



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.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

BY: **Attribution.** You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

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

Restricted Ambiguity of Erasing Morphisms

Daniel Reidenbach¹ and Johannes C. Schneider^{2*}

¹ Department of Computer Science, Loughborough University,
Loughborough, LE11 3TU, United Kingdom,
D.Reidenbach@lboro.ac.uk

² Fachbereich Informatik, Technische Universität Kaiserslautern,
Postfach 3049, 67653 Kaiserslautern, Germany,
jschneider@informatik.uni-kl.de

Abstract. A morphism h is called ambiguous for a string s if there is another morphism that maps s to the same image as h ; otherwise, it is called unambiguous. In this paper, we examine some fundamental problems on the ambiguity of erasing morphisms. We provide a detailed analysis of so-called ambiguity partitions, and our main result uses this concept to characterise those strings that have a morphism of strongly restricted ambiguity. Furthermore, we demonstrate that there are strings for which the set of unambiguous morphisms, depending on the size of the target alphabet of these morphisms, is empty, finite or infinite. Finally, we show that the problem of the existence of unambiguous erasing morphisms is equivalent to some basic decision problems for nonerasing multi-pattern languages.

1 Introduction

The research on the ambiguity of morphisms is based on the following, elementary questions: Given a string s and a morphism h , do there exist morphisms g with $g(s) = h(s)$, but $g(x) \neq h(x)$ for a symbol x in s ? If so, what properties do these morphisms g have? For example, let $s := \text{AABBCC}$, and let the morphism $h : \{\text{A}, \text{B}, \text{C}\}^* \rightarrow \{\text{a}, \text{b}\}^*$ be given by $h(\text{A}) := h(\text{C}) := \text{a}$ and $h(\text{B}) := \text{b}$. Then it can be easily verified that there is no morphism g satisfying $g(s) = \text{aabbba} = h(s)$ and $g(x) \neq h(x)$ for an $x \in \{\text{A}, \text{B}, \text{C}\}$. Therefore, we call h *unambiguous* for s . On the other, if we consider the morphism $h' : \{\text{A}, \text{B}, \text{C}\}^* \rightarrow \{\text{a}, \text{b}\}^*$, defined by $h'(\text{A}) := h'(\text{B}) := h'(\text{C}) := (\text{ab})^{10}$, then there are various other morphisms g that map s to $h'(s) = (\text{ab})^{60}$. Hence, h' is *ambiguous* for s . Furthermore, for every n with $0 \leq n \leq 30$ and for every symbol $x \in \{\text{A}, \text{B}, \text{C}\}$, there exists at least one morphism g satisfying $g(s) = h'(s)$ and $g(x) = (\text{ab})^n$. Thus, the ambiguity of h' for s is largely unrestricted. In the present paper, we wish to investigate this phenomenon, and we shall mainly focus on the question of whether, for any string, there *exists* a morphism with a restricted ambiguity. To this end, we distinguish between two types of restrictions: maximally restricted ambiguity (i. e., unambiguity) and so-called *moderate ambiguity*, a sophisticated yet natural concept to be introduced below.

* Corresponding Author

The existence of unambiguous and moderately ambiguous *nonerasing* morphisms has already been intensively studied (see, e. g., Freydenberger et al. [1], Reidenbach [9]), and characteristic criteria have been provided. These criteria reveal that the existence of such morphisms is alphabet-independent, i. e., for any string s over some alphabet \mathcal{A} and for any alphabets Σ, Σ' with at least two letters each, s has an unambiguous or moderately ambiguous nonerasing morphism $h : \mathcal{A}^* \rightarrow \Sigma^*$ if and only if there is a morphism $h' : \mathcal{A}^* \rightarrow \Sigma'^*$ with the equivalent property. In the present work, we study the ambiguity of all morphisms, including *erasing morphisms*, which map a symbol in s to the empty string. As pointed out by Schneider [13], here the existence of unambiguous erasing morphism does not only depend on the structure of the string, but also on the size of the target alphabet of the morphism, which turns the search for characteristic conditions into a rather intricate problem.

The examination of the ambiguity of morphisms is not only of intrinsic interest, but, due to the simplicity of the concept, also shows various connections to other topics in theoretical computer science and discrete mathematics. This primarily holds for those approaches where several morphisms are applied to one finite string, including pattern languages (see, e. g., Mateescu and Salomaa [8]) as well as equality sets (and, thus, the Post Correspondence Problem, cf. Harju and Karhumäki [2]). Particularly well understood are the relations to pattern languages, where several prominent problems have been solved using insights into the ambiguity of morphisms (see, e. g., Reidenbach [10]). Moreover, there are further connections of the ambiguity of morphisms to various concepts that involve morphisms such as fixed points of morphisms, avoidable patterns and word equations.

Our work is organised as follows: After giving some definitions and basic results, we provide a detailed analysis of *ambiguity partitions* (as introduced by Schneider [13]), which are a vital concept when investigating the ambiguity of erasing morphisms. In Section 4, we introduce and study moderate ambiguity, i. e., an important type of strongly restricted ambiguity. We characterise those strings for which there exist moderately ambiguous erasing morphisms, and this is the main result of our paper. In Section 5, we deal with unambiguous morphisms, and we study the number of such morphisms for certain strings. Finally, in Section 6, we reveal that the existence of unambiguous erasing morphisms can be characterised through basic decision problems for so-called nonerasing multi-pattern languages. This insight might be a worthwhile starting point for future research. Note that, due to space constraints, all proofs are omitted from this paper.

