University of Novi Sad
Faculty of Science
Department of Mathematics and Informatics

Bojan Bašić

Palindromes in finite and infinite words

-Ph.D. thesis-

Supervisor: Prof. Vojislav Petrović, Ph.D.

Novi Sad, 2012

Contents

4

Izvod

U tezi razmatramo aktuelne probleme u vezi s palindromskim podrečima i palindromskim faktorima konačnih i beskonačnih reči. Glavni pravac istraživanja jesu kriterijumi za određivanje koja od dve date reči je „palindromičnija" od druge, tj. određivanje stepena „palindromičnosti" date reči. Akcenat stavljamo na dva aktuelna pristupa: tzv. *MP-razmeru* i tzv. *palindromski defekt*, i odgovaramo na više otvorenih pitanja u vezi s njima.

Naime, u vezi sa MP-razmerom u literaturi je postavljeno više pitanja, intuitivno uverljivih, koja bi, u slučaju pozitivnog razrešenja, znatno pojednostavila izračunavanje MP-razmere. Ovim pitanjima dodajemo još jedno srodno, a zatim pokazujemo da, prilično neočekivano, sva ova pitanja imaju negativan odgovor.

U vezi s palindromskim defektom, glavni rezultat rada je konstrukcija beskonačne klase beskonačnih reči koje imaju više osobina za kojima je iskazana potreba u skorašnjim radovima iz ove oblasti. Među najzanimljivije spada činjenica da su sve one aperiodične reči konačnog pozitivnog defekta, i da im je skup faktora zatvoren za preokretanje – u nekim skorašnjim radovima konstrukcija makar jedne reči s ovim osobinama pokazala se kao prilično teška. Pomoću ovih reči, koje nazivamo *visokopotencijalne reči*, ispitujemo validnost više otvorenih hipoteza, i za više njih ustanovljavamo da nisu validne.

Abstract

In the thesis we are concerned with actual problems on palindromic subwords and palindromic factors of finite and infinite words. The main course of the research are the ways of determining which of two given words is "more palindromic" than the other one, that is, defining a measure for the degree of "palindromicity" of a word. Particularly, we pay attention to two actual approaches: the so-called MP-ratio and the so-called *palindromic defect*, and answer several open questions about them.

Namely, concerning the MP-ratio, a few plausible-looking question have been asked in the literature, which would have, if answered positively, made computations of MP-ratios significantly simpler. We add one more related question to these ones, and then show that, rather unexpectedly, all these questions have negative answer.

Concerning the palindromic defect, the main result of this work is a construction of an infinite class of infinite words that have several properties that were sought after in some recent works in this area. Among the most interesting facts is that that all these words are aperiodic words of a finite positive defect, having the set of factors closed under reversal—in some recent works, the construction of even a single such word turned out to be quite hard. Using these words, which we call *highly potential words*, we check the validity of several open conjectures, and for several of them we find out that they are false.

Preface

Combinatorics on words is a branch of mathematics having a very wide scope of applications. A similar thing can be said about palindromic words, that is, words that can be read indistinctly from left to right or from right to left. Namely, they play a major role in the study of so-called Sturmian sequences [22, 16], which in turn have applications in number theory, routing optimization, computer graphics and image processing, pattern recognition and more [1, Chapter 9]. Palindromes further have applications in seemingly unrelated fields such as quantum physics [18, 2, 14], molecular biology [21, 20] [23, Chapter 4] and recently even music theory [24, 13, 11].

Thus, a more detailed knowledge about the behavior of palindromes is of a growing importance. One of the questions arising is determining which of two given words (not necessarily palindromes) is "more palindromic" than the other one, that is, defining a measure for the degree of "palindromicity" of a word. Clearly, different approaches can be imagined, depending on the interpretation of "more palindromic". In this thesis we present two actual research directions, and answer several open questions related to them.

Holub and Saari [19] chose the following approach. Restricting themselves to binary words, they observed that each word w contains a palindromic subword of length at least $\left\lceil \frac{|w|}{2} \right\rceil$: a subword consisting of the dominant letter. On the basis of this observation, they called words w that do not contain palindromic subwords of length greater than $\left\lceil \frac{|w|}{2} \right\rceil$ *minimal-palindromic*: in-

tuitively, these are the least palindromic words. The degree of "palindromicity" of a word w is then measured by the so-called *MP-ratio*, defined with the following conception in mind: the word is more palindromic the harder it is to extend it to a minimal-palindromic word (a strict definition is given in the next chapter). In the end of their paper, Holub and Saari posed a few plausible-looking questions, which would have, if answered positively, made computations of MP-ratios significantly simpler.

Another approach to measuring the degree of "palindromicity" of words (not necessarily binary ones) depends on the notion of the so-called *palindromic defect*. Namely, by the result of Droubay, Justin and Pirillo [15], the number of palindromic factors of a given word does not exceed the length of the word increased by one. On the basis of this inequality, the notion of (palindromic) defect is introduced as the difference between those two values (therefore, the defect is always non-negative). By this approach, the words that are considered the most palindromic are those that have the defect equal to 0. The definition of defect can be naturally extended to infinite words. On the infinite words having the set of factors closed under reversal, Brlek and Reutenauer [9] defined a function that is in a way related to the palindromic defect, based on an inequality that connects the so-called *complexity* and the so-called *palindromic complexity* of a word, proved in [3]. They then conjectured an equality stating the connection between the defect, the palindromic complexity, and the (factor) complexity of an infinite word w, given that the set of factors of w is closed under reversal.

