Unambiguous 1-Uniform Morphisms

Hossein Nevisi[*]   Daniel Reidenbach

Department of Computer Science
Loughborough University
Loughborough, LE11 3TU, United Kingdom

H.Nevisi@lboro.ac.uk   D.Reidenbach@lboro.ac.uk

A morphism σ is unambiguous with respect to a word α if there is no other morphism τ that maps α to the same image as σ. In the present paper we study the question of whether, for any given word, there exists an unambiguous 1-uniform morphism, i.e., a morphism that maps every letter in the word to an image of length 1.

1 Introduction

If, for a morphism $\sigma : \Delta^* \to \Sigma^*$ (where Δ and Σ are arbitrary alphabets) and a word $\alpha \in \Delta^*$, there exists another morphism τ mapping α to $\sigma(\alpha)$, then σ is called *ambiguous* with respect to α; if such a τ does not exist, then σ is *unambiguous*. For example, the morphism $\sigma_0 : \{A,B,C\}^* \to \{a,b\}^*$ – given by $\sigma_0(A) := a$, $\sigma_0(B) := a$, $\sigma_0(C) := b$ – is ambiguous with respect to the word $\alpha_0 := ABCACB$, since the morphism τ_0 – defined by $\tau_0(A) := \varepsilon$ (i.e., τ_0 maps A to the empty word), $\tau_0(B) := a$, $\tau_0(C) := ab$ – satisfies $\tau_0(\alpha_0) = \sigma_0(\alpha_0)$ and, for a symbol X occuring in α, $\tau_0(X) \neq \sigma_0(X)$:

$$\sigma_0(\alpha_0) = \overbrace{\underbrace{a}_{\tau_0(B)}\ \underbrace{a\ \ b}_{\tau_0(C)}}^{\sigma_0(A)\ \ \sigma_0(B)\ \ \sigma_0(C)}\overbrace{\underbrace{a\ \ b}_{\tau_0(C)}\ \underbrace{a}_{\tau_0(B)}}^{\sigma_0(A)\ \ \sigma_0(C)\ \ \sigma_0(B)} = \tau_0(\alpha_0).$$

It can be verified with moderate effort that, e.g., the morphism $\sigma_1 : \{A,B,C\}^* \to \{a,b\}^*$ – given by $\sigma_1(A) := a$, $\sigma_1(B) := ab$, $\sigma_1(C) := b$ – is unambiguous with respect to α_0.

The potential ambiguity of morphisms is relevant to various concepts in the combinatorial theory of morphisms, such as pattern languages (see, e. g., Mateescu and Salomaa [9]), equality sets (see, e g., Harju and Karhumäki [6]) and word equations (see, e. g., Choffrut [2]). This relation is best understood for inductive inference of pattern languages, where it has been shown that a preimage can be computed from some of its morphic images if and only if these images have been generated by morphisms with a restricted ambiguity (see, e. g., Reidenbach [10]). Hence, intuitively speaking, unambiguous morphisms have a desirable, namely structure-preserving, property in such a context, and therefore previous literature on the ambiguity of morphisms mainly studies the question of the *existence* of unambiguous morphisms for arbitary words. In the initial paper, Freydenberger, Reidenbach and Schneider [5] show that there exists an unambiguous nonerasing morphism with respect to a word α if and only if α is not a *fixed point* of a nontrivial morphism, i. e., there is no morphism ϕ satisfying $\phi(\alpha) = \alpha$ and, for a symbol x in α, $\phi(x) \neq x$. Freydenberger and Reidenbach [4] study those sets of words with respect to which so-called segmented morphisms are unambiguous, and these results lead to a refinement of the techniques used in [5]. Schneider [13] and Reidenbach and Schneider [12] investigate the existence of unambiguous erasing morphisms – i. e., morphisms that may map symbols to the empty word. Finally, Freydenberger,

---

Nevisi and Reidenbach [3] study a definition of unambiguity that is completely restricted to nonerasing morphisms[1], and they provide a characterisation of those words with respect to which there exist unambiguous morphisms $\sigma : \Delta^+ \to \Sigma^+$ in such a context (this characterisation does not hold for binary target alphabets Σ, though).

In the present paper, we study the existence of unambiguous *1-uniform* morphisms for arbitrary words, i. e., just as our initial example σ_0, these morphisms map every symbol in the preimage to an image of length 1. In order to obtain unrestricted results, we wish to consider words over an unbounded alphabet Δ as morphic preimages. Therefore, we assume $\Delta := \mathbb{N}$; in accordance with the existing literature in the field, we call any word $\alpha \in \mathbb{N}^*$ a *pattern*, and we call any symbol $x \in \mathbb{N}$ occurring in α a *variable*. Thus, more formally, we wish to investigate the following problem:

Problem 1. *Let $\alpha \in \mathbb{N}^*$ be a pattern, and let Σ be an alphabet. Does there exists a 1-uniform morphism $\sigma : \mathbb{N}^* \to \Sigma^*$ that is unambiguous with respect to α, i. e., there is no morphism $\tau : \mathbb{N}^* \to \Sigma^*$ satisfying $\tau(\alpha) = \sigma(\alpha)$ and, for a variable x occurring in α, $\tau(x) \neq \sigma(x)$?*

There are two main reasons why we study this question: Firstly, any insight into the existence of unambiguous 1-uniform morphisms improves the construction by Freydenberger et al. [5], which provides comprehensive results on the existence of unambiguous nonerasing morphisms, but is based on morphisms that are often much more involved than required. This can be illustrated using our above example pattern α_0 (now interpreted as $\alpha_0 := 1 \cdot 2 \cdot 3 \cdot 1 \cdot 3 \cdot 2$ in order to fit with the definition of patterns as words over $\mathbb{N}$). Here, the unambiguous morphism σ_1 – which is not 1-uniform, but still of very limited complexity – produces a morphic image of length 8, whereas the unambiguous morphism for α_0 defined in [5] leads to a morphic image of length 162. This substantial complexity of known unambiguous morphisms has a severe effect on the runtime of inductive inference procedures for pattern languages, which, as mentioned above, are necessarily based on such morphisms. Thus, any insight into the existence of uncomplex unambiguous morphisms is not only of intrinsic interest, but is also important from a more applied point of view. Secondly, as shown by $\sigma_0(\alpha_0)$, the images under 1-uniform morphisms have a structure that is very close to that of their preimages. This is because, whenever the pattern contains more different variables than there are letters in the target alphabet, a 1-uniform morphism reduces the complexity of the preimage by mapping certain variables to the same image. Thus, such a *morphic simplification* and its potential ambiguity are a very basic phenomenon in the combinatorial theory of morphisms. Our studies shall suggest that Problem 1 is nevertheless a challenging question, and we shall demonstrate that it is related to a number of other concepts and problems in combinatorics on words.

