Periodicity Forcing Words☆,☆☆

Joel D. Day*, Daniel Reidenbach*

Department of Computer Science, Loughborough University, Loughborough Leicestershire, LE11 3TU, UK

Johannes C. Schneider*

DIaLOGIKa GmbH, Pascalschacht 1, 66125 Saarbrücken, Germany

Abstract

The Dual Post Correspondence Problem asks, for a given word α, if there exists a non-periodic morphism g and an arbitrary morphism h such that $g(\alpha) = h(\alpha)$. Thus α satisfies the Dual PCP if and only if it belongs to a non-trivial equality set. Words which do not satisfy the Dual PCP are called *periodicity forcing*, and are important to the study of word equations, equality sets and ambiguity of morphisms. In this paper, a 'prime' subset of periodicity forcing words is presented. It is shown that when combined with a particular type of morphism it generates exactly the full set of periodicity forcing words. Furthermore, it is shown that there exist examples of periodicity forcing words which contain any given factor/prefix/suffix. Finally, an alternative class of mechanisms for generating periodicity forcing words is developed, resulting in a class of examples which contrast those known already.

Keywords: Equality sets, Morphisms, Dual Post Correspondence Problem, Periodicity forcing sets, Periodicity forcing words, Ambiguity of morphisms

1. Introduction

The Dual Post Correspondence Problem (Dual PCP) is a decidable variation of the famous Post Correspondence Problem (see Post [12]). It was introduced by Culik II and Karhumäki in [1], where the authors make progress towards a characterisation of binary equality sets. A word is said to satisfy the Dual PCP if it belongs to an equality set $\mathrm{E}(g, h)$ for two morphisms g, h where at least

☆This work was supported by the London Mathematical Society, grant SC7-1112-02.

☆☆A preliminary version [4] of this work was presented at the conference WORDS 2013.

*Corresponding author.

Email addresses: `J.Day@lboro.ac.uk` (Joel D. Day), `D.Reidenbach@lboro.ac.uk` (Daniel Reidenbach), `johannes.schneider@dialogika.de` (Johannes C. Schneider)

one morphism is non-periodic. For example, the word $\mathtt{abba}$ belongs to $\mathrm{E}(g,h)$ where g, $h : \{\mathtt{a},\mathtt{b}\}^* \to \{\mathtt{a},\mathtt{b}\}^*$ are the morphisms given by:

$$g(x) := \begin{cases} \mathtt{aba} & \text{if } x = \mathtt{a}, \\ \mathtt{b} & \text{if } x = \mathtt{b}, \end{cases} \quad \text{and} \quad h(x) := \begin{cases} \mathtt{a} & \text{if } x = \mathtt{a}, \\ \mathtt{bab} & \text{if } x = \mathtt{b}. \end{cases}$$

Thus $\mathtt{abba}$ satisfies the Dual PCP; in other words, it is a non-trivial *equality word*. In contrast, the word $\mathtt{abaab}$ does not satisfy the Dual PCP, but this claim is much harder to verify. The latter is called a *periodicity forcing word* since it forces each pair of morphisms which agree on it to be periodic.

Identifying which words belong to non-trivial equality sets and which do not is of immediate significance to the Post Correspondence Problem, which is simply the emptiness problem for equality sets. It is well known that although the PCP is undecidable in general, it is decidable even in polynomial time in the binary case (see Halava, Holub [6]). It is therefore no surprise that, for binary words, the Dual PCP is relatively well understood.

This is due to both the original research by Culik II and Karhumäki [1], and from results on equality sets (e. g., Holub [7], Hadravova, Holub [5]) and word equations (e. g., Czeizler et al. [2], Karhumäki, Petre [9]). Much less, however, is known about the Dual PCP for larger alphabets.

One reason for this is that although the Dual PCP is known to be decidable, the proof (given by Culik II and Karhumäki [1]) relies on Makanin's algorithm for solving word equations [11]. While this algorithm demonstrates that the problem is computable in principle, the complexity is extremely high, and it provides little insight into the nature of words which do/do not satisfy the Dual PCP. It is worth noting that the decidability of the PCP for alphabet sizes 3 to 6 is a long-standing open problem, and therefore equality words over these alphabets are of particular interest.

In the present paper, we investigate the Dual PCP in the general case, specifically looking at periodicity forcing words. While examples of equality words are easily found, deciding on whether a word is periodicity forcing can be a particularly intricate task, and becomes even more so as the alphabet size increases. In [3], we overcome this problem by employing the use of morphisms to generate periodicity forcing words over arbitrary alphabets. Since it can be shown that many simple morphisms (such as $\varphi : \{\mathtt{a},\mathtt{b}\}^* \to \{\mathtt{a},\mathtt{b}\}^*$ given by $\varphi(\mathtt{a}) := \mathtt{a}$ and $\varphi(\mathtt{b}) := \mathtt{ab}$) preserve the property of being periodicity forcing, it is possible to span large parts of the set of periodicity forcing words (denoted by $\mathrm{DPCP}^\neg$) by applying such morphisms to existing examples.

In Section 3 of the present paper, we explore this phenomenon further. Specifically, $\mathrm{DPCP}^\neg$ is divided into those words which may be reached by a non-trivial morphism from other elements of the set, and those which cannot. The latter form a 'prime' subset of $\mathrm{DPCP}^\neg$ from which all periodicity forcing words may be generated using a specific class of morphisms characterised in [3]. In order to find examples of these prime words – therefore demonstrating that the subset is non-empty – it makes sense to consider the shortest periodicity forcing words. Thus, we also give bounds on the length of the shortest periodicity forcing words for any alphabet.

In Section 4, it is shown that there exist periodicity forcing words with arbitrary factors. This not only further demonstrates the complexity of the Dual PCP, but also provides another large, previously unknown class of periodicity forcing words and with it, further insight into their structure.

Finally, motivated by Section 3, we employ some alternative techniques for finding periodicity forcing words over larger alphabets, yielding insights into the set of 'prime' words.

2. Notation and Preliminary Results

An alphabet Σ is a set of symbols, or *letters*. A word over Σ is a concatenation of symbols from Σ. The empty word consisting of no symbols is ε. We denote by Σ^* the set of all words over Σ (including ε). Σ^+ is $\Sigma^* \backslash \{\varepsilon\}$. Let Σ be an alphabet. Let $u, v \in \Sigma^*$. Then v is a *factor* of u if there exist w_1, $w_2 \in \Sigma^*$ such that $u = w_1 v w_2$. A word $u \in \Sigma^*$ is *primitive* if $u = v^n$ for some $v \in \Sigma^*$ implies $n = 1$, otherwise u is *imprimitive*. If $u = v^n$ for some $n \in \mathbb{N}$ and v is primitive, then v is a *primitive root* of u; it is unique if and only if $u \neq \varepsilon$. Two words u, $v \in \Sigma^*$ *commute* if $uv = vu$. More generally, a set of words $\{u_1, u_2, \cdots, u_n\}$ commutes if for every i, j, $u_i u_j = u_j u_i$. For a set X, the notation $|X|$ refers to the cardinality of X, and for a word u, $|u|$ stands for the length of u. By $|u|_{\mathtt{a}}$, we denote the number of occurrences of the letter a in the word u. Let $u \in \{\mathtt{a}_1, \mathtt{a}_2, \cdots, \mathtt{a}_n\}^*$ be a word. The *Parikh vector of* u, denoted by $\mathrm{P}(u)$, is the vector $(|u|_{\mathtt{a}_1}, |u|_{\mathtt{a}_2}, \cdots, |u|_{\mathtt{a}_n})$. The result of dividing the Parikh vector by the greatest common divisor of its components is called the *basic Parikh vector*. A word $u \in \Sigma^*$ is *ratio-imprimitive* if there exist v, $w \in \Sigma^*$ such that $u = vw$ and v, w have the same basic Parikh vector. Otherwise u is ratio-primitive.

Let $\mathbb{N} := \{1, 2, \cdots\}$ be the set of natural numbers, and let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. We often use $\mathbb{N}$ as an infinite alphabet of symbols. In order to distinguish between a word over $\mathbb{N}$ and a word over a (possibly finite) alphabet Σ, we call the former a *pattern*. Given a pattern $\alpha \in \mathbb{N}^*$, we call symbols occurring in α *variables* and denote the set of variables in α by $\mathrm{var}(\alpha)$. Hence, $\mathrm{var}(\alpha) \subseteq \mathbb{N}$. Sometimes, for convenience, we will also use $\{x_1, x_2, \cdots\}$ to denote (possibly unknown) variables in $\mathbb{N}$. 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 patterns α and α', if α' may be obtained from α by deleting all occurrences of some variables in α, then α' is a *subpattern* of α. If $\mathrm{var}(\alpha) = \{1, 2, \cdots, n\}$ and the leftmost occurrence of each variable $x \in \mathbb{N}$ appears to the left of any variable y with $y > x$, then α is in *canonical form*.

Given arbitrary alphabets $\mathcal{A}, \mathcal{B}$, a *morphism* is a mapping $h : \mathcal{A}^* \to \mathcal{B}^*$ that is compatible with concatenation, i.e., for all $v, w \in \mathcal{A}^*$, $h(vw) = h(v)h(w)$. Hence, h is fully defined for all $v \in \mathcal{A}^*$ as soon as it is defined for all symbols in $\mathcal{A}$. A morphism h is called *periodic* if and only if there exists a $v \in \mathcal{B}^*$ such that $h(a) \in \{v\}^*$ for every $a \in \mathcal{A}$. The morphisms g, $h : \mathcal{A}^* \to \mathcal{B}^*$ are *distinct* if and only if there exists an $a \in \mathcal{A}$ such that $g(a) \neq h(a)$. For the *composition* of two morphisms $g, h : \mathcal{A}^* \to \mathcal{A}^*$, we write $g \circ h$, i. e., for every $w \in \mathcal{A}^*$, $g \circ h(w) = g(h(w))$. If $g(v) = h(v)$ for some $v \in \mathcal{A}^+$, then g and h *agree* on v. The set of all words on which g and h agree is called the *equality set* of g and h. A morphism $g : \mathcal{A}^* \to \mathcal{B}^*$ is called a *renaming morphism* if it is injective, and $|g(a)| = 1$ for every $a \in \mathcal{A}$. For words $u, v \in \mathcal{A}^+$, if there exists a renaming morphism g such that $v = g(u)$, then v is simply said to be a *renaming* of u.

