \langle \mathbf{x}, u_2 \rangle \dots \langle \mathbf{x}, u_r \rangle.$$

*Proof.* Induction on  $|w|$ . The assertion is trivially true for  $w = \lambda$  (where the only term in the sum is the product of  $r = 0$  factors equal to 1). Let  $|w| = n \geq 1$  and let the assertion be true for all words of length less than  $n$ . Then, from (1) of Proposition 2.8 we have  $\langle \mathbf{x}^{-1}, w \rangle = -\sum_{uv=w, v \neq \lambda} \langle \mathbf{x}^{-1}, u \rangle \langle \mathbf{x}, v \rangle$ . By applying the inductive hypothesis and to the terms with  $u \neq \lambda$  and renaming  $v$  to  $u_{r+1}$  we obtain

$$\begin{aligned} \langle \mathbf{x}^{-1}, w \rangle &= -\langle \mathbf{x}^{-1}, \lambda \rangle \langle \mathbf{x}, w \rangle \\ &\quad - \sum_{uv=w, v \neq \lambda} \sum_{u_1 u_2 \dots u_r = u, u_i \neq \lambda, r \geq 1} (-1)^r \langle \mathbf{x}, u_1 \rangle \dots \langle \mathbf{x}, u_r \rangle \langle \mathbf{x}, v \rangle \\ &= \sum_{uv=w, v \neq \lambda} \sum_{u_1 u_2 \dots u_r = u, u_i \neq \lambda, r \geq 1} (-1)^{r+1} \langle \mathbf{x}, u_1 \rangle \dots \langle \mathbf{x}, u_r \rangle \langle \mathbf{x}, v \rangle \\ &= (-1)^1 \langle \mathbf{x}, w \rangle \\ &\quad + \sum_{u_1 u_2 \dots u_r u_{r+1} = w, u_i \neq \lambda, r \geq 1} (-1)^{r+1} \langle \mathbf{x}, u_1 \rangle \dots \langle \mathbf{x}, u_r \rangle \langle \mathbf{x}, u_{r+1} \rangle \\ &= \sum_{u_1 u_2 \dots u_r = w, u_i \neq \lambda, r \geq 1} (-1)^r \langle \mathbf{x}, u_1 \rangle \dots \langle \mathbf{x}, u_r \rangle \end{aligned}$$

since the only term possible for the case  $r = 1$  is  $(-1)^1 \langle \mathbf{x}, w \rangle$ .  $\square$

**Example 2.10.** Consider the binary alphabet  $\Sigma = \{a, b\}$  and the series  $\mathbf{x} = 1 + a - 2ab \in \mathbb{Z}\langle\langle \Sigma \rangle\rangle$ . Since  $\langle \mathbf{x}, \lambda \rangle = 1$ , the inverse series  $\mathbf{x}^{-1}$  exists. Denote  $\mathbf{y} = 1 - \mathbf{x} = -a + 2ab$ . We can compute a few initial powers of  $\mathbf{y}$ :  $\mathbf{y}^0 = 1$ ,  $\mathbf{y}^1 = -a + 2ab$ ,  $\mathbf{y}^2 = aa - 2aab - 2aba + 4abab$ ,  $\mathbf{y}^3 = -aaa + 2aaab + 2aaba - 4aabab + 2abaa - 4abaab - 4ababa + 8ababab$ . All non-zero terms in  $\mathbf{y}^i$ ,  $i \geq 4$ , correspond to words of length at least 4. Hence the series  $\mathbf{z} = \mathbf{y}^0 + \mathbf{y}^1 + \mathbf{y}^2 + \mathbf{y}^3 = 1 - a + aa + 2ab - aaa - 2aab - 2aba + 2aaab + 2aaba + 2abaa + 4abab - 4aabab - 4abaab - 4ababa + 8ababab$  fully determines the values  $\langle \mathbf{x}^{-1}, w \rangle$  for  $|w| \leq 3$ . Therefore  $\mathbf{x}^{-1} = 1 - a + aa + 2ab - aaa - 2aab - 2aba + \dots$  where the remaining terms correspond to words of length at least 4. Using the formula from part (1) of Proposition 2.8, we obtain

$$\begin{aligned} \langle \mathbf{x}^{-1}, aba \rangle &= -(\langle \mathbf{x}^{-1}, \lambda \rangle \langle \mathbf{x}, aba \rangle + \langle \mathbf{x}^{-1}, a \rangle \langle \mathbf{x}, ba \rangle + \langle \mathbf{x}^{-1}, ab \rangle \langle \mathbf{x}, a \rangle) \\ &= -(0 + 0 + 2) = -2. \end{aligned}$$

Using the formula from Lemma 2.9, we obtain

$$\begin{aligned} \langle \mathbf{x}^{-1}, aba \rangle &= (-1)^2 \langle \mathbf{x}, a \rangle \langle \mathbf{x}, ba \rangle + (-1)^2 \langle \mathbf{x}, ab \rangle \langle \mathbf{x}, a \rangle + (-1)^3 \langle \mathbf{x}, a \rangle \langle \mathbf{x}, b \rangle \langle \mathbf{x}, a \rangle \\ &= 0 - 2 + 0 = -2. \end{aligned}$$

Assume now the factorial language  $L = \{\lambda, a, b, ab, ba, aba\}$ . Then  $\mathbf{x} \in \mathbb{Z}_L^{(1)}$  and one can easily verify that  $\pi_L(\mathbf{x}^{-1}) = \pi_L(\mathbf{z}) = 1 - a + 2ab - 2aba$  is the multiplicative inverse of  $\mathbf{x}$  in  $\mathbb{Z}_L^{(1)}$ .

### 3. THE GENERALIZATION OF THE PARIKH MAPPING

In this section we describe a framework for defining extensions of the Parikh mapping. The Parikh matrix [16] and its generalization from [23] contain information on counts of subword occurrences in the given word for subwords being factors of one fixed word only. Our idea is to associate with each word an element of  $\mathbb{Z}_L^{(1)}$  containing such counts for any word from a fixed finite language  $L$ . Such association will be given by a mapping based on the Magnus morphism assigning to each word  $\alpha$  a formal power series, reflecting the number of subword occurrences in  $\alpha$  of each word from  $\Sigma^*$ . The mapping  $\mathbf{\Pi}_L$ , described in the following text, reduces the formal series resulting from the Magnus morphism to a polynomial with a fixed size support given by the language  $L$ , thus making it a tool more suitable for practical computations, in particular if  $L$  is a factorial language. The Magnus morphism is used in Section 6.3 of [13] to prove orthogonality properties of generalized combinatorial numbers (the notation mentioned in Sect. 1) and identities of shuffle and infiltration product operations. The wider algebraic importance of the Magnus morphism is discussed there, as well. The results from Section 6.3 of [13], though dealing with subword counts, are not directly related to our current paper.

The Magnus (monoid) morphism  $\mu : \Sigma^* \rightarrow \mathbb{Z}\langle\langle\Sigma\rangle\rangle$  is defined, for  $i = 1, \dots, k$ , as  $\mu(s_i) = 1 + s_i$ . It can be expressed, for any  $\alpha \in \Sigma^*$ , as  $\mu(\alpha) = \sum_v |\alpha|_v v$ . We define the *extended Parikh mapping*  $\mathbf{\Pi}_L : \Sigma^* \rightarrow \mathbb{Z}_L$  as  $\mathbf{\Pi}_L = \pi_L \circ \mu$ . Hence  $\mathbf{\Pi}_L(\alpha) = \sum_{v \in L} |\alpha|_v v$ . This mapping has been considered for the case of a finite language  $L$  in [9]. If  $L$  is a factorial language then  $\mathbf{\Pi}_L(\Sigma^*) \subseteq \mathbb{Z}_L^{(1)}$  then  $\mathbf{\Pi}_L$  is a monoid morphism, since it is a composition of monoid morphisms (see Prop. 2.3).

We want to document by the following examples that the extended Parikh mapping  $\mathbf{\Pi}_L$  can be viewed as a generalization of any one of the following: the original Parikh mapping, the Parikh matrix mapping, the extended Parikh matrix mapping, and the p-matrix mapping, as shown in the following examples.

**Example 3.1.** The Parikh mapping [18] assigns to each word  $w$  the vector  $\mathbf{\Pi}(w) = (|w|_{s_1}, |w|_{s_2}, \dots, |w|_{s_k})$ . The Parikh mapping corresponds to  $\mathbf{\Pi}_{\mathcal{F}(\Sigma)} = \mathbf{\Pi}_{\Sigma \cup \{\lambda\}}$ .