2 Definitions and Basic Notes

In the present section we give some basic definitions and results. For notations not explained explicitly, we refer the reader to Rozenberg and Salomaa [12].

Let $\mathbb{N} := \{1, 2, \dots\}$ be the set of natural numbers. The *power set* of a set S is denoted by $\mathcal{P}(S)$. An *alphabet* \mathcal{A} is an enumerable set of symbols. A *string* (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 string X . The *empty* string ε is the unique sequence of symbols of length 0. For the *concatenation* of strings s, t we write $s \cdot t$ (or st for short). The string that results from the n -fold concatenation of a string s is denoted by s^n . The notation \mathcal{A}^* refers to the set of all strings over \mathcal{A} , i. e., more precisely, the *free monoid* generated by \mathcal{A} ; furthermore, $\mathcal{A}^+ := \mathcal{A}^* \setminus \{\varepsilon\}$. The number of occurrences of a symbol $x \in \mathcal{A}$ in a string $s \in \mathcal{A}^*$ is written as $|s|_x$. With regard to arbitrary strings $s, t \in \mathcal{A}^*$, we write $s = t \dots$ if there exists an $u \in \mathcal{A}^*$ such that $s = tu$, we write $s = \dots t$ if there exists an $u \in \mathcal{A}^*$ such that $s = ut$, and, finally, $s = \dots t \dots$ if there exist $u, v \in \mathcal{A}^*$ such that $s = utv$. We call t a *prefix*, *suffix* and *factor* of s , respectively. In contrast to this notation, if we omit some parts of a canonically given string, then we henceforth use the symbol $[\dots]$; e. g., $s = \dots ab[\dots]f$ means that s ends with the string $abcdef$.

We often use \mathbb{N} as an infinite alphabet of symbols. In order to distinguish between strings over \mathbb{N} and strings over a (possibly finite) alphabet Σ , we call the former *patterns*. Given a pattern $\alpha \in \mathbb{N}^*$, we call symbols occurring in α *variables* and denote the set of variables in α with $\text{var}(\alpha)$. Hence, $\text{var}(\alpha) \subseteq \mathbb{N}$. We use the symbol \cdot to separate the variables in a pattern, so that, for instance, $1 \cdot 1 \cdot 2$ is not confused with $11 \cdot 2$.

Given arbitrary alphabets \mathcal{A}, \mathcal{B} , a *morphism* is a mapping $h : \mathcal{A}^* \rightarrow \mathcal{B}^*$ 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} . We call h *erasing* if and only if $h(a) = \varepsilon$ for an $a \in \mathcal{A}$; otherwise, h is called *nonerasing*. If we call a morphism h (non)erasing with a certain input string s in mind, we only demand h to be (non)erasing for the symbols occurring in s .

A pattern $\alpha \in \mathbb{N}^+$ is called a *fixed point (of a morphism h)* if $h(\alpha) = \alpha$. A morphism $h : \mathbb{N}^* \rightarrow \mathbb{N}^*$ is said to be *nontrivial* if $h(x) \neq x$ for an $x \in \mathbb{N}$. Let $V \subseteq \mathbb{N}$. We call $h : \mathbb{N}^* \rightarrow \mathbb{N}^*$ *nontrivial for V* if $h(x) \neq x$ for an $x \in V$. The morphism $\pi_V : \mathbb{N}^* \rightarrow \mathbb{N}^*$ is given by $\pi_V(x) := x$ if $x \in V$ and $\pi_V(x) := \varepsilon$ if $x \notin V$.

For any alphabet Σ , for any morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ and for any pattern $\alpha \in \mathbb{N}^+$ with $\sigma(\alpha) \neq \varepsilon$, we call σ *unambiguous (for α)* if and only if there is no morphism $\tau : \mathbb{N}^* \rightarrow \Sigma^*$ satisfying $\tau(\alpha) = \sigma(\alpha)$ and, for some $x \in \text{var}(\alpha)$, $\tau(x) \neq \sigma(x)$. If σ is not unambiguous for α , it is called *ambiguous (for α)*. We extend this definition to any word $w \in \Sigma^*$ in the natural way, i. e., w is said to be *unambiguous (for α)* if there is an unambiguous morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ with $\sigma(\alpha) = w$, and w is called *ambiguous (for α)* if there is an ambiguous morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ satisfying $\sigma(\alpha) = w$. Furthermore, with regard to the E-pattern language of α to be introduced in the subsequent paragraph, we say that a word $w \in L_{E, \Sigma}(\alpha)$ is (un-)ambiguous if w is (un-)ambiguous for α .

Basically, the set of all images of a pattern $\alpha \in \mathbb{N}^+$ under morphisms $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$, where Σ is an arbitrary alphabet of so-called *terminal-symbols*, is called the *pattern language (generated by α)*. Formally, two main types of pattern languages of α are considered: its *E-pattern language* $L_{E, \Sigma}(\alpha) := \{\sigma(\alpha) \mid \sigma : \mathbb{N}^* \rightarrow \Sigma^* \text{ is a morphism}\}$ and its *NE-pattern language* $L_{NE, \Sigma}(\alpha) := \{\sigma(\alpha) \mid$

$\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ is a nonerasing morphism}. Note that, in literature, pattern languages as defined above are usually called *terminal-free*, since, in a more general understanding of the concept, a pattern may additionally contain terminal symbols. The morphisms $\sigma : (\mathbb{N} \cup \Sigma)^* \rightarrow \Sigma^*$ applied to such a pattern $\alpha \in (\mathbb{N} \cup \Sigma)^+$ when generating its pattern language must then be *terminal-preserving*, i. e., for any $a \in \Sigma$, $\sigma(a) = a$ must be satisfied.