Two words $u \in \mathcal{A}^+$, $v \in \mathcal{B}^+$ are *morphically coincident* if there exist morphisms $g : \mathcal{A}^* \to \mathcal{B}^*$ and $h : \mathcal{B}^* \to \mathcal{A}^*$ such that $g(u) = v$ and $h(v) = u$. A pattern $\alpha \in \mathbb{N}^+$ is *morphically imprimitive* if it is morphically coincident to some pattern β with $|\beta| < |\alpha|$. Otherwise α is *morphically primitive.* It is shown in [13] that if two patterns are morphically coincident, then they are either renamings of each other, or at least one is morphically imprimitive.

A morphism g is said to be *ambiguous* with respect to a pattern α if there exists another morphism h such that $g(\alpha) = h(\alpha)$ and g, h are distinct. Thus a pattern α satisfies the Dual PCP (see Section 1) if there exists an ambiguous non-periodic morphism with respect to α. In order to remain consistent with the notation in [3], we will often use σ and τ to denote morphisms when considering ambiguity and the Dual PCP, especially if they map patterns in $\mathbb{N}^*$ to words in Σ^*. We will normally use φ and ψ if we are mapping patterns to other patterns.

It is convenient, particularly in Section 3, to refer to the set of patterns which satisfy the Dual PCP and its complement. Thus we define the set: DPCP := $\{\alpha \in \mathbb{N}^+ \mid$ there exists a non-periodic morphism σ and an arbitrary morphism τ such that $\sigma(\alpha) = \tau(\alpha)$ and $\sigma(x) \neq \tau(x)$ for some $x \in \mathrm{var}(\alpha)\}$. We denote the complement of DPCP by DPCP$^\neg$. Note that DPCP$^\neg$ is exactly the set of periodicity forcing words (see Section 1).

We can extend periodicity forcing words to periodicity forcing sets in the natural way: a set of patterns is *periodicity forcing* if, whenever two distinct morphisms agree on all patterns in the set, they are periodic. A set of patterns T is said to be a *test set* of another set of patterns S if any two morphisms which agree on every pattern in T also agree on every pattern in S. Note that this means any test set of a periodicity forcing set must also be periodicity forcing.

For a set of unknowns $X := \{x_1, x_2, \cdots, x_n\}$, a word equation is an equation $\Phi = \Psi$ for some words $\Phi, \Psi \in X^+$. Its solutions, over some given alphabet Σ, are words $w_1, w_2, \cdots, w_n \in \Sigma^*$ such that substituting each w_i for x_i resolves the equation (it is equal on both sides). Thus solutions to the word equation may be expressed as morphisms $\sigma : X^* \to \Sigma^*$ such that $\sigma(\Phi)$ and $\sigma(\Psi)$ are equal. Unless otherwise specified, X is usually a set of variables, while Σ is a set of letters. As a result, word equations equate patterns, and their solutions are substitutions to terminal words (words which are not patterns). We will say that a set of words satisfies an equation if the associated substitution/morphism is a solution. We will use the following well known and fundamental result on word equations throughout the rest of the paper.

Lemma 1 (Lothaire [10]). *Non-trivial word equations in two unknowns have only periodic solutions.*

Thus, one obtains the following. For the third statement in the corollary to hold, it must be assumed that every primitive word is a primitive root of the empty word. Note that this fits with our definition given above.

Corollary 2. *Let u, v be words. The following conditions are equivalent:*

1. *u and v satisfy a non-trivial equation,*
2. *u and v commute, and*
3. *u, v have the same primitive root.*

In our investigation into the use of morphisms to generate periodicity forcing words in [3], we provide the following criterion. Any morphism φ which satisfies the criterion preserves the property of being periodicity forcing, and thus can be used to obtain new periodicity forcing words from known ones.

Lemma 3 ([3]). *Let Δ_1, Δ_2 be sets of variables. Let $\varphi : {\Delta_1}^* \to {\Delta_2}^*$ be a morphism such that, for every $x \in \Delta_2$, there exists a $y \in \Delta_1$ satisfying $x \in \mathrm{var}(\varphi(y))$, and*

(i) *for every non-periodic morphism $\sigma : {\Delta_2}^* \to \{\mathtt{a}, \mathtt{b}\}^*$, $\sigma \circ \varphi$ is non-periodic, and*

(ii) *for all distinct morphisms σ, $\tau : {\Delta_2}^* \to \{\mathtt{a}, \mathtt{b}\}^*$, where at least one is non-periodic, $\sigma \circ \varphi$ and $\tau \circ \varphi$ are distinct.*

Then, for any $\alpha \notin$ DPCP with $\mathrm{var}(\alpha) = \Delta_1$, $\varphi(\alpha) \notin$ DPCP.

Characterisations of morphisms which satisfy conditions (i) and (ii) of Lemma 3 are given in [3]. In particular, condition (ii) is satisfied if and only if the set $S := \{\varphi(x) \mid x \in \Delta_1\}$ is periodicity forcing.

Since a set of patterns commutes if and only if each pair of patterns in the set commutes, by Corollary 2, the morphism $\sigma \circ \varphi : \Delta_1 \to \{\mathtt{a}, \mathtt{b}\}^*$ is periodic if and only if the set $\{\sigma(\varphi(x)) \mid x \in \Delta_1\}$ commutes. Hence, condition (i) is satisfied if and only if the set S is *commutativity forcing*, that is, for every morphism σ for which the set $\{\sigma(\beta) \mid \beta \in S\}$ commutes, all images $\sigma(x)$, $x \in {\Delta_2}^*$ commute. This implies that σ is periodic.

Note that it follows from basic properties of morphisms that, for any periodicity forcing (resp. commutativity forcing) set $\{\beta_1, \beta_2, \cdots, \beta_n\}$, if a new pattern β_{n+1} is added which does not contain any 'new' variables (i.e., variables which do not appear in any β_i, $1 \leq i \leq n$), then the resulting set remains periodicity forcing (resp. commutativity forcing).

3. A 'Prime' Generating Subset of DPCP$^\neg$

A method of obtaining periodicity forcing words as the morphic images of previously known examples is developed in [3]. One consequence of the constructions given is the following:

Corollary 4 ([3]). *Let $\alpha \in \mathrm{DPCP}^{\neg}$. Then there exists a morphism $\varphi : \mathbb{N}^* \to \mathbb{N}^*$ which is not a renaming morphism, such that $\varphi(\alpha) \in \mathrm{DPCP}^{\neg}$.*

Although this statement is itself fairly easily obtained, and comes as no surprise, it is worth noting the richness and variety in such morphisms φ (which are characterised in [3]), and therefore also in the subsequent patterns $\varphi(\alpha)$ which can be obtained through the application of morphisms. Thus an obvious question arises: is every periodicity forcing word the morphic image of another?

Of course the answer is trivially affirmative if φ is permitted to be a renaming morphism (such as the identity), or if α can be unary (every pattern is a morphic image of $\alpha := 1$). However, if we restrict α and φ to avoid these trivial instances, the answer is no longer clear. In fact, a negative answer is provided by Proposition 9 below. Hence, the partition of periodicity forcing words into those which are morphic images of another, and those which are not, is non-trivial. We will call the latter *prime*. Moreover, it is reasonable to expect that these prime periodicity forcing words are sufficient, given the appropriate set of morphisms, to generate the full set. This is confirmed later by Theorem 10.

The proofs of these results rely on a lower bound for the length of periodicity forcing words, given relative to the alphabet size. This bound is achieved by considering patterns belonging to the equality sets of (pairs of) "nearly periodic morphisms" σ – of the form

$$\sigma(x) := \begin{cases} \mathtt{a}^r\ \mathtt{ba}^s & \text{if } x = y, \\ \mathtt{a}^{p_x} & \text{otherwise,} \end{cases}$$

where y is some fixed variable, and $r, s, p_x \in \mathbb{N}_0$. It is apparent that the equality set of two morphisms σ_1 and σ_2 of this type is determined by a system of linear Diophantine equations, and in the case that y is the same for both morphisms, it is possible to infer a strong sufficient condition for a pattern to belong to such an equality set. Since the morphisms are non-periodic, any such pattern is not periodicity forcing.

Proposition 5. *Let α be a pattern, and let $n := |\mathrm{var}(\alpha)|$. Suppose that $|\alpha|_x < n$ for some $x \in \mathrm{var}(\alpha)$. Then $\alpha \in \mathrm{DPCP}$.*

Proof. Consider a pattern α such that $\mathrm{var}(\alpha) = \{x_1, x_2, ..., x_n\}$, and $|\alpha|_{x_i} < n$ for some $i \leq n$. W.l.o.g. let $i := n$. Then there exists a $k \in \mathbb{N}$ such that $|\alpha|_{x_n} = n - k$, and α can be written as $\beta_1 \cdot x_n \cdot \beta_2 \cdot x_n \cdot ... \cdot \beta_{n-k} \cdot x_n \cdot \beta_{n-k+1}$ for some patterns $\beta_1, \beta_2, ..., \beta_{n-k+1} \in \{x_1, x_2, ..., x_{n-1}\}^*$.

