Loughborough
University

This item was submitted to Loughborough's Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

Morphic Primitivity and Alphabet Reductions

Hossein Nevisi* and Daniel Reidenbach

Department of Computer Science, Loughborough University,
Loughborough, Leicestershire, LE11 3TU, UK
{H.Nevisi,D.Reidenbach}@lboro.ac.uk

Abstract. An alphabet reduction is a 1-uniform morphism that maps a word to an image that contains a smaller number of different letters. In the present paper we investigate the effect of alphabet reductions on morphically primitive words, i.e., words that are not a fixed point of a nontrivial morphism. Our first main result answers a question on the existence of unambiguous alphabet reductions for such words, and our second main result establishes whether alphabet reductions can be given that preserve morphic primitivity. In addition to this, we study Billaud's Conjecture – which features a different type of alphabet reduction, but is otherwise closely related to the main subject of our paper – and prove its correctness for a special case.

Keywords: Combinatorics on words, Morphisms, Ambiguity, Morphic primitivity, Fixed points, Billaud's Conjecture

1 Introduction

In this paper, we study some fundamental combinatorial questions for a special type of morphisms which we call an *alphabet reduction*. Such morphisms are characterised by the fact that they map a word over some alphabet Δ_1 to a word over an alphabet Δ_2 that is a proper subset of Δ_1, and they are 1-uniform, i.e., they map every letter in Δ_1 to a word of length 1. Among all these morphisms, we are particularly interested in those that are the identity for every letter in Δ_2 and, in order to obtain unrestricted results, we assume Δ_1 to be a set of natural numbers, i.e., we consider morphisms $\phi : \Delta_1^* \to \Delta_2^*$, where $\Delta_2 \subset \Delta_1 \subseteq \mathbb{N}$. For example, the morphism $\phi : \{1, 2, 3, 4\}^* \to \{1, 2, 3\}^*$ with $\phi(1) = 1$, $\phi(2) = 2$, $\phi(3) = 3$ and $\phi(4) = 3$, is of the type we wish to investigate.

Due to reasons to be further explained below, we apply such morphisms to *morphically primitive* words over $\mathbb{N}$, i.e., words α for which there are no word β with $|\beta| < |\alpha|$ and morphisms $\phi, \psi : \mathbb{N}^* \to \mathbb{N}^*$ satisfying $\phi(\alpha) = \beta$ and $\psi(\beta) = \alpha$. Morphically primitive words have not only been studied by Reidenbach and Schneider [8], but they are also equivalent to those words α that are not a *fixed point* of a nontrivial morphism, which means that there is no morphism $\phi : \mathbb{N}^* \to \mathbb{N}^*$ such that $\phi(\alpha) = \alpha$ and for a letter x in α, $\phi(x) \neq x$.

---

* Corresponding author.

Since a word is a fixed point if and only if it is not morphically primitive (see [8] for additional explanations), we use these concepts interchangeably.

Our first question on alphabet reductions is concerned with their *ambiguity*. A morphism $\phi : \mathbb{N}^* \to \mathbb{N}^*$ is called *ambiguous* with respect to a word α if there exists another morphism ψ mapping α to $\phi(\alpha)$; if such a ψ does not exist, then ϕ is *unambiguous*. For example, the morphism $\phi_0 : \{1, 2, 3\}^* \to \{1, 2\}^*$ – given by $\phi_0(1) := 1$, $\phi_0(2) := 2$, $\phi_0(3) := 2$ – is ambiguous with respect to the word $\alpha_0 := 1 \cdot 2 \cdot 3 \cdot 3 \cdot 2 \cdot 1$ (where we separate the letters in a word by a dot), since the morphism ψ_0 – defined by $\psi_0(1) := 1$, $\psi_0(2) := \varepsilon$ (i.e., ψ_0 maps 2 to the empty word), $\psi_0(3) := 2 \cdot 2$ – satisfies $\psi_0(\alpha_0) = \phi_0(\alpha_0)$ and, for a letter x occurring in α, $\psi_0(x) \neq \phi_0(x)$:

$$\phi_0(\alpha_0) = \underbrace{\overbrace{1}^{\phi_0(1)}}_{\psi_0(1)} \underbrace{\overbrace{2}^{\phi_0(2)} \overbrace{2}^{\phi_0(3)}}_{\psi_0(3)} \underbrace{\overbrace{2}^{\phi_0(3)} \overbrace{2}^{\phi_0(2)}}_{\psi_0(3)} \underbrace{\overbrace{1}^{\phi_0(1)}}_{\psi_0(1)} = \psi_0(\alpha_0) \, .$$

It can be verified with moderate effort that, e. g., the morphism $\phi_1 : \{1, 2, 3\}^* \to \{1, 2\}^*$ – given by $\phi_1(1) := 1$, $\phi_1(2) := 1 \cdot 2$, $\phi_1(3) := 2$ – is unambiguous with respect to α_0.

The research on the ambiguity of morphisms was initiated by Freydenberger, Reidenbach and Schneider [3], and previous papers on this subject mainly focus on the question of whether unambiguous morphisms *exist* for a given word (see Schneider [10], Reidenbach and Schneider [9], Freydenberger, Nevisi and Reidenbach [2] and Nevisi and Reidenbach [7]). In [7] we have investigated this problem for 1-uniform morphisms, providing some first insights into it. In the first technical part of present paper we wish to continue this research, thus studying the following question:

Problem 1. Is it possible, for every morphically primitive word, to give an unambiguous alphabet reduction?

