Morphically Primitive Words ⋆

Daniel Reidenbach [a], Johannes C. Schneider [b,*]

[a] *Department of Computer Science, Loughborough University, Loughborough, LE11 3TU, United Kingdom*

[b] *Fachbereich Informatik, Technische Universität Kaiserslautern, Postfach 3049, 67653 Kaiserslautern, Germany*

Abstract

In the present paper, we introduce an alternative notion of the primitivity of words, that – unlike the standard understanding of this term – is not based on the power (and, hence, the concatenation) of words, but on morphisms. For any alphabet Σ, we call a word $w \in \Sigma^*$ *morphically imprimitive* provided that there are a shorter word v and morphisms $h, h' : \Sigma^* \to \Sigma^*$ satisfying $h(v) = w$ and $h'(w) = v$, and we say that w is *morphically primitive* otherwise. We explain why this is a well-chosen terminology, we demonstrate that morphic (im-)primitivity of words is a vital attribute in many combinatorial domains based on finite words and morphisms, and we study a number of fundamental properties of the concepts under consideration.

Key words: Combinatorics on Words, Primitivity, Morphisms, Fixed Points

1 Introduction

The definition of primitive words – i. e. of those words that are not a non-trivial power of another word – is well-established in combinatorics on words (cf. Lothaire [10]), and numerous elementary properties of words are based on this concept.

In the present paper, we wish to introduce another type of words that may be considered “primitive”: the morphically primitive words (over some alphabet

⋆ A preliminary version of this paper was presented at the conference WORDS 2007 (cf. [15]).

* Corresponding author.
Email addresses: D.Reidenbach@lboro.ac.uk (Daniel Reidenbach), jschneider@informatik.uni-kl.de (Johannes C. Schneider).

Σ). We designate a word v as *morphically primitive* if, for every word w with $|w| < |v|$, there do not exist morphisms $h, h' : \Sigma^* \to \Sigma^*$ satisfying $h(v) = w$ and $h'(w) = v$, and we call v *morphically imprimitive* if it is not morphically primitive. Since the properties of this concept are equivalent for all alphabets Σ, we assume Σ to be infinite; consequently, for the sake of convenience, we can choose $\Sigma := \mathbb{N}$. For instance, according to these basic definitions, the word $w = 1 \cdot 2 \cdot 2$ (where the symbol $\cdot$ refers to the concatenation and is used to avoid any confusion of "symbols" in $\mathbb{N}$) is morphically imprimitive, since the word $v := 1$, the morphism $h : \mathbb{N}^* \to \mathbb{N}^*$, given by $h(1) := 1 \cdot 2 \cdot 2$, and the morphism $h' : \mathbb{N}^* \to \mathbb{N}^*$, given by $h'(1) := 1$, $h'(2) := \varepsilon$ (where ε stands for the empty word), satisfy $|v| < |w|$, $h(v) = w$ and $h'(w) = v$. The word v, in turn, is obviously morphically primitive. We now consider a second and slightly more complex example, namely $w := 1 \cdot 2 \cdot 3 \cdot 3 \cdot 1 \cdot 2$. This word is morphically imprimitive, since $v := 1 \cdot 3 \cdot 3 \cdot 1$, $h : \mathbb{N}^* \to \mathbb{N}^*$, given by $h(1) := 1 \cdot 2$ and $h(3) := 3$, and $h' : \mathbb{N}^* \to \mathbb{N}^*$, given by $h'(1) := 1$, $h'(2) := \varepsilon$ and $h(3) := 3$, satisfy $|v| < |w|$, $h(v) = w$ and $h'(w) = v$. Just as in our first example, v is morphically primitive. The verification of this claim, however, is less straightforward; it is facilitated by some tools provided in the technical part of this paper.

In the subsequent sections, we show that this definition establishes a combinatorially rich theory, which, in particular, meets the common perception of "primitive" objects in mathematics and contains a number of very challenging basic problems. In addition to this, we point out that the concept of morphic primitivity is closely connected to several other topics in formal language theory and combinatorics on words, such as finite *fixed points* of morphisms (cf., e. g., Hamm and Shallit [6]), *pattern languages* (cf., e. g., Mateescu and Salomaa [11]) and the *ambiguity of morphisms* (cf. Freydenberger, Reidenbach and Schneider [5]). Thus, our approach does not only deal with a topic of intrinsic interest, but it also contributes nontrivial insights that are relevant for other active areas of research.

2 Basic Definitions

In the present section we establish some basic definitions and notations. For terms not defined explicitly, we refer the reader to Lothaire [10] and Rozenberg and Salomaa [17].

Let $\mathbb{N} := \{1, 2, \ldots\}$ be the set of natural numbers. An *alphabet* $\mathcal{A}$ is an enumerable set of symbols. In the subsequent sections, we largely use $\mathbb{N}$ as an infinite alphabet (see Section 1). A *word (over $\mathcal{A}$)* is a finite sequence of symbols taken from $\mathcal{A}$. By $|X|$ we denote the cardinality of a set X or the length of a word X. The *empty* word is the unique sequence of symbols of length 0;

we use ε for the empty word. For the *concatenation* of words v, w we write $v \cdot w$ (or vw for short). The notation $\mathcal{A}^*$ refers to the set of all words over $\mathcal{A}$, i. e., more precisely, the *free monoid* generated by $\mathcal{A}$; furthermore, $\mathcal{A}^+ := \mathcal{A}^* \setminus \{\varepsilon\}$. For any $n \in \mathbb{N}$, we define $\mathcal{A}^n := \{w \in \mathcal{A}^+ \mid |w| = n\}$. The number of occurrences of a symbol $x \in \mathcal{A}$ in a word $w \in \mathcal{A}^*$ is written as $|w|_x$. The term $\mathrm{symb}(w)$ stands for the set of symbols occurring in w; thus, e. g., for the word $w := 2 \cdot 5 \cdot 24 \cdot 24 \cdot 5 \in \mathbb{N}^*$ it is $\mathrm{symb}(w) = \{2, 5, 24\}$. Given $w \in \mathbb{N}^*$, we denote the minimal number of occurrences of a symbol in w by $\min_\#(w)$, i. e. $\min_\#(w) := \min\{|w|_x \mid x \in \mathrm{symb}(w)\}$. Thus, concerning our above example $w = 2 \cdot 5 \cdot 24 \cdot 24 \cdot 5$, we have $\min_\#(w) = |w|_2 = 1$. Moreover, we extend the operations *, $^+$ and the concatenation to sets of words in the usual manner; with regard to alphabets $\mathcal{A}$, $\mathcal{B}$, this means that, e. g., $(\mathcal{A}^*\mathcal{B})^+ = \{w \mid w = v_1 \cdot b_1 \cdot v_2 \cdot b_2 \cdot \ldots \cdot v_n \cdot b_n$ with $n \in \mathbb{N}$, $v_1, v_2, \ldots, v_n \in \mathcal{A}^*$, $b_1, b_2, \ldots, b_n \in \mathcal{B}\}$.

Within the scope of this paper, a *morphism* is a mapping $h : \mathcal{A}^* \to \mathcal{A}^*$ that is compatible with the concatenation, i. e., for all $v, w \in \mathcal{A}^*$, $h(vw) = h(v)h(w)$. Hence, h is fully defined for all $v \in \mathcal{A}^*$ as soon as it is defined for all symbols in $\mathcal{A}$. A morphism h is called *injective* if and only if, for all $v, w \in \mathcal{A}^*$, $h(v) = h(w)$ implies $v = w$. For any $v \in \mathcal{A}^*$, a morphism h is said to be *ambiguous (with respect to v)* if and only if there exists a morphism h' satisfying $h'(v) = h(v)$ and, for an $x \in \mathrm{symb}(v)$, $h'(x) \neq h(x)$. If h is not ambiguous with respect to v, it is called *unambiguous* with respect to v.

A morphism $r : \mathcal{A}^* \to \mathcal{A}^*$ is a *renaming* if and only if r is injective and, for every $x \in \mathcal{A}$, $|r(x)| = 1$. Given $v \in \mathbb{N}^*$, a morphism $r : \mathcal{A}^* \to \mathcal{A}^*$ is said to be a renaming *of* v if and only if r is injective on $\mathrm{symb}(v)$ and, for every $x \in \mathrm{symb}(v)$, $|r(x)| = 1$. Finally, for any words $v, w \in \mathcal{A}^*$, we call w a renaming of v provided that there is a renaming $r : \mathcal{A}^* \to \mathcal{A}^*$ mapping v onto w. Since we largely consider words as preimages of morphisms, most basic properties of words v, w to be studied in the subsequent sections are equivalent if w is a renaming of v. If we want to address one particular word among all renamings of a word w, we can choose the lexicographically minimal word, which is then called the *canonical form* [1] of w. For instance, $v := 1 \cdot 2 \cdot 2 \cdot 1 \cdot 3$ is the canonical form of the words $2 \cdot 3 \cdot 3 \cdot 2 \cdot 1$, $5 \cdot 21 \cdot 21 \cdot 5 \cdot 14$ and $1 \cdot 3 \cdot 3 \cdot 1 \cdot 4$ etc. Note that, throughout this paper, most examples are given in canonical form.

3 Morphic Primitivity

In the present section we introduce our notion of morphic primitivity, we describe basic properties that justify our terminology, and we point out similar

[1] This term is derived from the research on pattern languages, see, e. g., Angluin [1].

concepts in literature.