Consider the morphisms $\sigma, \tau : \{x_1, x_2, ..., x_n\}^* \to \{\mathtt{a},\mathtt{b}\}^*$ given by

$$\sigma(x_i) := \begin{cases} \mathtt{a}^{r_1}\ \mathtt{ba}^{s_1} & \text{if } i = n, \\ \mathtt{a}^{p_i} & \text{otherwise,} \end{cases} \quad \text{and } \tau(x_i) := \begin{cases} \mathtt{a}^{r_2}\ \mathtt{ba}^{s_2} & \text{if } i = n, \\ \mathtt{a}^{q_i} & \text{otherwise,} \end{cases}$$

for some $p_1, p_2, ..., p_n, q_1, q_2, ..., q_n, r_1, r_2, s_1, s_2 \in \mathbb{N}_0$. Clearly, σ and τ are non-periodic, provided $p_i \neq 0$ and $q_j \neq 0$ for some i, j respectively. For

$1 \leq i < n$ let $t_i := p_i - q_i$, let $r := r_2 - r_1$, and let $s := s_2 - s_1$. Then $\sigma(\alpha) = \tau(\alpha)$ if and only if the following system of equations is satisfied:

$$t_1|\beta_1|_{x_1} + t_2|\beta_1|_{x_2} + \cdots + t_{n-1}|\beta_1|_{x_{n-1}} = r$$

$$t_1|\beta_2|_{x_1} + t_2|\beta_2|_{x_2} + \cdots + t_{n-1}|\beta_2|_{x_{n-1}} = r + s$$

$$\vdots$$

$$t_1|\beta_{n-k}|_{x_1} + t_2|\beta_{n-k}|_{x_2} + \cdots + t_{n-1}|\beta_{n-k}|_{x_{n-1}} = r + s$$

$$t_1|\beta_{n-k+1}|_{x_1} + t_2|\beta_{n-k+1}|_{x_2} + \cdots + t_{n-1}|\beta_{n-k+1}|_{x_{n-1}} = s.$$

Since r, s, t_1, ..., $t_{n-1} \in \mathbb{N}_0$ depend on the definition of σ and τ, they may be chosen freely, and $|\beta_i|_{x_j}$ for $1 \leq i \leq n-k+1$ and $1 \leq j \leq n-1$ depend on α so they are fixed. Notice that σ and τ are distinct if and only if $s \neq 0$ or $r \neq 0$ or there exists an i such that $t_i \neq 0$. Thus, σ and τ can be chosen such that they are distinct, non-periodic and agree on α if there exists a non-trivial solution $(t_1, t_2, ..., t_{n-1})$.

Let $f_{i,j} := |\beta_i|_{x_j} - |\beta_1|_{x_j} - |\beta_{n-k+1}|_{x_j}$ for $1 \leq i < n$ and $2 \leq n-k$. Then our system can be written as follows:

$$t_1 f_{2,1} + t_2 f_{2,2} + \cdots + t_{n-1} f_{2,n-1} = 0$$

$$t_1 f_{3,1} + t_2 f_{3,2} + \cdots + t_{n-1} f_{3,n-1} = 0$$

$$\vdots$$

$$t_1 f_{n-k,1} + t_2 f_{n-k,2} + \cdots + t_{n-1} f_{n-k,n-1} = 0$$

This is a system of $n-k-1$ homogeneous equations in $n-1$ unknowns with integer coefficients, $k \geq 1$, and therefore there exists a non-trivial integer solution $(t_1, t_2, \ldots\ t_{n-1})$. Since r and s can be chosen freely, such a solution is always a solution to the first system for some integers r, s. Thus σ and τ can be chosen such that they are distinct, non-periodic and agree on α. Consequently, $\alpha \in \mathrm{DPCP}$ whenever $|\alpha|_x < n$ for some $x \in \mathrm{var}(\alpha)$. □

It follows that, for a periodicity forcing word with n variables, each variable must occur at least n times, implying the next corollary which provides a lower bound on the length of the shortest periodicity forcing word for any alphabet size.

Corollary 6. *Let $\alpha \notin \mathrm{DPCP}$, and let $n := |\mathrm{var}(\alpha)|$. Then $|\alpha| \geq n^2$.*

Since periodicity forcing words can be obtained as concatenations of words in a particular type of periodicity forcing set (see Section 5), it is possible to infer a corresponding upper bound from results by Holub, Kortelainen [8]. The authors provide a concise test set (containing at most $5n$ words, each of length n) for the set S_n consisting of all permutations of the word $x_1 \cdot x_2 \cdots x_n$. Although it is stated in [8] that S_n itself is not periodicity forcing, it can be verified using results from [8] and [1] that the augmented set ${S_n}' := S_n \cup \{x_1 \cdot x_1 \cdot x_2 \cdot x_2 \cdots x_n \cdot x_n\}$ is. Given a test set T_n for S_n, a test set for ${S_n}'$ is clearly $T_n \cup \{x_1 \cdot x_1 \cdot x_2 \cdot x_2 \cdots x_n \cdot x_n\}$. Thus, there exists a test set for ${S_n}'$ containing at most $5n$ words of length n and one word of length $2n$. The periodicity forcing word resulting from concatenating these words is at most $5n^2 + 2n$ letters long.

Proposition 7. *Let α_n be a shortest pattern not in* DPCP *such that $|\mathrm{var}(\alpha)| = n$. Then $n^2 \leq |\alpha| \leq 5n^2 + 2n$.*

The lower bounds are particularly useful when considering prime elements of $\mathrm{DPCP}^\neg$, which we define formally below.

Definition 8. *Let $\alpha \in \mathrm{DPCP}^\neg$ be a pattern with $|\mathrm{var}(\alpha)| \geq 2$. Then α is said to be a* prime element of $\mathrm{DPCP}^\neg$ *(or simply* prime*) if for every pattern $\beta \in \mathrm{DPCP}^\neg$ with $|\mathrm{var}(\beta)| > 1$, and every morphism $\varphi : \mathrm{var}(\beta)^* \to \mathrm{var}(\alpha)^*$, $\varphi(\beta) = \alpha$ implies that φ is a renaming morphism.*

Showing that a pattern satisfies Definition 8 is, in general, a highly non-trivial task, since all morphisms must be accounted for with respect to every pattern $\beta \in \mathrm{DPCP}^\neg$. However, due to Proposition 5, it is possible to provide a relatively simple example:

Proposition 9. *The pattern $\alpha := 1 \cdot 2 \cdot 1 \cdot 1 \cdot 2$ is a prime element of* $\mathrm{DPCP}^\neg$.

Proof. It is known from Culik II, Karhumäki [1] that α is periodicity forcing. Assume that $\beta \in \mathrm{DPCP}^\neg$ is a pattern, and that $\varphi : \mathrm{var}(\beta)^* \to \mathrm{var}(\alpha)^*$ is a morphism such that $\varphi(\beta) = \alpha$. Due to the fact that $|\alpha|_2 = 2$, there exists a variable $x \in \mathrm{var}(\beta)$ such that $|\beta|_x \leq 2$. Hence, by Proposition 5, $|\mathrm{var}(\beta)| = 2$. Since α is primitive, φ is non-erasing and thus $|\beta| \leq 5$. Furthermore, all periodicity forcing words of length at most 5 are given by Culik II, Karhumäki [1], so it is possible to determine by inspection that no non-renaming morphism exists which maps any of these patterns to α, and thus Definition 8 is satisfied. □

By the same argument, the patterns $1 \cdot 2 \cdot 1 \cdot 2 \cdot 2$, $1 \cdot 1 \cdot 2 \cdot 1 \cdot 2$ and $1 \cdot 2 \cdot 2 \cdot 1 \cdot 2$ are also prime.

As mentioned earlier, while Proposition 9 settles the question of whether every periodicity forcing word is the morphic image of another in a non-trivial way, the negative answer induces a second question: what is the smallest subset of $\mathrm{DPCP}^\neg$ required to span the full set via the application of morphisms? Clearly such a subset is strict (this follows from Corollary 4), and must be a superset of the set of prime elements of $\mathrm{DPCP}^\neg$.

In order to answer this question, it is necessary to determine whether there exist infinite chains of patterns

$$\cdots \to \beta_i \to \beta_{i+1} \to \beta_{i+2} \to \cdots \to \beta_{i+n} \to \cdots$$

where each β_i is the morphic image of β_{i-1}. By Corollary 4, all such chains can continue indefinitely in one direction. Theorem 10 below confirms that any such chain must terminate in the other. Note that for convenience when proving the theorem, the order of the indices of the patterns β_i has been reversed.

Theorem 10. *There does not exist an infinite sequence of periodicity forcing words $S := \beta_0, \beta_1, \beta_2, \cdots$ such that for every $i > 1$,*

- *there exists a morphism φ_i satisfying $\beta_{i-1} = \varphi_i(\beta_i)$ and*
- *φ_i is not a renaming morphism.*

PROOF. Assume to the contrary that a such a sequence S exists which is infinite. For any $i, j \in \mathbb{N}_0$ with $i > j$, let $\psi_{i,j} := \varphi_{j+1} \circ \varphi_{j+2} \circ \cdots \circ \varphi_i$, so that $\psi_{i,j}(\beta_i) = \beta_j$. We will need to use the following results from [13]: firstly that if two patterns are morphically coincident, then they are either the same (up to renaming) or at least one is morphically imprimitive and therefore not periodicity forcing, and secondly that if a pattern is fixed by a non-trivial morphism (not the identity), it is morphically imprimitive. We now prove some further preliminary claims.

Claim 1: No patterns β_i, β_j, $i \neq j$, in the sequence S are renamings of each other.

PROOF (CLAIM 1). Assume to the contrary that, for some $i, j \in \mathbb{N}_0$ with $i > j$, β_i is a renaming of β_j. Let σ be the renaming morphism such that $\sigma(\beta_j) = \beta_i$. If $i = j + 1$, then $\varphi_i(\beta_i) = \beta_j$. Thus, $\sigma \circ \varphi_i(\beta_i) = \beta_i$. However, since φ_i is not a renaming morphism, $\sigma \circ \varphi_i$ is not the identity, and β_i is morphically imprimitive. If $i > j + 1$, then $\varphi_i(\beta_i) = \beta_{i-1}$, and $\psi_{i-1,j}(\beta_{i-1}) = \beta_j$. This implies $\sigma \circ \psi_{i-1,j}(\beta_{i-1}) = \beta_i$. Thus, at least one of β_i, β_{i-1} is morphically imprimitive. □(Claim 1)

Our second claim provides a bound on the number of variables occurring in the patterns β_i.

Claim 2: There exists $n \in \mathbb{N}$ such that every pattern in S has at most n variables.