Note that this problem is restricted to morphically primitive patterns and alphabet reductions (instead of 1-uniform morphisms that may map a word α to an image that contains as many different letters as α), since it can be easily understood that *every* nonerasing morphism is ambiguous with respect to a word that is a fixed point of a nontrivial morphism, and that an unambiguous 1-uniform morphism with an arbitrary large target alphabet exists for a word α if and only if α is not a fixed point of a nontrivial morphism (see [3] for details).

The set of those words that have an unambiguous morphism has so far been characterised for general nonerasing morphisms (see [3] and, in a more restricted setting, [2]), and these characterisations show that the existence of such unambiguous morphisms is largely independent from the size of the target alphabet Δ_2 of the morphisms, In contrast to this, Schneider [10] shows that the set of words that have an unambiguous *erasing* morphism is different for every size of the target alphabet Δ_2, which implies that a characterisation of these sets needs to incorporate the size of Δ_2 and suggests that such a characterisation might be very difficult.

In [7] we have provided some results indicating that an equivalent phenomenon might hold with regard to Problem 1. Our first main result in the present paper shows that this indeed is true.

The second question we wish to study is more directly concerned with morphically primitive words. Since these words are equivalent to those that are not a fixed point of any nontrivial morphism, and since the latter words are known to have a number of important properties (see Reidenbach and Schneider [8]), they have been studied in a number of papers. Although these studies have provided, e.g., a characterisation (see Head [4]) and even a polynomial-time decision procedure (see Holub [5]), many fundamental properties and the actual fabric of morphically primitive words are not fully understood. This is epitomised by the fact that Billaud's Conjecture (see [1]), to be discussed in Section 4, is still largely unresolved.

In the present paper we shall investigate whether, for a given morphically primitive word α, there is an alphabet reduction ϕ such that $\phi(\alpha)$ is again morphically primitive:

Problem 2. Is it possible, for every morphically primitive word, to give an alphabet reduction that preserves morphic primitivity?

For example, let $\alpha := 1 \cdot 2 \cdot 3 \cdot 4 \cdot 1 \cdot 3 \cdot 2 \cdot 4$; if $\phi : \{1, 2, 3, 4\}^* \to \{1, 2, 4\}^*$ is a morphism with $\phi(1) := 1$, $\phi(2) := 2$, $\phi(3) := 2$ and $\phi(4) := 4$, then $\phi(\alpha)$ is not morphically primitive (i.e., it is *morphically imprimitive*). On the other hand, $\psi(\alpha)$, where $\psi : \{1, 2, 3, 4\}^* \to \{1, 3, 4\}^*$ is a morphism given by $\psi(1) := 1$, $\psi(2) := 1$, $\psi(3) := 3$ and $\psi(4) := 4$, is morphically primitive.

Problem 2 appears to be very similar to Billaud's Conjecture, but the latter features a different type of morphism (which, intuitively, still can be seen as an alphabet reduction). In Section 4, we solve Problem 2, and we prove the correctness of Billaud's Conjecture for a special case not studied in the literature so far.

Note that, due to space constraints, some proofs and some related examples are omitted from this paper.

2 Definitions

An *alphabet* $\mathcal{A}$ is a nonempty set of symbols and we call these symbols *letters*. A *word (over $\mathcal{A}$)* is a finite sequence of letters taken from $\mathcal{A}$. We denote the empty word by ε. The notation $\mathcal{A}^*$ refers to the set of all (empty and non-empty) words over $\mathcal{A}$, and $\mathcal{A}^+ := \mathcal{A}^* \setminus \{\varepsilon\}$. For the *concatenation* of two words α_1, α_2, we write $\alpha_1 \cdot \alpha_2$ or simply $\alpha_1 \alpha_2$. The word that results from n-fold concatenation of a word α is denoted by α^n. The notation $|x|$ stands for the size of a set x or the length of a word x. With regard to an arbitrary word α, $\mathrm{symb}(\alpha)$ denotes the set of all letters occurring in α. We call a word $\beta \in \mathcal{A}^*$ a *factor* of a word $\alpha \in \mathcal{A}^*$ if, for some $\gamma_1, \gamma_2 \in \mathcal{A}^*$, $\alpha = \gamma_1 \beta \gamma_2$; moreover, if β is a factor of α then we say that α *contains* β and denote this by $\beta \sqsubseteq \alpha$. If $\beta \neq \alpha$, then we say that β is a *proper* factor of α and denote this by $\beta \sqsubset \alpha$. If $\gamma_1 = \varepsilon$, then β is a prefix of α, and

if $\gamma_2 = \varepsilon$, then β is a suffix of α. For any words $\alpha, \beta \in \mathcal{A}^*$, $|\alpha|_\beta$ stands for the number of (possibly overlapping) occurrences of β in α. The symbol $[\ldots]$ is used to omit some canonically defined parts of a given word, e. g., $\alpha = 1 \cdot 2 \cdot [\ldots] \cdot 5$ stands for $\alpha = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5$.