**Example 3.2.** The Parikh matrix mapping [16] assigns to each word  $w \in \Sigma^*$  the upper-triangular  $(k + 1) \times (k + 1)$  matrix  $\Psi(w)$  where the main diagonal consists of 1's and, for  $1 \leq i \leq j \leq k$ , the  $(i, j + 1)$ -th element is  $|w|_{s_i s_{i+1} \dots s_j}$ . The monoid  $(\Psi(\Sigma^*), \cdot, \mathbf{1}_{k+1, k+1})$ , where  $\cdot$  is the usual matrix multiplication and  $\mathbf{1}_{k+1, k+1}$  is the  $(k + 1) \times (k + 1)$  identity matrix, is isomorphic to  $(\mathbb{Z}_L, \cdot_L, 1)$  where  $L$  consists of all factors of  $s_1 s_2 \dots s_k$ . The Parikh matrix mapping corresponds to  $\mathbf{\Pi}_{\mathcal{F}(s_1 s_2 \dots s_k)}$ . For example, if  $\Sigma = \{a, b, c\}$  then  $\mathcal{F}(abc) = \{\lambda, a, b, c, ab, bc, abc\}$ ,

$$\Psi(acbcba) = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and  $\mathbf{\Pi}_{\mathcal{F}(abc)}(acbcbac) = (1+a) \cdot_L (1+c) \cdot_L (1+b) \cdot_L (1+c) \cdot_L (1+b) \cdot_L (1+a) \cdot_L (1+c) = 1 + 2a + 2b + 3c + 2ab + 3bc + 3abc$ .

**Example 3.3.** Let  $a_1 \dots a_m$  be the symbol decomposition of a nonempty word  $u \in \Sigma^*$ . The generalized Parikh matrix mapping [23] assigns to each word  $w$  the upper-triangular  $(m + 1) \times (m + 1)$  matrix  $\Psi_u(w)$  where the main diagonal consists of 1's and, for  $1 \leq i \leq j \leq m$ , the  $(i, j + 1)$ -th elements is  $|w|_{a_i a_{i+1} \dots a_j}$ . The generalized Parikh matrix mapping corresponds to  $\mathbf{\Pi}_{\mathcal{F}(u)}$ .

For example, if  $\Sigma = \{a, b, \}$ ,  $u = abb$ , then  $\mathcal{F}(abb) = \{\lambda, a, b, ab, bb, abb\}$ ,

$$\Psi_u(abbaabb) = \begin{bmatrix} 1 & 3 & 8 & 8 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and  $\mathbf{\Pi}_{\mathcal{F}(abb)}(abbaabb) = (1+a) \cdot_L (1+b) \cdot_L (1+b) \cdot_L (1+a) \cdot_L (1+a) \cdot_L (1+b) \cdot_L (1+b) = 1 + 3a + 4b + 8ab + 6bb + 8abb$ .

Consider a further extension to the mapping  $\mathbf{\Pi}_{\mathcal{F}(u)}$ . Let  $\mathcal{M}_{m+1}$  denote the set of all upper-triangular  $(m + 1) \times (m + 1)$  matrices with elements from  $\mathbb{Z}$ , having all elements on the main diagonal equal to 1.  $\mathcal{M}_{m+1}$  with the usual multiplication is a group, since the determinant of every matrix is 1 (thus the inverse matrix always exists and consists of integer elements). Denote, for  $1 \leq i \leq j \leq m$ , by  $u_{i,j}$  the factor  $a_i \dots a_j$  of  $u$ . We define the mapping  $\phi : \mathbb{Z}_{\mathcal{F}(u)}^{(1)} \rightarrow \mathcal{M}_{m+1}$  where  $\phi(\mathbf{x})_{i,j+1} = \langle \mathbf{x}, u_{i,j} \rangle$ , for  $\mathbf{x} \in \mathbb{Z}_{\mathcal{F}(u)}^{(1)}$ ,  $1 \leq i \leq j \leq m$ .  $\phi$  is a group morphism. It is an isomorphism iff no factor occurs in  $u$  more than once. For a word  $\alpha$ ,  $\phi(\mathbf{\Pi}_{\mathcal{F}(u)}(\alpha))$  is the generalized Parikh matrix.

**Example 3.4.** The p-matrix mapping [8,10] assigns to each word  $w$  the  $k \times k$  matrix  $P(w)$  where, for  $1 \leq i \leq k$ ,  $P(w)_{i,i} = |w|_{s_i}$  and, for  $1 \leq i < j \leq k$ , the  $(i, j)$ -th element is  $|w|_{s_i s_j}$ . The “p-product” operation  $\circ$  of two  $k \times k$  matrices  $A, B$  is defined as  $(A \circ B)_{i,i} = A_{i,i} + B_{i,i}$ , for  $i = 1, \dots, k$ , and  $(A \circ B)_{i,j} = A_{i,j} + B_{i,j} + A_{i,i} B_{j,j}$ , for  $1 \leq i, j \leq k, i \neq j$ . The p-matrix mapping corresponds to  $\mathbf{\Pi}_L$  where  $L = \Sigma_2 - \{a^2 | a \in \Sigma\}$ . As follows from Proposition 1.1, in the case of a binary alphabet, the monoid of p-matrices is isomorphic to the monoid of Parikh matrices. For example, if  $\Sigma = \{a, b, c\}$  then  $L = \{\lambda, a, b, c, ab, ac, ba, bc, ca, cb\}$  and

$$P(acbcbac) = \begin{bmatrix} 2 & 2 & 4 \\ 2 & 2 & 3 \\ 2 & 3 & 3 \end{bmatrix}$$

and  $\mathbf{\Pi}_L(acbcbac) = (1+a) \cdot_L (1+c) \cdot_L (1+b) \cdot_L (1+c) \cdot_L (1+b) \cdot_L (1+a) \cdot_L (1+c) = 1 + 2a + 2b + 3c + 2ab + 4ac + 2ba + 3bc + 2ca + 3cb$ .

While the Parikh mapping (see Ex. 3.1) assigns the same vector to a word and its mirror image, the matrix generalizations of the Parikh mapping do not have an analogical property. Still, in the case of Parikh matrices and generalized Parikh



matrices, equality of the inverse matrix of a word and the so-called alternate matrix of the word's mirror image can be proved (see Thm. 16 in [23], being a generalization of a result from [16]). We will now show that a similar equality is valid in the general case of extended Parikh mappings for factorial languages. In the alternate matrix, the  $(i, j)$ th entry  $a_{i,j}$  of the original matrix is replaced by  $(-1)^{i+j}a_{i,j}$ . This concept can be rephrased in formal power series as follows. The *alternate series* of  $\mathbf{x} \in \mathbb{Z}\langle\langle\Sigma\rangle\rangle$  is defined as  $\bar{\mathbf{x}} = \sum_v (-1)^{|v|} \langle \mathbf{x}, v \rangle v$ . It is easy to check that the mapping  $\mathbf{x} \mapsto \bar{\mathbf{x}}$  is a ring morphism. The proof of the theorem within the framework of formal power series is shorter than those from [16] and [23] and gives a clearer clue why  $aa$  is a forbidden factor.

**Theorem 3.5.** *Let  $L$  be a factorial language, not containing any word  $aa$ , where  $a \in \Sigma$ . Let  $w \in \Sigma^*$ . Then, in  $\mathbb{Z}_L$ ,*

$$\mathbf{\Pi}_L(w)^{-1} = \overline{\mathbf{\Pi}_L(w^R)}.$$

*Proof.* The assertion is trivially true for  $w = \lambda$  and for  $w = a \in \Sigma - L$ . Let  $w = a \in \Sigma \cap L$ . Then  $\mathbf{\Pi}_L(w) \cdot_L \overline{\mathbf{\Pi}_L(w^R)} = \mathbf{\Pi}_L(a) \cdot_L \overline{\mathbf{\Pi}_L(a)} = (1+a) \cdot_L (1-a) = 1$ , since  $aa \notin L$ . Let us now assume that the assertion is true for some word  $w \in \Sigma^*$  and let  $a \in \Sigma$ . Then  $\mathbf{\Pi}_L(wa) \cdot_L \overline{\mathbf{\Pi}_L((wa)^R)} = \mathbf{\Pi}_L(wa) \cdot_L \overline{\mathbf{\Pi}_L(aw^R)} = \mathbf{\Pi}_L(w) \cdot_L \overline{\mathbf{\Pi}_L(a)} \cdot_L \overline{\mathbf{\Pi}_L(w^R)} = \mathbf{\Pi}_L(w) \cdot_L \overline{\mathbf{\Pi}_L(w^R)} = 1$ .  $\square$

Applying Lemma 2.9, (1) and (2) of Proposition 2.8, and (in the third description) Proposition 2.3, we obtain three equivalent descriptions of  $|w^R|_u$ .