PROOF (CLAIM 2). Let $n := |\beta_0|$. Let $i \in \mathbb{N}$ be arbitrary and consider the morphism $\psi_{i,0}$ mapping β_i to β_0. In particular, consider the subset of $\mathrm{var}(\beta_i)$ of variables which are not erased by $\psi_{i,0}$. Clearly the subset contains at least one variable x. Furthermore, $|\beta_i|_x \leq n$. By Proposition 5, it follows that $|\mathrm{var}(\beta_i)| \leq n$. □(Claim 2)

Note that we can replace any β_i with one of its renamings, and S will still satisfy the criteria of the theorem. Thus, by assuming that the patterns of the sequence are in canonical form, we can assume that there exists a finite alphabet Δ such that each $\beta_i \in \Delta^*$. We now give our final preliminary claim.

Claim 3: Any infinite subsequence of S also satisfies the conditions of the theorem.

PROOF (CLAIM 3). Let $S' = \beta_{p_0}, \beta_{p_1}, \beta_{p_2}, \ldots$ be an infinite subsequence of S. Then, for every $p_i > 1$, there exists a morphism φ'_{p_i} satisfying $\beta_{p_{i-1}} = \varphi'_{p_i}(\beta_{p_i})$ (simply take $\varphi' = \psi_{p_i, p_{i-1}}$). Furthermore, by Claim 1, each φ'_{p_i} cannot be a renaming morphism. Thus S' satisfies the conditions of the theorem.□(Claim 3)

$$
\begin{tikzcd}
\beta_0^{(k)} & \beta_1^{(k)} \arrow[l, "\varphi_0^{(k)}"'] & \beta_2^{(k)} \arrow[l, "\varphi_1^{(k)}"'] & \beta_3^{(k)} \arrow[l, "\varphi_2^{(k)}"'] & \beta_4^{(k)} \arrow[l, "\varphi_3^{(k)}"'] & \beta_5^{(k)} \arrow[l, "\varphi_4^{(k)}"'] \\
 & \delta_0^{(k)} \arrow[ul, dashed] & \delta_0^{(k)} & \delta_0^{(k)} & \delta_0^{(k)} & \delta_0^{(k)} \\
 & & \delta_1^{(k)} \arrow[uul, dashed] & \delta_1^{(k)} \arrow[uul, dashed] & \delta_1^{(k)} & \delta_1^{(k)} \\
 & & & \delta_2^{(k)} & \delta_2^{(k)} \arrow[uuul, dashed] & \delta_2^{(k)} \\
 & & & & \delta_3^{(k)} & \delta_3^{(k)} \arrow[uuuul, dashed] \\
 & & & & & \delta_4^{(k)}
\end{tikzcd}
$$

Figure 1: Depiction of the first 5 patterns of the sequence S_k. Each pattern $\beta_i^{(k)}$ has its subpatterns $\delta_j^{(k)}$ listed below. Solid arrows indicate the morphisms which are explicitly given in the definition of the sequence, while the dashed arrows represent the implicit non-erasing morphisms from the subpatterns. Note that for clarity, the dotted arrows are omitted for all but the leftmost occurrence of each $\delta_i^{(k)}$.

We are now ready to prove the theorem, which we do by deriving from S an infinite subsequence S_k which satisfies the conditions for the theorem whenever S does. Thus, by showing S_k does not satisfy the conditions, we obtain a contradiction and our assumption that S is infinite cannot hold.

Let $\delta_{i,0}$ be the subpattern of β_i whose variables are not erased by $\psi_{i,0}$. Since each $\delta_{i,0}$ contains only variables from a finite alphabet Δ, and must have length at most $|\beta_0|$, the set $\{\delta_{i,0} \mid i \in \mathbb{N}\}$ contains only finitely many different patterns. In particular, at least one such pattern $\delta_{i,0}$ must occur as a subpattern of infinitely many different patterns β_j. Let this pattern be δ_0. By Claim 3, the sequence S_0 obtained by removing all patterns after β_0 which do not have δ_0 as a subpattern still satisfies the criteria of the theorem. Note that S_0 is also still infinite. We will call the patterns of the modified sequence $\beta_0^{(0)}$, $\beta_1^{(0)}$, $\beta_2^{(0)}$ etc., and define the morphisms $\varphi_i^{(0)}$ and $\psi_{i,j}^{(0)}$ accordingly.

Similarly let $\delta_{i,1}^{(0)}$ be the subpattern of $\beta_i^{(0)}$ whose variables are not erased by $\psi_{i,1}^{(0)}$. By the same reasoning as above, there exists some infinitely occurring subpattern $\delta_1^{(0)}$, so we can produce an infinite subsequence S_1 of S_0 containing only the patterns $\beta_0^{(0)}$, $\beta_1^{(0)}$ and $\beta_i^{(0)}$ with $\delta_1^{(0)}$ as a subpattern when $i > 1$.

By repeating this process $k > 2^{|\Delta+1|}$ times, we have an infinite sequence S_k for which each pattern $\beta_i^{(k)}$, $i > k$ contains $\delta_0^{(k)}$, $\delta_1^{(k)}$, ..., $\delta_k^{(k)}$ as subpatterns (see Figure 1). Note that by definition, each $\beta_i^{(k)}$ is a (non-erasing) morphic image of $\delta_i^{(k)}$.

However, $\beta_i^{(k)}$ can only have finitely many (at most $2^{|\Delta|} - 1$) different, non-empty subpatterns. Thus there exist p, q, r such that $\delta_p^{(k)} = \delta_r^{(k)}$ for some $p > q > r$. Note that $\delta_r^{(k)}$ is a sub-pattern of $\beta_q^{(k)}$, since $q \geq r + 1$. Furthermore, there exists a morphism $\psi_{p,q}^{(k)}$ from $\beta_p^{(k)}$ to $\beta_q^{(k)}$. However, since $\delta_r^{(k)} (= \delta_p^{(k)})$ is a subpattern of $\beta_q^{(k)}$, there exists a morphism from $\beta_q^{(k)}$ to $\beta_p^{(k)}$ (see Figure 2). This implies they are morphically coincident, and since, by Claim 1, they are

$$\begin{tikzcd}
\beta_q^{(k)} \arrow[d] & & \beta_p^{(k)} \arrow[ll, "\psi_{p,q}^{(k)}"'] & \beta_{p+1}^{(k)} \arrow[l, dashed, "\varphi_p^{(k)}"'] \arrow[d, dashed] \\
\delta_r^{(k)} (= \delta_p^{(k)}) \arrow[urr] & & & \delta_p^{(k)} \arrow[ul, dashed]
\end{tikzcd}$$

Figure 2: Diagram showing morphic coincidence of $\beta_p^{(k)}$ and $\beta_q^{(k)}$. Morphisms are indicated by arrows, where the solid arrows indicate which morphisms responsible for the coincidence 'loop'.

not renamings of each other, at least one must be morphically imprimitive. This contradicts the assumption that all patterns are periodicity forcing, and thus completes the proof. □

Consequently every periodicity forcing word is either a prime element of $\mathrm{DPCP}^{\neg}$ or the morphic image of a prime element of $\mathrm{DPCP}^{\neg}$, and the set $\mathrm{DPCP}^{\neg}$ is spanned by one-sided infinite chains of the form

$$\beta_0 \to \beta_1 \to \cdots \beta_n \to \cdots$$

where each β_i is the morphic image of β_{i-1} and β_0 is prime.

Corollary 11. *Let α be a periodicity forcing word. Then α is either prime, or the morphic image of a prime periodicity forcing word.*

Since a characterisation of morphisms which map periodicity forcing words to periodicity forcing words is given in [3], Theorem 10 provides a strong insight into the structure of $\mathrm{DPCP}^{\neg}$.

By definition, it is not possible to use morphisms to generate prime periodicity forcing words, so alternative methods must be used to find them. This is is investigated in Section 5, where some additional insights are gained.

4. Patterns in $\mathrm{DPCP}^{\neg}$ with Arbitrary Factors

Section 3 and [3] present constructions for periodicity forcing words over any given alphabet. An immediate consequence is that we are also able to construct, for any pattern β, a periodicity forcing set containing β. For example, if $\beta' \in \mathrm{DPCP}^{\neg}$ and $\mathrm{var}(\beta) = \mathrm{var}(\beta')$, then $\{\beta, \beta'\}$ is periodicity forcing. More generally, the addition of a periodicity forcing word over an appropriate alphabet is sufficient to turn any finite set of patterns into a periodicity forcing set. Thus we have a high degree of freedom when producing sets which are periodicity forcing, and therefore also morphisms satisfying Lemma 3. In particular, we are able to construct, for any given pattern β, a morphism φ and pre-image α' such that the pattern $\alpha := \varphi(\alpha')$ is periodicity forcing and contains β as a factor, prefix or suffix.

In order to guarantee that φ satisfies the conditions given in Lemma 3, the set $\{\varphi(x) \mid x \in \mathrm{var}(\alpha)\}$ must not only be periodicity forcing, but also commutativity forcing – i.e. every morphism σ such that the words $\sigma(\varphi(x))$, $x \in \mathrm{var}(\alpha)$ commute is periodic. A construction satisfying this condition is given in the next proposition.