Note that, due to space constraints, this extended abstract contains just a few proofs, focussing on those that are reasonably short and suitable to illustrate our basic proof techniques.

2 Definitions and Preliminary Results

For the definitions of *patterns*, *variables*, *1-uniform morphisms*, *(un)ambiguous morphisms*, *fixed points* of nontrivial morphisms, and the symbol ε, Section 1 can be consulted.

Let A be an *alphabet*, i. e., an enumerable set of symbols. A *word (over A)* is a a finite sequence of symbols taken from A. The set A^* is the set of all words over A, and $A^+ := A^* \setminus \{\varepsilon\}$. For the *concatenation*

---

[1]Note that [5, 4] also deal with unambiguous *nonerasing* morphisms, but they use a stronger notion of unambiguity that is based on arbitrary monoid morphisms. Hence, they call a morphism σ unambiguous only if there is no other – erasing or nonerasing – morphism τ satisfying $\tau(\alpha) = \sigma(\alpha)$. In contrast to this, and in contrast to the present paper, [3] disregards erasing morphisms τ. Consequently, in the definition of unambiguity studied by [3], our initial example σ_0 is considered ("weakly") *unambiguous* with respect to α_0, since all morphisms τ with $\tau(\alpha_0) = \sigma_0(\alpha_0)$ are erasing morphisms.

of two words w_1, w_2, we write $w_1 \cdot w_2$ or simply $w_1 w_2$. The notion $|x|$ stands for the size of a set x or the length of a word x. For any word $w \in A^*$, the notation $|w|_x$ stands for the number of occurrences of the letter x in w. 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 $\alpha = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5$. We call a word $v \in A^*$ a *factor* of a word $w \in A^*$ if, for some $u_1, u_2 \in A^*$, $w = u_1 v u_2$; moreover, if v is a factor of w then we say that w contains v and denote this by $v \sqsubseteq w$ or $w = \cdots v \cdots$. If $v \neq w$, then we say that v is a *proper* factor of w and denote this by $v \sqsubset w$. If $u_1 = \varepsilon$, then v is a *prefix* of w, and if $u_2 = \varepsilon$, then v is a *suffix* of w. For every letter x in w, $L_x := \{y \in A \mid w = \cdots y \cdot x \cdots\} \cup L_x'$ and $R_x := \{y \in A \mid w = \cdots x \cdot y \cdots\} \cup R_x'$, where $L_x' = \{\varepsilon\}$ if $w = x \cdots$ and $L_x' = \emptyset$ if $w \neq x \cdots$, and $R_x' = \{\varepsilon\}$ if $w = \cdots x$ and $R_x' = \emptyset$ if $w \neq \cdots x$. We refer to the sets L_x and R_x as *neighbourhood sets*.

For alphabets A, B, a mapping $h : A^* \to B^*$ is a *morphism* if h is compatible with the concatenation, i. e., for all $v, w \in A^*$, $h(v) \cdot h(w) = h(vw)$. We call B the *target alphabet* of h. The morphism h is said to be *nonerasing* if, for every $x \in A$, $h(x) \neq \varepsilon$. A morphism is called a *renaming* if it is injective and 1-uniform. We additionally call any word v a renaming of a word w if there is a morphism h that is a renaming and satisfies $h(w) = v$. A word $w \in A^*$ is said to be *in canonical form* if it is lexicographically minimal (with regard to any fixed order on A) among all its renamings in A^*.

With regard to an arbitrary pattern $\alpha \in \mathbb{N}^*$, $\mathrm{var}(\alpha)$ denotes the set of all variables occurring in α. If we say that a pattern is in canonical form, then this shall always refer to the usual order on $\mathbb{N}$, i. e., $1 < 2 < 3 < \ldots$.

The question of whether a pattern α is a fixed point of a nontrivial morphism (which can be decided in polynomial time, see Holub [7]) is equivalent to a number of other concepts in combinatorics on words. More precisely, α is a fixed point of a nontrivial morphism iff α is *prolix* iff α is *morphically imprimitive* iff there exist a certain characteristic factorisation of α; these equivalences are explained by Reidenbach and Schneider [11] in more detail. Results on unambiguous morphisms have been stated using any of these concepts. In the present paper, our presentation shall focus on the notion of fixed points. Therefore, we can now paraphrase a simple yet fundamental insight by Freydenberger et al. [5] – which implies that an answer to Problem 1 is trivial for those patterns that are fixed points of nontrivial morphisms – as follows:

Theorem 1 (Freydenberger et al. [5])**.** *Let $\alpha \in \mathbb{N}^*$ be a fixed point of a nontrivial morphisms, and let Σ be any alphabet. Then every nonerasing morphism $\sigma : \mathbb{N}^* \to \Sigma^*$ is ambiguous with respect to α.*

Hence, we can safely restrict our subsequent considerations to those patterns that are not fixed points.