A *morphism* is a mapping that is compatible with concatenation, i. e., for any alphabets $\mathcal{A}, \mathcal{B}$, $\phi : \mathcal{A}^* \to \mathcal{B}^*$ is a morphism if it satisfies $\phi(\alpha \cdot \beta) = \phi(\alpha) \cdot \phi(\beta)$ for all words $\alpha, \beta \in \mathcal{A}^*$. We call $\mathcal{B}$ the *target alphabet* of ϕ. A morphism $\phi : \mathcal{A}^* \to \mathcal{B}^*$ is called *nonerasing* provided that, for every $i \in \mathcal{A}$, $\phi(i) \neq \varepsilon$. If ϕ is nonerasing, then we often indicate this by writing $\phi : \mathcal{A}^+ \to \mathcal{B}^+$. A morphism ϕ is *1-uniform* if, for every $i \in \mathcal{A}$, $|\phi(i)| = 1$.

3   Unambiguous Alphabet Reductions

In the present section, we investigate Problem 1. Our main result in this section strengthens our results in [7] regarding the existence of unambiguous alphabet reductions $\phi : \mathbb{N}^* \to \Delta^*$, $\Delta \subset \mathbb{N}$, for a *fixed* target alphabet (i. e., the size of Δ does not depend on the number of letters in the preimage). The overall goal of most of the papers on unambiguous morphisms is to characterise the set of words that have an unambiguous morphism, and this goal has so far been accomplished for general nonerasing morphisms in two different settings (see Section 1). These results benefit from the fact that, regarding such types of morphisms, the size of Δ does not have a major impact on the sets of words to be characterised. Before we explain whether the same phenomenon is true for 1-uniform morphisms, we give a definition of a morphism that is not only vital for the proof of Theorem 4 below, but also for our considerations in Section 4.

Definition 3. *Let $\alpha \in \mathbb{N}^*$. For any $i, j \in \mathbb{N}$ with $i \neq j$ and, for every $x \in \mathbb{N}$, let the morphism $\phi_{i,j} : \mathrm{symb}(\alpha)^* \to \mathrm{symb}(\alpha)^*$ be given by*

$$\phi_{i,j}(x) := \begin{cases} i, & if\ x = j\,, \\ x, & if\ x \neq j \end{cases}$$

and let $\alpha_{i,j} := \phi_{i,j}(\alpha)$.

Using the above concepts of $\phi_{i,j}$ and $\alpha_{i,j}$, we can now prove that, unfortunately, it is impossible to give a characteristic condition on those words that have an unambiguous alphabet reduction if this condition does not incorporate the size of the target alphabet Δ of the alphabet reduction:

Theorem 4. *For every $k \in \mathbb{N}$ and for every alphabet $\Delta \subset \mathbb{N}$ with $|\Delta| \leq k$, there exist an $\alpha_k \in \mathbb{N}^+$ and an alphabet $\Delta' \subset \mathbb{N}$ with $k < |\Delta'| < |\mathrm{symb}(\alpha_k)|$ such that*

- *there is no 1-uniform morphism $\psi : \mathbb{N}^* \to \Delta^*$ that is unambiguous with respect to α_k and*
- *there is a 1-uniform morphism $\psi : \mathbb{N}^* \to \Delta'^*$ that is unambiguous with respect to α_k.*

Proof. In order to support the understanding of the elements of the proof, Example 5 can be consulted.

Let

$$\alpha_1 := 1 \cdot 2^2 \cdot 3^2 \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 2^2 \,,$$

and, for every $k \geq 2$,

$$\alpha_k := x_k \cdot (x_k + 1)^2 \cdot (x_k + 2)^2 \cdot [\ldots] \cdot (x_k + k)^2 \cdot \alpha_{k-1} \cdot$$
$$x_k \cdot (x_k + 1)^2 \cdot (x_k + 2)^2 \cdot [\ldots] \cdot (x_k + k)^2 \cdot \alpha_{k-1} \cdot (x_k + 1)^2 \,,$$

where $x_k = \max(\text{symb}(\alpha_{k-1}))+1$, and all superscripts refer to the concatenation.

We now show that every 1-uniform morphism $\psi : \mathbb{N}^* \to \Delta^*$ is ambiguous with respect to α_k. For the sake of a convenient reasoning, we define $x_0 = 2$ and $x_1 := 1$. The ambiguity of all such ψ almost directly results from the following fact:

Claim 1. Let $k \in \mathbb{N}$, let Δ be an alphabet, and let $\psi : \mathbb{N}^* \to \Delta^*$ be a morphism. If there exist distinct letters $y, z \in \{x_0+1, x_1+1, \ldots, x_k+1\}$ satisfying $\psi(y) = \psi(z)$, then ψ is ambiguous with respect to α_k.

Let now $\psi : \mathbb{N}^* \to \Delta^*$ be any 1-uniform morphism. As stated by the Theorem, Δ consists of at most k letters. On the other hand, the set $\{x_0+1, x_1+1, \ldots, x_k+1\}$ consists of $k+1$ distinct letters. Hence, ψ must map at least two of these letters to the same image. According to Claim 1, this means that ψ is ambiguous with respect to α_k.

We now give the proof of the second statement of Theorem 4. Hence we need to find an alphabet Δ' with $k < |\Delta'| < |\text{symb}(\alpha_k)|$ and a 1-uniform morphism $\psi_k : \mathbb{N}^* \to \Delta'^*$ that is unambiguous with respect to α_k.