**Corollary 3.6.** *Let  $w \in \Sigma^*$ . If a word  $u \in \Sigma^*$  does not contain any factor of the form  $aa$ , where  $a \in \Sigma$ , then*

$$\begin{aligned} (1) \quad |w^R|_u &= (-1)^{|u|} \sum_{u_1 u_2 \dots u_r = u, u_i \neq \lambda} (-1)^r |w|_{u_1} |w|_{u_2} \dots |w|_{u_r} \\ (2) \quad |w^R|_u &= (-1)^{|u|} \left\langle \sum_{r=0}^{|u|} (1 - \mathbf{\Pi}_{\mathcal{F}(u)}(w))^r, u \right\rangle \\ (3) \quad |w^R|_u &= (-1)^{|u|} \left\langle \sum_{r=0}^{|u|} \pi_{\mathcal{F}(u)}((1 - \mathbf{\Pi}_{\mathcal{F}(u)}(w))^r), u \right\rangle. \end{aligned}$$

**Example 3.7.** Let  $\Sigma = \{a, b, c\}$ ,  $L = \{\lambda, a, b, c, ab, ac\}$ . Then  $\mathbf{\Pi}_L(abbcab) = 1 + 2a + 3b + c + 4ab + ac$ ,  $\mathbf{\Pi}_L(w^R) = 1 + 2a + 3b + c + 2ab + ac$ ,  $\mathbf{\Pi}_L(w)^{-1} = 1 - 2a - 3b - c + 2ab + ac$ . For  $w = abbcab$ ,

$$\begin{aligned} |w^R|_{acb} &= |bacbba|_{acb} \\ &= (-1)^3((-1)^3|w|_a|w|_c|w|_b + (-1)^2|w|_a|w|_{cb} + (-1)^2|w|_{ac}|w|_b \\ &\quad + (-1)^1|w|_{acb}) \\ &= -(-2 \cdot 1 \cdot 3 + 2 \cdot 1 + 1 \cdot 3 - 1) \\ &= 2 \\ &= (-1)^3(0 - 1 + 5 - 6) \end{aligned}$$

since

$$\begin{aligned}\pi_{\mathcal{F}(acb)}(1 - \mathbf{\Pi}_{\mathcal{F}(acb)}(w))^0 &= 1 \\ \pi_{\mathcal{F}(acb)}(1 - \mathbf{\Pi}_{\mathcal{F}(acb)}(w))^1 &= -2a - 3b - c - ac - cb - acb \\ \pi_{\mathcal{F}(acb)}(1 - \mathbf{\Pi}_{\mathcal{F}(acb)}(w))^2 &= 2ac + 3cb + 5acb \\ \pi_{\mathcal{F}(acb)}(1 - \mathbf{\Pi}_{\mathcal{F}(acb)}(w))^3 &= -6acb.\end{aligned}$$

#### 4. AMIABLE WORDS

One of the principal questions on Parikh matrices (as described in Ex. 3.2), widely studied in literature, is the problem of ambiguity of the Parikh matrix mapping [2–4, 12, 15]. When does the Parikh matrix uniquely determine the original word? Two words yielding the same Parikh matrix are called amiable [4]. We will consider a more general concept of  $L$ -amiability for an arbitrary factorial language  $L$ , based on our framework introduced in Section 3. In particular, we will focus on the problem of  $\Sigma_p$ -amiability,  $p \geq 2$ , which in the case  $p = 2$  coincides with the original Parikh-matrix amiability. The motivation for our research comes from mid 19th century investigations of Eugène Prouhet [19]<sup>2</sup> of a problem, currently known as Prouhet-Tarry-Escott (*PTE*) problem (see, *e.g.*, [7]). His solution can be obtained observing positions of symbol occurrences in the building blocks of the Thue-Morse sequence (therefore, this sequence is sometimes called Prouhet-Thue-Morse sequence – see [1]). These blocks can be described by iteration of the monoid morphism  $h : \{a, b\}^* \rightarrow \{a, b\}^*$ ,  $h(a) = ab$ ,  $h(b) = ba$ . The blocks  $h^p(a)$  and  $h^p(b)$  are  $\{a, b\}_p$ -amiable, for  $p \geq 1$ . Our aim is to show that, in fact, any pair of  $\Sigma_p$ -amiable words induces a solution the *PTE*-problem of order  $p$ . (We will call a pair of words inducing such solution to be *PTE* <sub>$p$</sub> -equivalent.) We proceed as follows. First we introduce the general concept of  $L$ -amiability and its basic property in Proposition 4.1. Then we define *PTE* <sub>$n$</sub> -equivalence of words (Def. 4.7) and list some of its properties in Proposition 4.8. The definition is based on the mapping  $\sigma_a^{(q)}$ , which assigns to each word  $\alpha$  the sum of the  $q$ -th powers of all positions in  $\alpha$ , where the particular symbol  $a$  occurs. Properties of  $\sigma_a^{(q)}$ , provided in Lemma 4.3 – Lemmas 4.6, 4.10, and Proposition 4.11, are used in later proofs. Our main result is expressed in two theorems. Theorem 4.12 states, that for a binary alphabet  $\Sigma$ , amiability in the original sense of Parikh matrix mapping, which is easy seen to be equivalent to  $\Sigma_2$ -amiability, is identical to *PTE*<sub>2</sub>-equivalence. As illustrated in Example 4.13, *PTE* <sub>$p$</sub> -equivalence is not a sufficient condition for  $\Sigma_p$ -amiability if  $|\Sigma| \geq 3$  or  $p \geq 3$ . However, it is a necessary condition for any  $|\Sigma| \geq 2$  and  $p \geq 2$ , as stated in Theorem 4.19. The inductive proof of Theorem 4.19 is based on the alphabetical morphism  $h_{a,b}$  investigated in Proposition 4.14, Lemmas 4.15 and 4.18. Lemmas 4.16 and 4.17 provide rather technical combinatorial tools used to prove Lemma 4.18. One may wonder, which condition, if added to the necessary condition of *PTE* <sub>$p$</sub> -equivalence, will make it

<sup>2</sup>More details can be found in [5].

sufficient for  $\Sigma_p$ -amiability. Lemma 4.21 provides such a condition in the case  $|\Sigma| = 3$ . However, Example 4.22 shows, that there is no straightforward generalization for larger alphabets. Renaming a symbol (as it happens by application of the morphism  $h_{a,b}$ ) does not change positions of other symbols and does not influence the values of  $\sigma_c^{(q)}$  for  $c \neq a, b$ . On the other hand, erasing all occurrences of some symbol causes shifts in positions of other symbols and thus changes the value of  $\sigma_c^{(q)}$ . Theorem 4.25 states, that  $PTE_p$ -equivalence still remains valid, if such arising is applied to a pair of  $\Sigma_p$ -amiable words. Our final Proposition 4.26 contains observation on  $PTE$ -equivalence and the values of  $\sigma_a^{(q)}$  for so-called fair words.

Let  $L$  be a factorial language, such that  $\Sigma \subseteq L$ . Two words  $\alpha, \beta \in \Sigma^*$  are called  $L$ -amiable (denoted as  $\alpha \sim_L \beta$ ), if  $\mathbf{\Pi}_L(\alpha) = \mathbf{\Pi}_L(\beta)$ . Since  $\mathbf{\Pi}_L$  is a monoid morphism, we have

**Proposition 4.1.** *The relation  $\sim_L$  is an equivalence relation being both left and right congruence with respect to concatenation.*

For  $p \geq 1$ ,  $\Sigma_p$  is a factorial language. Our interest in investigation of  $\Sigma_p$ -amiability, for values of  $p \geq 2$  and/or  $|\Sigma| \geq 2$ , is based on the following fact.

**Remark 4.2.** Proposition 1.1 implies that two words are amiable (in the sense of the original Parikh matrix mapping) iff they are  $\Sigma_2$ -amiable.

We will consider  $PTE$ -equivalence of strings, based on the Prouhet-Tarry-Escott ( $PTE$ ) problem [7]: find, for a given degree  $p \geq 1$ , two equally-sized sets of integers  $A, B$  satisfying  $\sum_{a \in A} a^q = \sum_{b \in B} b^q$  for  $0 \leq q < p$ . Usually, the sets  $A, B$  are required to be disjoint; we will relax this condition here. To each word and each symbol, one can assign a  $PTE$ -like sum. Assume the symbol decomposition  $\alpha = a_1 \dots a_n \in \Sigma^*$ . For  $a \in \Sigma$ ,  $q \geq 0$  we denote  $\sigma_a^{(q)}(\alpha) = \sum_{a_i=a} i^q$ . Observe that, for  $p = 0$ ,  $\sigma_a^{(0)}(\alpha) = |\alpha|_a$ .