We conclude the definitions in this section with a partition of the set of all patterns subject to the following criterion:

Definition 1. *Let $\alpha \in \mathbb{N}^+$. We call α prolix if and only if there exists a factorisation $\alpha = \beta_0 \gamma_1 \beta_1 \gamma_2 \beta_2 \dots \gamma_n \beta_n$ with $n \geq 1$, $\beta_i \in \mathbb{N}^*$, $0 \leq i \leq n$, and $\gamma_i \in \mathbb{N}^+$, $1 \leq i \leq n$, such that*

1. *for every $i \in \{1, 2, \dots, n\}$, $|\gamma_i| \geq 2$,*
2. *for every $i \in \{0, 1, \dots, n\}$, for every $j \in \{1, 2, \dots, n\}$, $\text{var}(\beta_i) \cap \text{var}(\gamma_j) = \emptyset$,*
3. *for every $i \in \{1, 2, \dots, n\}$, there exists an $y_i \in \text{var}(\gamma_i)$ such that y_i occurs exactly once in γ_i and, for every $i' \in \{1, 2, \dots, n\}$, if $y_i \in \gamma_{i'}$ then $\gamma_i = \gamma_{i'}$.*

We call $\alpha \in \mathbb{N}^+$ succinct if and only if it is not prolix.

A succinct pattern is the shortest generator of its respective E-pattern language, i. e., for any Σ , $|\Sigma| \geq 2$, and any succinct pattern α , there is no pattern β with $|\beta| < |\alpha|$ and $L_{E, \Sigma}(\beta) = L_{E, \Sigma}(\alpha)$. Furthermore, the set of prolix patterns exactly corresponds to the class of finite fixed points of nontrivial morphisms (cf. Head [3]). Note that set of succinct patterns is also equivalent to the set of *morphically primitive* words (as introduced by Reidenbach and Schneider [11]).

Regarding the unambiguity of *nonerasing* morphisms, the classification of patterns into succinct and prolix patterns is vital:

Theorem 1 (Freydenberger, Reidenbach, and Schneider [1]). *Let $\alpha \in \mathbb{N}^+$, let Σ be an alphabet, $|\Sigma| \geq 2$. There exists an unambiguous nonerasing morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ for α if and only if α is succinct.*

According to this result, for any prolix pattern α , every nonerasing morphism is ambiguous. In contrast to this negative insight, there are prolix patterns that have unambiguous *erasing* morphisms (as pointed out by Schneider [13]). However, this is not a universal property of prolix patterns; thus, certain prolix patterns do not have any unambiguous morphism at all. This phenomenon is the main topic of our paper.

3 Ambiguity Partitions

Previous results show that *ambiguity partitions* as introduced by Schneider [13] are a crucial notion when investigating the ambiguity of erasing morphisms, and the main result of our paper, given in Section 4, further illustrates their importance. In the present section, we therefore study some fundamental properties of this concept.

Definition 2. We inductively define an ambiguity partition (for any $\alpha \in \mathbb{N}^+$):

- (i) $(\emptyset, \text{var}(\alpha))$ is an ambiguity partition for α .
- (ii) If (E, N) is an ambiguity partition for α and there exists a morphism $h : \mathbb{N}^* \rightarrow \mathbb{N}^*$ that is nontrivial for N and satisfies $h(\alpha) = \pi_N(\alpha)$, then (E', N') is an ambiguity partition with $E' := E \cup \{x \in N \mid h(x) = \varepsilon\}$, $N' := \{x \in N \mid h(x) \neq \varepsilon\}$.

According to [13], Definition 2 permits a number of fundamental insights into the ambiguity of erasing morphisms to be established. They directly or indirectly result from the following, slightly technical fact:

Theorem 2 (Schneider [13]). Let Σ be an alphabet. Let $\alpha \in \mathbb{N}^+$ and let (E, N) be an ambiguity partition for α . Then every morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ satisfying $\sigma(x) \neq \varepsilon$ for an $x \in E$ is ambiguous for α .

Consequently, for any pattern α , an ambiguity partition (E, N) for α gives us valuable information on the set S of variables in α which must be erased by unambiguous morphisms, since $S \supseteq E$. Thus, the larger the set E becomes, the more information we get. Therefore, we name ambiguity partitions with a set E of maximal size in the following definition:

Definition 3. Let $\alpha \in \mathbb{N}^+$. An ambiguity partition (E, N) for α is called maximal if and only if every ambiguity partition (E', N') for α satisfies $|E'| \leq |E|$ and $|N'| \geq |N|$.

This definition supports some of our proofs, and we can use it to express vital statements on the (non-)existence of morphisms with a restricted ambiguity.