Proposition 12. *Let α_0 be a pattern, and let $n := \lceil \log_2(|\mathrm{var}(\alpha_0)|) \rceil$. There exist patterns α_1, α_2, ..., α_n with $\mathrm{P}(\alpha_0) = \mathrm{P}(\alpha_1) = \cdots = \mathrm{P}(\alpha_n)$ such that $\{\alpha_0, \alpha_1, \cdots, \alpha_n\}$ is commutativity forcing.*

Proof. Consider the case that $|\mathrm{var}(\alpha_0)| = 2^n$. The case that this is not true may easily be adapted. W.l.o.g. let α_0 be in canonical form, and note that this implies that α_0 can be expressed as $\gamma_1 \cdot \gamma_2 \cdots \gamma_m$ where $m = |\mathrm{var}(\alpha_0)|$, and $\gamma_i := i \cdot \beta_i$ for some pattern $\beta_i \in \{1, 2,..., i\}^*$. For $i \leq k$, let α_i be the pattern obtained from α_0 by 'swapping' adjacent factors consisting of 2^{i-1} consecutive patterns γ_j, i.e.,

$$\alpha_1 = \gamma_2 \cdot \gamma_1 \cdot \gamma_4 \cdot \gamma_3 \cdots \gamma_{m-1} \cdot \gamma_m$$

$$\alpha_2 = \gamma_3 \cdot \gamma_4 \cdot \gamma_1 \cdot \gamma_2 \cdots \gamma_{m-1} \cdot \gamma_m \cdot \gamma_{m-3} \cdot \gamma_{m-2}$$

$$\vdots$$

$$\alpha_k = \gamma_{\frac{m}{2}+1} \cdot \gamma_{\frac{m}{2}+2} \cdots \gamma_m \cdot \gamma_1 \cdot \gamma_2 \cdots \gamma_{\frac{m}{2}}$$

Note that $\mathrm{P}(\alpha_0) = \mathrm{P}(\alpha_1) = \cdots = \mathrm{P}(\alpha_k)$, so for any morphism σ, we have that

$$|\sigma(\alpha_0)| = |\sigma(\alpha_1)| = \cdots = |\sigma(\alpha_n)|.$$

Thus, the system of word equations

$$\alpha_i \alpha_j = \alpha_j \alpha_i$$

for all i, j with $0 \leq i < j \leq n$ is equivalent to the simpler system

$$\alpha_0 = \alpha_1 = \cdots = \alpha_n.$$

It is now shown that all solutions to the above system of word equations are periodic. Let $\sigma : \{1, 2, ..., n\}^* \to \{\mathtt{a}, \mathtt{b}\}^*$ be an arbitrary solution, and consider the equality $\alpha_0 = \alpha_1$. This is equivalent to

$$\sigma(\gamma_1) \cdot \sigma(\gamma_2) \cdots \sigma(\gamma_m) = \sigma(\gamma_2) \cdot \sigma(\gamma_1) \cdots \sigma(\gamma_m) \cdot \sigma(\gamma_{m-1}).$$

By comparing the prefix of length $|\sigma(\gamma_1)| + |\sigma(\gamma_2)|$ on either side, $\sigma(\gamma_1)\sigma(\gamma_2) = \sigma(\gamma_2)\sigma(\gamma_1)$. By Corollary 2, it follows that there exists a primitive word $w_1 \in \{\mathtt{a}, \mathtt{b}\}^*$ such that $\sigma(\gamma_1)$, $\sigma(\gamma_2) \in \{w_1\}^*$. A similar argument may be made for the next, and indeed every pair of patterns γ_j, γ_{j+1} where $j < m$ is odd. Thus, for $1 \leq i \leq \frac{m}{2}$, there exists a primitive word $w_i \in \{\mathtt{a}, \mathtt{b}\}^*$ such that $\sigma(\gamma_{2i-1})$, $\sigma(\gamma_{2i}) \in \{w_i\}^*$. Moreover, by the equation $\alpha_1 = \alpha_2$, it is possible to employ the same argument to determine that for $1 \leq i \leq \frac{m}{4}$, the words w_{2i-1} and w_{2i} are equal. By continuing this argument for each successive equality $\alpha_j = \alpha_{j+1}$, it follows that $w_1 = w_2 = \cdots = w_{\frac{m}{2}}$, so there exists a primitive word $w \in \{\mathtt{a}, \mathtt{b}\}^*$ such that $\sigma(\gamma_i) \in \{w\}^*$ for all $1 \leq i \leq m$.

Since $\gamma_1 \in 1^+$, this implies $\sigma(1) \in \{w\}^*$. Assume that $\sigma(1)$, $\sigma(2)$, ..., $\sigma(r) \in \{w\}^*$ for some $1 \leq r < m$. Then since $\sigma(\gamma_{r+1}) \in \{w\}^*$ and $\gamma_{r+1} \in \{1, 2, ..., r+1\}^+$ and $r+1 \in \mathrm{var}(\gamma_{r+1})$, by Lemma 1, $\sigma(r+1) \in \{w\}^*$. Thus, by induction, $\sigma(x) \in \{w\}^*$ for all $1 \leq x \leq m$, and σ is periodic. □

It is now possible to show that for any given pattern β, there exists a periodicity forcing word with β as a factor.

Theorem 13. *For any pattern $\beta \in \mathbb{N}^+$, there exists a pattern $\alpha \notin \mathrm{DPCP}$ such that β is a factor/prefix/suffix of α.*

Proof. It is known from [3] that there exists a pattern $\beta_1 \notin \mathrm{DPCP}$ such that $\mathrm{var}(\beta) = \mathrm{var}(\beta_1)$. By Proposition 12, there exist patterns β_2, β_3, ..., β_n with $\mathrm{P}(\beta) = \mathrm{P}(\beta_2) = \cdots = \mathrm{P}(\beta_n)$ such that the set $\{\beta, \beta_2, \cdots, \beta_n\}$ is commutativity forcing. Since $\mathrm{var}(\beta) = \mathrm{var}(\beta_i)$ for $1 \leq i \leq n$, it follows that the augmented set $\{\beta, \beta_1, \beta_2, \cdots, \beta_n\}$ is commutativity forcing. Furthermore, since β_1 is periodicity forcing, the set is also periodicity forcing. Thus the morphism $\varphi : \{1, 2, ..., n+1\}^* \to \mathrm{var}(\beta)^*$ given by $\varphi(i) := \beta_i$ for $1 \leq i \leq n$ and $\varphi(n+1) := \beta$ satisfies both conditions of Lemma 3. From [3], there exists a pattern $\alpha' \notin \mathrm{DPCP}$ such that $\mathrm{var}(\alpha') = \{1, 2, ..., n+1\}$, and by Lemma 3, $\alpha := \varphi(\alpha') \notin \mathrm{DPCP}$. Since $\beta = \varphi(n+1)$ and $n+1 \in \mathrm{var}(\alpha')$, β is a factor of α as required. The case that β is a prefix (resp. suffix) of α can be shown simply by using renamings of α' for which $n+1$ occurs at as a prefix (resp. suffix). □

Example 14 demonstrates how φ, and therefore α may be constructed in the case that $\beta = 1 \cdot 1 \cdot 2 \cdot 3$.

Example 14. *Let $\beta := 1 \cdot 1 \cdot 2 \cdot 3$. Let $\beta_1 := 1 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 3 \cdot 1 \cdot 1 \cdot 3 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 3 \cdot 1 \cdot 1 \cdot 3 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1$. By [3] (Proposition 32), $\beta_1 \notin \mathrm{DPCP}$. By Proposition 12, there exist patterns β_2, β_3 such that $\mathrm{P}(\beta) = \mathrm{P}(\beta_2) = \mathrm{P}(\beta_3)$, and $\{\beta, \beta_2, \beta_3\}$ is a commutativity forcing set. In particular, using the construction given in the proof of Proposition 12 we obtain $\beta_2 := 2 \cdot 3 \cdot 1 \cdot 1$, and $\beta_3 := 3 \cdot 1 \cdot 1 \cdot 2$. It is easy to verify that these patterns satisfy the condition, as any morphism σ will map β, β_2 and β_3 to words of the same length. Thus,*

$$\begin{aligned} \sigma(311)\sigma(2) = \sigma(2)\sigma(311) & \\ &= \sigma(11)\sigma(23) = \sigma(23)\sigma(11) \\ &= \sigma(112)\sigma(3) = \sigma(3)\sigma(112) \end{aligned}$$

implying that $\sigma(1)$, $\sigma(2)$ and $\sigma(3)$ commute, and hence σ is periodic. It follows that the extended set $\{\beta_1, \beta, \beta_2, \beta_3\}$ is commutativity forcing. Hence the morphism $\varphi : \{1, 2, 3, 4\}^* \to \{1, 2, 3\}^*$ given by $\varphi(i) := \beta_i$ for $1 \leq i \leq 3$ and $\varphi(4) := \beta$ satisfies Condition (ii) of Lemma 3. Since $\beta_1 \in \mathrm{DPCP}^\neg$, the *set of patterns* $\{\beta, \beta_1, \beta_2, \beta_3\}$ *is also periodicity forcing and thus* φ *satisfies Condition (ii) of the Lemma.*

Let α' *be a pattern in* $\mathrm{DPCP}^{\neg}$ *with* $\mathrm{var}(\alpha) = \{1, 2, 3, 4\}$. *Then by Lemma 3,* $\alpha := \varphi(\alpha) \in \mathrm{DPCP}^{\neg}$. *Moreover it is clear that* β *appears as a factor of* α.

5. An Alternative Means of Finding Patterns not in DPCP

While Section 3 provides motivation for the further study of generating periodicity forcing words with morphisms, it also demonstrates the need for other methods, since prime patterns can clearly not be obtained in this way. In [1], Culik II and Karhumäki show that this may be done using periodicity forcing sets. Indeed, patterns not in DPCP are essentially periodicity forcing sets with a cardinality of 1. However, it is generally easier to construct periodicity forcing sets with higher cardinalities, as more patterns result in a more restricted class of pairs of morphisms which agree on every pattern. This is precisely the advantage gained when using morphisms to generate periodicity forcing words.

It follows from their basic properties that the agreement of two morphisms on a ratio-imprimitive pattern can be reduced to the agreement of those morphisms on a set of two (or more) shorter patterns. In particular, if $\alpha = \beta_1 \cdot \beta_2 \cdot ... \cdot \beta_n$, where $\mathrm{P}(\beta_1) = \mathrm{P}(\beta_2) = \cdots = \mathrm{P}(\beta_n)$, then $\alpha \notin \mathrm{DPCP}$ if and only if $\{\beta_1, \beta_2, ..., \beta_n\}$ is a periodicity forcing set.

Hence, given a periodicity forcing set of patterns with the same basic Parikh vector, it is possible to construct periodicity forcing words by concatenating all the patterns in the set. It is the focus of the present section to investigate periodicity forcing sets which have this additional property and use them to obtain periodicity forcing words which may be prime.