3   Fixed Target Alphabets

In the the present section, we describe a number of conditions on the existence of unambiguous 1-uniform morphisms $\sigma : \mathbb{N}^* \to \Sigma^*$ with a *fixed* target alphabet Σ, i. e., the size of Σ does not depend on the number of variables occurring in α. While the main result by Freydenberger et al. [5] demonstrates that the set of patterns with an unambiguous *nonerasing* morphisms is independent of the size of Σ (provided that $|\Sigma| \geq 2$), our initial example α_0 and all patterns $\alpha_m := 1 \cdot 1 \cdot 2 \cdot 2 \cdot [\ldots] \cdot m \cdot m$ with $m \geq 4$ do not have an unambiguous *1-uniform* morphism $\sigma : \mathbb{N}^* \to \Sigma^*$ for binary alphabets Σ. In contrast to this, such morphisms can be given for ternary (and, thus, larger) alphabets:

Theorem 2. *Let $m \in \mathbb{N}$, $m \geq 4$, let Σ be an alphabet, and let $\alpha_m := 1 \cdot 1 \cdot 2 \cdot 2 \cdot [\ldots] \cdot m \cdot m$. There exists a 1-uniform morphism $\sigma : \mathbb{N}^* \to \Sigma^*$ that is unambiguous with respect to α_m if and only if $|\Sigma| \geq 3$.*

Proof. Since squares cannot be avoided over unary and binary alphabets, it can be shown with very limited effort that there is no unambiguous 1-uniform morphism $\sigma : \mathbb{N}^* \to \Sigma^*$ with respect to any α_m if Σ does not contain at least three letters.

According to Thue [14], there exists an infinite square-free word over a ternary alphabet. Let this word be w. Thus,

$$w = abcacbabcbacabcacbaca\cdots .$$

We define the word w' by repeating every letter of w twice. Consequently,

$$w' = aabbccaaccbbaabbccbbaaccaabbccaaccbbaaccaa\cdots .$$

We now define a 1-uniform morphism $\sigma : \mathbb{N}^* \to \{a,b,c\}^*$ such that $\sigma(\alpha_m)$ is a prefix of w'. Since w is square-free, the only square factors of w' are aa, bb and cc. Hence, it can be easily verified that σ is unambiguous with respect to α_m. □

Thus – and just as for the equivalent problem on unambiguous *erasing* morphisms (see Schneider [13]) – any characteristic condition on the existence of unambiguous 1-uniform morphisms needs to incorporate the size of Σ, which suggests that such criteria might be involved. Therefore, our results in this section are restricted to *sufficient* conditions on the existence of unambiguous 1-uniform morphisms.

Our first criterion is based on (un)avoidable patterns and is, thus, related to the above-mentioned property of the patterns α_m:

Theorem 3. *Let* $n \in \mathbb{N}$, $\beta := r_1 \cdot r_2 \cdot [\ldots] \cdot r_{\lceil n/2 \rceil}$ *and* $\alpha := 1^{r_1} \cdot 2^{r_1} \cdot 3^{r_2} \cdot 4^{r_2} \cdot [\ldots] \cdot n^{(r_{\lceil n/2 \rceil})}$ *with* $r_i \geq 2$ *for every* i, $1 \leq i \leq \lceil n/2 \rceil$. *If* β *is square-free, then there exists a 1-uniform morphism* $\sigma : \mathbb{N}^* \to \{a,b\}^*$ *that is unambiguous with respect to* α.

Our second criterion again holds for binary (and, thus, all larger) alphabets Σ. It features a rather restricted class of patterns, which, however, are minimal with regard to their length.

Theorem 4. *Let* $n \in \mathbb{N}$, $n \geq 2$. *If n is even, let*

$$\alpha := 1 \cdot 2 \cdot [\ldots] \cdot n \cdot (n/2+1) \cdot 1 \cdot (n/2+2) \cdot 2 \cdot [\ldots] \cdot n \cdot n/2,$$

and if n is odd, let

$$\alpha := 1 \cdot 1 \cdot 2 \cdot 3 \cdot [\ldots] \cdot n \cdot (\lceil n/2 \rceil + 1) \cdot 2 \cdot (\lceil n/2 \rceil + 2) \cdot 3 \cdot [\ldots] \cdot n \cdot \lceil n/2 \rceil.$$

Then α *is a shortest pattern with* $|\mathrm{var}(\alpha)| = n$ *that is not a fixed point of a nontrivial morphism, and there exists a 1-uniform morphism* $\sigma : \mathbb{N}^* \to \{a,b\}^*$ *that is unambiguous with respect to* α.

The following examples illustrates Theorem 4 and its proof: For $n := 6$, $\alpha := 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 4 \cdot 1 \cdot 5 \cdot 2 \cdot 6 \cdot 3$, and the 1-uniform morphism $\sigma : \mathbb{N}^* \to \{a,b\}^*$ with $\sigma(1) := \sigma(2) := \sigma(3) := a$ and $\sigma(4) := \sigma(5) := \sigma(6) := b$ is unambiguous with respect to α. For $n := 5$, $\alpha := 1 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 4 \cdot 2 \cdot 5 \cdot 3$, and the respective unambiguous morphism is given by $\sigma(1) := \sigma(2) := \sigma(3) := a$ and $\sigma(4) := \sigma(5) := b$.

From Theorem 4 we can conclude that patterns α with unambiguous 1-uniform morphisms using a binary target alphabet exist for every cardinality of $\mathrm{var}(\alpha)$ and that corresponding examples can be given where every variable occurs just twice.

4   Variable Target Alphabets

In order to continue our examination of Problem 1, we now relax one of the requirements of Section 3: We no longer investigate criteria on the existence of unambiguous 1-uniform morphisms for a fixed target alphabet Σ, but we permit Σ to depend on the number of variables in the pattern α in question. Regarding this question, we conjecture the following statement to be true:

Conjecture 1. *Let α be a pattern with $|\mathrm{var}(\alpha)| \geq 4$. There exists an alphabet Σ satisfying $|\Sigma| < |\mathrm{var}(\alpha)|$ and a 1-uniform morphism $\sigma : \mathbb{N}^* \to \Sigma^*$ that is unambiguous with respect to α if and only if α is not a fixed point of a nontrivial morphism.*

This conjecture would be trivially true if we allowed Σ to satisfy $|\Sigma| \geq |\mathrm{var}(\alpha)|$. That explains why we exclusively study the case where the number of letters in the target alphabet is smaller than the number of variables in the pattern. From Theorem 2, it directly follows that an analogous conjecture would not be true if we considered fixed binary target alphabets (as is done in Section 3), since none of the patterns α_m is a fixed point of a nontrivial morphism – this can be easily verified using tools discussed by Reidenbach and Schneider [11] and Holub [7]. Hence, characteristic criteria must necessarily look different in such a context. It can also be effortlessly understood that Conjecture 1 would be incorrect if we dropped the condition that α needs to contain at least 4 distinct variables, since not only σ_0, but all 1-uniform morphisms $\sigma : \mathbb{N}^* \to \Sigma^*$ with $|\Sigma| \leq 2$ are ambiguous with respect to our example pattern $\alpha_0 = 1 \cdot 2 \cdot 3 \cdot 1 \cdot 3 \cdot 2$ discussed in Section 1.