The definition of morphic primitivity is based on a morphic relation between words which is given as follows:

Definition 1 *Let $v, w \in \mathbb{N}^*$. We call v and w* morphically coincident *if and only if there exist morphisms $h, h' : \mathbb{N}^* \to \mathbb{N}^*$ such that $h(v) = w$ and $h'(w) = v$. If v and w are morphically coincident then we write $v \equiv_* w$ for short.*

For instance, the words $w_1 := 1 \cdot 2 \cdot 2 \cdot 3 \cdot 3 \cdot 1 \cdot 2 \cdot 2$ and $w_2 := 1 \cdot 1 \cdot 2 \cdot 3 \cdot 3 \cdot 1 \cdot 1 \cdot 2$ are morphically coincident since there exist morphisms $h, h' : \mathbb{N}^* \to \mathbb{N}^*$ – given by $h(1) := 1 \cdot 1 \cdot 2$, $h(2) := \varepsilon$, $h(3) := 3$ and $h'(1) := \varepsilon$, $h'(2) = 1 \cdot 2 \cdot 2$, $h'(3) := 3$ – satisfying $h(w_1) = w_2$ and $h'(w_2) = w_1$. Contrary to this, if we additionally consider the word $w_3 = 1 \cdot 2 \cdot 2 \cdot 3 \cdot 1 \cdot 2 \cdot 2$ then we can make use of the following necessary condition for morphic coincidence, which can be verified by straightforward considerations:

Proposition 2 *Let $v, w \in \mathbb{N}^*$. If $v \equiv_* w$ then $\min_\#(v) = \min_\#(w)$.*

Consequently, referring to our above example, $\min_\#(w_1) = \min_\#(w_2) = 2$ (due to the fact that the symbols 1 and 3 have two occurrences in w_1 and the symbols 2 and 3 have two occurrences in w_2), whereas $\min_\#(w_3) = 1$ (because of the single occurrence of the symbol 3 in w_3). Thus, neither w_1 and w_3 nor w_2 and w_3 can be morphically coincident.

If follows by definition that $\equiv_*$ describes an equivalence relation on $\mathbb{N}^*$. Given any word v, we can consider the corresponding equivalence class of v, i. e. all words w satisfying $v \equiv_* w$, and we designate the shortest words in this equivalence class as morphically primitive:

Definition 3 *Let $v \in \mathbb{N}^*$. We call v* morphically primitive *if and only if there exists no $v' \in \mathbb{N}^*$ such that $v' \equiv_* v$ and $|v'| < |v|$. If v is not morphically primitive, it is called* morphically imprimitive.

For any $w \in \mathbb{N}^*$, we call a morphically primitive $v \in \mathbb{N}^*$ satisfying $v \equiv_* w$ a morphic root *(of w).*

For instance, with regard to the example words $w_4 := 1 \cdot 2 \cdot 1 \cdot 2 \cdot 3 \cdot 3$ and $v_1 := 1 \cdot 1 \cdot 2 \cdot 2$, it can be easily seen that $v_1 \equiv_* w_4$, so that w_4 is morphically imprimitive. Due to Proposition 2, we can also verify with little effort that there is no word v' with $|v'| < |v_1|$ and $v' \equiv_* v_1$. Consequently, v_1 is morphically primitive and, hence, a morphic root of w_4. Furthermore, Proposition 2 and Definition 3 imply that a word v with $\min_\#(v) = 1$ is morphically primitive if and only if $|v| = 1$. Therefore, our above example $w_3 = 1 \cdot 2 \cdot 2 \cdot 3 \cdot 1 \cdot 2 \cdot 2$ is morphically imprimitive and, e. g., $v_2 := 1$ is a morphic root of w_3.

Before we examine some important basic properties of morphically primitive words, we note a small, but vital technical lemma:

Lemma 4 *Let $v \in \mathbb{N}^*$ be morphically primitive and let $h : \mathbb{N}^* \to \mathbb{N}^*$ be a morphism such that $h(v) \equiv_* v$. Then, for every $x \in \mathrm{symb}(v)$, $h(x) \neq \varepsilon$.*

PROOF. Assume to the contrary that there exists an $x \in \mathrm{symb}(v)$ with $h(x) = \varepsilon$. We define the morphism $e : \mathbb{N}^* \to \mathbb{N}^*$ by, for every $y \in \mathbb{N}$,

$$e(y) := \begin{cases} \varepsilon, & y = x, \\ y, & \text{else.} \end{cases}$$

Since $h(v) \equiv_* v$, there is a morphism g with $g(h(v)) = v$. Furthermore, the definition of e and h implies $g(h(e(v))) = v$ or, in other words, $v \equiv_* e(v)$. As $|e(v)| < |v|$, this contradicts the morphic primitivity of v. Consequently, there is no $x \in \mathrm{symb}(v)$ satisfying $h(x) = \varepsilon$, which proves the lemma. □

Note that a number of similar lemmata is given by Reidenbach [14].

The subsequent first remark on morphic primitivity states that our approach strongly differs from the well-established understanding of primitivity of words (as briefly mentioned in Section 1 and extensively described by Lothaire [10]):

Proposition 5 *The set of primitive words over $\mathbb{N}$ and the set of morphically primitive words over $\mathbb{N}$ are incomparable.*

PROOF. It can be easily verified that, e. g., $1 \cdot 2$ is primitive, but not morphically primitive, and $1 \cdot 1 \cdot 2 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 2$ is morphically primitive, but not primitive. □

Clearly, this implies that the set of imprimitive words and that of morphically imprimitive words are incomparable, too.

In various fields of mathematics, it is an intrinsic property of "prime" or "primitive" objects to be indecomposable with respect to the relevant operation under consideration (e. g. multiplication for prime numbers), i. e. these elements have no decomposition into further, smaller elements. In our case, however, we have to state that morphically primitive words are not indecomposable in terms of the concatenation:

Proposition 6 *There are morphically imprimitive words $v, w \in \mathbb{N}^*$ such that vw is morphically primitive.*

PROOF. Let $v := 1{\cdot}2{\cdot}2$, $w := 1{\cdot}2$. These words are morphically imprimitive, and $vw = 1 \cdot 2 \cdot 2 \cdot 1 \cdot 2$ is morphically primitive. $\square$

In spite of Proposition 5 and Proposition 6, we consider our terminology well-chosen since there exist two basic facts demonstrating that Definition 3 introduces a type of words that shows the usual traits of a "primitive" object. Both of these facts are concerned with the question to which extent (and how) the morphically primitive words can span a free monoid. We discuss this topic by means of the following definition:

Definition 7 *Let $v \in \mathbb{N}^*$ be morphically primitive. Then we call the set* $\mathrm{IMPRIM}(v) := \{w \in \mathbb{N}^* \mid |w| > |v| \text{ and } w \equiv_* v\}$ *the* imprimitive hull *(of v).*

Referring to this term, our first observation on the relation between a word monoid and its morphically primitive words states that we evidently can generate *all* words by applying suitable morphisms to morphically primitive words:

Proposition 8 *Let* PRIM *be the set of all morphically primitive words over $\mathbb{N}$. Then $\bigcup_{v \in \mathrm{PRIM}}(v \cup \mathrm{IMPRIM}(v)) = \mathbb{N}^*$.*

PROOF. Directly from Definition 3. $\square$

Our second (and less obvious) observation implies that, on the other hand, the imprimitive hulls of morphically primitive words v, v' can only have a common element provided that v' is a renaming of v:

Proposition 9 *Let $v, v' \in \mathbb{N}^*$ be morphically primitive words. Then* $\mathrm{IMPRIM}(v) \cap \mathrm{IMPRIM}(v') = \emptyset$ *if and only if v' is not a renaming of v.*

PROOF. We first prove the *if* part (and we do so by contraposition). If there exist a $w \in \mathrm{IMPRIM}(v) \cap \mathrm{IMPRIM}(v')$ then $v \equiv_* w \equiv_* v'$. Consequently, we can conclude that there exist morphisms h, h' satisfying $h(v) = v'$ and $h'(v') = v$. As v and v' are morphically primitive, we can conclude from Lemma 4 that, for every $x \in \mathrm{symb}(v)$, $h(x) \neq \varepsilon$ and, for every $x' \in \mathrm{symb}(v')$, $h'(x') \neq \varepsilon$. Thus, h is a renaming of v (and h' is a renaming of v', of course).

We now consider the *only if* part (again by contrapositon). To this end, let v' be a renaming of v. Then for every $w \in \mathrm{IMPRIM}(v)$, it is $w \equiv_* v \equiv_* v'$ and, for every $w' \in \mathrm{IMPRIM}(v')$, we have $w' \equiv_* v' \equiv_* v$. Thus, $\mathrm{IMPRIM}(v) = \mathrm{IMPRIM}(v')$, which trivially implies that $\mathrm{IMPRIM}(v) \cap \mathrm{IMPRIM}(v') \neq \emptyset$. $\square$

Consequently, for any set PRIMBASE $\subset \mathbb{N}^*$ of morphically primitive words, the words in PRIMBASE can only generate the full free monoid $\mathbb{N}^*$ by morphisms ensuring morphic coincidence if, for every morphically primitive word v, PRIMBASE contains at least one renaming of v. We therefore feel that Propositions 8 and 9 back our terminology.

Moreover, we can directly conclude from Proposition 9 that a morphic root of a word is unique (up to renaming), which is a desirable property of primitive objects in various theories:

Corollary 10 *Let $w \in \mathbb{N}^*$. If v, v' are primitive roots of w then v' is a renaming of v.*

PROOF. Directly from Proposition 9. $\square$

We now demonstrate that morphic primitivity of words is not only a natural concept (as explained by Propositions 8 and 9), but also an important one:

Theorem 11 *Let $v \in \mathbb{N}^*$. The following statements are equivalent:*

(1) v is morphically primitive.
(2) v is not a fixed point [2] *of a nontrivial morphism $h : \mathbb{N}^* \to \mathbb{N}^*$.*
(3) v is a succinct pattern [3].
(4) There is an unambiguous injective morphism $\sigma : \mathbb{N}^* \to \{\mathtt{a}, \mathtt{b}\}^*$ with respect to v.

PROOF. We first prove that statement 1 implies statement 2 (and we do so by contraposition). Hence, let v be a fixed point of a nontrivial morphism h, i. e. $h(v) = v$ and, for some $x \in \text{symb}(v)$, $h(x) \neq x$. It can be straightforward verified these properties of h entail the existence of a variable $y \in \text{symb}(v)$ with $h(y) = \varepsilon$. Thus, by Lemma 4 (which can be applied since $h(v) = v$ evidently implies $h(v) \equiv_* v$), v is morphically imprimitive.

[2] A word $w \in \mathbb{N}^*$ is said to be a *fixed point of (a nontrivial morphism) $h : \mathbb{N}^* \to \mathbb{N}^*$* provided that $h(w) = w$ and, for some $x \in \text{symb}(w)$, $h(x) \neq x$. Further information on fixed points of morphisms is provided by, e. g., Hamm and Shallit [6], Levé and Richomme [9].