Our reasoning is based on the following observation:

Claim 2. For any $k \geq 1$, α_k is not a fixed point of a nontrivial morphism.

We now consider the morphism $\psi_k : \mathbb{N}^* \to \mathbb{N}^*$ that is given by $\psi_k := \phi_{x_k, x_k+1}$ (see Definition 3), i.e., $\psi_k(x_k + 1) := x_k$, and ψ_k is the identity otherwise. Consequently $\psi_k(\alpha_k) = x_k^3 \beta_1 x_k^3 \beta_2 x_k^2$ with $\beta_1, \beta_2 \in (\mathbb{N} \setminus \{x_k\})^+$. We note that ψ_k maps the word α_k to a word over an alphabet Δ' satisfying $k < |\Delta'| = |\text{symb}(\alpha_k)| - 1 < |\text{symb}(\alpha_k)|$, and we shall demonstrate that ψ_k is unambiguous with respect to α_k.

We begin with an observation that imposes some restrictions on any morphism ψ that maps α_k to the same image as ψ_k:

Claim 3. Let ψ be a morphism that satisfies $\psi(\alpha_k) = \psi_k(\alpha_k)$. Then $\psi(x_k) = x_k = \psi(x_k + 1)$.

We now assume to the contrary that there exists a morphism $\psi : \mathbb{N}^* \to \Delta'^*$ satisfying $\psi(\alpha_k) = \psi_k(\alpha_k)$ and, for an $x \in \text{symb}(\alpha_k)$, $\psi(x) \neq \psi_k(x)$. From Claim 3, we know that $x \notin \{x_k, x_k + 1\}$. If $k = 1$, then we immediately obtain a contradiction, since

 − α_1 contains just three different letters,

- x_k and $x_k + 1$ satisfy $\psi(x_k) = \psi_k(x_k)$ and $\psi(x_k + 1) = \psi_k(x_k + 1)$, and
- if there is an x with $\psi(x) \neq \psi_k(x)$, then obviously there must also be an x' with $x' \neq x$ and $\psi(x') \neq \psi_k(x')$.

For $k \geq 2$, we define a morphism $\phi : \mathbb{N}^* \to \mathbb{N}^*$ by

$$\phi(x) := \begin{cases} \psi(x), & x \in \mathrm{symb}(\alpha_k) \setminus \{x_k, x_k + 1\}, \\ x & \text{else.} \end{cases}$$

Due to Claim 3 and due to $\psi(\alpha_k) = \psi_k(\alpha_k)$, we can conclude that

$$\phi((x_k + 2)^2 \cdot [\ldots] \cdot (x_k + k)^2 \cdot \alpha_{k-1}) = (x_k + 2)^2 \cdot [\ldots] \cdot (x_k + k)^2 \cdot \alpha_{k-1}. \quad (1)$$

Because of $\psi(x) \neq \psi_k(x)$ for an $x \in \mathrm{symb}(\alpha_k) \setminus \{x_k, x_k + 1\}$, and since ψ_k for all these letters is the identity, we know that ϕ is nontrivial. Furthermore, (1) implies that $\phi(\alpha_k) = \alpha_k$. Consequently, if ψ_k is ambiguous with respect to α_k, then α_k is a fixed point of a nontrivial morphism, and this contradicts Claim 2. Therefore, the second statement of Theorem 4 is correct. $\qquad \square$

The following example shows the structure of α_k in the above theorem:

Example 5. The words α_k, $2 \leq k$, look as follows:

$$\alpha_2 := 4 \cdot 5^2 \cdot 6^2 \cdot \overbrace{1 \cdot 2^2 \cdot 3^2 \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 2^2}^{\alpha_1} \cdot$$
$$4 \cdot 5^2 \cdot 6^2 \cdot \overbrace{1 \cdot 2^2 \cdot 3^2 \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 2^2}^{\alpha_1} \cdot 5^2,$$
$$\alpha_3 := 7 \cdot 8^2 \cdot 9^2 \cdot 10^2 \cdot \overbrace{4 \cdot 5^2 \cdot 6^2 \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 2^2}^{\text{prefix of } \alpha_2} \cdot$$
$$\overbrace{4 \cdot 5^2 \cdot 6^2 \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 2^2 \cdot 5^2}^{\text{suffix of } \alpha_2} \cdot$$
$$7 \cdot 8^2 \cdot 9^2 \cdot 10^2 \cdot \overbrace{4 \cdot 5^2 \cdot 6^2 \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 2^2}^{\text{prefix of } \alpha_2} \cdot$$
$$\overbrace{4 \cdot 5^2 \cdot 6^2 \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 2^2 \cdot 5^2}^{\text{suffix of } \alpha_2} \cdot 8^2,$$

and so on. The symbols x_0, x_1, x_2, x_3, and x_4 stand for the letters 2, 1, 4, 7, and 11, respectively. $\qquad \diamond$

Let 1-UNAMB$_k$ be the set of all words that have an unambiguous 1-uniform morphism $\psi : \mathbb{N}^* \to \Delta_k^*$ with $|\Delta_k| = k$. Using this concept, we now describe the above mentioned consequence of Theorem 4 that needs to be accounted for when studying a characterisation of those words that have an unambiguous 1-uniform morphism:

Corollary 6. *For every $k \in \mathbb{N}$ there exists a $k' \in \mathbb{N}$ with $k' > k$ such that 1-UNAMB$_k \subset$ 1-UNAMB$_{k'}$.*

Hence, similarly to Schneider's [10] insight into the existence of unambiguous *erasing* morphisms, any characteristic condition on those words that have unambiguous 1-uniform morphisms needs to distinguish between infinitely many different sizes of the target alphabet Δ_k. In this regard, the condition must therefore be more complex than the said main results by Freydenberger et al. [3, 2].