From Definition 2, it is not obvious whether or not a maximal ambiguity partition for a pattern α is unique. However, it can be shown that, for any pattern, there is exactly one maximal ambiguity partition:

Theorem 3. Let $\alpha \in \mathbb{N}^+$ and (E, N) be a maximal ambiguity partition for α . Then (E, N) is unique.

Evidently, the uniqueness of the maximal ambiguity partition (E, N) of a pattern α is a nontrivial property only if $E \neq \text{var}(\alpha)$. On the other hand, if $(\text{var}(\alpha), \emptyset)$ is the maximal ambiguity partition of α , then it is known that the following statement on the existence of unambiguous morphisms holds true:

Corollary 1 (Schneider [13]). Let Σ be an alphabet, and let $\alpha \in \mathbb{N}^+$. If $(\text{var}(\alpha), \emptyset)$ is an ambiguity partition for α , then every morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ is ambiguous for α .

While Corollary 1, in the case of arbitrary alphabets Σ , uses ambiguity partitions $(\text{var}(\alpha), \emptyset)$ to establish a *sufficient* criterion on the nonexistence of unambiguous morphisms (note that [13] gives examples demonstrating that this criterion is not characteristic), an even stronger result is known for infinite Σ :

Theorem 4 (Schneider [13]). Let Σ be an infinite alphabet, and let $\alpha \in \mathbb{N}^+$. Then $(\text{var}(\alpha), \emptyset)$ is an ambiguity partition for α if and only if every morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ is ambiguous for α .

Thus, when investigating the existence of unambiguous erasing morphisms, the question of whether or not $(\text{var}(\alpha), \emptyset)$ is an ambiguity partition for α leads to an important (and sometimes even characteristic) partition of \mathbb{N}^+ . Therefore, we now introduce a new terminology reflecting this question:

Definition 4. *Let $\alpha \in \mathbb{N}^+$. We call α morphically erasable if and only if $(\text{var}(\alpha), \emptyset)$ is an ambiguity partition for α . Otherwise, α is called morphically unerasurable.*

Thus, referring to Definition 4, Corollary 1 demonstrates that, for finite alphabets Σ , the search for patterns with unambiguous morphisms can be narrowed down to the morphically unerasurable ones. Therefore, and since our main result in Section 4 again is based on this property, we now give a nontrivial characterisation of such patterns. To this end, we use a condition that is based on the inclusion of E-pattern languages, which is a well-investigated problem (see Jiang et al. [6]).

Condition 1. A pattern $\alpha \in \mathbb{N}^+$ satisfies Condition 1 if and only if there exists a set $N \subseteq \text{var}(\alpha)$ such that, for every $M \subseteq \text{var}(\alpha)$ with $M \not\supseteq N$ and for any alphabet Σ with $|\Sigma| \geq 2$, $L_{E,\Sigma}(\pi_M(\alpha)) \not\supseteq L_{E,\Sigma}(\pi_N(\alpha))$.

Lemma 1. *A pattern $\alpha \in \mathbb{N}^+$ satisfies Condition 1 if and only if α is morphically unerasurable.*

Summarising the above statements, we can note the following sufficient condition on the nonexistence of unambiguous erasing morphisms, that is equivalent to Corollary 1:

Theorem 5. *Let Σ be an alphabet. If an $\alpha \in \mathbb{N}^+$ does not satisfy Condition 1, then every morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ with $\sigma(\alpha) \neq \varepsilon$ is ambiguous for α .*

The original motivation for investigating the ambiguity of morphisms is derived from inductive inference of E-pattern languages – i. e., the problem of computing a pattern from the words in its pattern languages –, which strongly depends on the inclusion relation between E-pattern languages. In this context, certain morphisms with a restricted ambiguity are known to generate words that contain reliable and algorithmically usable information about their generating pattern (cf. Reidenbach [10]) and, thus, are a vital input to any inference procedure. Theorem 5 further illustrates this close connection between the two topics.

The techniques used in [10] are based on the notion of an ambiguity of specific *nonerasing* morphisms that is restricted in a particular manner. We now introduce and study an equivalent concept for *erasing* morphisms.

4 Moderate Ambiguity

Theorem 4 shows that, in case of an infinite alphabet Σ , the property of a pattern α being morphically unerasurable is characteristic for the existence of a morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ that is unambiguous for α . However, concerning finite

target alphabets Σ , there are morphically unerasable patterns for which there exists no unambiguous morphism (see the examples given by Schneider [13]). Although we are, hence, not able to achieve unambiguity for every morphically unerasable pattern, we shall demonstrate below that a certain restricted ambiguity is possible, which can be interpreted as unambiguity of a morphism with regard to particular factors of $\sigma(\alpha)$. As briefly mentioned above, a similar property of nonerasing morphisms is used for many fundamental results on inductive inference of E-pattern languages, and an extensive analysis of this phenomenon is provided by Reidenbach [9].