We will give constructions (Theorem 17 and Theorem 21) which allow new periodicity forcing sets to be formed from existing ones. In particular, since strong sufficient conditions are known for a set of patterns over two variables to be periodicity forcing (see, e.g., Holub [7]), we will provide constructions which increase the alphabet size. We take the following concise example from [1] which will be used later on.

Lemma 15 (Culik II, Karhumäki [1]). *The set* $\{1{\cdot}2, 1{\cdot}1{\cdot}2{\cdot}2\}$ *is periodicity forcing.*

Note that by the reasoning above, we can infer that the patterns $1{\cdot}2{\cdot}1{\cdot}1{\cdot}2{\cdot}2$ and $1 \cdot 1 \cdot 2 \cdot 2 \cdot 1 \cdot 2$ are periodicity forcing.

Our constructions are based on the substitution of individual variables with patterns. For example, consider the set $\{\alpha \cdot \beta, \alpha \cdot \alpha \cdot \beta \cdot \beta\}$ for some patterns α, β. We can immediately conclude for any σ, τ which agree on both patterns of the set, that they are either identical over α and β (i.e. $\sigma(\alpha) = \tau(\alpha)$ and $\sigma(\beta) = \tau(\beta)$), or they are periodic over α and β (i.e., $\sigma(\alpha)$, $\tau(\alpha)$, $\sigma(\beta)$, $\tau(\beta) \in \{w\}^*$ for some word w). Since any morphic image of α (resp. β) is also a morphic image of 1 (resp. 2), the existence of σ and τ not adhering to one of these cases would be in direct contradiction to Lemma 15.

Note however that the set $\{\alpha \cdot \beta,\, \alpha \cdot \alpha \cdot \beta \cdot \beta\}$ is not necessarily periodicity forcing. For example, it may be the case that a morphism σ is periodic over α and β, but not their individual variables. In general, additional patterns will be required in order to achieve to turn the original set into a periodicity forcing one. These additional patterns will be formed by splitting a pattern $\gamma = \gamma_1 \cdot \gamma_2$ and inserting some other pattern δ, obtaining $\gamma_1 \cdot \delta \cdot \gamma_2$. Thus in the case described above, we have that $\sigma(\gamma_1 \cdot \delta \cdot \gamma_2)$ is of the form $w^{k_1} \cdot u \cdot w^q \cdot v \cdot w^{k_2}$ where $uv = w$. Thus, we will use the following technical lemma when considering the agreement of two such morphisms on $\gamma_1 \cdot \delta \cdot \gamma_2$.

Lemma 16. *Let w be a primitive word, and let u, u', v, v' be words such that u,$v \neq \varepsilon$ and $u \cdot v = u' \cdot v' = w$. Then for any k_1, k_2, k_3, k_4, q_1, $q_2 \in \mathbb{N}_0$ with $q_1 \neq 0$ or $q_2 \neq 0$, the equation*

$$w^{k_1} \cdot u \cdot w^{q_1} \cdot v \cdot w^{k_2} = w^{k_3} \cdot u' \cdot w^{q_2} \cdot v' \cdot w^{k_4} \tag{1}$$

only has solutions in the case that $k_1 = k_3$, $k_2 = k_4$, $q_1 = q_2$, $u = u'$ and $v = v'$.

Proof. Firstly, suppose that $q_1 = 0$. Then equality (1) can be reduced to $w^{(k_1+k_2+1)-(k_3+k_4)} = u' \cdot w^{q_2} \cdot v'$. In this case is well known and easily proved that u, v and w commute and thus that the statement of the Lemma holds. Hence we assume $q_1 \neq 0$. Symmetrically, we can also assume that $q_2 \neq 0$, and by the same reasoning, that u', $v' \neq \varepsilon$.

W. l. o. g. let $|u| \geq |u'|$. Then since $u \cdot v = u' \cdot v'$, there exist words c, d, e such that $u = cd$, $v = e$, $u' = c$ and $v' = de$. Note that this implies $w = cde$. Hence equality (1) can be expressed as

$$(cde)^{k_1} \cdot cd \cdot (cde)^{q_1} \cdot e \cdot (cde)^{k_2} = (cde)^{k_3} \cdot c \cdot (cde)^{q_2} \cdot de \cdot (cde)^{k_4}.$$

If $d = \varepsilon$, then unless $k_1 = k_3$, $k_2 = k_4$ and $q_1 = q_2$, the equation is non-trivial and in two unknowns – namely c and e, so by Lemma 1, c and e commute and w is imprimitive. Hence $c \neq \varepsilon$, $d \neq \varepsilon$ and $e \neq \varepsilon$.

The equation can be divided into three distinct cases, according to the sign of $k_1 - k_3$. In each case, it is shown that whenever the equation is non-trivial, w must be imprimitive, which is a contradiction.

If $k_1 > k_3$, by comparing the prefix of each side of length $(k_3 + 1)|cde| + |c|$,

$$(cde)^{k_3} \cdot (cde) \cdot c = (cde)^{k_3} \cdot c \cdot (cde).$$

Therefore

$$(cde) \cdot c = c \cdot (cde).$$

so c and cde commute. Since c, d, $e \neq \varepsilon$, $|w| > |c|$. Thus, w is imprimitive, which is a contradiction.

If $k_1 < k_3$, by comparing the prefix of length $(k_3 + q_2)|cde| + |c| + |d|$, there exist n, $m \in \mathbb{N}_0$ such that

$$(cde)^{k_1} \cdot cd \cdot (cde)^{q_1 - n} \cdot (ecd)^m = (cde)^{k_3} \cdot c \cdot (cde)^{q_2} \cdot d,$$

and therefore

$$cd \cdot (cde)^{q_1-n} \cdot (ecd)^m = (cde)^{k_3-k_1} \cdot c \cdot (cde)^{q_2} \cdot d,$$

where $m \leq k_2$, $0 \leq n < q_1$ and $m = 0$ if $n \neq 0$. Notice that $k_3 - k_1 \geq 1$. If $m = 0$ then by comparing the suffix of length $|d| + |e|$ of either side, d and e commute. By Corollary 2, equality (1) becomes a non-trivial equation in two unknowns, so c, d, e commute. If $m \geq 2$, then by comparing the suffix of length $2|d| + |c| + |e|$, $decd = cded$. Thus $dec = cde$ and de, c commute. If $m = 1$, then

$$cd \cdot (cde)^{q_1} \cdot ec = (cde)^{k_3-k_1} \cdot c \cdot (cde)^{q_2};$$

so $q_1 \geq q_2$ (since $k_3 - k_1 \geq 1$). It follows that

$$(cde)^{q_2} \cdot ec = ec \cdot (cde)^{q_2},$$

and hence cde, ec commute. Since $|cde| > |ec|$, it follows that $cde^r = ec^s$ for some $r > s > 0$. Thus if $k_1 < k_3$, w is not primitive, which is a contradiction.

If $k_1 = k_3$, then

$$d \cdot (cde)^{q_1} \cdot e \cdot (cde)^{k_2} = (cde)^{q_2} \cdot de \cdot (cde)^{k_4},$$

so w is imprimitive, providing a contradiction as required. □

We now present our first of two constructions for producing new periodicity forcing sets from existing ones. Note that both constructions can easily be used to produce sets of patterns which share the same basic Parikh vector. Thus we can use the following theorems to generate periodicity forcing words which are not necessarily obtainable using the methods from [3]. The construction relies on 'splitting' one variable y into two (so each occurrence of y becomes, e. g., y_1y_2) in each pattern. New patterns are then introduced to force the periodicity of y_1 and y_2. Although the theorem appears very technical, it is relatively simple to apply, as Example 18 shall demonstrate.

Theorem 17. *Let $\Delta := \{x_1, x_2, ..., x_n\}$ be a set of variables, and let $y \notin \Delta$ be a variable. Let $\Pi := \{\alpha_1, \alpha_2, ..., \alpha_m\}$ be a periodicity forcing set such that $\bigcup_{i=1}^{m} \mathrm{var}(\alpha_m) = \Delta$. Let $\varphi : \Delta^* \to (\Delta \cup \{y\})^*$ be the morphism given by $\varphi(x_n) := x_n \cdot y$ and $\varphi(x_i) := x_i$ for $1 \leq i < n$. Let $t \in \mathbb{N}$, and for $1 \leq i \leq t$, let $\beta_i := x_n \cdot \gamma_i \cdot y$ for some pattern γ_i. Let $\beta_{t+1} := x_1 \cdot x_1 \cdot x_2 \cdot x_2 \cdots x_n \cdot x_n \cdot y \cdot y$. If*

- *(i) $\gamma_1, \gamma_2, ..., \gamma_t$ are patterns such that $\mathrm{var}(\gamma_1) = \mathrm{var}(\gamma_2) = \cdots = \mathrm{var}(\gamma_t) = \Delta \backslash \{x_n\}$, and*
- *(ii) the set $\{\gamma_1, \gamma_2, \cdots, \gamma_t\}$ is commutativity forcing,*

then the set $\{\varphi(\alpha_1), \varphi(\alpha_2), ..., \varphi(\alpha_m), \beta_1, \beta_2, ..., \beta_{t+1}\}$ is periodicity forcing.

Proof. Let $\sigma, \tau : (\Delta \cup \{y\})^* \to \{\mathtt{a}, \mathtt{b}\}^*$ be two distinct morphisms which agree on the set $\varphi(\alpha_2), ..., \varphi(\alpha_m)\}$. Then since $\{\alpha_1, \alpha_2, ..., \alpha_m\}$ is a periodicity forcing set, we have one of the following cases:

(1) $\sigma(\varphi(x_i)) = \tau(\varphi(x_i))$ for $1 \leq i \leq n$, or

(2) there exists a primitive word $w \in \{\mathtt{a},\mathtt{b}\}^*$ such that $\sigma(\varphi(x_i))$, $\tau(\varphi(x_i)) \in \{w\}^*$ for $1 \leq i \leq n$.