Still, Theorem 4 is somewhat weaker than the result by Schneider [10], who shows that for any $k \in \mathbb{N}$, the set of words that have an unambiguous erasing morphism with a target alphabet of size k is a proper subset of those words that have an unambiguous erasing morphism with a target alphabet of size $k + 1$. If Theorem 4 is meant to be strengthened (hence stating 1-UNAMB$_k \subset$ 1-UNAMB$_{k+1}$), then a number of possibly complex technical challenges arise. For example, the morphic images of the letters $x_k + 1, x_k + 2, \ldots, x_k + k$ must be carefully chosen in order to avoid squares, and this choice, in turn, might facilitate more complex types of unambiguity.

As mentioned before, Theorem 4 and Corollary 6 suggest that, for fixed target alphabets, Problem 1 might be extremely hard. In contrast to this, for variable target alphabets (i. e., the size of the target alphabet depends on the number of letters in the given word), [7] conjectures that the problem has a nice and compact solution:

Conjecture 7 (Nevisi and Reidenbach [7]). Let α be a word with $|\text{symb}(\alpha)| \geq 4$. There exist $i, j \in \text{symb}(\alpha)$, $i \neq j$, such that $\phi_{i,j}$ (see Definition 3) is unambiguous with respect to α if and only if α is morphically primitive.

While we are unable to prove or refute this conjecture, we can point out that it shows some connections to Problem 2. These shall be discussed in the next section.

4 Alphabet Reductions Preserving Morphic Primitivity

We now turn our attention to Problem 2, i. e., we study whether there exists an alphabet reduction that maps a morphically primitive word to a morphically primitive word.

We start with a general observation, that links the research on ambiguity morphism to the question of whether a morphic image is morphically primitive:

Proposition 8. *Let $\alpha \in \mathbb{N}^+$. If $\phi : \mathbb{N}^* \to \mathbb{N}^*$ is unambiguous with respect to α then $\phi(\alpha)$ is morphically primitive.*

In general, the converse of the above proposition does not hold true. For example, let $\alpha := 1 \cdot 2 \cdot 3 \cdot 4 \cdot 4 \cdot 3 \cdot 1 \cdot 2$. Thus, $\phi_{1,2}(\alpha) = 1 \cdot 1 \cdot 3 \cdot 4 \cdot 4 \cdot 3 \cdot 1 \cdot 1$ which is morphically primitive. However, $\phi_{1,2}$ is ambiguous with respect to α, because we can define a morphism φ satisfying $\varphi(\alpha) = \phi_{1,2}(\alpha)$ by $\varphi(1) := \phi_{1,2}(1) \cdot \phi_{1,2}(1)$, $\varphi(2) := \varepsilon$, $\varphi(3) := \phi_{i,j}(3)$ and $\varphi(4) := \phi_{i,j}(4)$.

If Conjecture 7 is correct, then Problem 2 can be answered in the affirmative. This is a direct consequence of the following application of Proposition 8:

Corollary 9. *Let $\alpha \in \mathbb{N}^+$ and assume that there exist $i, j \in \mathrm{symb}(\alpha)$, $i \neq j$, such that $\phi_{i,j}$ is unambiguous with respect to α. Then, $\alpha_{i,j}$ is morphically primitive.*

Hence, if Conjecture 7 is correct, then it is stronger than Proposition 8.

The above approach does not only facilitate a direct application of our results in [7] on the existence of unambiguous 1-uniform morphisms to Problem 2, but it also has the advantage of providing a chance of a constructive method that might reveal which letters to map to the same image in an alphabet reduction that preserves morphic primitivity. However, since we are unable to prove Conjecture 7, we now present in Theorem 12 below a non-constructive answer to Problem 2. This is based on two lemmata, the first of which is a basic insight into fixed points of nontrivial morphisms:

Lemma 10. *Let α be a fixed point of a nontrivial morphism. Then there exists a nontrivial morphism $\phi : \mathrm{symb}(\alpha)^* \to \mathrm{symb}(\alpha)^*$ such that $\phi(\alpha) = \alpha$ and, for every $x \in \mathrm{symb}(\alpha)$, if $\phi(x) \neq \varepsilon$, then $x \sqsubseteq \phi(x)$.*

Using Lemma 10, we can now prove the following technical observation on the pattern $\alpha_{i,j}$ as introduced in Definition 3, which is required in the proof of Theorem 12:

Lemma 11. *Let α be a word that is not a fixed point of a nontrivial morphism. For any $i, j \in \mathrm{symb}(\alpha)$, $i \neq j$, if $\alpha_{i,j}$ is a fixed point of a nontrivial morphism $\phi : \mathrm{symb}(\alpha)^* \to \mathrm{symb}(\alpha)^*$, then $\phi(i) = \varepsilon$.*