In accordance with [9], we call the said type of ambiguity *moderate ambiguity*. Intuitively, it can be understood as follows: A morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ is called *moderately ambiguous* for a pattern α if, for every variable position j of a variable x in α with $\sigma(x) \neq \varepsilon$, there exists a certain factor w_j of $\sigma(\alpha)$ at a certain position (between the l_j th and r_j th letter in $\sigma(\alpha)$) such that *every* morphism $\tau : \mathbb{N}^* \rightarrow \Sigma^*$ with $\tau(\alpha) = \sigma(\alpha)$ maps the variable x at position j to a word which covers at least the factor w_j at this particular position. We illustrate this type of ambiguity in the following example:

Example 1. Let $\Sigma := \{\mathbf{a}, \mathbf{b}\}$ and $\alpha := i_1 \cdot i_2 \cdot i_3 \cdot i_4 \cdot i_5 \cdot i_6 \cdot i_7 \cdot i_8 \cdot i_9 \cdot i_{10} := 1 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 3 \cdot 1 \cdot 3$. Let $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ be a morphism defined by $\sigma(1) := \varepsilon$, $\sigma(2) := \mathbf{aba}$, $\sigma(3) := \mathbf{abb}$. The morphism σ is ambiguous for α since $\tau : \mathbb{N}^* \rightarrow \Sigma^*$, defined by $\tau(1) := \mathbf{a}$, $\tau(2) := \mathbf{b}$, $\tau(3) := \mathbf{bb}$, satisfies $\tau(\alpha) = \sigma(\alpha)$. Hence, the situation looks as follows:

$$\sigma(\alpha) = \underbrace{\underbrace{\mathbf{a}}_{\tau(1)} \underbrace{\mathbf{b}}_{\tau(2)} \underbrace{\mathbf{a}}_{\tau(1)}}_{\sigma(2)} \underbrace{\underbrace{\mathbf{a}}_{\tau(1)} \underbrace{\mathbf{b}}_{\tau(2)} \underbrace{\mathbf{a}}_{\tau(1)}}_{\sigma(2)} \underbrace{\underbrace{\mathbf{a}}_{\tau(1)} \underbrace{\mathbf{bb}}_{\tau(3)}}_{\sigma(3)} \underbrace{\underbrace{\mathbf{a}}_{\tau(1)} \underbrace{\mathbf{bb}}_{\tau(3)}}_{\sigma(3)} = \tau(\alpha).$$

However, we call σ *moderately ambiguous* since all morphism τ' with $\tau'(\alpha) = \sigma(\alpha)$ map every variable i_k with $\sigma(i_k) \neq \varepsilon$ to a certain factor w_k of $\sigma(i_k)$ at a particular position. In this example, we have $w_2 = w_5 = \mathbf{b}$ and $w_8 = w_{10} = \mathbf{bb}$ – and the only morphisms τ' with $\tau'(\alpha) = \sigma(\alpha)$ are σ itself and τ which satisfy $\sigma(i_k) = \dots w_k \dots = \tau(i_k)$ for $k = 2, 5, 8, 10$.

We now formalise moderate ambiguity. As explained above, we consider this a very natural way of slightly relaxing the requirement of unambiguity, and the importance of this approach has been demonstrated in the context of inductive inference of pattern languages. Nevertheless, our definition is quite involved, since we do not only postulate that, for a given pattern α and for every $x \in \text{var}(\alpha)$, there exists a string w_x such that, for every morphism τ with $\tau(\alpha) = \sigma(\alpha)$, $\tau(x)$ contains w_x as a factor (which could be called *factor-preserving ambiguity*), but we also demand that these factors are located at fixed positions for all τ . This means that we need to identify and mark the positions of the factors.

Definition 5. Let Σ be an alphabet, let $\alpha = i_1 \cdot i_2 \cdot [\dots] \cdot i_n$ with $n, i_1, i_2, \dots, i_n \in \mathbb{N}$, and let $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ be a morphism satisfying $\sigma(\alpha) \neq \varepsilon$. Then σ is called moderately ambiguous (for α) provided that there exist $l_2, l_3, \dots, l_n, r_1, r_2, \dots, r_{n-1} \in \mathbb{N} \cup \{0\}$ such that, for every morphism $\tau : \mathbb{N}^* \rightarrow \Sigma^*$ with $\tau(\alpha) = \sigma(\alpha)$,

- (i) if $\sigma(i_1) \neq \varepsilon$ then $r_1 \geq 1$,
- (ii) if $\sigma(i_n) \neq \varepsilon$ then $l_n \leq |\sigma(\alpha)|$,
- (iii) for every $k \in \{2, 3, \dots, n-1\}$ with $\sigma(i_k) \neq \varepsilon$, $l_k \leq r_k$,
- (iv) for every k with $1 \leq k \leq n-1$, $|\tau(i_1 \cdot i_2 \cdot [\dots] \cdot i_k)| < l_{k+1}$, and
- (v) for every k with $1 \leq k \leq n-1$, $|\tau(i_1 \cdot i_2 \cdot [\dots] \cdot i_k)| \geq r_k$.

We call σ strongly ambiguous (for α) if and only if it is not moderately ambiguous (for α).

In the definition, for any pattern α and any moderately ambiguous morphism σ for α , a pair (l_k, r_k) for some $i_k \in \text{var}(\alpha)$ with $\sigma(i_k) \neq \varepsilon$ “marks” the factor w_k from position l_k to r_k in $\sigma(\alpha)$. This factor must be covered by the image of i_k under every morphism τ with $\tau(\alpha) = \sigma(\alpha)$ – this is guaranteed by the conditions (iv) and (v). Considering Example 1, we choose the following markers l_i, r_k : Let $r_1 := 0$, $(l_2, r_2) := (2, 2)$, $(l_5, r_5) := (5, 5)$, $(l_8, r_8) := (8, 9)$, $l_{10} := 11$ and finally $(l_k, r_k) := (|\sigma(\alpha)| + 1, 0)$ for $k \in \{1, 3, 4, 6, 7, 9\}$ since, for these k , $\sigma(i_k) = \varepsilon$, and, thus, no factor has to be marked. It can be verified that these values of l_j , $2 \leq j \leq n$, and r_k , $1 \leq k \leq n-1$ meet the requirements (i)–(v) of Definition 5.