[3] The *E-pattern language* $L(\alpha)$ of a word $\alpha \in \mathbb{N}^+$ – which in this context is not called a "word", but a *(terminal-free) pattern* – is the set of all morphic images of α in an arbitrarily chosen free monoid. The pattern α is said to be *succinct* if and only if it is a shortest generator of its E-pattern language, i. e., for every $\beta \in \mathbb{N}^*$, if $L(\beta) = L(\alpha)$ then $|\beta| \geq |\alpha|$. A survey on pattern languages is given by, e. g., Mateescu and Salomaa [11], recent results are presented by Reidenbach [14].

We now show that statement 2 implies statement 1: If v is not a fixed point of a nontrivial morphism then, for every morphism $h : \mathbb{N}^* \to \mathbb{N}^*$ with $h(v) = v$ and for every $x \in \mathrm{symb}(v)$, $h(x) = x$. Consequently, for any two morphisms g, g' with $g'(g(v)) = v$ and for every $x \in \mathrm{symb}(x)$, $g(x) \neq \varepsilon$. Thus, for every morphism g satisfying $g(v) \equiv_* v$, $|g(v)| \geq |v|$. By definition, this means that v is morphically primitive.

The equivalence of the statements 2, 3 and 4 is explained by Freydenberger, Reidenbach and Schneider [5]. □

Consequently, due to the fundamental equivalences noted in Theorem 11, we can benefit from well-known results in literature when analysing morphic primitivity (and we mainly do so in Section 5). This particularly holds for the analogies between our approach and the field of fixed points of morphisms, since any word v which, for some word v' and morphisms h, h', satisfies $h(v) = v'$ and $h'(v') = v$ is just a fixed point of the morphism $h' \circ h$. Our point of view, in turn, allows to precisely address the so far hardly examined relation between a morphic root and its imprimitive hull. In the subsequent sections, we study this topic in more detail.

4 The Imprimitive Hull

In the present section, we wish to gain a deeper understanding of those morphisms that are used for generating the imprimitive hull of any given morphically primitive word. We designate such morphisms as follows:

Definition 12 *Let $v \in \mathbb{N}^*$ be morphically primitive. Then a morphism $h : \mathbb{N}^* \to \mathbb{N}^*$ is called an* imprimitivity morphism (for v) *provided that $|h(v)| > |v|$ and $h(v) \equiv_* v$.*

It follows by definition that, for every morphically primitive word v and every imprimitivity morphism h for v, $h(v)$ is a morphically imprimitive word and v is a morphic root of $h(v)$. Furthermore, for every morphically imprimitive word w and every morphic root v of w, there exists an imprimitivity morphism h satisfying $h(v) = w$.

We now characterise the imprimitivity morphisms:

Theorem 13 *Let $v \in \mathbb{N}^*$ be morphically primitive. Then a morphism $h : \mathbb{N}^* \to \mathbb{N}^*$ is an imprimitivity morphism for v if and only if*

(i) *for every $x \in \mathrm{symb}(v)$ there exists an $x_h \in \mathrm{symb}(h(x))$ such that $|h(x)|_{x_h} = 1$ and $|h(y)|_{x_h} = 0$, $y \in \mathrm{symb}(v) \setminus \{x\}$, and*

(ii) *there exists an* $x \in \mathrm{symb}(v)$ *with* $|h(x)| \geq 2$.

PROOF. We first prove the *if* direction. Hence, let h be a morphism satisfying the conditions (i) and (ii). We define the morphism $g : \mathbb{N}^* \to \mathbb{N}^*$ by, for every $y \in \mathbb{N}$,

$$g(y) := \begin{cases} x, & y = x_h, \\ \varepsilon, & \text{else.} \end{cases}$$

Then conditions (i) and (ii) of Theorem 13 imply $|h(v)| > |v|$ and, because of condition (i), it is $g(h(v)) = v$. Thus $h(v) \equiv_* v$.

Conversely, let h be a morphism that does not satisfy condition (i) or that does not satisfy condition (ii). If h does not satisfy condition (ii) then $|v| \geq |h(v)|$ and therefore h is not an imprimitivity morphism. If h does not satisfy condition (i) then there exists an $x \in \mathrm{symb}(v)$ such that, for every $y \in \mathrm{symb}(h(x))$, $|h(v)|_y \neq |v|_x$ or, more precisely, $|h(v)|_y > |v|_x$. Now assume to the contrary that there exists a morphism g with $g(h(v)) = v$. Then necessarily $g(h(x)) \neq x$. It can be straightforward verified that this implies the existence of a symbol $x' \in \mathrm{symb}(v)$ with $g \circ h(x') = \varepsilon$; furthermore, the assumption $g \circ h(v) = v$ by definition implies $g \circ h(v) \equiv_* v$. Due to Lemma 4, this contradicts the morphic primitivity of v. Consequently, there is no morphism g satisfying $g(h(v)) = v$. Thus, v and $h(v)$ are not morphically coincident, and this implies that h is no imprimitivity morphism for v. □

Theorem 13 has an immediate consequence on the relation between any word and its morphic roots:

Corollary 14 *Let* $v, w \in \mathbb{N}^*$ *such that* v *is a morphic root of* w*. Then* $|\mathrm{symb}(v)| \leq |\mathrm{symb}(w)|$*. If* w *is morphically imprimitive then* $|\mathrm{symb}(v)| < |\mathrm{symb}(w)|$.

PROOF. Directly from Theorem 13. □

Corollary 14 explains why we consider it more appropriate to regard an infinite alphabet for our studies: If we restrict ourselves to words over a finite alphabet $\Sigma \subset \mathbb{N}$ then, for any morphically primitive word $v \in \Sigma^*$ with $\mathrm{symb}(v) = |\Sigma|$, the set $\mathrm{IMPRIM}(v) \cap \Sigma^*$ is empty. Hence, although v is morphically primitive, it is not a morphic root of any morphically imprimitive word in Σ^*. We feel that this phenomenon does not completely meet our notion of morphic primitivity; nevertheless, all results of this work also hold for finite alphabets or can be adapted to this case with little effort.

The following example demonstrates the effect of imprimitivity morphisms:

Example 15 *Let* $v := 1 \cdot 2 \cdot 3 \cdot 4 \cdot 1 \cdot 4 \cdot 3 \cdot 5 \cdot 5 \cdot 2$*, and let the morphism* h *be given by* $h(1) := 1 \cdot 6$*,* $h(2) := 2$*,* $h(3) := 3 \cdot 6$*,* $h(4) := 4$*,* $h(5) := 5 \cdot 6$*. Due to the length and complexity of* v*, our preliminary insights (such as Proposition 2 and Corollary 14) presented so far do not allow to easily discuss the question of whether* v *is morphically primitive. Therefore we state without proof that* v *is morphically primitive, and, for a verification of this claim, we refer to Section 5 (and, in particular, Corollary 19), which provides some appropriate tools. Furthermore, we can verify the conditions of Theorem 13 by* $x_h := x$ *for every* $x \in \mathrm{symb}(v)$ *and* $|h(1)| \geq 2$*. Hence,* h *is an imprimitivity morphism for* v *and leads to* $h(v) = 1 \cdot 6 \cdot 2 \cdot 3 \cdot 6 \cdot 4 \cdot 1 \cdot 6 \cdot 4 \cdot 3 \cdot 6 \cdot 5 \cdot 6 \cdot 5 \cdot 6 \cdot 2$*. Additionally, we see* $|\mathrm{symb}(v)| < |\mathrm{symb}(h(v))|$ *as stated by Corollary 14.*

Note that another imprimitivity morphism h', given by $h'(1) := 1$, $h'(2) := 6 \cdot 2$, $h'(3) := 3$, $h'(4) := 6 \cdot 4$, $h'(5) := 6 \cdot 5$, leads to $h'(v) = h(v)$. Hence, imprimitivity morphisms can be ambiguous with respect to "their" morphic root. This is caused by a certain structure of the example word v which can be generalised as follows:

Definition 16 *We say that a word* $w \in \mathbb{N}^*$ *has an* $SCRN$-factorisation *if and only if there exist pairwise disjoint sets* $S, C, R, N \subseteq \mathrm{symb}(w)$ *such that* $w \in (N^* S C^* R)^+ N^*$*.*

Definition 16 is derived from the research on the ambiguity of morphisms (cf. Freydenberger and Reidenbach [4]), where it is of major importance. Concerning the word v given in Example 15, we have $S = \{1, 3\}$, $C = \{5\}$, $R = \{2, 4\}$ and $N = \emptyset$.

We now wish to demonstrate that this structure characterises the ambiguity of imprimitivity morphisms:

Theorem 17 *Let* $v \in \mathbb{N}^*$ *be morphically primitive. There exist imprimitivity morphisms* $h, h' : \mathbb{N}^* \to \mathbb{N}^*$ *for* v *such that*

(i) $h(v) = h'(v)$ *and,*
(ii) *for an* $x \in \mathrm{symb}(v)$*,* $h(x) \neq h'(x)$

if and only if v *has an* $SCRN$*-factorisation.*

PROOF. Assume that v has an $SCRN$-factorisation. Let $s \in \mathbb{N} \setminus \mathrm{symb}(v)$. It can be easily verified that the imprimitivity morphisms $h : \mathbb{N}^* \to \mathbb{N}^*$,

$$h(x) := \begin{cases} x \cdot s, & x \in S \cup C, \\ x, & x \in R \cup N, \end{cases}$$

$x \in \mathbb{N}$, and $h' : \mathbb{N}^* \to \mathbb{N}^*$,

$$h'(x) := \begin{cases} s \cdot x, & x \in C \cup R, \\ x, & x \in S \cup N, \end{cases}$$

$x \in \mathbb{N}$, satisfy $h(v) = h'(v)$ and $h(x) \neq h'(x)$ for any symbol x contained in one of the (by definition nonempty) sets S or R.

We now prove the *only if* part. Hence, let h, h' be imprimitivity morphisms for v satisfying the conditions (i) and (ii). Due to Theorem 13, we know that, for every $x \in \mathrm{symb}(v)$, there exist an $x_h \in \mathbb{N}$ and $u, w \in (\mathbb{N} \setminus \{x_h\})^*$ such that $h(x) = u\, x_h\, w$. Our subsequent argumentation is based on the fact that, additionally, x_h also occurs in $h'(x)$ exactly once:

Claim. For every $x \in \mathrm{symb}(v)$, there exist $u', w' \in (\mathbb{N} \setminus \{x_h\})^*$ such that $h'(x) = u'\, x_h\, w'$.