We now provide a comprehensive and affirmative answer to Problem 2 for all alphabets that have at least six distinct letters. As mentioned above, our corresponding proof is non-constructive, which means that it does not provide any direct insights into the character of alphabet reductions that preserve morphic primitivity. On the other hand, the applicability of our technique to Billaud's Conjecture (see below) can therefore easily be examined, and the fact that it is not applicable allows some conclusions to be drawn on the complexity of that Conjecture.
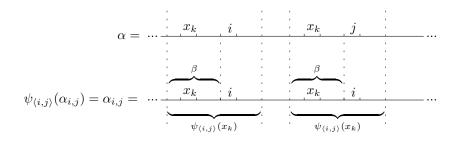
Theorem 12. *Let α be a word with $|\mathrm{symb}(\alpha)| > 5$. If α is morphically primitive, then there exist $i, j \in \mathrm{symb}(\alpha)$, $i \neq j$, such that $\alpha_{i,j}$ is morphically primitive.*

Proof. Assume to the contrary that, for every $i, j \in \mathrm{symb}(\alpha)$, $\alpha_{i,j}$ is morphically imprimitive, or in other words, $\alpha_{i,j}$ is a fixed point of a nontrivial morphism. Therefore, due to Lemma 10, for every i, j, there exists a nontrivial morphism $\psi_{\langle i,j \rangle} : \mathrm{symb}(\alpha)^* \to \mathrm{symb}(\alpha)^*$ satisfying $\psi_{\langle i,j \rangle}(\alpha_{i,j}) = \alpha_{i,j}$ and, for every $x \in \mathrm{symb}(\alpha_{i,j})$, if $\psi_{\langle i,j \rangle}(x) \neq \varepsilon$, then $x \sqsubseteq \psi_{\langle i,j \rangle}(x)$. On the other hand, it results from Lemma 11 that $\psi_{\langle i,j \rangle}(i) = \varepsilon$. Consequently, for every occurrence of i in $\alpha_{i,j}$, there exists a letter $x \in \mathrm{symb}(\alpha_{i,j}) \setminus \{i\}$ with $i \sqsubseteq \psi_{\langle i,j \rangle}(x)$ and $x \sqsubseteq \psi_{\langle i,j \rangle}(x)$. We assume that there exist m different letters x in $\alpha_{i,j}$ and we denote them by $x_1, x_2, [...], x_m$. Since α is not a fixed point of a nontrivial morphism, for every k, $1 \leq k \leq m$, $|\alpha_{i,j}|_{x_k} \geq 2$. As a result, for every k, $1 \leq k \leq m$, $|\psi_{\langle i,j \rangle}(\alpha_{i,j})|_{\psi_{\langle i,j \rangle}(x_k)} \geq 2$.

Claim. There exists an x_k, $1 \leq k \leq m$, with at least two occurrences of $\psi_{\langle i,j \rangle}(x_k)$ in $\psi_{\langle i,j \rangle}(\alpha_{i,j})$ such that

- one of them contains an occurrence of i as nth letter, $1 \leq n \leq |\psi_{\langle i,j \rangle}(x_k)|$, which is at the same position in $\alpha_{i,j}$ as an occurrence of i in α, and
- the other one contains an occurrence of i as nth letter, which is at the same position in $\alpha_{i,j}$ as an occurrence of j in α.

We illustrate the Claim in the following diagram, where β is a prefix of $\psi_{\langle i,j \rangle}(x_k)$ with length $(n-1)$.

Proof(Claim). We denote those occurrences of i in $\alpha_{i,j}$ that are at the same positions as j in α with i_j. We assume to the contrary that there does not exist any x_k, $1 \leq k \leq m$, with at least two occurrences of $\psi_{\langle i,j \rangle}(x_k)$ in $\psi_{\langle i,j \rangle}(\alpha_{i,j})$ satisfying the following conditions:

- one of them contains an occurrence of i as nth letter, $1 \leq n \leq |\psi_{\langle i,j \rangle}(x_k)|$, and
- the other one contains an occurrence of i_j as nth letter.

Let X_j be a set of those letters $q \in \mathrm{symb}(\alpha_{i,j}) \setminus \{i\}$ satisfying $|\psi_{\langle i,j \rangle}(q)| \geq 2$ and $i_j \sqsubset \psi_{\langle i,j \rangle}(q)$. Due to the above conditions, there does not exist any $q' \in X_j$ with at least two occurrences of $\psi_{\langle i,j \rangle}(q')$ in $\psi_{\langle i,j \rangle}(\alpha_{i,j})$ such that one of them contains an occurrence of i at the same position as an occurrence of i_j in the other one. Therefore, we can define a nontrivial morphism $\phi : \mathrm{symb}(\alpha)^* \to \mathrm{symb}(\alpha)^*$ over α by, for every $y \in \mathrm{symb}(\alpha)$,

$$\phi(y) := \begin{cases} \varepsilon, & y = j, \\ \varphi_{\langle i,j \rangle}(\psi_{\langle i,j \rangle}(y)), & y \in X_j, \\ \psi_{\langle i,j \rangle}(y), & \text{else}, \end{cases}$$

where $\varphi_{\langle i,j \rangle} : \mathbb{N}^* \to \mathbb{N}^*$ is a morphism with, for every $y' \in \mathrm{symb}(\alpha_{i,j})$,

$$\varphi_{\langle i,j \rangle}(y') = \begin{cases} j, & y' = i_j, \\ y', & \text{else}. \end{cases}$$