The following lemma is useful when studying moderate ambiguity since, in certain cases, it circumvents a check of the minutiae of Definition 5.

Lemma 2. *Let Σ be an alphabet, $\alpha \in \mathbb{N}^+$ and $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ be a morphism. If there exists a morphism $\tau : \mathbb{N}^* \rightarrow \Sigma^*$ such that $\tau(\alpha) = \sigma(\alpha)$, but $\tau(x) = \varepsilon \neq \sigma(x)$ for an $x \in \text{var}(\alpha)$, then σ is not moderately ambiguous for α .*

As suggested by the definitions and further substantiated by Example 1, for any given morphism, the requirement of being moderately ambiguous is less strict than that of being unambiguous:

Proposition 1. *Let Σ be an alphabet, let $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ be a morphism, and let $\alpha \in \mathbb{N}^+$. If σ is unambiguous for α , then σ is moderately ambiguous for α . In general, the converse does not hold.*

This directly implies that if there exists no moderately ambiguous morphism for a pattern α , then there exists no unambiguous morphism for α and, thus, every morphism is strongly ambiguous for α .

With these new terms of ambiguity, we can give a stronger version of Theorem 2:

Theorem 6. *Let Σ be an alphabet. Let $\alpha \in \mathbb{N}^+$ and let (E, N) be an ambiguity partition for α . Then every morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ satisfying $\sigma(x) \neq \varepsilon$ for an $x \in E$ is strongly ambiguous for α .*

The main result of our paper characterises those patterns that have a moderately ambiguous morphism. More precisely, it states that moderate ambiguity can be achieved if and only if the pattern is morphically unerasable:

Theorem 7. *Let Σ be an alphabet, $|\Sigma| \geq 2$, let $\alpha \in \mathbb{N}^+$. There exists a morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ that is moderately ambiguous for α if and only if α is morphically unerasable.*

In addition to the facts that Theorem 7 provides an algorithmically verifiable characteristic condition on a vital problem regarding the existence of morphisms with a restricted ambiguity and, furthermore, implies the equivalent result for the weaker requirement of factor-preserving ambiguity, we consider two other aspects of it quite remarkable. Firstly, it confirms that ambiguity partitions are indeed a crucial tool when investigating the ambiguity of erasing morphisms, since they cannot only be used to give sufficient criteria on the subject (cf. Corollary 1) and characteristic criteria for special cases (cf. Theorem 4), but are also capable of expressing a key phenomenon in this field of study.

Secondly, it establishes a quite remarkable and counter-intuitive difference between the ambiguity of erasing and nonerasing morphisms. As demonstrated by Freydenberger et al. [1], the existence of a moderately ambiguous nonerasing morphism σ for a pattern implies the existence of an unambiguous nonerasing morphism σ' . More technically, it can be shown that σ can be turned into σ' by applying some minor yet sophisticated changes that depend on the structure of the pattern in question (see Reidenbach [9] for a detailed discussion of this topic). It is also important to note that the morphism σ and σ' both use a binary target alphabet; hence, the existence of such morphisms – which characterises the succinct patterns, cf. Theorem 1 – exclusively depends on the pattern and not on the size of Σ (provided that Σ contains at least two letters). In contrast to these observations, Theorem 7 demonstrates that the existence of moderately ambiguous erasing morphisms does not imply the existence of unambiguous erasing morphisms:

Corollary 2. *Let Σ be an alphabet. There exists an $\alpha \in \mathbb{N}^+$ and a morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ such that σ is moderately ambiguous for α , but no morphism is unambiguous for α .*

Hence, the main result of our paper also shows that the technical concepts used by Freydenberger et al. [1] to turn a moderately ambiguous morphism into an unambiguous one necessarily fail for erasing morphisms. Since this insight is rather unexpected, it is also surprising that Theorem 7 is alphabet-independent, whereas any characterisation of the set of those patterns that have an unambiguous erasing morphism must depend on the size of Σ (as shown by Schneider and to be further addressed by Section 5).

We wish to conclude this section with an insight into the complexity of the problem of deciding on the existence of moderately ambiguous morphisms:

Corollary 3. *Let Σ be an alphabet, $|\Sigma| \geq 2$. The problem of deciding, for any given $\alpha \in \mathbb{N}^+$, on whether there is an erasing morphism $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ that is moderately ambiguous for α , is NP-complete.*

This nicely contrasts with the recent result by Holub [4], which implies that there is a polynomial-time procedure deciding on the existence of unambiguous nonerasing morphisms.

As briefly mentioned above, we now study another fundamental property of those patterns that can be used to prove Corollary 2.