Proof (Claim). Evidently, for every $x \in \mathrm{symb}(v)$, $|h'(x)|_{x_h} \leq 1$, since otherwise $|h'(v)|_{x_h} > |v|_x = |h(v)|_{x_h}$, which contradicts the condition $h'(v) = h(v)$. We now show that $|h'(x)|_{x_h} \neq 0$. Assume to the contrary that there is an $x \in \mathrm{symb}(v)$ such that $x_h \notin \mathrm{symb}(h'(x))$. It can be verified by straightforward combinatorial considerations that this implies the existence of a variable $x' \in \mathrm{symb}(v)$ such that, for *every* $y \in \mathrm{symb}(v)$, $y_h \notin \mathrm{symb}(h'(x'))$. We now consider the morphism g as introduced in the proof of Theorem 13, and we define a second morphism $e : \mathbb{N}^* \to \mathbb{N}^*$ by, for every $y \in \mathbb{N}$,

$$e(y) := \begin{cases} \varepsilon, & y = x', \\ y, & \text{else.} \end{cases}$$

Then, for all $y \in \mathrm{symb}(v)$, $h'(e(v))$ still contains the symbols y_h in the same order and number as specified by $h'(v)$ (which equals $h(v)$). Therefore, the fact $g(h(v)) = v$ implies that $g(h'(e(v))) = v$ and, hence, $g \circ h' \circ e(v) \equiv_* v$. On the other hand, it is $g \circ h' \circ e(x') = \varepsilon$, which, by Lemma 4, contradicts the morphic primitivity of v. Consequently, for each $x \in \mathrm{symb}(v)$, it is $x_h \in \mathrm{symb}(h'(x))$. This proves the claim. $\square$ (Claim)

Thus, for every $x \in \mathrm{symb}(v)$ there exist $u, w, u', w' \in (\mathbb{N} \setminus \{x_h\})^*$ with $h(x) = u\, x_h\, w$ and $h'(x) = u'\, x_h\, w'$. Hence, we can define S, C, R, N by considering four different cases:

$$\begin{aligned}
S &:= \{x \in \mathrm{symb}(v) \mid \exists u, w, u', w' \in \mathbb{N}^* : h(x) = u x_h w, h'(x) = u' x_h w', u = u', w \neq w'\},\\
C &:= \{x \in \mathrm{symb}(v) \mid \exists u, w, u', w' \in \mathbb{N}^* : h(x) = u x_h w, h'(x) = u' x_h w', u \neq u', w \neq w'\},\\
R &:= \{x \in \mathrm{symb}(v) \mid \exists u, w, u', w' \in \mathbb{N}^* : h(x) = u x_h w, h'(x) = u' x_h w', u \neq u', w = w'\},\\
N &:= \{x \in \mathrm{symb}(v) \mid \exists u, w, u', w' \in \mathbb{N}^* : h(x) = u x_h w, h'(x) = u' x_h w', u = u', w = w'\}.
\end{aligned}$$

Obviously, $S \cup C \cup R \cup N = \mathrm{symb}(v)$ and $S \cap C \cap R \cap N = \emptyset$. We verify that the definition of S, C, R, N implies that v has an $SCRN$-factorisation by examining v from left to right. Due to condition (ii), there must be a leftmost symbol x such that $h(x) \neq h'(x)$. All symbols to the left of this symbol belong to N. Clearly, $h(x)$ and $h'(x)$ can only differ on symbols to the right of the occurrence of x_h in $h(x)$ and $h'(x)$ since otherwise $h(v)$ would not equal $h'(v)$. Thus, x belongs to S. To satisfy condition (i), symbols from C may follow, but necessarily, at some point, followed by a symbol from R. Consequently, the so far considered prefix v_l of v has the shape $v_l \in N^*SC^*R$ and satisfies $h(v_l) = h'(v_l)$. We can continue with the same argumentation and, thus, receive $v \in (N^*SC^*R)^+N^*$. Consequently, v has an $SCRN$-factorisation. $\square$

This result concludes our examination of basic properties of those morphisms that map a given morphically primitive word onto a morphically imprimitive word. In the subsequent section, we mainly turn our attention to the relation between a given morphically imprimitive word and its morphic roots.

5 The Morphic Roots

In the present section, we primarily examine the combinatorial properties of morphisms mapping morphically imprimitive words onto morphically coincident shorter words. In addition to its intrinsic interest, this topic is evidently motivated by the elementary algorithmic problems of how we can decide on whether a given word $w \in \mathbb{N}^*$ is morphically primitive or morphically imprimitive and, in the latter case, of how we can find a morphic root of w. From an algorithmic point of view, we can immediately state that there exists a simple decision procedure for the morphic primitivity of words which, moreover, automatically leads to a procedure that computes a morphic root of a morphically imprimitive word w: we can simply test, for all words v with $|v| < |w|$ and $\mathrm{symb}(v) \subseteq \mathrm{symb}(w)$, whether $v \equiv_* w$. To this end, however, we have to check the existence of morphisms between words – which is an NP-complete problem (cf. Ehrenfeucht and Rozenberg [3]) – and we have to do so exponentially many times. Hence, such a procedure is extremely unsatisfactory, and therefore our subsequent results do not only contribute to the understanding of the relation between words and their morphic roots, but they also lay the foundations of more efficient respective algorithms.

Our considerations are largely based on the following factorisation of words, which, e. g., can be derived from the research on fixed points of morphisms (cf. Hamm and Shallit [6], Levé and Richomme [9]):

Definition 18 *Let* $w \in \mathbb{N}^*$*. An* imprimitivity factorisation (of w) *is a mapping* $f : \mathbb{N}^+ \to \mathbb{N}^n \times (\mathbb{N}^+)^n$*,* $n \in \mathbb{N}$*, such that, for* $f(w) = (x_1, x_2, \ldots, x_n; v_1, v_2, \ldots, v_n)$*, there exist* $u_0, u_1, \ldots, u_n \in \mathbb{N}^*$ *satisfying* $w = u_0 v_1 u_1 v_2 u_2 \ldots v_n u_n$ *and*

(i) *for every* $i \in \{1, 2, \ldots, n\}$*,* $|v_i| \geq 2$*,*
(ii) *for every* $i \in \{0, 1, \ldots, n\}$ *and for every* $j \in \{1, 2, \ldots, n\}$*,* $\mathrm{symb}(u_i) \cap \mathrm{symb}(v_j) = \emptyset$*,*
(iii) *for every* $i \in \{1, 2, \ldots, n\}$*,* $|v_i|_{x_i} = 1$ *and if* $x_i \in \mathrm{symb}(v_{i'})$*,* $i' \in \{1, 2, \ldots, n\}$*, then* $v_i = v_{i'}$ *and* $x_i = x_{i'}$*.*

It is a well-known fact that the existence of an imprimitivity factorisation characterises the fixed points of nontrivial morphisms (cf. Head [7]); furthermore, Reidenbach [14] introduces the equivalent characterisation for the *prolix* (i. e. non-succinct) patterns. Hence, referring to Theorem 11, we may immediately conclude that the existence of an imprimitivity factorisation also characterises the morphically imprimitive words:

Corollary 19 *A word* $w \in \mathbb{N}^*$ *is morphically primitive if and only if there exists no imprimitivity factorisation of* w*.*

PROOF. Directly from Head [7] and Theorem 11. □

The following example illuminates Definition 18 and Corollary 19:

Example 20 *Let* $w := 1 \cdot 2 \cdot 3 \cdot 3 \cdot 4 \cdot 3 \cdot 5 \cdot 5 \cdot 4 \cdot 3 \cdot 1 \cdot 2 \cdot 3 \cdot 3 \cdot 5$*. A possible imprimitivity factorisation of* w *is* $f(w) = (2, 4, 4, 2; 2 \cdot 3 \cdot 3, 4 \cdot 3, 4 \cdot 3, 2 \cdot 3 \cdot 3)$*, which can be illustrated as follows:*

$$w = 1 \cdot \underbrace{2 \cdot 3 \cdot 3}_{v_1} \cdot \underbrace{4 \cdot 3}_{v_2} \cdot 5 \cdot 5 \cdot \underbrace{4 \cdot 3}_{v_3} \cdot 1 \cdot \underbrace{2 \cdot 3 \cdot 3}_{v_4} \cdot 5.$$

Consequently, w *is morphically imprimitive.*

Thus, in order to see whether or not a word w is morphically primitive, we can just as well search for an imprimitivity factorisation of w. If such a factorisation exists, we do not only know that w is morphically imprimitive, but (as to be shown below) we also obtain a morphism that maps w onto a shorter word v satisfying $v \equiv_* w$. This morphism is defined as follows:

Definition 21 *Let $w \in \mathbb{N}^*$ be a morphically imprimitive word and $f(w) = (x_1, x_2, \ldots, x_n; v_1, v_2, \ldots, v_n)$, $n \in \mathbb{N}$, an imprimitivity factorisation of w. We define a morphism $\varphi_{f(w)} : \mathbb{N}^* \to \mathbb{N}^*$ by, for every $x \in \mathbb{N}$,*

$$\varphi_{f(w)}(x) := \begin{cases} \varepsilon, & \text{if } x \in \left(\bigcup_{i\in\{1,2,\ldots,n\}} \operatorname{symb}(v_i)\right) \setminus \{x_1, x_2, \ldots, x_n\}, \\ x, & \text{else.} \end{cases}$$

Additionally, we define a morphism $\psi_{f(w)} : \mathbb{N}^* \to \mathbb{N}^*$ by, for every $x \in \mathbb{N}$,

$$\psi_{f(w)}(x) := \begin{cases} v_i, & \text{if } x = x_i \text{ for an } i \in \{1, 2, \ldots, n\}, \\ x, & \text{else.} \end{cases}$$

Referring to Example 20, the following example illustrates Definition 21:

Example 22 *We consider w and $f(w)$ from Example 20. Then*

$$\varphi_{f(w)}(w) = 1 \cdot 2 \cdot 4 \cdot 5 \cdot 5 \cdot 4 \cdot 1 \cdot 2 \cdot 5.$$

With regard to $\psi_{f(w)}$, we can observe that it inverts the effect of $\varphi_{f(w)}$:

Theorem 23 *Let $w \in \mathbb{N}^*$ be a morphically imprimitive word and $f(w)$ an imprimitivity factorisation of w. Then $\psi_{f(w)}(\varphi_{f(w)}(w)) = w$.*

PROOF. An analogous argumentation to that in the proof of Theorem 13 can be used since one can easily see that $\psi_{f(w)}$ satisfies conditions (i) and (ii) of Theorem 13. □

Note that the morphism $\psi_{f(w)} \circ \varphi_{f(w)}$ is nontrivial, since $|\varphi_{f(w)}(w)| < |w|$. Additionally, w is a fixed point of $\psi_{f(w)} \circ \varphi_{f(w)}$. While Theorem 11 only claims the *existence* of a nontrivial fixed point morphism for a morphically imprimitive word, we can now directly specify such a morphism.