Technically, many of our subsequent technical considerations are based on the following generic morphisms:

Definition 1. *Let Σ be an infinite alphabet, and let $\sigma : \mathbb{N}^* \to \Sigma^*$ be a renaming. For any $i, j \in \mathbb{N}$ with $i \neq j$ and for every $x \in \mathbb{N}$, let the morphism $\sigma_{i,j}$ be given by*

$$\sigma_{i,j}(x) := \begin{cases} \sigma(i), & \text{if } x = j, \\ \sigma(x), & \text{if } x \neq j. \end{cases}$$

Thus, $\sigma_{i,j}$ maps exactly two variables to the same image, and therefore, for any pattern α with at least two different variables, $\sigma_{i,j}(\alpha)$ is a word over $|\mathrm{var}(\alpha)| - 1$ distinct letters. Using this definition, we can now state a more specific version of Conjecture 1:

Conjecture 2. *Let α be a pattern with $|\mathrm{var}(\alpha)| \geq 4$. There exist $i, j \in \mathrm{var}(\alpha)$, $i \neq j$, such that $\sigma_{i,j}$ is unambiguous with respect to α if and only if α is not a fixed point of a nontrivial morphism.*

As a side note, we consider it worth mentioning that Conjecture 2 shows connections to another conjecture from the literature. In order to state the latter, we define, for any $i \in \mathbb{N}$, the morphism $\delta_i : \mathbb{N}^* \to \mathbb{N}^*$ by $\delta_i(i) := \varepsilon$ and, for every $j \in \mathbb{N} \setminus \{i\}$, $\delta_i(j) := j$.

Conjecture 3 (Billaud [1], Levé and Richomme [8])**.** *Let α be a pattern with $|\mathrm{var}(\alpha)| \geq 3$. If, for every $i \in \mathrm{var}(\alpha)$, $\delta_i(\alpha)$ is a fixed point of a nontrivial morphism, then α is a fixed point of a nontrivial morphism.*

In general, the correctness of Conjecture 3 has not been established yet. The problem is intensively studied by Levé and Richomme [8], where it is shown to be correct for certain subclasses of $\mathbb{N}^*$.

Due to Theorem 1, the *only if* directions of Conjectures 1 and 2 hold true immediately. In the remainder of this section, we shall therefore exclusively study those patterns that are not fixed points. Our corresponding results yield large classes of such patterns that have an unambiguous 1-uniform morphism, but we have to leave the overall correctness of our conjectures open.

Conjecture 2 suggests that the examination of the existence of unambiguous 1-uniform morphisms for a pattern α may be reduced to finding suitable variables i and j such that $\sigma_{i,j}$ is unambiguous with respect to α. In this regard, one particular choice can be ruled out immediately:

Proposition 1. *Let α be a pattern, and let $i, j \in \mathrm{var}(\alpha)$, $i \neq j$. If $\sigma_{i,j}(\alpha)$ is a fixed point of a nontrivial morphism, then $\sigma_{i,j}$ is ambiguous with respect to α.*

For example, if we consider the pattern $\alpha_1 := 1 \cdot 2 \cdot 3 \cdot 4 \cdot 1 \cdot 4 \cdot 3 \cdot 2$ (which is not a fixed point) and define $\Sigma := \{a, b, c\}$, then $\sigma_{2,4}(\alpha_1)$ equals *abcbabcb* (or any renaming thereof), which is a fixed point of the morphism ϕ given by $\phi(a) := abcb$ and $\phi(b) := \phi(c) := \varepsilon$. Thus, $\sigma_{2,4}$ is ambiguous with respect to α_1. However, Proposition 1 does not provide a characteristic condition on the ambiguity of $\sigma_{i,j}$, since $\sigma_{2,3}(\alpha_1) = abbcacbb$ is not a fixed point, but still $\sigma_{2,3}$ is ambiguous with respect to α_1. Furthermore, while the ambiguity of $\sigma_{2,3}$ results from the fact that α_1 contains the factors $2 \cdot 3$ and $3 \cdot 2$, and is therefore easy to comprehend, there are more difficult examples of morphisms $\sigma_{i,j}$ that are ambiguous although they do not lead to a morphic image that is a fixed point. This is illustrated by the example $\alpha_2 := 1 \cdot 2 \cdot 3 \cdot 3 \cdot 4 \cdot 4 \cdot 1 \cdot 2 \cdot 3 \cdot 3 \cdot 4 \cdot 4 \cdot 2$. Here, $\sigma_{2,4}(\alpha_1) = abccbbabccbbb$ again is not a fixed point, but $\sigma_{2,4}$ is nevertheless ambiguous with respect to α_2, since the morphism τ given by $\tau(1) := abccb$, $\tau(2) := b$ and $\tau(3) := \tau(4) := \varepsilon$ satisfies $\tau(\alpha_2) = \sigma_{2,4}(\alpha_2)$. We therefore conclude that it seems not to be a straightforward task to find amendments that could turn Proposition 1 into a characteristic condition.