Consider first Case 1. It follows from the definition of φ that $\sigma(x_n \cdot y) = \tau(x_n \cdot y)$, and $\sigma(x_i) = \tau(x_i)$ for $1 \leq i < n$. Furthermore, $\sigma(\beta_{t+1}) = \tau(\beta_{t+1})$. Then σ and τ must agree on $x_n \cdot x_n \cdot y \cdot y$. However, by Lemma 15 $\{x_n \cdot y,\, x_n \cdot x_n \cdot y \cdot y\}$ is a periodicity forcing set, so there exists a $w \in \{\mathtt{a},\mathtt{b}\}^*$ and k_1, k_2, k_3, $k_4 \in \mathbb{N}_0$ such that $\sigma(x_n) = w^{k_1}$, $\tau(x_n) = w^{k_3}$, $\sigma(y) = w^{k_2}$, $\tau(y) = w^{k_4}$. Due to the fact that $\sigma(\beta_i) = \tau(\beta_i)$ for $1 \leq i \leq t$,

$$w^{k_1} \cdot \sigma(\gamma_1) \cdot w^{k_2} = w^{k_3} \cdot \tau(\gamma_1) \cdot w^{k_4}$$
$$w^{k_1} \cdot \sigma(\gamma_2) \cdot w^{k_2} = w^{k_3} \cdot \tau(\gamma_2) \cdot w^{k_4}$$
$$\vdots$$
$$w^{k_1} \cdot \sigma(\gamma_t) \cdot w^{k_2} = w^{k_3} \cdot \tau(\gamma_t) \cdot w^{k_4}.$$

Note that since $\sigma(\varphi(x_i)) = \tau(\varphi(x_i))$ for $1 \leq i \leq n$ and $\gamma_i \in \{x_1, x_2, ..., x_{n-1}\}^*$ for $1 \leq i \leq t$, it follows that $\sigma(\gamma_i) = \tau(\gamma_i)$ for $1 \leq i \leq t$. Unless $k_1 = k_3$, and $k_2 = k_4$ (in which case σ and τ are not distinct), each equation is non-trivial and in two variables (w and $\sigma(\gamma_i)$), so by Lemma 1, $\sigma(\gamma_i) \in \{w\}^*$ for $1 \leq i \leq t$. Thus the words $\sigma(\gamma_i)$ commute. However, by Condition (ii) of the proposition, this implies that there exists a primitive word $w' \in \{\mathtt{a},\mathtt{b}\}^*$ such that $\sigma(x_i) \in \{w'\}^*$ for $1 \leq i < n$. It follows from Lemma 1 that $w' = w$, so σ is periodic. The same holds for τ.

Consider Case 2. Then there exist k_1, k_2, ... k_n, l_1, l_2, ... $l_n \in \mathbb{N}_0$ and a word $w \in \{\mathtt{a},\mathtt{b}\}^+$ such that $\sigma(x_i) = w^{k_i}$ and $\tau(x_i) = w^{l_i}$ for $1 \leq i < n$, and $\sigma(x_n \cdot y) = w^{k_n}$, $\tau(x_n \cdot y) = w^{l_n}$. If $k_n = l_n = 0$, then σ and τ are periodic. Otherwise there exist u, v, u', v' and q_1, q_2, q_3, $q_4 \in \mathbb{N}_0$ such that $\sigma(x_n) = w^{q_1} \cdot u$, $\sigma(y) = v \cdot w^{q_2}$, $\tau(x_n) = w^{q_3} \cdot u'$ and $\tau(y) = v' \cdot w^{q_4}$, with $uv = u'v' = w$. Note that if $u = \varepsilon$ or $v = \varepsilon$, σ is periodic. Since $\sigma(\beta_1) = \tau(\beta_1)$,

$$\sigma(x_n) \cdot \sigma(\gamma_1) \cdot \sigma(y) = \tau(x_n) \cdot \tau(\gamma_1) \cdot \tau(y),$$

so

$$w^{q_1} \cdot u \cdot w^{s_1} \cdot v \cdot w^{q_2} = w^{q_3} \cdot u' \cdot w^{s_2} \cdot v' \cdot w^{q_4}$$

for some s_1, $s_2 \in \mathbb{N}_0$. If $s_1 = s_2 = 0$, then $\sigma(x_n \cdot y) = \tau(x_n \cdot y)$, and if $\sigma(\beta_{t+1}) = \tau(\beta_{t+1})$, $\sigma(x_n \cdot x_n \cdot y \cdot y) = \tau(x_n \cdot x_n \cdot y \cdot y)$, so by Lemma 15, σ and τ must be periodic over $\{x_n, y\}$ (i.e., $\sigma(x)$, $\sigma(y)$, $\tau(x)$, $\tau(y) \in \{z\}^*$ for some word $z \in \{\mathtt{a}, \mathtt{b}\}^*$). Since they are empty over all other variables, they are periodic over Δ. Otherwise, by Lemma 16, u, v, u', $v' \in \{w\}^*$, which again implies that σ and τ are periodic.

Thus, there exist no two non-periodic morphisms which agree on the set $\{\varphi(\alpha_1), ..., \varphi(\alpha_m), \beta_1, \beta_2, ..., \beta_{t+1}\}$. Hence it is periodicity forcing as required. $\square$

Example 18. *Let* $\Delta := \{1, 2\}$*,* $y := 3$*, and* $\Pi := \{1 \cdot 2, 1 \cdot 1 \cdot 2 \cdot 2\}$*. Then* $\varphi : \{1, 2\}^* \to \{1, 2, 3\}^*$ *is the morphism given by* $\varphi(1) = 1$ *and* $\varphi(2) = 2 \cdot 3$*. Let* $\gamma_1 := 1$*,* $\beta_1 := 2 \cdot 1 \cdot 3$ *and* $\beta_2 := 1 \cdot 1 \cdot 2 \cdot 2 \cdot 3 \cdot 3$*. Then by Theorem 17, we have that the set* $\Pi' := \{1 \cdot 2 \cdot 3,\ 1 \cdot 1 \cdot 2 \cdot 3 \cdot 2 \cdot 3,\ 2 \cdot 1 \cdot 3,\ 1 \cdot 1 \cdot 2 \cdot 2 \cdot 3 \cdot 3\}$ *is periodicity forcing. Since all the patterns have the same basic Parikh vector, we can conclude that, for example, the pattern* $1 \cdot 2 \cdot 3 \cdot 1 \cdot 1 \cdot 2 \cdot 3 \cdot 2 \cdot 3 \cdot 2 \cdot 1 \cdot 3 \cdot 1 \cdot 1 \cdot 2 \cdot 2 \cdot 3 \cdot 3$ *is periodicity forcing.*

We can then use Π' *to again apply the theorem. This time we have* $y := 4$ *and* $\Delta := \{1, 2, 3\}$*. By Proposition 12, possible choices for* γ_1 *and* γ_2 *are* $1 \cdot 2$ *and* $2 \cdot 1$*. Thus, by applying the theorem, we can conclude that the set* $\Pi'' := \{1 \cdot 2 \cdot 3 \cdot 4,\ 1 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 2 \cdot 3 \cdot 4,\ 2 \cdot 1 \cdot 3 \cdot 4,\ 1 \cdot 1 \cdot 2 \cdot 2 \cdot 3 \cdot 4 \cdot 3 \cdot 4,\ 3 \cdot 1 \cdot 2 \cdot 4,\ 3 \cdot 2 \cdot 1 \cdot 4,\ 1 \cdot 1 \cdot 2 \cdot 2 \cdot 3 \cdot 3 \cdot 4 \cdot 4\}$ *is periodicity forcing, and again we can concatenate the patterns to form a periodicity forcing word.*

Our second method relies on inserting a new variable repeatedly into occurrences of a single pattern not in DPCP. It is relatively simple to establish a set of patterns with the same basic Parikh vectors in this way. The following definition is given to provide a notation for inserting a new variable x at a specified place in a pattern α.

Definition 19. *Let* α *be a pattern and let* $x \in \mathrm{var}(\alpha)$ *be a variable. Let* $\mathrm{pre}_x(\alpha)$ *be the prefix of* α *up to, and including the first occurrence of* x*. Let* $\mathrm{suf}_x(\alpha)$ *be the suffix of* α *starting after (not including) the first occurrence of* x*.*

Note that $\mathrm{pre}_x(\alpha) \cdot \mathrm{suf}_x(\alpha) = \alpha$, so the pattern $\mathrm{pre}_x(\alpha) \cdot y \cdot \mathrm{suf}_x(\alpha)$ is the pattern obtained by inserting the variable y into the pattern α directly after the first occurrence of x.

The following lemma produces periodicity forcing sets which will form the basis of our construction. Although the patterns in these sets do not have the same basic Parikh vectors, it is expanded in Theorem 21 to provide a construction with patterns that do, and thus can be used to produce periodicity forcing words.

Lemma 20. *Let* $\alpha \notin \mathrm{DPCP}$ *be a pattern, and let* $x \notin \mathrm{var}(\alpha)$ *be a variable. Let* β_z *denote the pattern* $\mathrm{pre}_z(\alpha) \cdot x \cdot \mathrm{suf}_z(\alpha)$ *for any* $z \in \mathrm{var}(\alpha)$*. Then the set* $\{\alpha, x\} \cup \{\beta_y \mid y \in \mathrm{var}(\alpha)\}$ *is periodicity forcing.*

Proof. Let $\sigma, \tau : (\mathrm{var}(\alpha) \cup \{x\})^* \to \{\mathtt{a},\mathtt{b}\}^*$ be distinct morphisms, let y be arbitrary, and consider the equation $\sigma(\beta_y) = \tau(\beta_y)$. If $\sigma(\alpha) = \tau(\alpha)$, by properties of DPCP, there must exist a word $w \in \{\mathtt{a},\mathtt{b}\}^*$ such that $\sigma(z) \in \{w\}^*$ for every $z \in \mathrm{var}(\alpha)$. Therefore, there exist $p, q, r, s \in \mathbb{N}_0$ such that $\sigma(\beta_y) = \tau(\beta_y)$ if and only if

$$w^p \cdot \sigma(x) \cdot w^q = w^r \cdot \tau(x) \cdot w^s.$$

Note that y can be chosen such that $p \neq r$ whenever σ, τ are distinct, by taking the leftmost variable such that $\sigma(y) \neq \tau(y)$. Furthermore, because $\sigma(x) = \tau(x) = u$ for some word $u \in \{\mathtt{a},\mathtt{b}\}^*$, by Lemma 1, u and w must commute, so σ and τ must be periodic to agree on every pattern in $\{\alpha, x\} \cup \{\beta_y \mid y \in \mathrm{var}(\alpha)\}$ as required. $\square$

Note that in the following theorem, the set $\{x,\, \alpha\}$ from Lemma 20 is replaced with a set containing patterns with the same basic Parikh vector as the others. More specifically, the new set is formed by substituting the variables 1 and 2 in the example from Lemma 15 for x and α. Using the set from Lemma 15 is not the only possibility, however. The construction is easily generalised to use any periodicity forcing set of patterns with the appropriate basic Parikh vector.