Keeping in mind our overall goal of finding a morphic root of a morphically imprimitive word, we can now summarise the so far gained results, according to which the morphism $\varphi_{f(w)}$ indeed maps w onto a shorter and morphically coincident word:

Corollary 24 *Let $w \in \mathbb{N}^*$ be a morphically imprimitive word and $f(w)$ an imprimitivity factorisation of w. Then $\varphi_{f(w)}(w) \equiv_* w$ and $|\varphi_{f(w)}(w)| < |w|$.*

PROOF. Directly from Theorem 23 and Definition 21. □

Unfortunately, $\varphi_{f(w)}(w)$ does not need to be morphically primitive (as demonstrated by $\varphi_{f(w)}(w)$ in Example 22, which is morphically imprimitive). Still, Corollary 24 allows to implement an iterative procedure: Given a word w, we can repeat the process of searching for an imprimitivity factorisation $f(w)$ of w, applying $\varphi_{f(w)}$ to w and defining $w := \varphi_{f(w)}(w)$ until no further imprimitivity factorisation is found. The resulting w is then a morphic root of the initially given word w.

However, this approach does not discuss the nature of those imprimitivity factorisations that lead to a morphic root. Therefore, we now introduce a more elegant and instructive method to find a morphic root of a given word. To this end, we need the following technical definition:

Definition 25 *Let $w \in \mathbb{N}^*$ be a morphically imprimitive word, and let $f(w)$ be an imprimitivity factorisation of w. We call $f(w)$ a* maximal imprimitivity factorisation *if and only if, for every imprimitivity factorisation $f'(w) = (x'_1, x'_2, \ldots, x'_{n'}; v'_1, v'_2, \ldots, v'_{n'})$ of w, $n' \in \mathbb{N}$, one of the following statements holds true:*

(i) $\sum_{i=1}^{n'} |v'_i| < \sum_{i=1}^{n} |v_i|$ *or*
(ii) $\sum_{i=1}^{n'} |v'_i| = \sum_{i=1}^{n} |v_i|$ *and $n' \geq n$.*

Thus, an imprimitivity factorisation is not maximal if there exists another imprimitivity factorisation that consists of larger v_i (cond. (i)) or that consists of less v_i of the same total length (cond. (ii)). We consider these criteria to be natural and intuitive when comparing different imprimitivity factorisations. The following example illustrates Definition 25:

Example 26 *We regard w from Example 20. Then $f_{max}(w) = (1, 4, 4, 1; 1 \cdot 2 \cdot 3 \cdot 3, 4 \cdot 3, 4 \cdot 3, 1 \cdot 2 \cdot 3 \cdot 3)$ is a maximal imprimitivity factorisation of w:*

$$w = \underbrace{1 \cdot 2 \cdot 3 \cdot 3}\cdot\underbrace{4 \cdot 3}\cdot 5 \cdot 5 \cdot \underbrace{4 \cdot 3}\cdot\underbrace{1 \cdot 2 \cdot 3 \cdot 3}\cdot 5.$$

Note that $f_{max2}(w) = (2, 4, 4, 2; 1 \cdot 2 \cdot 3 \cdot 3, 4 \cdot 3, 4 \cdot 3, 1 \cdot 2 \cdot 3 \cdot 3)$ is a maximal imprimitivity factorisation of w, too.

The subsequent main result of the present section demonstrates that a maximal imprimitivity factorisation $f(w)$ entails a morphism $\varphi_{f(w)}$ that maps a morphically imprimitive word w onto a morphically coincident word of minimum length:

Theorem 27 *Let $w \in \mathbb{N}^*$ be a morphically imprimitive word and $f(w)$ a maximal imprimitivity factorisation of w. Then $\varphi_{f(w)}(w)$ is morphically primitive.*

PROOF. Let $f(w) = (x_1, x_2, \ldots, x_n; v_1, v_2, \ldots, v_n)$. Assume to the contrary that $\varphi_{f(w)}(w)$ is morphically imprimitive. Then there exists an imprimitivity factorisation $f_\star$ of $\varphi_{f(w)}(w)$ and the morphism $\varphi_{f_\star(\varphi_{f(w)}(w))}$. We define some abbreviatory notations:

$$\begin{aligned}
&\varphi := \varphi_{f(w)}, \qquad\quad \psi := \psi_{f(w)},\\
&\eta := \varphi_{f_\star(\varphi_{f(w)}(w))},\ \chi := \psi_{f_\star(\varphi_{f(w)}(w))},\\
&s := \eta(\varphi(w))\,.
\end{aligned}$$

Hence, we have the following relations:

$$w \overset{\varphi}{\mapsto} \varphi(w) \overset{\eta}{\mapsto} s \text{ and } s \overset{\chi}{\mapsto} \varphi(w) \overset{\psi}{\mapsto} w\,.$$

In addition to this, it follows by Definition 21 that

$$\mathrm{symb}(w) \supset \mathrm{symb}(\varphi(w)) \supset \mathrm{symb}(s).$$

We consider the following subsets of $\mathrm{symb}(s)$:

$$\begin{aligned}
&U := \{x \in \mathrm{symb}(s) \mid \psi(\chi(x)) = x\},\\
&V := \{x \in \mathrm{symb}(s) \mid \psi(\chi(x)) = s'xs'',\ s', s'' \in \mathbb{N}^* \text{ with } |s's''| \geq 1\},
\end{aligned}$$

Hence, for suitable $n' \in \mathbb{N}$, $x'_1, x'_2, \ldots, x'_{n'} \in V$ and $u'_0, u'_1, \ldots, u'_{n'} \in U^*$, we have

$$s = u'_0\, x'_1\, u'_1\, x'_2\, u'_2 \ldots x'_{n'}\, u'_{n'}.$$

This definition has two immediate consequences:

Claim 1. $U \cap V = \emptyset$.

Claim 2. $U \cup V = \mathrm{symb}(s)$.

Furthermore, the following claim can be verified by Definition 21 with a bit of effort:

Claim 3. For all $x \in V$ with $\psi(\chi(x)) = s'x\,s''$, $\mathrm{symb}(s's'') \cap \mathrm{symb}(s) = \emptyset$.

We now show that the morphism $\psi \circ \chi$, applied to the symbols in V, induces an imprimitivity factorisation of w. For this purpose, let $f' : \mathbb{N}^+ \to \mathbb{N}^{n'} \times (\mathbb{N}^+)^{n'}$ and

$$f'(w) = (x'_1, x'_2, \ldots, x'_{n'}; v'_1, v'_2, \ldots v'_{n'})$$

with $v'_i := \psi(\chi(x'_i))$, $i \in \{1, 2, \ldots, n'\}$. $f'(w)$ is an imprimitivity factorisation since the conditions (i)–(iii) of Definition 18 are satisfied: (i) holds by definition of V, (ii) follows from Claims 1, 2, 3 and (iii) is satisfied since $|\psi(\chi(x'_i))|_{x'_i} = 1$, and, as $\psi \circ \chi$ is well-defined, $x'_i = x'_j$ implies $\psi(\chi(x'_i)) = \psi(\chi(x'_j))$.

It remains to show that the existence of $f'(w)$ contradicts the assumption that $f(w)$ is a maximal imprimitivity factorisation. To this end, we consider the lengths of $\varphi(w)$ and s:

$$
\begin{aligned}
|\varphi(w)| &= |w| - \sum_{i=1}^{n}(|v_i| - 1) = |w| - \left(\sum_{i=1}^{n}|v_i|\right) + n\,,\\
|s| &= |w| - \sum_{i=1}^{n'}(|v'_i| - 1) = |w| - \left(\sum_{i=1}^{n'}|v'_i|\right) + n'\,.
\end{aligned}
$$

According to Corollary 24 it is $|s| = |\eta(\varphi(w))| < |\varphi(w)|$, hence

$$|s| = |w| - \left(\sum_{i=1}^{n'}|v'_i|\right) + n' < |w| - \left(\sum_{i=1}^{n}|v_i|\right) + n = |\varphi(w)|.$$

This leads to

$$\left(\sum_{i=1}^{n'}|v'_i|\right) - n' > \left(\sum_{i=1}^{n}|v_i|\right) - n.$$

We now consider three cases:

Case 1. $\sum_{i=1}^{n'}|v'_i| < \sum_{i=1}^{n}|v_i|$. This case cannot occur since by definition of V, for all x_i, there exists an x'_j such that x_i is in $\chi(x'_j)$; hence every v_i is contained in the corresponding $\psi(\chi(x'_j))$ and, thus, $\sum_{i=1}^{n'}|\psi(\chi(x'_i))| \geq \sum_{i=1}^{n}|v_i|$.

Case 2. $\sum_{i=1}^{n'}|v'_i| > \sum_{i=1}^{n}|v_i|$. In this case, $f(w)$ does not satisfy Condition (i) of Definition 25.

Case 3. $\sum_{i=1}^{n'}|v'_i| = \sum_{i=1}^{n}|v_i|$. It immediately follows that $n' < n$, and therefore $f(w)$ does not satisfy Condition (ii) of Definition 25.

Consequently, each case contradicts the maximality of $f(w)$. $\square$

Hence, instead of investigating the existence of an arbitrary imprimitivity factorisation, we can directly seek for a maximal imprimitivity factorisation, which then – with its morphism φ – immediately leads to a morphic root of the morphically imprimitive word under consideration. We again illustrate this effect by our standard example:

Example 28 *We consider w from Example 20 and $f_{max}(w)$ from Example 26. Then*

$$\varphi_{f_{max}(w)}(w) = 1 \cdot 4 \cdot 5 \cdot 5 \cdot 4 \cdot 1 \cdot 5,$$

which is morphically primitive.