**Lemma 4.3.** *Let  $\alpha, \beta \in \Sigma^*$ ,  $a \in \Sigma$ ,  $q \geq 0$ . Then*

$$\sigma_a^{(q)}(\alpha\beta) = \sigma_a^{(q)}(\alpha) + \sum_{r=0}^q \binom{q}{r} |\alpha|^{q-r} \sigma_a^{(r)}(\beta).$$

*Proof.* Let  $\beta = b_1 \dots b_n$  be the symbol decomposition of  $\beta$ . If symbol  $a$  occurs at the  $i$ -th position in  $\beta$  then it occurs at position  $i + |\alpha|$  in  $\alpha\beta$ . Therefore

$$\sigma_a^{(q)}(\alpha\beta) = \sigma_a^{(q)}(\alpha) + \sum_{b_i=a} (i + |\alpha|)^q.$$

Using the binomial theorem and changing the order of summation in the following step yields

$$\begin{aligned}\sigma_a^{(q)}(\alpha\beta) &= \sigma_a^{(q)}(\alpha) + \sum_{b_i=a} \sum_{r=0}^q \binom{q}{r} |\alpha|^{q-r} i^r \\ &= \sigma_a^{(q)}(\alpha) + \sum_{r=0}^q \binom{q}{r} |\alpha|^{q-r} \sum_{b_i=a} i^r \\ &= \sigma_a^{(q)}(\alpha) + \sum_{r=0}^q \binom{q}{r} |\alpha|^{q-r} \sigma_a^{(r)}(\beta). \quad \square\end{aligned}$$

**Corollary 4.4.** *Let  $\alpha, \beta \in \Sigma^*$ ,  $a \in \Sigma$ . Then*

$$\sigma_a^{(1)}(\alpha\beta) = \sigma_a^{(1)}(\alpha) + \sigma_a^{(1)}(\beta) + |\alpha||\beta|_a.$$

**Lemma 4.5.** *Let  $\alpha \in \Sigma^*$ ,  $a \in \Sigma$ . Then  $\sigma_a^{(1)}(\alpha) + \sigma_a^{(1)}(\alpha^R) = (|\alpha| + 1)|\alpha|_a$ .*

*Proof.* Assume the symbol decomposition  $\alpha = a_1 \dots a_n$ . Then  $\sigma_a^{(1)}(\alpha) + \sigma_a^{(1)}(\alpha^R) = \sum_{a_i=a} i + \sum_{a_i=a} (|\alpha| + 1 - i) = \sum_{a_i=a} (|\alpha| + 1) = (|\alpha| + 1)|\alpha|_a$ .  $\square$

**Lemma 4.6.** *Let  $q \geq 0$ ,  $\alpha, \beta \in \{a, b\}^*$ ,  $|\alpha| = |\beta|$ . Then  $\sigma_a^{(q)}(\alpha) = \sigma_a^{(q)}(\beta)$  iff  $\sigma_b^{(q)}(\alpha) = \sigma_b^{(q)}(\beta)$ .*

*Proof.* If  $\sigma_a^{(q)}(\alpha) = \sigma_a^{(q)}(\beta)$  then  $\sigma_b^{(q)}(\alpha) = \sum_{i=1}^{|\alpha|} i^q - \sigma_a^{(q)}(\alpha) = \sum_{i=1}^{|\beta|} i^q - \sigma_a^{(q)}(\beta) = \sigma_b^{(q)}(\beta)$ . The role of  $a, b$  can be interchanged.  $\square$

**Definition 4.7.** Two words  $\alpha, \beta \in \Sigma^*$  are called  $PTE_p$ -equivalent,  $p \geq 1$  (denoted as  $\alpha \approx_p \beta$ ), if  $\sigma_a^{(q)}(\alpha) = \sigma_a^{(q)}(\beta)$  for each  $a \in \Sigma$  and each  $0 \leq q < p$ .

In the following proposition, we summarize some basic properties of the  $PTE_p$ -equivalence  $\approx_p$  ( $\mathbf{\Pi}$  denotes there the Parikh mapping, as described in Ex. 3.1). The properties (1) and (2) are quite straightforward, property (3) follows from Lemma 4.3 using the fact (from property (1)) that two  $PTE_p$ -equivalent words are of the same length.

**Proposition 4.8.** *Let  $\alpha, \beta \in \Sigma^*$ ,  $p \geq 1$ .*

- (1)  $\alpha \approx_p \beta$  implies  $\mathbf{\Pi}(\alpha) = \mathbf{\Pi}(\beta)$  and, consequently,  $|\alpha| = |\beta|$ .
- (2)  $\alpha \approx_1 \beta$  iff  $\mathbf{\Pi}(\alpha) = \mathbf{\Pi}(\beta)$ .
- (3)  $\approx_p$  is a left and right congruence with respect to concatenation.

**Example 4.9.**  $\alpha = aaaabb$ ,  $\beta = ababba$  are words with  $\mathbf{\Pi}(\alpha) \neq \mathbf{\Pi}(\beta)$  satisfying  $\sigma_a^{(1)}(\alpha) = \sigma_a^{(1)}(\beta) = 10$ ,  $\sigma_b^{(1)}(\alpha) = \sigma_b^{(1)}(\beta) = 11$ . In this case,  $15 = \sigma_a(a\alpha) \neq \sigma_a(a\beta) = 14$ . Thus the equivalence based on  $\sigma^{(1)}$  alone is not a congruence.

**Lemma 4.10.** *Let  $\alpha, \beta \in \Sigma^*$ ,  $a \in \Sigma$ ,  $p \geq 1$ . Assume  $\alpha \approx_p \beta$ . If  $\sigma_a^{(p)}(\alpha^R) = \sigma_a^{(p)}(\beta^R)$  then  $\sigma_a^{(p)}(\alpha) = \sigma_a^{(p)}(\beta)$ .*

*Proof.* Assume the symbol decomposition  $\alpha = a_1 \dots a_n$ . If a symbol occurs at position  $i$  in  $\alpha$  then it occurs at position  $|\alpha| + 1 - i$  in  $\alpha^R$ . Then  $\sigma_a^{(p)}(\alpha^R) = \sum_{a_i=a} (|\alpha| + 1 - i)^p$ . Using the binomial theorem and changing the order of summation in the following step yields

$$\begin{aligned} \sigma_a^{(p)}(\alpha^R) &= \sum_{a_i=a} \sum_{j=0}^p (-1)^j \binom{p}{j} (|\alpha| + 1)^{p-j} i^j \\ &= \sum_{j=0}^p (-1)^j \binom{p}{j} (|\alpha| + 1)^{p-j} \sum_{a_i=a} i^j \\ &= \sum_{j=0}^p (-1)^j \binom{p}{j} (|\alpha| + 1)^{p-j} \sigma_a^{(j)}(\alpha). \end{aligned}$$

From here we can express  $\sigma_a^{(p)}(\alpha)$  being the term for  $j = p$  in the right-hand side sum

$$\sigma_a^{(p)}(\alpha) = (-1)^p \left[ \sigma_a^{(p)}(\alpha^R) - \sum_{j=0}^{p-1} (-1)^j \binom{p}{j} (|\alpha| + 1)^{p-j} \sigma_a^{(j)}(\alpha) \right].$$

A similar equation is valid for  $\sigma_a^{(p)}(\beta)$ . The equivalence  $\alpha \approx_p \beta$  implies  $\sigma_a^{(j)}(\alpha) = \sigma_a^{(j)}(\beta)$  for  $0 \leq j \leq p - 1$  and together with the additional assumption  $\sigma_a^{(p)}(\alpha^R) = \sigma_a^{(p)}(\beta^R)$  it implies  $\sigma_a^{(p)}(\alpha) = \sigma_a^{(p)}(\beta)$  (since  $|\alpha| = |\beta|$ ) by (1) of Proposition 4.8.  $\square$

**Proposition 4.11.** *Let  $\alpha, \beta \in \Sigma^*$ ,  $p \geq 1$ . Then  $\alpha \approx_p \beta$  iff  $\alpha^R \approx_p \beta^R$ .*

*Proof.* It is enough to prove the if-condition. We proceed by induction on  $p$ . The assertion is true for  $p = 1$ . Assume it is true for some  $p - 1 \geq 1$ . Let  $\alpha^R \approx_p \beta^R$ . Then  $\sigma_a^{(p-1)}(\alpha^R) = \sigma_a^{(p-1)}(\beta^R)$  for every  $a \in \Sigma$ , and, by the inductive hypothesis,  $\alpha \approx_{p-1} \beta$ . Lemma 4.10 implies  $\alpha \approx_p \beta$ .  $\square$

$\{a, b\}_2$ -amiable words – words over a binary alphabet yielding identical Parikh matrix – have been characterized in several different ways in [2,3]. Our following theorem provides one more characterization by  $PTE_2$ -equivalence.