5 Patterns with Finitely Many Unambiguous Morphisms

Once the existence of morphisms with a restricted ambiguity has been established for a given pattern, it is a natural problem to investigate the number of such morphisms. Since the existence of one moderately ambiguous morphism for a given pattern immediately implies an infinite number of such morphisms (the morphism used to prove Theorem 7 can easily be generalised), we now study the above-mentioned topic with regard to a maximal restriction of ambiguity, i. e. unambiguity. To this end, we introduce the following notation:

Definition 6. *Let Σ be an alphabet and $\alpha \in \mathbb{N}^+$. Then $\text{UNAMB}_{\Sigma}(\alpha)$ is the set of all $\sigma(\alpha)$, where $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ is a morphism that is unambiguous for α , and $\text{UNAMB}_{\text{NE},\Sigma}(\alpha)$ is the set of all $\sigma(\alpha)$, where $\sigma : \mathbb{N}^* \rightarrow \Sigma^*$ is a morphism that is nonerasing and unambiguous for α .*

We wish to point out that the sets $\text{UNAMB}_{\Sigma}(\alpha)$ and $\text{UNAMB}_{\text{NE},\Sigma}(\alpha)$ do not consist of morphisms, but of morphic images. This makes sure that all unambiguous morphisms indirectly collected by these sets necessarily differ on variables that are contained in $\text{var}(\alpha)$.

We first consider the case of *nonerasing* morphisms.

Theorem 8. *Let $\alpha \in \mathbb{N}^+$. Then either, for all alphabets Σ with $|\Sigma| \geq 2$, $\text{UNAMB}_{\text{NE},\Sigma}(\alpha)$ is empty or, for all alphabets Σ with $|\Sigma| \geq 2$, $\text{UNAMB}_{\text{NE},\Sigma}(\alpha)$ is infinite.*

If we study the equivalent question for the ambiguity of *erasing* morphisms, we can observe a novel phenomenon that establishes a further difference to the case of nonerasing morphisms. More precisely, for certain patterns α , the cardinality of $\text{UNAMB}_{\Sigma}(\alpha)$ can be finite, and this essentially depends on the size of Σ :

Theorem 9. *Let $k \in \mathbb{N}$. Let $\Sigma_k, \Sigma_{k+1}, \Sigma_{k+2}$ be alphabets with $k, k+1, k+2$ letters, respectively. There exists an $\alpha_k \in \mathbb{N}^+$ such that $|\text{UNAMB}_{\Sigma_k}(\alpha_k)| = 0$, $|\text{UNAMB}_{\Sigma_{k+1}}(\alpha_k)| = m$ for an $m \in \mathbb{N}$, and $\text{UNAMB}_{\Sigma_{k+2}}(\alpha_k)$ is an infinite set.*

6 Connections to NE-pattern Languages

In this final main section of our paper we wish to study a topic that, after the particularly strong result in Theorem 7, remains as the most fundamental open problem on erasing morphisms with a restricted ambiguity, namely a characterisation of those patterns that have an unambiguous erasing morphism. As a matter of fact, the main result of the present section can be understood as such a characterisation, but the immediate usefulness of the result is limited. Nevertheless, our examinations reveal some enlightening and rather counter-intuitive insights that might be useful for further investigations.

While the existence of a relation between the ambiguity of erasing morphisms and certain properties of E-pattern languages (as, e. g., demonstrated by Condition 1 and Theorem 5) is by no means surprising, our characterisation shall demonstrate likewise deep connections between the main subject of our paper and vital properties of NE-pattern languages. It reads as follows:

Theorem 10. *Let Σ be an alphabet, and let $\alpha \in \mathbb{N}^+$. For any partition (U, V) of $\mathcal{P}(\text{var}(\alpha)) \setminus \{\emptyset\}$, let $L_{\alpha, U, V} := \bigcup_{u \in U} L_{\text{NE}, \Sigma}(\pi_u(\alpha)) \cap \bigcup_{v \in V} L_{\text{NE}, \Sigma}(\pi_v(\alpha))$. There is no unambiguous word in $L_{\text{E}, \Sigma}(\alpha) \setminus \{\varepsilon\}$ if and only if there is no unambiguous word in $L_{\text{E}, \Sigma}(\alpha) \setminus (\{\varepsilon\} \cup L_{\alpha, U, V})$.*

It is a noteworthy property of Theorem 10 that it covers the ambiguity of both erasing and nonerasing morphisms and, hence, allows a unified view on both topics. However, for the latter case, Theorem 1 already gives a definite answer, indirectly stating that, for every *succinct* pattern α , there is *no* partition (U, V) of $\mathcal{P}(\text{var}(\alpha)) \setminus \{\emptyset\}$ such that every word in $L_{\text{E}, \Sigma}(\alpha) \setminus (\{\varepsilon\} \cup L_{\alpha, U, V})$ is ambiguous for α . Thus, we can completely concentrate on prolix patterns when investigating applicability and consequences of Theorem 10.

From a practical point of view, Theorem 10 is not too helpful yet, as it merely reduces the number of words that need to be examined with regard to their ambiguity. Thus, it cannot be seen as an applicable characterisation of those patterns that have an unambiguous erasing morphism. On the other hand, it constitutes a promising starting point for further research on that topic, asking how U and V have to be chosen such that $L_{\alpha, U, V}$ has maximal size and what a maximal $L_{\alpha, U, V}$ looks like for a given α . In this regard, it is worth mentioning that example patterns α and sets $U, V \subseteq \mathcal{P}(\text{var}(\alpha)) \setminus \{\emptyset\}$ can be given where $L_{\alpha, U, V}$ is a nonempty subset of $L_{\text{E}, \Sigma}(\alpha)$ or even equals $L_{\text{E}, \Sigma}(\alpha) \setminus \{\varepsilon\}$.