We conclude this section by some remarks on the ambiguity of imprimitivity factorisations. As shown by Example 26, a maximal imprimitivity factorisation of a morphically imprimitive word w does not need to be unique. Additionally, there even exist non-maximal imprimitivity factorisations of w that lead to a morphic root of w. Consequently, the maximality of an imprimitivity factorisation $f(w)$ is a sufficient, but not a necessary condition for receiving a morphism $\varphi_{f(w)}$ that maps w onto a morphically primitive word:

Example 29 *Let* $w := 1 \cdot 4 \cdot 4 \cdot 2 \cdot 1 \cdot 4 \cdot 4 \cdot 3 \cdot 4 \cdot 4 \cdot 3 \cdot 4 \cdot 4 \cdot 2$. *The imprimitivity factorisations* $f(w) = (1, 1, 3, 3; 1 \cdot 4 \cdot 4, 1 \cdot 4 \cdot 4, 3 \cdot 4 \cdot 4, 3 \cdot 4 \cdot 4)$ *and* $f'(w) = (1, 2, 1, 3, 3, 2; 1 \cdot 4, 4 \cdot 2, 1 \cdot 4, 4 \cdot 3 \cdot 4, 4 \cdot 3 \cdot 4, 4 \cdot 2)$, *that can be illustrated as follows*

$$w = 1 \cdot 4 \cdot 4 \cdot 2 \cdot 1 \cdot 4 \cdot 4 \cdot 3 \cdot 4 \cdot 4 \cdot 3 \cdot 4 \cdot 4 \cdot 2,$$

satisfy $\varphi_{f(w)}(w) = \varphi_{f'(w)}(w) = 1 \cdot 2 \cdot 1 \cdot 3 \cdot 3 \cdot 2$, *which is morphically primitive, but* $f(w)$ *is not maximal.*

At first glance, we consider this insight a bit surprising since the conditions of the definition of a maximal imprimitivity factorisation seem to be the natural foundations of a morphism mapping a morphically imprimitive word onto a shortest morphically coincident word. On the other hand, however, if we recall the results presented in the previous section on the ambiguity of imprimitivity morphisms then we can observe that the primitive root of w presented in Example 29 is $SCRN$-partitionable, which leads to an ambiguous imprimitivity morphism (cf. Theorem 17). Consequently, within the scope of the above example, we implicitly make use of the imprimitivity morphisms $\psi_{f(w)}$ and $\psi_{f'(w)}$, that lead to different imprimitivity factorisations. Additionally, the phenomenon described by Example 29 requires that at least one of the underlying imprimitivity morphisms maps certain symbols in the corresponding morphically primitive word onto words of length greater than or equal to 3. We feel certain that these insights and the tools provided by Section 4 can be used to give a nontrivial characterisation of those imprimitivity factorisations which lead to a morphic root of a morphically imprimitive word. Nevertheless, we expect this to be a rather cumbersome task.

Note that, given a morphically imprimitive word w, yet another approach is possible in order to find an imprimitivity factorisation $f(w)$ such that $\varphi_{f(w)}(w)$ is a morphic root of w: Since $|\varphi_{f(w)}(w)|$ is minimal if and only if the number of symbols in w erased by $\varphi_{f(w)}$ is maximal, Definition 21 directly implies that $\varphi_{f(w)}(w)$ is morphically primitive if and only if $\sum_{i=1}^{n}(|v_i|-1)$ is maximal. From a practical point of view it makes no difference whether to check this criterion or to verify if $f(w)$ is a maximal imprimitivity factorisation, since both criteria have to deal with all possible imprimitivity factorisations of w.

6 An Application

In the present section, we apply some of the insights gained so far to the following natural question on the *invertibility* of morphic mappings:

Problem 30 *Given a word w and a morphism h, does there exist a morphism g such that $g(h(w)) = w$?*

Pritykin [13] states that the decidability of Problem 30 is open if infinite words w are considered. With regard to the case $w \in \mathbb{N}^*$ (i. e. w is finite), $h : \mathbb{N}^* \to \mathbb{N}^*$ and $g : \mathbb{N}^* \to \mathbb{N}^*$, the problem is obviously decidable since it is decidable for *any* two finite words v, v' whether there exists a morphism mapping v onto v'. However, as mentioned in Section 5, this general problem is NP-complete, and therefore it implies an unsatisfactory decision procedure when applied to the special question raised by Problem 30. Furthermore, this statement on the decidability of Problem 30 for finite words does not provide any insights into the *nature* of those words w and morphisms h with respect to which the problem can be answered in the affirmative.

In contrast to this, our theory of morphically primitive words yields an elegant and more instructive solution:

Proposition 31 *Let $w \in \mathbb{N}^*$, and let $h : \mathbb{N}^* \to \mathbb{N}^*$ be a morphism. Then there exists a morphism $g : \mathbb{N}^* \to \mathbb{N}^*$ satisfying $g(h(w)) = w$ if and only if, for a morphic root v of w and a morphism ψ with $\psi(v) = w$, $h \circ \psi$ is an imprimitivity morphism for v or a renaming of v.*

PROOF. Let $\varphi : \mathbb{N}^* \to \mathbb{N}^*$ be a morphism such that $\varphi(w) = v$. (Note that by definition, given w and one of its morphic roots v, there always exist such morphisms φ and ψ since $v \equiv_* w$.)

We first show the "only if"-part: Let $g : \mathbb{N}^* \to \mathbb{N}^*$ be a morphism satisfying $g(h(w)) = w$. Thus, $w \equiv_* h(w)$. The morphic relation between v, w and $h(w)$ can be illustrated as follows:

$$\begin{tikzcd}[row sep=huge, column sep=huge]
w \arrow[r, shift right, "h"'] \arrow[d, shift right, "\varphi"'] & h(w) \arrow[l, shift right, "g"'] \\
v \arrow[u, shift right, "\psi"'] \arrow[ur, dotted, bend right, "h \circ \psi"'] &
\end{tikzcd}$$

We now consider three cases:

Case 1. $|h(\psi(v))| = |v|$. It follows from Lemma 4 that $|h(\psi(x))| = 1$ for every $x \in \mathrm{symb}(v)$. Assume to the contrary that there exist $x, y \in \mathrm{symb}(v)$ with $x \neq y$ and $h \circ \psi(x) = h \circ \psi(y)$. Since $\varphi(g(h(\psi(v)))) = v$, it follows from straightforward combinatorial considerations that there exists a $z \in \mathrm{symb}(h \circ \psi(v))$ with $\varphi \circ g(z) = \varepsilon$. Consequently, v is a fixed point of the nontrivial morphism $\varphi \circ g \circ h \circ \psi$, and, hence, not morphically primitive (cf. Theorem 11), which contradicts the choice of v. Thus, $h \circ \psi(x) \neq h \circ \psi(y)$ for every $x, y \in \mathrm{symb}(v)$ with $x \neq y$. Hence, $h \circ \psi$ is a renaming of v.

Case 2. $|h(\psi(v))| > |v|$. With $\varphi(g(h(\psi(v)))) = v$ it follows that $v \equiv_* h(\psi(v))$ and, thus, the morphism $h \circ \psi$ is an imprimitivity morphism (cf. Definition 12).

Case 3. $|h(\psi(v))| < |v|$. This contradicts the choice of v to be a morphic root of w since $h(w) \equiv_* w$ and $|h(w)| < |v|$.

We proceed with the "if"-part of the statement: If $h \circ \psi$ is a renaming of v or an imprimitivity morphism for v, it follows by definition that $h(\psi(v)) \equiv_* v$ and, thus, that there exists a morphism g' such that $g'(h(\psi(v))) = v$. Consequently, the morphism $g := \psi \circ g'$ satisfies $g(h(w)) = \psi(g'(h(w))) = \psi(g'(h(\psi(v)))) = \psi(v) = w$. $\square$

Note that we can efficiently check whether or not $h \circ \psi$ is a renaming of v (by testing if $|h(x)| = 1$ and $h(x) \neq h(y)$ for every $x, y \in \mathrm{symb}(v)$, $x \neq y$) and whether or not $h \circ \psi$ is an imprimitivity morphism (by testing the conditions of Theorem 13). Hence, if there exists an efficient procedure to find a morphic root of a word w (and a suitable morphism ψ), then the characterisation in Proposition 31 even leads to a polynomial decision procedure regarding Problem 30 (restricted to finite words w).

7 The Number of Morphically Primitive Words

In this section, we study the *number* of morphically primitive words for any fixed length n. Of course, since we regard an infinite alphabet, there are infinitely many such words, and therefore it is necessary to impose some respective restrictions. A natural corresponding choice is to only deal with words w in canonical form (cf. Section 2). In this case, the *total* number of words to be considered corresponds to the nth *Bell number*, i. e. the number of partitions of a set of size n into nonempty subsets (see, e. g., Rota [16]):

Proposition 32 *For each $n \in \mathbb{N}$, the number of words in $\mathbb{N}^*$ of length n that are in canonical form equals the number of all partitions of a set S with $|S| = n$ into nonempty subsets.*

PROOF. W. l. o. g., we assume $S := \{1, 2, \ldots, n\}$. Let $\mathbb{N}^n_{\text{can}} := \{w \in \mathbb{N}^n \mid w \text{ is in canonical form}\}$ be the set of all words of length n in canonical form; furthermore, we define $\text{PAR}(S) := \{\{S_1, S_2, \ldots, S_m\} \mid m \in \mathbb{N},\ S_1, S_2, \ldots, S_m \neq \emptyset,\ S_i \cap S_{i'} = \emptyset,\ 1 \leq i < i' \leq m,\ \bigcup_{1 \leq i \leq m} S_i = S\}$, i. e. $\text{PAR}(S)$ is the set of all partitions of S into nonempty subsets. We show that that there exists a bijection $b : \text{PAR}(S) \to \mathbb{N}^n_{\text{can}}$; this directly implies the correctness of Proposition 32.