**Theorem 4.12.** *Let  $\alpha, \beta \in \{a, b\}^*$ . The following three conditions are equivalent: (1)  $\alpha, \beta$  are amiable, (2)  $\alpha \sim_{\{a,b\}_2} \beta$ , (13)  $\alpha \approx_2 \beta$ .*

*Proof.* The equivalence of (1) and (2) follows from Remark 4.2. To prove the equivalence of (1) and (3), observe that each of these two conditions implies  $\mathbf{\Pi}(\alpha) = \mathbf{\Pi}(\beta)$  (see Prop. 4.8). Assume therefore that  $\mathbf{\Pi}(\alpha) = \mathbf{\Pi}(\beta)$  holds. Then  $|\alpha| = |\beta|$  and, following Lemma 4.6,  $\sigma_a^{(1)}(\alpha) = \sigma_a^{(1)}(\beta)$  iff  $\sigma_b^{(1)}(\alpha) = \sigma_b^{(1)}(\beta)$ . Denote  $r = |\alpha|_b = |\beta|_b$ . Then  $\alpha = a^{x_0} b a^{x_1} b \dots b a^{x_{r-1}} b a^{x_r}$  and  $\beta = a^{y_0} b a^{y_1} b \dots b a^{y_{r-1}} b a^{y_r}$  for some integers  $x_0, \dots, x_r, y_0, \dots, y_r \geq 0$ . The value  $\sigma_b^{(1)}(\alpha)$  can be expressed as  $\sigma_b^{(1)}(\alpha) = \sum_{i=0}^{r-1} \sum_{j=0}^i (x_j + 1)$ . In this sum the term  $x_i$  occurs exactly  $r - i$

times. Each symbol in the  $i$ -th group of  $a$ 's participates in  $r - i$  subwords  $ab$  of  $\alpha$ . Therefore  $\sigma_b^{(1)}(\alpha) = \sum_{i=0}^{r-1} (r - i)x_i + r(r + 1)/2 = |\alpha|_{ab} + r(r + 1)/2$ . Similarly,  $\sigma_b^{(1)}(\beta) = |\beta|_{ab} + r(r + 1)/2$ . Thus  $\sigma_b^{(1)}(\alpha) = \sigma_b^{(1)}(\beta)$  iff  $|\alpha|_{ab} = |\beta|_{ab}$ . The latter condition is equivalent to amiability of  $\alpha$  and  $\beta$ , since we assumed  $\mathbf{\Pi}(\alpha) = \mathbf{\Pi}(\beta)$ .  $\square$

If  $|\Sigma| = k > 2$  or  $p > 2$ , the conditions  $\alpha \sim_{\Sigma_p} \beta$  and  $\alpha \approx_p \beta$  need not be equivalent, as illustrated by Example 4.13. However, the sequence of assertions following Example 4.13 will lead to our main result (Thm. 4.19) stating that the former equivalence implies the latter.

**Example 4.13.**

- (1) In the case  $k = 2$ ,  $p = 3$ , consider the words

$$\alpha = ababaabab, \quad \beta = baaaabbbba.$$

Here  $\mathbf{\Pi}(\alpha) = \mathbf{\Pi}(\beta) = (5, 4)$ ,  $\sigma_a^{(1)}(\alpha) = \sigma_a^{(1)}(\beta) = 23$ ,  $\sigma_b^{(1)}(\alpha) = \sigma_b^{(1)}(\beta) = 22$  and  $\sigma_a^{(2)}(\alpha) = \sigma_a^{(2)}(\beta) = 135$ ,  $\sigma_b^{(2)}(\alpha) = \sigma_b^{(2)}(\beta) = 150$ , hence  $\alpha \approx_3 \beta$ . However,  $|\alpha|_{aab} = 17 \neq 18 = |\beta|_{aab}$ , therefore  $\alpha \not\approx_{\{a,b\}_3} \beta$ .

- (2) In the case  $k = 3$ ,  $p = 2$ , consider the words

$$\alpha = abbabcabbacabcabaab, \quad \beta = bbbcbaaabaaacbbbcbaba.$$

Here  $\mathbf{\Pi}(\alpha) = \mathbf{\Pi}(\beta) = (8, 10, 3)$  and  $\sigma_a^{(1)}(\alpha) = \sigma_a^{(1)}(\beta) = 94$ ,  $\sigma_b^{(1)}(\alpha) = \sigma_b^{(1)}(\beta) = 103$ ,  $\sigma_c^{(1)}(\alpha) = \sigma_c^{(1)}(\beta) = 34$ , hence  $\alpha \approx_2 \beta$ . However,  $|\alpha|_{ab} = 35 \neq 34 = |\beta|_{ab}$ , therefore  $\alpha \not\approx_{\{a,b,c\}_2} \beta$ .

Let  $a, b \in \Sigma$ . We define a morphism  $h_{a,b} : \Sigma^* \rightarrow (\Sigma - a)^*$  replacing the symbol  $a$  by  $b$ . The morphism is given as  $h_{a,b}(a) = b$  and  $h_{a,b}(d) = d$ , for  $d \in \Sigma - a$ . An easy observation yields

**Proposition 4.14.** *Let  $a, b \in \Sigma$  be distinct symbols, and let  $\alpha, \beta \in \Sigma^*$ .*

- (1) *Let  $q \geq 0$ . Then  $\sigma_b^{(q)}(h_{a,b}(\alpha)) = \sigma_a^{(q)}(\alpha) + \sigma_b^{(q)}(\alpha)$ . If  $c \in \Sigma - \{a, b\}$  then  $\sigma_c^{(q)}(h_{a,b}(\alpha)) = \sigma_c^{(q)}(\alpha)$ .*  
(2) *Let  $p \geq 1$ . If  $\alpha \approx_p \beta$  then  $h_{a,b}(\alpha) \approx_p h_{a,b}(\beta)$ .*

A similar property as (2) in Proposition 4.14 holds for  $\Sigma_p$ -amiability.

**Lemma 4.15.** *Let  $\alpha, \beta \in \Sigma^*$ ,  $p \geq 1$  and  $a, b \in \Sigma$ . If  $\alpha \sim_{\Sigma_p} \beta$  then  $h_{a,b}(\alpha) \sim_{(\Sigma - a)_p} h_{a,b}(\beta)$ .*

*Proof.* For  $u \in (\Sigma - a)_p$ , consider the usual notation  $h_{a,b}^{-1}(u) = \{v \in \Sigma^* | h_{a,b}(v) = u\} \subseteq \Sigma^{|u|} \subseteq \Sigma_p$ . Then  $|\alpha|_u = \sum_{v \in h_{a,b}^{-1}(u)} |\alpha|_v = \sum_{v \in h_{a,b}^{-1}(u)} |\beta|_v = |\beta|_u$ .  $\square$

Before proceeding to Lemma 4.18, which is crucial for proving our main result (Thm. 4.19), we need two more technical lemmas. We assume that, by default, an empty sum is equal to 0, an empty product is equal to 1, and  $\binom{u}{t} = 0$  for  $t > u \geq 0$ .

**Lemma 4.16.** *Let  $t, u \geq 0$  be integers. Then*

$$t! \binom{u}{t} = (u - t + 1)(t - 1)! \binom{u}{t - 1}.$$

*Proof.* Both sides are equal to 0 if  $u < t$ . Let  $u \geq t$ . Then

$$\begin{aligned} (u - t + 1)(t - 1)! \binom{u}{t - 1} &= (u - t + 1)(t - 1)! \binom{u}{t - 1} \\ &= (u - t + 1)(t - 1)! \frac{u!}{(t - 1)!(u - t + 1)!} \\ &= t(t - 1)! \frac{u!}{t!(u - t)!} \\ &= t! \binom{u}{t}. \end{aligned} \quad \square$$