Theorem 21. *Let $\alpha \notin \mathrm{DPCP}$ and let $x \notin \mathrm{var}(\alpha)$. Then the set $\Pi := \{x \cdot \alpha,\, x \cdot x \cdot \alpha \cdot \alpha\} \cup \{\mathrm{pre}_y(\alpha) \cdot x \cdot \mathrm{suf}_y(\alpha) \mid y \in \mathrm{var}(\alpha)\}$ is periodicity forcing.*

Proof. Let $\sigma, \tau : (\mathrm{var}(\alpha) \cup \{x\})^* \to \{\mathtt{a,b}\}^*$ be distinct morphisms which agree on every pattern in Π. Then they agree on $x \cdot \alpha$ and $x \cdot x \cdot \alpha \cdot \alpha$, so by Lemma 15, either

(1) $\sigma(x) = \tau(x)$ and $\sigma(\alpha) = \tau(\alpha)$, or

(2) $\sigma(x),\, \tau(x),\, \sigma(\alpha),\, \tau(\alpha) \in \{w\}^*$ for some primitive word w.

If Case 1 holds, then σ and τ agree on $\Pi \cup \{\alpha,\, x\}$. Since this is a superset of the set $\{x,\, \alpha\} \cup \{\mathrm{pre}_y(\alpha) \cdot x \cdot \mathrm{suf}_y(\alpha) \mid y \in \mathrm{var}(\alpha)\}$, which by Lemma 20 is periodicity forcing, σ and τ are periodic. Consider Case 2 and assume to the contrary that σ is non-periodic. Then there exists a $y \in \mathrm{var}(\alpha)$ such that $\sigma(y) \notin \{w\}^*$. Let y be the first such variable to occur in α, and consider the equation

$$\sigma(\mathrm{pre}_y(\alpha) \cdot x \cdot \mathrm{suf}_y(\alpha)) = \tau(\mathrm{pre}_y(\alpha) \cdot x \cdot \mathrm{suf}_y(\alpha)).$$

Clearly, $\sigma(\mathrm{pre}_y(\alpha)) = w^{k_1} \cdot u$ for some word $u \notin \{w\}^*$ and $k_1 \in \mathbb{N}_0$. It follows that $\sigma(\mathrm{suf}_y(\alpha)) = v \cdot w^{k_2}$ for some word $v \notin \{w\}^*$ and $k_2 \in \mathbb{N}_0$ with $u \cdot v = w$. Furthermore, there exist words u', v' such that $\tau(\mathrm{pre}_y(\alpha)) = w^{k_3} \cdot u'$ and $\tau(\mathrm{suf}_y(\alpha)) = v' \cdot w^{k_4}$ for some $k_3,\, k_4 \in \mathbb{N}_0$ with $u' \cdot v' = w$. Let $\sigma(x) = w^{q_1}$ and $\tau(x) = w^{q_2}$ for some numbers $q_1,\, q_2$. Then

$$w^{k_1} \cdot u \cdot w^{q_1} \cdot v \cdot w^{k_2} = w^{k_3} \cdot u' \cdot w^{q_2} \cdot v' \cdot w^{k_4}.$$

Note that if both q_1 and q_2 are 0, then $\sigma(x) = \tau(x) = \varepsilon$, meaning $\sigma(\alpha) = \tau(\alpha)$; so σ must be periodic, which is a contradiction. Thus it is assumed that $q_1 > 0$ or $q_2 > 0$, and by Lemma 16, $k_1 = k_3$, $k_2 = k_4$, $q_1 = q_2$, $u = u'$, and $v = v'$. Therefore σ and τ are not distinct, which is a contradiction. A symmetrical argument can be made for when τ is non-periodic. Thus σ and τ must be periodic to agree on every element in Π, so Π is a periodicity forcing set. $\square$

By applying Theorem 21 to $\alpha := 1 \cdot 2 \cdot 1 \cdot 1 \cdot 2$ and $x := 3$, and concatenating the patterns in the resulting set, we obtain, for example, the periodicity forcing word

$$3 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 3 \cdot 3 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 3 \cdot 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 2 \cdot 3 \cdot 1 \cdot 1 \cdot 2,$$

which appears to be a good candidate for being prime. We can also conclude the following from Theorem 21:

Proposition 22. *Let $\beta = \alpha^k$ for some pattern α and number $k \geq |\mathrm{var}(\alpha)| + 3$. Then β is not a prime element of* $\mathrm{DPCP}^\neg$.

Proof. Let α be a pattern and let $x \notin \mathrm{var}(\alpha)$. By Theorem 21, the set $\Pi := \{x \cdot \alpha,\ x \cdot x \cdot \alpha \cdot \alpha\} \cup \{\mathrm{pre}_y(\alpha) \cdot x \cdot \mathrm{suf}_y(\alpha) \mid y \in \mathrm{var}(\alpha)\}$ is periodicity forcing. Furthermore, every pattern in Π has the same basic Parikh vector. Thus any concatenation of patterns in Π such that every pattern is included at least once is not in DPCP. Let $\beta = \gamma_1 \cdot \gamma_2 \cdot \ldots \cdot \gamma_k$ be such a pattern with $\gamma_i \in \Pi$ for $1 \leq i \leq k$. Notice that $k \geq |\Pi|$, and $|\Pi| = 3 + |\mathrm{var}(\alpha)|$. Let $\varphi : (\mathrm{var}(\alpha) \cup \{x\})^* \to \mathrm{var}(\alpha)^*$ be the morphism given by $\varphi(x) := \varepsilon$ and $\varphi(y) := y$ for every $y \in \mathrm{var}(\alpha)$. Clearly $\varphi(\gamma_i) = \alpha$ for $1 \leq i \leq k$, so $\varphi(\beta) = \alpha^k$, and $\beta = \alpha^k$ is not prime as required. □

This is an interesting result since the properties associated with the Dual PCP are, due to the nature of morphisms, generally consistent for repetitions of the same word. It can also be interpreted that, as a result of the proposition, the majority of periodicity forcing words are not prime.

6. Conclusion

In a recent paper [3], we began an analysis of the Dual PCP in the context of larger alphabets, complementing the existing research which has so far been focused on the better-understood binary case. In the present paper, we have continued this analysis by focusing specifically on those words which do not satisfy the Dual PCP.

In Section 3 we have introduced a prime subset of $\mathrm{DPCP}^\neg$, allowing the set to be described as chains of morphic images. We have shown that this subset is non-empty, and thus that $\mathrm{DPCP}^\neg$ can be exactly generated by the set of prime periodicity forcing words. In Section 4, we have given a construction for periodicity forcing words containing any given factor/prefix/suffix. This not only produces a rich class of new examples, but demonstrates a previously unknown level of generality within the seemingly very restrictive set. In Section 5, motivated by the study of the prime periodicity forcing words introduced earlier, we have examined alternative methods for generating periodicity forcing words. The results give examples of periodicity forcing words which contrast those known so far, and provide further insights into the prime words considered earlier in the paper. As a by-product of results from this paper and existing literature, it has been possible to give tight bounds on the length of the shortest periodicity forcing word over a given alphabet.

Acknowledgements

The authors wish to thank the anonymous referees of the conference version [4] of this paper for their helpful remarks and suggestions which have provided a useful additional reference and a construction which has produced a stronger form of Proposition 7. The helpful suggestions of the referees of the full version of this paper are also gratefully acknowledged.

[1] K. Culik II and J. Karhumäki. On the equality sets for homomorphisms on free monoids with two generators. *Theoretical Informatics and Applications (RAIRO)*, 14:349–369, 1980.

[2] E. Czeizler, Š. Holub, J. Karhumäki, and M. Laine. Intricacies of simple word equations: An example. *International Journal of Foundations of Computer Science*, 18:1167–1175, 2007.

[3] J.D. Day, D. Reidenbach, and J.C. Schneider. On the Dual Post Correspondence problem. In *Proc. 17th International Conference on Developments in Language Theory, DLT 2013*, 2013.

[4] J.D. Day, D. Reidenbach, and J.C. Schneider. Periodicity forcing words. In *Proc. 9th International Conference on Words, WORDS 2013*, volume 8079 of *Lecture Notes in Computer Science*, pages 107–118, 2013.

[5] J. Hadravova and Š. Holub. Large simple binary equality words. *International Journal of Foundations of Computer Science*, 23:1385–1403, 2012.

[6] V. Halava and Š. Holub. Reduction tree of the binary generalized post correspondence problem. *International Journal of Foundations of Computer Science*, 22:473–490, 2011.

[7] Š. Holub. Binary equality sets are generated by two words. *Journal of Algebra*, 259:1 – 42, 2003.

[8] Š. Holub and J. Kortelainen. Linear size test sets for certain commutative languages. *Theoretical Informatics and Applications (RAIRO)*, 35:453–475, 2001.

[9] J. Karhumäki and E. Petre. On some special equations on words. In *Technical Report 584, Turku Centre for Computer Science, TUCS*, 2003.

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

[11] G.S. Makanin. The problem of solvability of equations in a free semi-group. *Soviet Mathematics Doklady*, 18:330–334, 1977.

[12] E.L. Post. A variant of a recursively unsolvable problem. *Bulletin of the American Mathematical Society*, 52:264–268, 1946.

[13] D. Reidenbach and J.C. Schneider. Morphically primitive words. *Theoretical Computer Science*, 410:2148–2161, 2009.