We now show that Conjecture 2 is correct for several types of patterns. To this end, we need the following simple sufficient condition on a pattern being a fixed point:

Lemma 1. *Let $\alpha \in \mathbb{N}^+$. If there exists a variable $i \in \mathrm{var}(\alpha)$ such that*

  1. *$\varepsilon \notin L_i$ and, for every $k \in L_i$, $R_k = \{i\}$, or*

  2. *$\varepsilon \notin R_i$ and, for every $k \in R_i$, $L_k = \{i\}$,*

then α is a fixed point of a nontrivial morphism.

Using this lemma, we can now establish a class of patterns for which Conjecture 2 holds true. All variables in these patterns have the same number of occurrences and satisfy some additional conditions:

Theorem 5. *Let $m \in \mathbb{N}$, $m \geq 2$. Let $\alpha \in \mathbb{N}^+$ be a pattern that is not a fixed point of a nontrivial morphism and satisfies, for every $x \in \mathrm{var}(\alpha)$, $|\alpha|_x = m$. If there are $i, j \in \mathrm{var}(\alpha)$, $i \neq j$, such that*

  - *there is no $k \in \mathrm{var}(\alpha)$ with $\{i, j\} \subseteq L_k$ or $\{i, j\} \subseteq R_k$, and*

  - *$\alpha \neq \alpha_1 \cdot \mathbf{i} \cdot \mathbf{j} \cdot \alpha_2 \cdot \mathbf{j} \cdot \mathbf{i} \cdot \alpha_3$, $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{N}^*$,*

then $\sigma_{i,j}$ is unambiguous with respect to α.

Proof. Assume to the contrary that $\sigma_{i,j}$ is ambiguous. So, there exists a morphism $\tau : \mathbb{N}^+ \to \Sigma^*$ satisfying $\tau(\alpha) = \sigma_{i,j}(\alpha)$ and, for some $x \in \mathrm{var}(\alpha)$, $\tau(x) \neq \sigma_{i,j}(x)$. Since $\sigma_{i,j}$ is a 1-uniform morphism, there exists a $k \in \mathrm{var}(\alpha)$ with $|\tau(k)| \geq 2$. Let $uv \sqsubseteq \tau(k)$, $u, v \in \Sigma$. Due to the fact that k occurs m times in α, $\sigma_{i,j}(\alpha) = \tau(\alpha) = w_1 \cdot uv \cdot w_2 \cdot uv \cdot [\ldots] \cdot w_m \cdot uv \cdot w_{m+1}$ with, for every q, $1 \leq q \leq m+1$, $w_q \in \Sigma^*$. We now consider the following cases:

  - $\sigma_{i,j}(i) \neq u$ and $\sigma_{i,j}(i) \neq v$. This implies that there exist the variables $x_1, x_2 \in \mathrm{var}(\alpha)$, $x_1, x_2 \neq i$ and $x_1, x_2 \neq j$, such that $\alpha = \alpha_1 \cdot x_1 x_2 \cdot \alpha_2 \cdot x_1 x_2 \cdot [\ldots] \cdot \alpha_m \cdot x_1 x_2 \cdot \alpha_{m+1}$, for every q, $1 \leq q \leq m+1$, $\alpha_q \in \mathbb{N}^*$, and $\sigma_{i,j}(x_1) = u$ and $\sigma_{i,j}(x_2) = v$. Due to $|\alpha|_{x_1} = |\alpha|_{x_2} = m$, the variables x_1, x_2 satisfy, for every q with $1 \leq q \leq m+1$, $x_1, x_2 \not\sqsubseteq \alpha_q$. This implies that $R_{x_1} = \{x_2\}$ and $L_{x_2} = \{x_1\}$. Then, according to Lemma 1, α is a fixed point of a nontrivial morphism, which is a contradiction to the assumption of the theorem.

- $\sigma_{i,j}(i) = \sigma_{i,j}(j) = u$, and $u \neq v$. So, we assume that $\alpha = \alpha_1 \cdot x_1 x' \cdot \alpha_2 \cdot x_2 x' \cdot [\ldots] \cdot \alpha_m \cdot x_m x' \cdot \alpha_{m+1}$ with, $x' \in \text{var}(\alpha)$ and, for every q, $1 \leq q \leq m+1$, $x_q \in \text{var}(\alpha)$, $\alpha_q \in \mathbb{N}^*$, and $\sigma_{i,j}(x_q) = u$ and $\sigma_{i,j}(x') = v$. Additionally, since $\sigma_{i,j}(x') = v$ and $u \neq v$, we can conclude that $x' \neq i$ and $x' \neq j$. We now consider the following cases:

  1. For every q, $1 \leq q \leq m$, $x_q = i$. This implies, using the same reasoning as above, that α is a fixed point of a nontrivial morphism which is a contradiction.
  2. There exists q, q', $1 \leq q, q' \leq m$ and $q \neq q'$, such that $x_q = i$ and $x_{q'} = j$. This means that $\{i, j\} \subseteq L_{x_2}$, which contradicts the first condition of the theorem.

- $\sigma_{i,j}(i) = v$, and $u \neq v$. The reasoning is analogous to that in the previous case.

- $\sigma_{i,j}(i) = \sigma_{i,j}(j) = u$ and $v = u$. Hence, we may assume that $\alpha = \alpha_1 \cdot x_1 x'_1 \cdot \alpha_2 \cdot x_2 x'_2 \cdot [\ldots] \cdot \alpha_m \cdot x_m x'_m \cdot \alpha_{m+1}$ with, for every q, $1 \leq q \leq m+1$, $\alpha_q \in \mathbb{N}^*$, $x_q, x'_q \in \text{var}(\alpha)$ and $\sigma_{i,j}(x_q) = \sigma_{i,j}(x'_q) = u$. Due to the conditions of the theorem, the factors $i \cdot i \cdot j$, $i \cdot j \cdot j$, $j \cdot i \cdot i$ and $j \cdot j \cdot i$ cannot be factors of α. Moreover, it must be noticed that $u \cdot u \cdot u \not\sqsubseteq \tau(k)$; otherwise, since $\tau(\alpha) = \sigma_{i,j}(\alpha)$, then $|\alpha|_i > m$ or $\alpha_j > m$. This implies that $i \cdot j \cdot i$ and $j \cdot i \cdot j$ are not factors of α. We now consider the following cases:

  1. For every q, $1 \leq q \leq m$, $x_q = i$ and $x'_q = j$. As a result, $R_i = \{j\}$ and $L_j = \{i\}$. According to Lemma 1, α is a fixed point of a nontrivial morphism.
  2. For every q, $1 \leq q \leq m$, $x_q = j$ and $x'_q = i$. Thus, $R_j = \{i\}$ and $L_i = \{j\}$, which, due to Lemma 1, again implies that α is a fixed point of a nontrivial morphism.
  3. There exists a q, q', $1 \leq q, q' \leq m$ and $q \neq q'$, such that $x_q \cdot x'_q = i \cdot j$ and $x_{q'} \cdot x'_{q'} = j \cdot i$. This case contradicts the second condition of the theorem.
  4. There exists a q, q', $1 \leq q, q' \leq m$ and $q \neq q'$, such that $x_q \cdot x'_q = i \cdot j$ and, $x_{q'} \cdot x'_{q'} = i \cdot i$ or $x_{q'} \cdot x'_{q'} = j \cdot j$. This means that $\{i, j\} \subseteq R_i$ or $\{i, j\} \subseteq L_j$, which is a contradiction to the first condition of the theorem.
  5. There exists a q, q', $1 \leq q, q' \leq m$ and $q \neq q'$, such that $x_q \cdot x'_q = j \cdot i$ and, $x_{q'} \cdot x'_{q'} = i \cdot i$ or $x_{q'} \cdot x'_{q'} = j \cdot j$. This implies that $\{i, j\} \subseteq L_i$ or $\{i, j\} \subseteq R_j$, which contradicts the first condition of the theorem.
  6. There exist q, q', $1 \leq q, q' \leq m$, $q' \neq q$, such that $x_q \cdot x'_q = i \cdot i$ and $x_{q'} \cdot x'_{q'} = j \cdot j$. Since $uu \sqsubseteq \tau(k)$ and due to the conditions of the theorem, it follows from $\tau(\alpha) = \sigma_{i,j}(\alpha)$ that $k \neq i$ and $k \neq j$. In other words, $\tau(i) \neq uu$ and $\tau(j) \neq uu$; otherwise, $|\tau(\alpha)|_u > |\sigma_{i,j}(\alpha)|_u$. Moreover, it must be noticed that if $\sigma_{i,j}(k) \sqsubseteq \tau(k)$, then this implies that there exists $x \in \text{var}(\alpha) \setminus \{i, j\}$, with $\{i, j\} \subseteq L_x$ or $\{i, j\} \subseteq R_x$, which is a contradiction. Thus, $\sigma_{i,j}(k) \not\sqsubseteq \tau(k)$. Since $\tau(\alpha) = \sigma_{i,j}(\alpha)$, there must be a $k' \in \text{var}(\alpha)$, $k' \neq k, i, j$, such that $\sigma_{i,j}(k) \sqsubseteq \tau(k')$, which means that $|\tau(k')| \geq 2$, or we can extend the reasoning to other variables. Consequently, since $\tau(\alpha) = \sigma(\alpha)$, this discussion implies the existence of a $k'' \in \text{var}(\alpha)$, $k'' \neq k, i, j$, such that $|\tau(k'')| \geq 2$, which, according to the above cases, leads to a contradiction.

$\square$

We wish to point out that Theorem 5 does not only demonstrate the correctness of Conjecture 2 for the given class of patterns, but additionally provides an efficient way of finding an unambiguous morphism $\sigma_{i,j}$. For example, we can immediately conclude from it that $\sigma_{1,4}$ is unambiguous with respect to our above example pattern α_1. Furthermore, the theorem also holds for patterns with less than four different variables.

We now consider those patterns that are not a fixed point and, moreover, contain all of their variables exactly twice (note that some of these "shortest" patterns that are not fixed points are also studied in Theorem 4). We wish to demonstrate that Theorem 5 implies the existence of an unambiguous $\sigma_{i,j}$ for *every* such pattern. This insight is based on the following lemma:

Lemma 2. *Let $\alpha \in \mathbb{N}^+$ be a pattern with $|\operatorname{var}(\alpha)| > 6$ and, for every $x \in \operatorname{var}(\alpha)$, $|\alpha|_x = 2$. Then there exist $i, j \in \operatorname{var}(\alpha)$, $i \neq j$, such that*

- *there is no $k \in \operatorname{var}(\alpha)$ with $\{i, j\} \subseteq L_k$ or $\{i, j\} \subseteq R_k$, and*

- $\alpha \neq \alpha_1 \cdot \mathbf{i} \cdot \mathbf{j} \cdot \alpha_2 \cdot \mathbf{j} \cdot \mathbf{i} \cdot \alpha_3$, $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{N}^*$.

Proof. Let $n := |\operatorname{var}(\alpha)|$. Since every variable occurs exactly twice in α, it directly follows that, for every $x \in \operatorname{var}(\alpha)$, $|R_x| \leq 2$ and $|L_x| \leq 2$. By omitting the neighbourhood sets containing ε, we have at most $2n - 2$ sets of size 2. Besides, it can be verified with little effort that α contains at most $n - 1$ different factors $i \cdot j$, $i, j \in \operatorname{var}(\alpha)$, $i \neq j$, such that $j \cdot i \sqsubseteq \alpha$ (e. g., for $n := 4$, $\alpha := 1 \cdot 2 \cdot 3 \cdot 4 \cdot 4 \cdot 3 \cdot 2 \cdot 1$ has 3 different factors $i \cdot j$, $i, j \in \operatorname{var}(\alpha)$, $i \neq j$, satisfying $j \cdot i \sqsubseteq \alpha$). Assume to the contrary that, for every $i, j \in \operatorname{var}(\alpha)$, one of the following cases is satisfied:

- there exists a $k \in \operatorname{var}(\alpha)$ with $\{i, j\} \subseteq L_k$ or $\{i, j\} \subseteq R_k$, or

- $\alpha = \alpha_1 \cdot \mathbf{i} \cdot \mathbf{j} \cdot \alpha_2 \cdot \mathbf{j} \cdot \mathbf{i} \cdot \alpha_3$, $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{N}^*$.