**Lemma 4.17.**

(1) *For each integer  $t \geq 0$ , there exists a sequence of integers  $\{a_{t,m}\}_{m=1}^{t-1}$  such that, for any integer  $u \geq 0$ ,*

$$u^t = t! \binom{u}{t} + \sum_{m=1}^{t-1} a_{t,m} u^m. \tag{4.1}$$

(2) *Let  $r \geq 0$  be a fixed integer. Then for each integer  $s \geq 0$  there exists a sequence of integers  $\{b_{s,j}\}_{j=0}^{s-1}$  (depending on  $r$ ) such that, for any integer  $0 \leq i \leq r$ ,*

$$i^s = (-1)^s s! \binom{r-i}{s} + \sum_{j=0}^{s-1} b_{s,j} i^j. \tag{4.2}$$

*Proof.* (1) Induction on  $t$ . For  $t = 0, 1$ , the assertion is satisfied by the empty sequence, and for  $t = 2$  by the sequence with the only element  $a_{2,1} = 1$ . Let there be a sequence  $\{a_{t-1,m}\}_{m=1}^{t-2}$  for some  $t - 1 \geq 2$ , satisfying (4.1). Choose  $a_{t,1} = -(t - 1)a_{t-1,1}$ ,  $a_{t,m} = a_{t-1,m-1} - (t - 1)a_{t-1,m}$  for  $1 \leq m \leq t - 2$ , and  $a_{t,t-1} = a_{t-1,t-2} + t - 1$ . We express  $u^t$  as  $u^t = u \cdot u^{t-1} = (t - 1)u^{t-1} + (u - t + 1)u^{t-1}$ . Now we use the inductive hypothesis and, consequently, Lemma 4.16:

$$\begin{aligned} u^t &= (t - 1)u^{t-1} + (u - t + 1) \left[ (t - 1)! \binom{u}{t - 1} + \sum_{m=1}^{t-2} a_{t-1,m} u^m \right] \\ &= (t - 1)u^{t-1} + t! \binom{u}{t} + \sum_{m=1}^{t-2} a_{t-1,m} u^{m+1} - (t - 1) \sum_{m=1}^{t-2} a_{t-1,m} u^m \\ &= t! \binom{u}{t} + \sum_{m=1}^{t-1} a_{t,m} u^m. \end{aligned}$$

(2) Induction on  $s$ . For  $s = 0$  the assertion is satisfied by the empty sequence and for  $s = 1$  by the sequence with the only element  $b_{1,0} = r$ . Let there be a sequence

$\{b_{s-1,j}\}_{j=0}^{s-2}$  for some  $s-1 \geq 1$ , satisfying (4.2). Choose  $b_{s,0} = -(r-s+1)b_{s-1,0}$ ,  $b_{s,j} = b_{s-1,j-1} - (r-s+1)b_{s-1,j}$  for  $1 \leq j \leq s-2$ , and  $b_{s,s-1} = b_{s-1,s-2} + r-s+1$ . We express  $i^s$  as  $i^s = i \cdot i^{s-1} = (r-s+1)i^{s-1} - (r-i-s+1)i^{s-1}$ . Now we use the inductive hypothesis and, consequently, Lemma 4.16 (choosing  $u = r-i$  and  $t = s$ )

$$\begin{aligned} i^s &= (r-s+1)i^{s-1} \\ &\quad - (r-i-s+1) \left[ (-1)^{s-1}(s-1)! \binom{r-i}{s-1} + \sum_{j=0}^{s-2} b_{s-1,j} i^j \right] \\ &= (r-s+1)i^{s-1} + (-1)^s s! \binom{r-i}{s} \\ &\quad + \sum_{j=0}^{s-2} b_{s-1,j} i^{j+1} - (r-s+1) \sum_{j=0}^{s-2} b_{s-1,j} i^j \\ &= (-1)^s s! \binom{r-i}{s} + \sum_{j=0}^{s-1} b_{s,j} i^j. \quad \square \end{aligned}$$

**Lemma 4.18.** *Let  $\alpha, \beta \in \{a, b\}^*$ , and  $p \geq 2$ . Assume  $\alpha \sim_{\{a,b\}_p} \beta$ ,  $\alpha = a^{x_1} b a^{x_2} b a^{x_3} \dots a^{x_r} b a^{x_{r+1}}$  and  $\beta = a^{y_1} b a^{y_2} b a^{y_3} \dots a^{y_r} b a^{y_{r+1}}$ ,  $r \geq 1$ . Denote, for  $1 \leq i \leq r$ ,  $X_i = \sum_{j=1}^i x_j$  and  $Y_i = \sum_{j=1}^i y_j$ . If  $s, t \geq 0$  and  $s+t \leq p-1$  then  $\sum_{i=1}^r i^s X_i^t = \sum_{i=1}^r i^s Y_i^t$ .*

*Proof.* Induction on  $s+t$ . The assertion is trivial for  $s+t=0$ . Let it be true for all  $s, t \geq 0$  satisfying  $s+t \leq z-1$  for some  $1 \leq z \leq p-2$ . Let  $s+t = z$ . Using 4.1 with  $u = X_i$  we obtain

$$\begin{aligned} \sum_{i=1}^r i^s X_i^t &= \sum_{i=1}^r i^s \left[ t! \binom{X_i}{t} + \sum_{m=1}^{t-1} a_{t,m} X_i^m \right] \\ &= t! \sum_{i=1}^r i^s \binom{X_i}{t} + \sum_{m=1}^{t-1} a_{t,m} \sum_{i=1}^r i^s X_i^m \end{aligned}$$

where in the second term of the latter equation we changed the order of summation. Now we apply (4.2):

$$\begin{aligned} \sum_{i=1}^r i^s X_i^t &= t! \sum_{i=1}^r \left[ (-1)^s s! \binom{r-i}{s} + \sum_{j=0}^{s-1} b_{s,j} i^j \right] \binom{X_i}{t} \\ &\quad + \sum_{m=1}^{t-1} a_{t,m} \sum_{i=1}^r i^s X_i^m \\ &= (-1)^s s! t! \sum_{i=1}^r \binom{r-i}{s} \binom{X_i}{t} \\ &\quad + \sum_{j=0}^{s-1} b_{s,j} \sum_{i=1}^r i^j t! \binom{X_i}{t} \\ &\quad + \sum_{m=1}^{t-1} a_{t,m} \sum_{i=1}^r i^s X_i^m. \end{aligned}$$



The second term was obtained by changing the order of summation. Applying (4.1) with  $u = X_i$  again we get

$$\begin{aligned} \sum_{i=1}^r i^s X_i^t &= (-1)^s s! t! \sum_{i=1}^r \binom{r-i}{s} \binom{X_i}{t} \\ &\quad + \sum_{j=0}^{s-1} b_{s,j} \sum_{i=1}^r i^j \left[ X_i^t - \sum_{m=1}^{t-1} a_{t,m} X_i^m \right] \\ &\quad + \sum_{m=1}^{t-1} a_{t,m} \sum_{i=1}^r i^s X_i^m. \end{aligned}$$

The second term can be now expressed as a sum of two parts. The latter one of them, after changing the order of summation, will be joined with the last term.

$$\begin{aligned} \sum_{i=1}^r i^s X_i^t &= (-1)^s s! t! \sum_{i=1}^r \binom{r-i}{s} \binom{X_i}{t} \\ &\quad + \sum_{j=0}^{s-1} b_{s,j} \sum_{i=1}^r i^j X_i^t \\ &\quad + \sum_{m=1}^{t-1} a_{t,m} \left[ \sum_{i=1}^r i^s X_i^m - \sum_{j=0}^{s-1} b_{s,j} \sum_{i=1}^r i^j X_i^m \right]. \end{aligned} \tag{4.3}$$

A similar equality is valid for  $\sum_{i=1}^r i^s Y_i^t$ . Using the inductive hypothesis, we may replace each occurrence of  $X_i$  in the last two terms of (4.3) by  $Y_i$ . The assumption  $\alpha \sim_{\{a,b\}_p} \beta$ , implies  $|\alpha|_{a^t b^{s+1}} = |\beta|_{a^t b^{s+1}}$ . Now observe that  $\sum_{i=1}^r \binom{r-i}{s} \binom{X_i}{t} = |\alpha|_{a^t b^{s+1}}$ , since there are exactly  $\binom{r-i}{s} \binom{X_i}{t}$  occurrences of the subword  $a^t b^{s+1}$  in  $\alpha$  with the first  $b$  at position  $i$ . In a similar way,  $\sum_{i=1}^r \binom{r-i}{s} \binom{Y_i}{t} = |\beta|_{a^t b^{s+1}}$ . Hence  $X_i$  may be replaced by  $Y_i$  in the term  $(-1)^s s! t! \sum_{i=1}^r \binom{r-i}{s} \binom{X_i}{t}$ , as well. Consequently,  $\sum_{i=1}^r i^s X_i^t = \sum_{i=1}^r i^s Y_i^t$ .  $\square$

We are ready to prove the main result of this section.

**Theorem 4.19.** *Let  $|\Sigma| \geq 2, p \geq 1$  and  $\alpha, \beta \in \Sigma^*$ . Then  $\alpha \sim_{\Sigma_p} \beta$  implies  $\alpha \approx_p \beta$ .*

*Proof.* Induction on  $k = |\Sigma|$ .