Due to $\psi_{\langle i,j \rangle}(i) = \varepsilon$, because of the definition of $\varphi_{\langle i,j \rangle}$, and since there does not exist any x_k, $1 \leq k \leq m$, satisfying the above mentioned conditions, it can be verified that $\phi(\alpha) = \alpha$, which contradicts the fact that α is not a fixed point of a nontrivial morphism. Therefore, the Claim holds true. *q.e.d.(Claim)*

Henceforth, we denote those occurrences of i in $\psi_{\langle i,j \rangle}(x_k)$ satisfying the conditions of the Claim by i'. Consequently, according to the Claim, there exists an x_k, $1 \leq k \leq m$, with $i' \sqsubseteq \psi_{\langle i,j \rangle}(x_k)$. Furthermore, if we wish to refer to the relation between x_k on the one hand and the letters i, j on the other hand as described by the Claim, we say that x_k is *responsible for* the pair $\langle i,j \rangle$.

We now study the following question: Is x_k responsible for any pair of letters of α except $\langle i,j \rangle$ (we do not distinguish between the pairs $\langle i,j \rangle$ and $\langle j,i \rangle$, in other words, $\langle i,j \rangle$ and $\langle j,i \rangle$ are the same pairs)? If the answer is yes, for how many pairs can this happen?

In order to answer this question, we consider the following cases:

1. The letter i' occurs to the right of x_k in $\psi_{\langle i,j \rangle}(x_k)$. So, we can assume that $\alpha = ... \cdot \alpha_1 \cdot x_k \cdot \alpha_2 \cdot i \cdot \alpha_3 \cdot ... \cdot \alpha_4 \cdot x_k \cdot \alpha_5 \cdot j \cdot \alpha_6 \cdot ...$, where, for every k', $1 \leq k' \leq 6$, $\alpha_{k'} \in \mathrm{symb}(\alpha)^*$, and $\psi_{\langle i,j \rangle}(x_k) := \beta_1 \cdot x_k \cdot \beta_2 \cdot i' \cdot \beta_3$, $\beta_1, \beta_2, \beta_3 \in \mathrm{symb}(\alpha_{i,j})^*$.

We now examine the mentioned question for the pair $\langle l,r \rangle$, $l, r \in \mathrm{symb}(\alpha)$ and $\langle l,r \rangle \neq \langle i,j \rangle$, by assuming that $\alpha_{l,r}$ is a fixed point of a nontrivial morphism $\psi_{\langle l,r \rangle}$. According to our discussion for $\langle i,j \rangle$, if x_k is responsible for $\langle l,r \rangle$, we need to have l' (defined analogously to i') in $\psi_{\langle l,r \rangle}(x_k)$.
We assume that l' occurs to the right of x_k in $\psi_{\langle l,r \rangle}(x_k)$. Therefore, one of the following cases needs to be satisfied:

- l' occurs to the right of i'. As a result, due to $\langle l,r \rangle \neq \langle i,j \rangle$, in one occurrence of $\psi_{\langle l,r \rangle}(x_k)$ in $\psi_{\langle l,r \rangle}(\alpha_{l,r})$, we have an occurrence of i, and in the other occurrence of $\psi_{\langle l,r \rangle}(x_k)$ at the same position as i, we have j, which is a contradiction.
- l' occurs in β_2. Then, because of $\langle l,r \rangle \neq \langle i,j \rangle$, there exists an occurrence of $\psi_{\langle i,j \rangle}(x_k)$ in $\psi_{\langle i,j \rangle}(\alpha_{i,j})$ such that its β_2 factor is different from the factor β_2 of the other occurrences of $\psi_{\langle i,j \rangle}(x_k)$ in $\psi_{\langle i,j \rangle}(\alpha_{i,j})$, which is again a contradiction.
- l' occurs at the same position as i'. However, this contradicts the fact that $\langle l,r \rangle \neq \langle i,j \rangle$.

Consequently, x_k can be responsible for $\langle l,r \rangle$ iff l' occurs to the left of x_k in $\psi_{\langle l,r \rangle}(x_k)$. By investigating the responsibility of x_k for any other pair of letters $\langle q,z \rangle$, $q, z \in \mathrm{symb}(\alpha)$, $\langle q,z \rangle \neq \langle i,j \rangle$ and $\langle q,z \rangle \neq \langle l,r \rangle$, we can conclude with the same reasoning as above that q' cannot occur to the right of x_k in $\psi_{\langle q,z \rangle}(x_k)$. Also, by assuming that l' occurs to the left of x_k in $\psi_{\langle l,r \rangle}(x_k)$, an analogous reasoning as above leads to the fact that q' cannot

occur to the left of x_k in $\psi_{\langle q,z \rangle}(x_k)$. Consequently, x_k cannot be responsible for any other pairs $\langle q,z \rangle$, $q,z \in \mathrm{symb}(\alpha)$, $\langle q,z \rangle \neq \langle i,j \rangle$ and $\langle q,z \rangle \neq \langle l,r \rangle$.