As mentioned above, the maximum number of pairs that are covered by the first case is $2n - 2$, and for the second case it is $n - 1$. On the other hand, since $|\operatorname{var}(\alpha)| = n$, there exist $\binom{n}{2}$ different pairs of variables. However, for $n > 6$, we have $\binom{n}{2} > (2n - 2) + (n - 1)$, which contradicts the assumption. $\square$

Hence, whenever a pattern α is not a fixed point, the conditions of Theorem 5 are automatically satisfied if α contains at least seven distinct variables and all of its variables occur exactly twice. Using a less elegant reasoning than the one on Lemma 2, we can extend this insight to all such patterns over at least four distinct variables. This yields the following result:

Theorem 6. *Let $\alpha \in \mathbb{N}^+$ be a pattern with $|\operatorname{var}(\alpha)| > 3$ and, for every $x \in \operatorname{var}(\alpha)$, $|\alpha|_x = 2$. If α is not a fixed point of a nontrivial morphism, then there exist $i, j \in \operatorname{var}(\alpha)$, $i \neq j$, such that $\sigma_{i,j}$ is unambiguous with respect to α.*

Theorem 6 does not only directly prove the correctness of Conjecture 2 for all patterns that contain all their variables exactly twice, but it also allows a large set of patterns to be constructed for which the Conjecture holds true as well. This construction is specified as follows:

Theorem 7. *Let $\alpha := \alpha_1 \cdot \beta \cdot \alpha_2$ and $\gamma := \alpha_1 \cdot \alpha_2$ be patterns with $\alpha_1, \alpha_2, \beta \in \mathbb{N}^*$, such that*

- γ *and* β *are not a fixed point of a nontrivial morphism,*

- $|\operatorname{var}(\gamma)| > 3$ *and, for every $x \in \operatorname{var}(\gamma)$, $|\gamma|_x = 2$, or $|\operatorname{var}(\beta)| > 3$ and, for every $x \in \operatorname{var}(\beta)$, $|\beta|_x = 2$, and*

- $\operatorname{var}(\gamma) \cap \operatorname{var}(\beta) = \emptyset$.

Then there exist $i, j \in \operatorname{var}(\alpha)$, $i \neq j$, such that $\sigma_{i,j}$ is unambiguous with respect to α.

In the remainder of this section, we shall not directly address the morphism $\sigma_{i,j}$ any longer. Hence, we focus on Conjecture 1, and we use an approach that differs quite significantly from those above: We consider words that cannot be morphic images of a pattern under any ambiguous 1-uniform morphism, and we construct suitable morphic preimages from these words. This method yields another major set of patterns for which Conjecture 1 is satisfied.

Our corresponding technique is based on the well-known concept of *de Bruijn sequences*. Since de Bruijn sequences are cyclic, which does not fit with our subject, we introduce a non-cyclic variant:

Definition 2. *A* non-cyclic De Bruijn sequence *(of order n) is a word over a given alphabet* Σ *(of size k) for which all possible words of length n in* Σ^* *appear exactly once as factors of this sequence. We denote the set of all non-cyclic De Bruijn sequences of order n by* $B'(k,n)$.

For example, the word $w_0 :=$ *aabacbbcca* is a non-cyclic de Bruijn sequence in $B'(3,2)$ if we assume $\Sigma := \{a,b,c\}$.

It can now be easily understood that a non-cyclic de Bruijn sequence cannot be a morphic image of any pattern under ambiguous 1-uniform morphisms:

Theorem 8. *Let* Σ *be an alphabet, and let* $\alpha \in \mathbb{N}^+$ *be a pattern satisfying, for every* $x \in \mathrm{var}(\alpha)$, $|\alpha|_x \geq 2$. *Let* $\sigma : \mathbb{N}^* \to \Sigma^*$ *be a 1-uniform morphism such that, for every* $u_1 u_2 \sqsubseteq \sigma(\alpha)$, $u_1, u_2 \in \Sigma$, *the factor* $u_1 u_2$ *occurs in* $\sigma(\alpha)$ *exactly once. Then* σ *is unambiguous with respect to* α.

This insight implies that every pattern that can be mapped by a 1-uniform morphism to a de Bruijn sequence necessarily is not a fixed point, and thus, fits with Conjecture 1:

Corollary 1. *Let* Σ *be an alphabet, and let* $\alpha \in \mathbb{N}^+$ *be a pattern satisfying, for every* $x \in \mathrm{var}(\alpha)$, $|\alpha|_x \geq 2$. *Let* $\sigma : \mathbb{N}^* \to \Sigma^*$ *be a 1-uniform morphism such that, for every* $u_1 u_2 \sqsubseteq \sigma(\alpha)$, $u_1, u_2 \in \Sigma$, *the factor* $u_1 u_2$ *occurs in* $\sigma(\alpha)$ *exactly once. Then* α *is not a fixed point of a nontrivial morphism.*

We now show how we can construct patterns that fit with the requirements of Theorem 8 and Corollary 1:

Definition 3. *Let* $\Sigma := \{a_1, a_2, \ldots, a_k\}$. *Let* $B'(k,2)$ *be the set of non-cyclic de Bruijn sequences of order 2 over* Σ. *Then* $\Pi_{DB}(k) \subseteq \mathbb{N}^*$ *is the set of all patterns that can be constructed as follows: For every* $w \in B'(k,2)$ *and every letter* a_j *in* w, *all* n_j *occurrences of* a_j *are replaced by* $\lfloor n_j/2 \rfloor$ *different variables from a set* $N_j := \{x_{j_1}, x_{j_2}, \ldots, x_{j_{\lfloor n_j/2 \rfloor}}\} \subseteq \mathbb{N}$, *such that the following conditions are satisfied:*

- *for every* $x \in N_j$, $|\alpha|_x > 1$,
- *for all* i, i', $1 \leq i, i' \leq k$, *with* $i \neq i'$, $N_i \cap N_{i'} = \emptyset$, *and*
- *for all* i, $1 \leq i \leq k$, *the variables in* N_i *are assigned to occurrences of* a_i *in a way such that the resulting pattern is in canonical form.*

For instance, with regard to our above example word $w_0 = aabacbbcca \in B'(3,2)$, Definition 3 says that, e. g., the pattern $1 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 2 \cdot 2 \cdot 4 \cdot 4 \cdot 3$ is contained in $\Pi_{DB}(3)$.