We introduce b as follows: Intuitively, b interprets the elements of any subset S_i of S as the *positions* of those symbols in a word that equal i. Formally, we define an order $\leq_{\min}$ on subsets of $\mathbb{N}$ by, for any $A, B \subseteq \mathbb{N}$, $A \leq_{\min} B$ if and only if $\min A \leq \min B$. W. l. o. g., we now may assume that each $\{S_1, S_2, \ldots, S_m\} \in \text{PAR}(S)$ satisfies $S_1 <_{\min} S_2 <_{\min} \ldots <_{\min} S_m$. If we consider any $w \in \mathbb{N}^n_{\text{can}}$ – i. e. $w := x_1 x_2 \ldots x_n$ with $x_j \in \mathbb{N}$, $1 \leq j \leq n$ – then $b(\{S_1, S_2, \ldots, S_m\}) := x_1 x_2 \ldots x_n$, where x_j, $1 \leq j \leq n$, equals i if and only if $j \in S_i$, $1 \leq i \leq m$. For example, regarding $n := 5$, $b(\{\{1, 4\}, \{2, 3\}, \{5\}\}) := 1 \cdot 2 \cdot 2 \cdot 1 \cdot 3$.

In order to show that b is a *bijection* $b : \text{PAR}(S) \to \mathbb{N}^n_{\text{can}}$, we now demonstrate that,

(1) for every $p \in \text{PAR}(S)$, $b(p) \in \mathbb{N}^n_{\text{can}}$,
(2) for every $p, q \in \text{PAR}(S)$, $p \neq q$ implies $b(p) \neq b(q)$, and
(3) for every $w \in \mathbb{N}^n_{\text{can}}$ there exists a $p \in \text{PAR}(S)$ satisfying $b(p) = w$.

Ad 1: Since the length of $b(p)$ obviously equals n, we only have to show that $b(p)$ is in canonical form. In other words, we need to demonstrate that $b(p)$ is lexicographically smaller than every of its renamings.

Let $p := \{S_1, S_2, \ldots, S_m\}$, $m \in \mathbb{N}$, and let $r : \mathbb{N}^* \to \mathbb{N}^*$ be any renaming that satisfies $r(b(p)) \neq b(p)$. If, for every $y \in \text{symb}(b(p))$, $r(y) \geq y$, then $b(p)$ is evidently lexicographically smaller than $r(b(p))$. Hence, we merely have to consider the case that there is a symbol z' in $b(p)$ with $r(z') < z'$. Assume to the contrary that $r(y) \leq y$ holds for each $y \in \mathrm{symb}(b(p)) \setminus \{z'\}$. Since r is a renaming – i. e. it is injective and maps every symbol onto a word of length 1 – and, by definition, $\mathrm{symb}(b(p)) = \{1, 2, \ldots, m\}$, the existence of z' and the assumption on all other symbols in $b(p)$ implies that $|\mathrm{symb}(r(b(p)))| < |\mathrm{symb}(b(p))|$. Evidently, this contradicts the definition of a renaming. Consequently, there is also a symbol z in $b(p)$ with $r(z) > z$. W. l. o. g. we assume z to be the smallest symbol in $\mathrm{symb}(b(p))$ showing this property, i. e. $r(y) \leq y$ for each $y < z$. We now regard the prefix vz of $b(p)$ where $v \in \mathbb{N}^*$ satisfies $z \notin \mathrm{symb}(v)$; evidently, this prefix is unique. According to our above assumption, p satisfies $S_1 <_{\min} S_2 <_{\min} \ldots <_{\min} S_m$. Consequently, the definition of b implies that $\mathrm{symb}(v) = \{1, 2, \ldots, z-1\}$. Thus, $r(y) \leq y$ for each $y \in \mathrm{symb}(v)$. Now assume to the contrary that there exists a $y' \in \mathrm{symb}(v)$ satisfying $r(y') < y'$. Then, using an analogous reasoning to that given above on the existence of z, $|\mathrm{symb}(r(v))| < |\mathrm{symb}(v)|$, which again contradicts the definition of a renaming. Therefore, $r(y) = y$ for each $y \in \mathrm{symb}(v)$. Thus, $b(p)$ has the prefix vz and $r(b(p))$ has the prefix $r(v)r(z) = v\,r(z)$. Since $r(z) > z$, $b(p)$ is lexicographically smaller than $r(b(p))$.

Ad 2: For some $m, m' \in \mathbb{N}$, let $p =: \{S_1, S_2, \ldots, S_m\}$ and $q =: \{T_1, T_2, \ldots, T_{m'}\}$. Recall our above assumption according to which $S_1 <_{\min} S_2 <_{\min} \ldots <_{\min} S_m$ and $T_1 <_{\min} T_2 <_{\min} \ldots <_{\min} T_{m'}$. Therefore, $p \neq q$ implies that, for some i, i' with $i \neq i'$, there exists a $j \in S$ with $j \in S_i \cap T_{i'}$. Let $b(p) =: x_1 x_2 \ldots x_n$ and $b(q) =: y_1 y_2 \ldots y_n$. Then the definition of b leads to $x_j = i$ and $y_j = i'$. Since $i \neq i'$, we therefore can conclude $b(p) \neq b(q)$.

Ad 3: For any $w \in \mathbb{N}^n_{\mathrm{can}}$ there exists an $m \in \mathbb{N}$ such that $\mathrm{symb}(w) = \{1, 2, \ldots, m\} = S$, since otherwise we could give a renaming w' of w that is lexicographically smaller than w. Let $w =: x_1 x_2 \ldots x_n$ with $x_j \in \mathbb{N}$, $1 \leq j \leq n$. Then, for every $i \in \mathrm{symb}(w)$, we define $S_i := \{j \mid x_j = i, 1 \leq j \leq n\}$. Obviously, this implies that $S_1 \cup S_2 \cup \ldots \cup S_m = S$; furthermore, for every i, i', S_i and $S_{i'}$ are disjoint. Hence, $p := \{S_1, S_2, \ldots, S_m\}$ is a partition of S. An application of b to p directly shows that $b(p) = w$.

This concludes the proof. $\square$

From now on, we write $B(n)$ for the nth Bell number. Hence, there exist $B(n)$ different words $w \in \mathbb{N}^+$ of length n in canonical form. Referring to Rota [16], we can compute $B(n)$ using, e. g., the recurrence equation $B(n+1) = \sum_{k=0}^{n} \binom{n}{k} B(k)$ (with $B(0) := 1$).

Since there exists an effective procedure to enumerate all words w in canonical form as well as a decision algorithm to decide whether or not a given word w is

	$n = 1$	2	3	4	5	6	7	8
$B(n)$	1	2	5	15	52	203	877	4140
$mPrim(n)$	1	1	1	3	11	32	152	625
$Twice(n)$	0	1	1	4	11	41	162	715
$mPrim(n)/B(n)$	1.0000	0.5000	0.2000	0.2000	0.2115	0.1576	0.1733	0.1510
$Twice(n)/B(n)$		0.5000	0.2000	0.2667	0.2115	0.2020	0.1847	0.1727
$B(n-1)/B(n)$		0.5000	0.4000	0.3333	0.2885	0.2562	0.2315	0.2118

	$n = 9$	10	11	12	...	50	...	100	...
$B(n)$	21147	115975	678570	4213597	...	$\sim 10^{47}$	...	$\sim 10^{115}$	...
$mPrim(n)$	3152	16154	90993	539181	...	?	...	?	...
$Twice(n)$	3425	17722	98253	580317	...	$\sim 10^{46}$	...	$\sim 10^{114}$	...
$mPrim(n)/B(n)$	0.1491	0.1393	0.1341	0.1280	...	?	...	?	...
$Twice(n)/B(n)$	0.1620	0.1528	0.1448	0.1377	...	0.0546	...	0.0329	...
$B(n-1)/B(n)$	0.1958	0.1823	0.1709	0.1610	...	0.0578	...	0.0340	...

Table 1
The number of morphically primitive words.

morphically primitive (cf. Section 5), we can effectively count the morphically primitive words in canonical form for any fixed length $n \in \mathbb{N}$; we use the term $mPrim(n)$ to denote the number of these words. However, this is an expensive task because $B(n)$ words have to be tested – a number that grows exponentially – and, additionally, a polynomial test for morphic primitivity is not known so far, although the methods and results in Section 5 might help to develop such an algorithm.

Rows 1, 2 and 4 of Table 1 show some example values for $B(n)$, $mPrim(n)$ and the ratio $mPrim(n)/B(n)$. It can immediately be seen that the number of morphically primitive words is small when compared to the number of all words. This once again backs the choice of the term "primitive". Furthermore, it seems that the ratio $mPrim(n)/B(n)$ is strictly monotonic decreasing for $n \geq 7$.

Although $mPrim(n)/B(n)$ is becoming smaller for larger n, it is clear that, for any $n \in \mathbb{N}$, there always exist morphically primitive words of length n, such as the words 1^n and $1^{\lfloor n/2 \rfloor} \cdot 2^{\lceil n/2 \rceil}$ $(n \geq 4)$. The following proposition gives a lower bound for $mPrim(n)$:

Proposition 33 *Let $n \in \mathbb{N}$. Then*

$$B(\lfloor n/2 \rfloor) \leq mPrim(n).$$

PROOF. Let $n \in \mathbb{N}$. We define a mapping $b : \mathbb{N}^{\lfloor n/2 \rfloor} \to \mathbb{N}^n$ in the following way: If n is even, then, for $w = x_1 x_2 \dots x_{\lfloor n/2 \rfloor}$, $x_i \in \mathbb{N}$, $1 \leq i \leq \lfloor n/2 \rfloor$, $b(w) := x_1 x_1 x_2 x_2 \dots x_{\lfloor n/2 \rfloor} x_{\lfloor n/2 \rfloor}$. If n is odd, then, for $w = x_1 x_2 \dots x_{\lfloor n/2 \rfloor}$, $x_i \in \mathbb{N}$, $1 \leq i \leq \lfloor n/2 \rfloor$, $b(w) := x_1 x_1 x_1 x_2 x_2 x_3 x_3 \dots x_{\lfloor n/2 \rfloor} x_{\lfloor n/2 \rfloor}$. It immediately follows from the construction of b that $b(w)$ is in canonical form if w is in canonical form.