Assume first  $k = 2$  and  $\Sigma = \{a, b\}$ . We use induction on  $p$ . The assertion is trivial for  $p = 1$ . Let us assume it is true for some  $p - 1 \geq 1$ . We will prove its validity for  $p$ . Let  $\alpha \sim_{\{a,b\}_p} \beta$ . Then, by the inductive hypothesis,  $\alpha \approx_{p-1} \beta$ . Using the notation from Lemma 4.18, we obtain  $\sigma_b^{(p-1)}(\alpha) = \sum_{i=1}^r (X_i + i)^{p-1} = \sum_{i=1}^r \sum_{s+t=p-1} \binom{p-1}{s} i^s X_i^t = \sum_{s+t=p-1} \binom{p-1}{s} \sum_{i=1}^r i^s X_i^t$ . In the same way one obtains the equality  $\sigma_b^{(p-1)}(\beta) = \sum_{s+t=p-1} \binom{p-1}{s} \sum_{i=1}^r i^s Y_i^t$ . Lemma 4.18 implies  $\sigma_b^{(p-1)}(\alpha) = \sigma_b^{(p-1)}(\beta)$ . The equality  $\sigma_a^{(p-1)}(\alpha) = \sigma_a^{(p-1)}(\beta)$  follows from Lemma 4.6.

Let us assume now that the implication is true for some  $k - 1 \geq 2$ . Let  $p \geq 1$ . Take any  $c \in \Sigma$  and any two symbols  $a, b \in \Sigma$ , distinct from  $c$ . Consider  $\alpha, \beta \in \Sigma^*$  and assume  $\alpha \sim_{\Sigma_p} \beta$ . Lemma 4.15 together with (1) of Proposition 4.14 and the inductive hypothesis implies  $\sigma_c^q(\alpha) = \sigma_c^q(h_{a,b}(\alpha)) = \sigma_c^q(h_{a,b}(\beta)) = \sigma_c^q(\beta)$  for  $0 \leq q < p$ .  $\square$

**Remark 4.20.** An inspection of the Proof of Lemma 4.18 and the case  $k = 2$  of the Proof of Theorem 4.19 reveals, that in the case  $\Sigma = \{a, b\}$ , the assumption  $\alpha \sim_{\{a,b\}_p} \beta$  can be replaced by a weaker assumption  $\alpha \sim_{a^*b^*\cap\{a,b\}_p} \beta$ . We do not know whether some similar weaker assumption is sufficient in the case of a larger alphabet.

Our next lemma documents that, though, generally, the  $PTE_2$ -equivalence is not sufficient for  $\Sigma_2$ -amiability, an additional condition can make it sufficient in case  $k = 3$ . However, this is not true for  $k \geq 4$ , as shown in Example 4.22.

**Lemma 4.21.** *Let  $|\Sigma| = 3$  and let  $\alpha, \beta \in \Sigma^*$ , such that  $\alpha \approx_2 \beta$ . If, for two distinct symbols  $a, b \in \Sigma$ ,  $|\alpha|_{ab} = |\beta|_{ab}$  then  $\alpha \sim_{\Sigma_2} \beta$ .*

*Proof.* Assume  $\Sigma = \{a, b, c\}$ . Then  $h_{c,b}(\alpha) \approx_2 h_{c,b}(\beta)$  by (2) of Proposition 4.14. Therefore, by Theorem 4.12,  $|h_{c,b}(\alpha)|_{ab} = |h_{c,b}(\beta)|_{ab}$ . Thus  $|\alpha|_{ab} + |\alpha|_{ac} = |\beta|_{ab} + |\beta|_{ac}$  implying  $|\alpha|_{ac} = |\beta|_{ac}$ . The equality  $|\alpha|_{bc} = |\beta|_{bc}$  can be proved in a similar way, using the morphism  $h_{c,a}$  and Proposition 1.1. The same Proposition 1.1 then implies similar equalities for the remaining three words from  $\Sigma_2$ . Therefore  $\alpha \sim_{\Sigma_2} \beta$ .  $\square$

**Example 4.22.** Consider the words

$$\alpha = dbbabcbabbacdbcbabaab, \beta = bbbcbaddbaaacbbbcbaba.$$

Then  $\mathbf{\Pi}(\alpha) = \mathbf{\Pi}(\beta) = (6, 10, 3, 2)$  and  $\sigma_a^{(1)}(\alpha) = \sigma_a^{(1)}(\beta) = 79$ ,  $\sigma_b^{(1)}(\alpha) = \sigma_b^{(1)}(\beta) = 103$ ,  $\sigma_c^{(1)}(\alpha) = \sigma_c^{(1)}(\beta) = 34$ ,  $\sigma_d^{(1)}(\alpha) = \sigma_d^{(1)}(\beta) = 15$  hence  $\alpha \approx_2 \beta$ . Furthermore,  $|\alpha|_{ab} = 22 = |\beta|_{ab}$ ,  $|\alpha|_{ac} = 7 \neq 8 = |\beta|_{ac}$ ,  $|\alpha|_{ad} = 3 \neq 2 = |\beta|_{ad}$ ,  $|\alpha|_{bc} = 17 \neq 16 = |\beta|_{bc}$ ,  $|\alpha|_{bd} = 7 \neq 8 = |\beta|_{bd}$ ,  $|\alpha|_{cd} = 2 = |\beta|_{cd}$  therefore  $\alpha \not\sim_{\{a,b,c,d\}_2} \beta$ . Hence out of every three distinct symbols one can choose two symbols  $x, y$  such that  $|\alpha|_{xy} = |\beta|_{xy}$  and still  $\alpha \not\sim_{\{a,b,c,d\}_2} \beta$ .

After replacing of all occurrences of a symbol  $a$  by a symbol  $b$  using the morphism  $h_{a,b}$ , the value of  $\sigma_c^{(q)}$  remains unchanged for  $c$  distinct from  $a, b$  ((1) of Prop. 4.14). We now introduce a morphism  $h_a : \Sigma^* \rightarrow (\Sigma - a)^*$  erasing all occurrences of a symbol  $a \in \Sigma$ . The morphism is defined as  $h_a(a) = \lambda$  and  $h_a(d) = d$ , for  $d \in \Sigma - a$ . By applying the morphism, the value of  $\sigma_c^{(q)}$ , for  $c \neq a$ , may, and mostly will, change. We will mention just a few facts about the effect of the morphism  $h_a$  on the value of  $\sigma_c^{(q)}$  and  $PTE$ -equivalence of words.

**Proposition 4.23.** *Let  $\alpha, \beta \in \Sigma^*$ . Then  $\alpha \sim_{\Sigma_2} \beta$  implies  $h_a(\alpha) \sim_{(\Sigma-a)_2} h_a(\beta)$ .*

**Lemma 4.24.** *Let  $a, b \in \Sigma$ ,  $a \neq b$ , and  $\alpha \in \Sigma^*$ . Then  $\sigma_b^{(1)}(\alpha) = \sigma_b^{(1)}(h_a(\alpha)) + |\alpha|_{ab}$ .*

*Proof.* If  $b$  occurs in  $\alpha$  at position  $i$ , then the image of this symbol occurs in  $h_a(\alpha)$  at position  $i - m$ , where  $m$  is the number of symbols  $a$  occurring in the prefix of  $\alpha$  of length  $i - 1$ . This occurrence of  $b$  participates in exactly  $m$  subwords  $ab$  of  $\alpha$ .  $\square$

**Theorem 4.25.** *Let  $|\Sigma| \geq 2$  and  $\alpha, \beta \in \Sigma^*$ . Then  $\alpha \sim_{\Sigma_2} \beta$  iff  $h_a(\alpha) \approx_2 h_a(\beta)$  for every  $a \in \Sigma$ .*

*Proof.* For  $|\Sigma| = 2$ , the assertion is trivial. Let us assume  $|\Sigma| \geq 3$ . The only-if part follows from Lemma 4.24 and Theorem 4.19. To prove the if part, assume  $\alpha, \beta \in \Sigma^*$  such that  $h_a(\alpha) \approx_2 h_a(\beta)$  for every  $a \in \Sigma$ . Proposition 4.8 implies  $\mathbf{\Pi}(\alpha) = \mathbf{\Pi}(\beta)$ . Let  $a, b$  be distinct symbols from  $\Sigma$ . We will prove that  $|\alpha|_{ab} = |\beta|_{ab}$ . Lemma 4.24 implies  $\sigma_a^{(1)}(\alpha) = \sigma_a^{(1)}(h_b(\alpha)) + |\alpha|_{ba}$  and  $\sigma_b^{(1)}(\alpha) = \sigma_b^{(1)}(h_a(\alpha)) + |\alpha|_{ab}$ . Therefore  $\sigma_a^{(1)}(\alpha) + \sigma_b^{(1)}(\alpha) = \sigma_a^{(1)}(h_b(\alpha)) + \sigma_b^{(1)}(h_a(\alpha)) + |\alpha|_a |\alpha|_b$ , where we applied (1.1). From  $h_a(\alpha) \approx_2 h_a(\beta)$  and  $h_b(\alpha) \approx_2 h_b(\beta)$  it follows (using Prop. 4.8), that

$$\sigma_a^{(1)}(\alpha) + \sigma_b^{(1)}(\alpha) = \sigma_a^{(1)}(\beta) + \sigma_b^{(1)}(\beta). \tag{4.4}$$