Since, for any pattern α , $L_{\text{E}, \Sigma}(\alpha)$ is equivalent to a finite union of NE-pattern languages (see Theorem 2.1 by Jiang et al. [5]), Theorem 10 shows that the existence of unambiguous *erasing* morphisms strongly depends on equivalence and inclusion of certain finite unions of NE-pattern languages (or *nonerasing multi-pattern languages*, as they are called by Kari et al. [7]). This is not only a rather counter-intuitive insight, but it also gives an idea of how difficult the problem of the existence of unambiguous erasing morphisms might be. More precisely, even the decidability of the inclusion problem for ordinary terminal-free NE-pattern languages is open and includes some prominent open problems on pattern avoidability (cf. [5]). The inclusion of terminal-free NE-pattern languages is also known to depend on the size of the target alphabet, which fits very well with what is known for the subject of our paper (see, e. g., Theorem 9).

The following sufficient condition illustrates how Theorem 10 can be used to find criteria on the nonexistence of unambiguous erasing morphisms:

Corollary 4. *Let Σ be an alphabet, and let $\alpha \in \mathbb{N}^+$. If there exists a partition (U, V) of $\mathcal{P}(\text{var}(\alpha)) \setminus \{\emptyset\}$ with $L_{\text{E}, \Sigma}(\alpha) \setminus \{\varepsilon\} = \bigcup_{u \in U} L_{\text{NE}, \Sigma}(\pi_u(\alpha)) = \bigcup_{v \in V} L_{\text{NE}, \Sigma}(\pi_v(\alpha))$, then there is no unambiguous word in $L_{\text{E}, \Sigma}(\alpha) \setminus \{\varepsilon\}$.*

We finally wish to mention that Theorem 10 and Corollary 4 do not need to be based on a *partition* (U, V) of $\mathcal{P}(\text{var}(\alpha)) \setminus \{\emptyset\}$. Alternatively, they could refer to *arbitrary* disjoint subsets U and V of $\mathcal{P}(\text{var}(\alpha)) \setminus \{\emptyset\}$.

7 Conclusion and Open Problems

Concerning the ambiguity of erasing morphisms, the partition of patterns into morphically unerasing and erasing patterns (introduced and studied in Sec-

tion 3) has similar importance as the partition into succinct and prolix patterns regarding the ambiguity of nonerasing morphisms: Both partitions characterise a vital property of strings, namely the (non-)existence of moderately ambiguous morphisms (cf. Theorem 7 and Reidenbach [9]). While, in the case of nonerasing morphisms, this restricted ambiguity can additionally be turned into unambiguity, this does not hold for erasing morphisms since their ambiguity essentially depends on the size of the target alphabet (cf. Corollary 2 and, featuring a rather unexpected insight, Theorem 9).

A characterisation of those patterns that have an unambiguous erasing morphism is the main remaining open problem on the subject of the present paper, and even its mere decidability is still unresolved. Due to the insights summarised above, it seems evident that any solution to it requires concepts that significantly differ from the techniques used regarding moderate ambiguity. Section 6 reveals fundamental and quite surprising connections between the ambiguity of erasing morphisms and decision problems for nonerasing multi-pattern languages. An examination of these topics might be a helpful starting point for future studies.

References

- [1] 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.
- [2] T. Harju and J. Karhumäki. Morphisms. In [12], chapter 7, pages 439–510.
- [3] T. Head. Fixed languages and the adult languages of OL schemes. *International Journal of Computer Mathematics*, 10:103–107, 1981.
- [4] Š. Holub. Polynomial-time algorithm for fixed points of nontrivial morphisms. *Discrete Mathematics*, 309:5069–5076, 2009.
- [5] T. Jiang, E. Kinber, A. Salomaa, K. Salomaa, and S. Yu. Pattern languages with and without erasing. *International Journal of Computer Mathematics*, 50:147–163, 1994.
- [6] T. Jiang, A. Salomaa, K. Salomaa, and S. Yu. Decision problems for patterns. *Journal of Computer and System Sciences*, 50:53–63, 1995.
- [7] L. Kari, A. Mateescu, G. Păun, and A. Salomaa. Multi-pattern languages. *Theoretical Computer Science*, 141:253–268, 1995.
- [8] A. Mateescu and A. Salomaa. Patterns. In [12], chapter 4.6, pages 230–242.
- [9] D. Reidenbach. *The Ambiguity of Morphisms in Free Monoids and its Impact on Algorithmic Properties of Pattern Languages*. Logos Verlag, Berlin, 2006.
- [10] D. Reidenbach. Discontinuities in pattern inference. *Theoretical Computer Science*, 397:166–193, 2008.
- [11] D. Reidenbach and J.C. Schneider. Morphically primitive words. *Theoretical Computer Science*, 410:2148–2161, 2009.
- [12] G. Rozenberg and A. Salomaa. *Handbook of Formal Languages*, volume 1. Springer, Berlin, 1997.
- [13] J.C. Schneider. Unambiguous erasing morphisms in free monoids. *Theoretical Informatics and Applications (RAIRO)*, to appear. DOI: 10.1051/ita/2009020.