From this construction, it follows that Conjecture 1 holds true for every pattern in $\Pi_{DB}(k)$:

Theorem 9. *Let* $\Sigma := \{a_1, a_2, \ldots, a_k\}$, $k \geq 3$. *Then, for every* $\alpha \in \Pi_{DB}(k)$,

- $\mathrm{var}(\alpha)$ *contains at least* $k+1$ *elements, and*
- *there exists a 1-uniform morphism* $\sigma : \mathbb{N}^* \to \Sigma^*$ *that is unambiguous with respect to* α.

Proof. We begin this proof with the first statement of the theorem: It is obvious that there are k^2 different words of length 2 over Σ. The shortest word that contains k^2 factors of length 2 has length $k^2 + 1$, which means that this is the length of any word $w \in B'(k,2)$. Thus, there must be at least one letter in w that has at least $\lceil (k^2 + 1)/k \rceil$ occurrences. Since we assume $k \geq 3$, this means that this letter has at least 4 occurrences. From Definition 3 it then follows that this letter is replaced by at least two different variables when a pattern $\alpha \in \Pi_{DB}(k)$ is generated from w. Since all other letters in w must be replaced by at least one variable, this shows that $|\mathrm{var}(\alpha)| \geq k+1$.

Concerning the second statement, we define σ by, for every j, $1 \leq j \leq k$, and for every $x \in N_j$, $\sigma(x) := a_j$. Thus, σ is 1-uniform, and $\sigma(\alpha) \in B'(k,2)$. This implies that, for every $u_1 u_2 \sqsubseteq \sigma(\alpha)$, $u_1, u_2 \in \Sigma$, the factor $u_1 u_2$ occurs in $\sigma(\alpha)$ exactly once. Consequently, according to Theorem 8, σ is unambiguous with respect to α. $\square$

We conclude this paper with a statement on the cardinality of $\Pi_{DB}(k)$, demonstrating that the use of de Bruijn sequences indeed leads to a rich class of patterns α with unambiguous 1-uniform morphisms, and that these morphisms, in general, can even have a target alphabet of size much less than $\mathrm{var}(\alpha) - 1$ (as featured by Theorem 9):

Theorem 10. *Let $k \in \mathbb{N}$. Then $|\Pi_{DB}(k)| \geq k!^{(k-1)}$, and, for every $\alpha \in \Pi_{DB}(k)$,*

$$|\mathrm{var}(\alpha)| = (k-1)\lfloor k/2 \rfloor + \lfloor (k+1)/2 \rfloor.$$

References

[1] M. Billaud (1993): *A problem with words*. Letter in Newsgroup Comp.theory.

[2] C. Choffrut (1997): *Equations in Words*. In M. Lothaire, editor: *Combinatorics on Words*, chapter 9, Cambbridge University Press, pp. 164–183.

[3] D.D. Freydenberger, H. Nevisi & D. Reidenbach (2011): *Weakly Unambiguous Morphisms*. In: *Proc. 28th Annual Symposium on Theoretical Aspects of Computer Science, STACS 2011*, LIPIcs 9, pp. 213–224, doi:10.4230/LIPIcs.STACS.2011.213.

[4] D.D. Freydenberger & D. Reidenbach (2009): *The Unambiguity of Segmented Morphisms*. Discrete Applied Mathematics 157, pp. 3055–3068, doi:10.1016/j.dam.2009.06.009.

[5] D.D. Freydenberger, D. Reidenbach & J.C. Schneider (2006): *Unambiguous Morphic Images of Strings*. International Journal of Foundations of Computer Science 17, pp. 601–628, doi:10.1142/S0129054106004017.

[6] T. Harju & J. Karhumäki (1997): *Morphisms*. In G. Rozenberg & A. Salomaa, editors: *Handbook of Formal Languages*, chapter 7, 1, Springer, pp. 439–510.

[7] S. Holub (2009): *Polynomial-time algorithm for fixed points of nontrivial morphisms*. Discrete Mathematics 309, pp. 5069–5076, doi:10.1016/j.disc.2009.03.019.

[8] F. Levé & G. Richomme (2005): *On a conjecture about finite fixed points of morphisms*. Theoretical Computer Science 339, pp. 103–128, doi:10.1016/j.tcs.2005.01.011.

[9] A. Mateescu & A. Salomaa (1997): *Patterns*. In G. Rozenberg & A. Salomaa, editors: *Handbook of Formal Languages*, chapter 4.6, 1, Springer, pp. 230–242.

[10] D. Reidenbach (2008): *Discontinuities in pattern inference*. Theoretical Computer Science 397, pp. 166–193, doi:10.1016/j.tcs.2008.02.029.

[11] D. Reidenbach & J.C. Schneider (2009): *Morphically primitive words*. Theoretical Computer Science 410, pp. 2148–2161, doi:10.1016/j.tcs.2009.01.020.

[12] D. Reidenbach & J.C. Schneider (2011): *Restricted Ambiguity of Erasing Morphisms*. Theoretical Computer Science 412, pp. 3510–3523, doi:10.1016/j.tcs.2011.02.041.

[13] J.C. Schneider (2010): *Unambiguous Erasing Morphisms in Free Monoids*. Theoretical Informatics and Applications (RAIRO) 44, pp. 193–208, doi:10.1051/ita/2009020.

[14] A. Thue (1906): *Über unendliche Zeichenreihen*. Kra. Vidensk. Selsk. Skrifter. I Mat. Nat. Kl. 7.