Consider a third distinct symbol  $c \in \Sigma$ . Equalities similar to (4.4) are valid for the pairs  $a, c$  and  $b, c$ , as well. These three equalities then imply  $\sigma_b(\alpha) = \sigma_b(\beta)$  and the equality  $|\alpha|_{ab} = |\beta|_{ab}$  follows from Lemma 4.24.  $\square$

We conclude by observing the values of the mapping  $\sigma_a^{(1)}$  on fair words [10,21]. A word  $\alpha \in \Sigma$  is called *fair* if for each pair of symbols  $a, b \in \Sigma$ ,  $|\alpha|_{ab} = |\alpha|_{ba}$ . Every palindrome (a word  $\alpha$  satisfying  $\alpha = \alpha^R$ ) is a fair word,  $ab^3a^2b$  is an example of a fair word not being a palindrome. In the following proposition we use the notation  $\delta_{a,b} = 1$  if  $a = b$  then 0 else and the fact that every palindrome can be expressed as  $\alpha b \alpha^R$  for some  $\alpha \in \Sigma^*$  and  $b \in \Sigma \cup \lambda$ .

**Proposition 4.26.** *Let  $\alpha \in \Sigma^*$ ,  $a, b \in \Sigma$ .*

- (1) *If  $\alpha$  is fair, then  $\alpha \sim_{\Sigma_2} \alpha^R$  and  $\alpha \approx_2 \alpha^R$ .*
- (2) *If  $\alpha$  is fair, then  $\sigma_a^{(1)}(\alpha) = (|\alpha| + 1)|\alpha|_a/2$ . If  $|\Sigma| = 2$  and  $\sigma_a^{(1)}(\alpha) = (|\alpha| + 1)|\alpha|_a/2$ , then  $\alpha$  is fair.*
- (3)  *$\sigma_a^{(1)}(\alpha\alpha^R) = (2|\alpha| + 1)|\alpha|_a$ ,  $\sigma_a^{(1)}(\alpha b \alpha^R) = (|\alpha| + 1)(2|\alpha|_a + \delta_{a,b})$ .*

*Proof.* (1) If  $\alpha$  is fair then, for  $a, b \in \Sigma$ ,  $a \neq b$ ,  $|\alpha|_{ab} = |\alpha|_{ba} = |\alpha^R|_{ab}$ . Thus  $\alpha \sim_{\Sigma_2} \alpha^R$  and Theorem 4.19 implies  $\alpha \approx_2 \alpha^R$ . (2) Following Part (1),  $\sigma_a^{(1)}(\alpha^R) = \sigma_a^{(1)}(\alpha)$ . The first assertion follows from Lemma 4.5. The second assertion is obtained using Lemmas 4.5, 4.6, and Theorem 4.12. (3) Using Corollary 4.4 and Lemma 4.5 we obtain  $\sigma_a^{(1)}(\alpha\alpha^R) = \sigma_a^{(1)}(\alpha) + \sigma_a^{(1)}(\alpha^R) + |\alpha| |\alpha^R|_a = (|\alpha| + 1)|\alpha^R|_a + |\alpha| |\alpha|_a = (2|\alpha| + 1)|\alpha|_a$ , since  $|\alpha^R|_a = |\alpha|_a$ , and  $\sigma_a^{(1)}(\alpha b \alpha^R) = \sigma_a^{(1)}(\alpha b) + \sigma_a^{(1)}(\alpha^R) + |\alpha b| |\alpha^R|_a = \sigma_a^{(1)}(\alpha) + (|\alpha| + 1)\delta_{a,b} + \sigma_a^{(1)}(\alpha^R) + (|\alpha| + 1)|\alpha|_a = (|\alpha| + 1)(2|\alpha|_a + \delta_{a,b})$ .  $\square$

Hence the symbols in a fair word  $\alpha$  are “evenly” distributed, since the average position number of any symbol is  $(|\alpha| + 1)/2$ .

**Example 4.27.** The word  $\alpha\beta^R$ , where  $\alpha, \beta$  are the words from (2) of Example 4.13, satisfies the condition  $\sigma_d^{(1)}(\alpha\beta^R) = (|\alpha\beta^R| + 1)|\alpha\beta^R|_d/2$  for  $d \in \{a, b, c\}$ . However,  $|\alpha\beta^R|_{ab} = 81 \neq 79 = |\alpha\beta^R|_{ba}$ . The second assertion from (2) Proposition 4.26 cannot be generalized for  $|\Sigma| \geq 3$ .

## REFERENCES

- [1] J.-P. Allouche and J.O. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, in *Sequences and Their Applications, Proceedings of SETA '98*, edited by C. Ding, T. Helleseht and H. Niederreiter. Springer-Verlag (1999) 1–16.
- [2] A. Atanasiu, Binary amiable words. *Int. J. Found. Comput. Sci.* **18** (2007) 387–400.
- [3] A. Atanasiu, R. Atanasiu and I. Petre, Parikh matrices and amiable words. *Theoret. Comput. Sci.* **390** (2008) 102–109.
- [4] A. Atanasiu, C. Martín-Vide and A. Mateescu, On the injectivity of the Parikh matrix mapping. *Fund. Inform.* **49** (2002) 289–299.
- [5] J. Berstel and D. Perrin, The origins of combinatorics on words. *Eur. J. Combin.* **28** (2007) 996–1022.
- [6] G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra*. MacMillan, New York, 4th edn. (1977).
- [7] P. Borwein and C. Ingalls, The Prouhet-Tarry-Escott problem revisited. *Enseign. Math.* **40** (1994) 3–27.
- [8] A. Černý, On fairness of D0L systems. *Discrete Appl. Math.* **155** (2007) 1769–1773.
- [9] A. Černý, On subword symmetry of words. *Int. J. Found. Comput. Sci.* **19** (2008) 243–250.
- [10] A. Černý, On fair words. *J. Autom. Lang. Comb.* **14** (2009) (to appear).
- [11] Ö. Eğecioğlu and O.H. Ibarra, A matrix q-analogue of the Parikh mapping, in *IFIP TCS*, edited by J.-J. Lévy, E.W. Mayr and J.C. Mitchell. Kluwer (2004) 125–138.
- [12] S. Fossé and G. Richomme, Some characterizations of Parikh matrix equivalent binary words. *Inform. Process. Lett.* **92** (2004) 77–82.
- [13] M. Lothaire, *Combinatorics on words*. Cambridge University Press (1997).
- [14] A. Mateescu, Algebraic aspects of Parikh matrices, in *Theory Is Forever*, edited by J. Karhumäki, H.A. Maurer, G. Păun and G. Rozenberg. *Lect. Notes Comput. Sci.* **3113** (2004) 170–180.
- [15] A. Mateescu and A. Salomaa, Matrix indicators for subword occurrences and ambiguity. *Int. J. Found. Comput. Sci.* **15** (2004) 277–292.
- [16] A. Mateescu, A. Salomaa, K. Salomaa and S. Yu, A sharpening of the Parikh mapping. *RAIRO-Theor. Inf. Appl.* **35** (2001) 551–564.
- [17] A. Mateescu, A. Salomaa and Sheng Yu, Subword histories and Parikh matrices. *J. Comput. System Sci.* **68** (2004) 1–21.
- [18] R.J. Parikh, On context-free languages. *J. ACM* **13** (1966) 570–581.
- [19] E. Prouhet, Mémoire sur quelques relations entre les puissances des nombres. *C.R. Acad. Sci. Paris* **33** (1851) 255.
- [20] A. Salomaa, Independence of certain quantities indicating subword occurrences. *Theoret. Comput. Sci.* **362** (2006) 222–231.
- [21] A. Salomaa, Comparing subword occurrences in binary D0L sequences. *Int. J. Found. Comput. Sci.* **18** (2007) 1395–1406.
- [22] A Salomaa, Subword balance in binary words, languages and sequences. *Fund. Inform.* **75** (2007) 469–482.
- [23] T.-F. Şerbănuţă, Extending Parikh matrices. *Theoret. Comput. Sci.* **310** (2004) 233–246.
- [24] Wikipedia. Rings. [http://en.wikipedia.org/wiki/Ring\\_\(mathematics\)](http://en.wikipedia.org/wiki/Ring_(mathematics)).

Communicated by P. Weil.

Received April 9, 2009. Accepted October 12, 2009.