2. The letter i' occurs to the left of x_k in $\psi_{\langle i,j \rangle}(x_k)$. An analogous reasoning to that in the previous case implies that, firstly, x_k can be responsible for another pair of letters $\langle l,r \rangle$, $\langle l,r \rangle \neq \langle i,j \rangle$, iff l' occurs to the right of x_k in $\psi_{\langle l,r \rangle}(x_k)$. Secondly, x_k is not responsible for any other pairs $\langle q,z \rangle$, $q,z \in \mathrm{symb}(\alpha)$, $\langle q,z \rangle \neq \langle i,j \rangle$ and $\langle q,z \rangle \neq \langle l,r \rangle$.

Consequently, due to the above cases, we can conclude that every letter $x \in \alpha$ can at most be responsible for two pairs of letters. On the other hand, if $|\mathrm{symb}(\alpha)| = n$, the number of pairs of letters of α is $\binom{n}{2}$. Referring to the assumption of the theorem, $n > 5$. Therefore,

$$\binom{n}{2} > 2 * n.$$

This implies that there is a word $\alpha_{i,j}$, $i,j \in \mathrm{symb}(\alpha)$ such that there does not exist any letter $x \in \mathrm{symb}(\alpha_{i,j}) \setminus \{i\}$ that is responsible for the pair $\langle i,j \rangle$, which is a contradiction to the Claim. Thus, there exist letters $i,j \in \mathrm{symb}(\alpha)$ such that $\alpha_{i,j}$ is morphically primitive. $\qquad\square$

Since morphically primitive words are equivalent to those words that are not a fixed point of a nontrivial morphism, Theorem 12 shows that the structural property of a word α that eliminates the existence of a nontrivial morphism ψ satisfying $\psi(\alpha) = \alpha$ is strong enough to also eliminate the existence of a nontrivial morphism ψ' satisfying $\psi'(\phi_{i,j}(\alpha)) = \phi_{i,j}(\alpha)$ for an appropriate choice of the alphabet reduction $\phi_{i,j}$ (see Definition 3). However, if we consider a different notion of an alphabet reduction, namely a morphism $\delta_i : \mathbb{N}^* \to \mathbb{N}^*$ defined by $\delta_i(i) := \varepsilon$ and $\delta_i(x) := x$ for $x \in \mathbb{N} \setminus \{i\}$, then Theorem 12 and its proof are not sufficient to establish a result that is equivalent to Theorem 12. Hence, we have to study Billaud's Conjecture separately:

Conjecture 13 (Billaud [1]). Let α be a word with $|\mathrm{symb}(\alpha)| \geq 3$. If α is not a fixed point of a nontrivial morphism, then there exists an $i \in \mathrm{symb}(\alpha)$ such that $\delta_i(\alpha)$ is not a fixed point of a nontrivial morphism.

Levé and Richomme [6] provide a confirmation of the contraposition of Conjecture 13 for a special case, but, apart from that, little is known about this problem. The final result of our paper shall demonstrate that Conjecture 13 is correct if words are considered that contain each of their letters exactly twice:

Theorem 14. *Let α be a word with $|\mathrm{symb}(\alpha)| \geq 3$ that is not a fixed point of a nontrivial morphism. If, for every $x \in \mathrm{symb}(\alpha)$, $|\alpha|_x = 2$, then there exists an $i \in \mathrm{symb}(\alpha)$ such that $\delta_i(\alpha)$ is not a fixed point of a nontrivial morphism.*

We expect that even a moderate extension of Theorem 14 would require a substantially more involved reasoning. We therefore conclude that the actual nature of morphically primitive words, despite our almost comprehensive result in Theorem 12 and the strong insights that are due to Head [4] and Holub [5], is not

really understood. This view is further substantiated by the fact that another property of morphically primitive words, namely their frequency, is largely unresolved as well (see Reidenbach and Schneider [8]).

References

1. M. Billaud. A problem with words. Letter in Newsgroup Comp.Theory, 1993. `https://groups.google.com/d/topic/comp.theory/V_xDDtoR9a4/discussion`.
2. D.D. Freydenberger, H. Nevisi, and D. Reidenbach. Weakly unambiguous morphisms. *Theoretical Computer Science*. To appear.
3. D.D. Freydenberger, D. Reidenbach, and J.C. Schneider. Unambiguous morphic images of strings. *International Journal of Foundations of Computer Science*, 17:601–628, 2006.
4. T. Head. Fixed languages and the adult languages of 0L schemes. *International Journal of Computer Mathematics*, 10:103–107, 1981.
5. S. Holub. Polynomial-time algorithm for fixed points of nontrivial morphisms. *Discrete Mathematics*, 309:5069–5076, 2009.
6. F. Levé and G. Richomme. On a conjecture about finite fixed points of morphisms. *Theoretical Computer Science*, 339:103–128, 2005.
7. H. Nevisi and D. Reidenbach. Unambiguous 1-uniform morphisms. In *Proc. 8th International Conference on Words, WORDS 2011*, volume 63 of *EPTCS*, pages 158–167, 2011.
8. D. Reidenbach and J.C. Schneider. Morphically primitive words. *Theoretical Computer Science*, 410:2148–2161, 2009.
9. D. Reidenbach and J.C. Schneider. Restricted ambiguity of erasing morphisms. *Theoretical Computer Science*, 412:3510–3523, 2011.
10. J.C. Schneider. Unambiguous erasing morphisms in free monoids. *RAIRO – Theoretical Informatics and Applications*, 44:193–208, 2010.