Now, let U be the set of all words of length $\lfloor n/2 \rfloor$ in canonical form and $V := \{b(u) \mid u \in U\}$. It can be easily verified that b is a bijection between U and V. Hence, we know that V contains exactly $B(\lfloor n/2 \rfloor)$ many words in canonical form of length n. Thus, in order to prove the statement of the proposition, it remains to be shown that every $v \in V$ is morphically primitive.

Let $v \in V$. Assume to the contrary that v is morphically imprimitive. Then there exists an imprimitivity factorisation $f(v) = (x_1, x_2, \dots, x_n; v_1, v_2, \dots, v_k)$ of v, $k \in \mathbb{N}$, such that $v = u_0 v_1 u_1 v_2 u_2 \dots v_k u_k$ for some $u_0, u_1, \dots, u_k \in \mathbb{N}^*$ (cf. Definition 18). Now consider v_i for some $1 \leq i \leq k$: Due to conditions (i) and (iii) of Definition 18, we have $|v_i| \geq 2$ and $|v_i|_{x_i} = 1$. Because of the structure of v, v_i must either begin or end with x_i, otherwise x_i would occur twice in v_i. We assume that v_i begins with x_i (the other case leads to an analogous reasoning). Thus,

$$v = \dots x_i \underbrace{x_i v_i'}_{v_i} \dots .$$

Consequently, the occurrence of x_i to the left of v_i either belongs to u_{i-1} or to v_{i-1} (if $i = 1$, only the former case is possible), which contradicts condition (ii) or condition (iii) of Definition 18, respectively. Hence, there is no imprimitivity factorisation of v, and therefore v is morphically primitive. $\square$

We now establish an upper bound for $mPrim(n)$. To this end, we introduce the following number: For any $n \in \mathbb{N}$, let

$$Twice(n) := |\{w \in \mathbb{N}^n \text{ in canonical form} \mid \text{for all } x \in \mathrm{symb}(w) : |w|_x \geq 2\}|$$

Referring to this definition, we can note the following observation:

Proposition 34 *Let $n \in \mathbb{N}$, $n > 1$. Then*

$$mPrim(n) \leq Twice(n).$$

PROOF. It can be easily verified that every word $w_{\text{once}} \in \mathbb{N}^+$ of length greater than 1 that contains one symbol only once is morphically imprimitive, since $w_{\text{once}} \equiv_* 1$. Thus, it is a necessary condition for any morphically primitive word $w \in \mathbb{N}^*$, $|w| > 1$, that every symbol in $\mathrm{symb}(w)$ occurs at least twice in w. $\square$

Using an approximation for $Twice(n)$ to be further examined below, we can express both the lower and an upper bound of $mPrim(n)$ by suitable Bell numbers, and, thus, paraphrase Propositions 33 and 34 as follows:

Corollary 35 *Let $n \in \mathbb{N}$, $n > 1$. Then*

$$B(\lfloor n/2 \rfloor) \leq mPrim(n) \leq B(n-1).$$

PROOF. The lower bound is given in Proposition 33. For the upper bound, we first observe that every word $w \in \mathbb{N}^*$ in canonical form corresponds to a rhyming scheme for a stanza of $n := |w|$ lines, interpreted in the following way: If $w = x_1 x_2 \ldots x_n$ then $x_i = x_j$ means that line i and j rhyme, $1 \leq i, j \leq n$. Hence, $Twice(n)$ corresponds to the number of *complete* rhyming schemes for a stanza of n lines since the condition $|w|_x \geq 2$ for all $x \in \mathrm{symb}(w)$ guarantees that every line rhymes with at least one other line. Thus, we can make use of the results by Becker [2], according to which $Twice(n) = B(n-1) - Twice(n-1)$. This proves the corollary. □

Table 1 shows values of $Twice(n)$ and $B(n-1)$ for some n, as well as the ratios $Twice(n)/B(n)$ and $B(n-1)/B(n)$. Concerning these ratios, we can conclude from Proposition 34 and Corollary 35 that

$$mPrim(n)/B(n) \leq Twice(n)/B(n) \leq B(n-1)/B(n).$$

Table 1 suggests that the latter upper bound for the ratio $mPrim(n)/B(n)$ is already quite tight; therefore, we can use it to compute close approximations even for larger n.

Since $Twice(n)$ seems to be a good upper bound for $mPrim(n)$, it is obvious that $B(n) - Twice(n)$ can be considered a good lower bound for the size of the set of all morphically *im*primitive words of length n in canonical form. Note that this number can also be derived from Table 1 since $B(n) - Twice(n) = Twice(n+1)$ (cf. proof of Corollary 35). The, when compared to $B(n) - Twice(n)$, relatively small number $Twice(n) - mPrim(n)$ suggests that, remarkably and perhaps also surprisingly, the major part of the set of all morphically imprimitive words consists of those words that contain at least one symbol only once. Such a word w_{once} with $|w|_x = 1$ for an $x \in \mathrm{symb}(w)$ has very special properties in terms of different theories (cf. Theorem 11): With regard to morphic primitivity, w_{once} has the morphic root x and the very simple imprimitivity factorisation $f(w) = (x; w_{\text{once}})$. Concerning finite fixed points of morphism, w_{once} is a fixed point of the simple morphism that maps x onto w_{once} and erases all other symbols. In terms of erasing pattern languages, where words (then called patterns) generate languages over a fixed alphabet Σ, w_{once} generates the full monoid Σ^*.

Finally, note that the well-known formula for the number of standard primitive words of length n over a k-ary alphabet (as described in, e. g., Lothaire [10]) can be easily adapted to our case of regarding words in canonical form over an infinite alphabet. More precisely, if we refer to the primitive words of length n in canonical form by $\mathit{Prim}(n)$, we have

$$\mathit{Prim}(n) = \sum_{c,d \in \mathbb{N} \,|\, cd=n} \mu(c) B(d),$$

where μ refers to the Möbius function, for the number of primitive words of length n in canonical form. A list of $\mathit{Prim}(n)$ for some n can be found in [12]. Since $\mathit{Prim}(n)$ depends on the number of divisors of n, it is manifest that $\mathit{Prim}(n)$ is nonmonotonic and, thus, very different from $\mathit{mPrim}(n)$. Due to Proposition 5, this observation is by no means surprising.

8 Open Problems

From our point of view, there exist two outstanding open problems on morphic primitivity of words.

Evidently, as explained by Section 5, we are interested in the time complexity of the *morphic primitivity problem*, i. e. the problem of whether or not a given word is morphically primitive. Due to the fact that the morphically primitive words are equivalent to those words that are not a fixed point of a nontrivial endomorphism, to the succinct terminal-free patterns and to those words in $\mathbb{N}^*$ for which there exists an unambiguous injective morphism $\sigma : \mathbb{N}^* \to \{\mathtt{a}, \mathtt{b}\}^*$ (cf. Theorem 11), this question is not only a manifest, but also a very important one. In addition to this, any (related) procedure that computes a morphic root of a morphically imprimitive pattern can immediately be turned into a decision procedure for the *equivalence problem* for *terminal-free E-pattern language* (cf. [8]), which is crucial, e. g., for many considerations on such languages within the scope of *algorithmic learning theory*. While the abovementioned result by Ehrenfeucht and Rozenberg [3] suggests that the morphic primitivity problem is NP-complete, we feel that our insights presented in Section 5 raise hope of a polynomial time procedure, since they reduce the morphic primitivity problem to the problem of finding an imprimitivity factorisation.

The second subject we consider particularly worth to be further examined is the *number* of morphically primitive words for any fixed length $n \in \mathbb{N}$: A precise (recurrence) equation is still missing, and the upper and particularly the lower bound given in Section 7 can certainly be improved.

Acknowledgements

The authors wish to thank Stepan Holub for his valuable comments on an earlier version of this paper.

References

[1] D. Angluin, Finding patterns common to a set of strings, Journal of Computer and System Sciences 21 (1980) 46–62.

[2] H. Becker, Solution to problem E 461, American Mathematical Monthly 48 (1941) 701–702.

[3] A. Ehrenfeucht, G. Rozenberg, Finding a homomorphism between two words is NP-complete, Information Processing Letters 9 (1979) 86–88.

[4] D. Freydenberger, D. Reidenbach, The unambiguity of segmented morphisms, in: Proc. 11th International Conference on Developments in Language Theory, DLT 2007, Lecture Notes in Computer Science, 2007.

[5] D. Freydenberger, D. Reidenbach, J. Schneider, Unambiguous morphic images of strings, International Journal of Foundations of Computer Science 17 (2006) 601–628.

[6] D. Hamm, J. Shallit, Characterization of finite and one-sided infinite fixed points of morphisms on free monoids, Tech. Rep. CS-99-17, Dep. of Computer Science, University of Waterloo, `http://www.cs.uwaterloo.ca/~shallit/papers.html` (1999).

[7] T. Head, Fixed languages and the adult languages of 0L schemes, International Journal of Computer Mathematics 10 (1981) 103–107.

[8] T. Jiang, A. Salomaa, K. Salomaa, S. Yu, Decision problems for patterns, Journal of Computer and System Sciences 50 (1995) 53–63.

[9] F. Levé, G. Richomme, On a conjecture about finite fixed points of morphisms, Theoretical Computer Science 339 (2005) 103–128.

[10] M. Lothaire, Combinatorics on Words, Addison-Wesley, Reading, MA, 1983.

[11] A. Mateescu, A. Salomaa, Patterns, in: G. Rozenberg, A. Salomaa (eds.), Handbook of Formal Languages, vol. 1, chap. 4.6, Springer, 1997, pp. 230–242.

[12] On-Line Encyclopedia of Integer Sequences, The number of primitive (aperiodic) word structures of length n using an infinite alphabet, `http://www.research.att.com/~njas/sequences/A082951`.

[13] Y. Pritykin, Information in infinite words, in: Proc. 6th International Conference on Words, WORDS 2007, 2007.

[14] D. Reidenbach, Discontinuities in pattern inference, Theoretical Computer Science 397 (2008) 166–193.

[15] D. Reidenbach, J. C. Schneider, Morphically primitive words, in: Proc. 6th International Conference on Words, WORDS 2007, 2007.

[16] G.-C. Rota, The number of partitions of a set, American Mathematical Monthly 71 (1964) 498–504.

[17] G. Rozenberg, A. Salomaa, Handbook of Formal Languages, vol. 1, Springer, Berlin, 1997.