Brlek and Reutenauer proved their conjecture in the case of periodic words, and observed that, on the basis of some earlier results, the conjecture also holds for words of defect 0. They further tested the conjecture for some words of infinite defect, namely: the Thue-Morse word, the paperfolding sequences and the generalized Rudin-Shapiro sequences, and the results were positive. The next logical step would be to find more evidence for the conjecture by testing it for aperiodic words of a finite positive defect, at least for a few examples, but it turned out that the authors were unable to find even a single such word.

In the same paper, Brlek and Reutenauer recalled the conjecture of Blondin-Massé et al. [7], stating that there does not exists an aperiodic word of a finite positive defect that is a fixed point of some primitive morphism. Brlek and Reutenauer showed that, under a stronger conjecture that there does not exists an aperiodic word of a finite positive defect that is a fixed point of any non-identical morphism, their conjecture holds for fixed points

of non-identical morphisms. The assumed conjecture remained open.

Balková, Pelantová and Starosta [4] proved the Brlek-Reutenauer conjecture for uniformly recurrent words. Apart from this proof, they gave a few other related theorems, one of which turns out to be incorrect.

In this thesis we intend to shed more light on these topics. The work is organized as follows.

In Chapter 1 we present the notation and necessary definitions, as well as all the known results that our research builds onto. All of the previous results are given with a reference, and for the most of them the proof is also included. The chapter is divided into three sections, where the first one has a general character, the second one introduces the MP-ratio and related notions, and the third one introduces the palindromic defect and related notions.

Chapters 2 and 3 present the fully original work.

In Chapter 2 we pay attention to the MP-ratio. The main results of this chapter are answers to the three questions posed by Holub and Saari, and also one further question of a similar kind. Rather surprisingly, all the answers turn out to be negative. The results from this chapter have been published in [6].

In Chapter 3 we pay attention to the palindromic defect. The chapter is divided into five sections. In Section 3.1 we construct a counterexample to the abovementioned theorem of Balková, Pelantová and Starosta. The results from this section have been published in [5].

After that, we introduce a class of words related to all the problems discussed above. The construction of this class of words is defined in Section 3.2. Since they seem to have a high potential to serve as examples and counterexamples in various problems on words, we dub them *highly potential words*. We observe that each highly potential word has its set of factors closed under reversal, that it is aperiodic, recurrent, but not uniformly recurrent. We prove that each highly potential word has a finite positive defect.

In Section 3.3 we prove that the Brlek-Reutenauer conjecture indeed holds for highly potential words. Note that, since highly potential words are not uniformly recurrent, this result does not follow from the result of Balková, Pelantová and Starosta.

In Section 3.4 we show that highly potential words are counterexamples to the statement of a theorem by Balková, Pelantová and Starosta. Since there was only one counterexample presented in Section 3.1, having a rather pathological flavor, the value of this section is the fact that there are more of them, that constitute a less artificial family.

In Section 3.5 we construct a highly potential word that is a fixed point of a non-identical morphism. Since highly potential words are aperiodic words of a finite positive defect, this construction disproves Brlek and Reutenauer's strengthening of the conjecture by Blondin-Massé et al.

I would like to express my gratitude toward some persons without whom this thesis would not come into existence. These are, in the first place, my parents, who were raising and supporting me from 1986 onwards.

Further, there are many people responsible for my love toward mathematics and for everything I know about it, namely, many teachers from the elementary school, the high school and the faculty. It would be impossible to name them all, but there are five that have a direct connection to this thesis. I am very grateful to them for accepting to supervise this thesis, respectively to be in the Defend board, and for many useful comments that improved the overall work. They are: Prof. Vojislav Petrović, the supervisor, and Prof. Petar Marković, Prof. Igor Dolinka, Prof. Ivica Bošnjak and Prof. Rade Doroslovački, members of the Defend board.

Novi Sad, May 2012   Bojan Bašić

1

Introduction

1.1 *Preliminaries*

Let us start with the notation and necessary definitions. All these notions are mainly standard, and can be found in, for example, [12].

Given a set Σ called the *alphabet*, we call its elements *letters*, and finite, respectively infinite, sequences of letters are called *words*, respectively *infinite words*. Let Σ^* denote the set of all finite words over Σ, and let Σ^∞ denote the set of all finite or infinite words over Σ. For words $w = a_1 a_2 \ldots a_n$ and $u = b_1 b_2 \ldots b_m$ (where $a_1, \ldots, a_n, b_1, \ldots, b_m \in \Sigma$), with wu we denote the concatenation of words w and u, that is, $wu = a_1 a_2 \ldots a_n b_1 b_2 \ldots b_m$. Given a word w and $k \in \mathbb{N}_0$ (where $\mathbb{N}_0$ denotes the set of non-negative integers), we write w^k for $\underbrace{ww \ldots w}_{k \text{ times}}$ (called the *k-th power* of a word w), and we write w^∞ for the infinite word $www \ldots$.

If a is a letter, we write a^* for the set $\{a^k : k \geqslant 0\}$, and if b is an additional letter, we write $a^* b^*$ for the set $\{a^k b^l : k, l \geqslant 0\}$.

The length of a word w is denoted with $|w|$. Notation $|w|_a$, where a is a letter, stands for the total number of occurrences of a in w. The unique word of length equal to 0, called the *empty word*, is denoted with ε.

Definition 1.1. We define the following basic relations between words:

- A word $v = a_1 a_2 \ldots a_n$ is a *subword* of a word w if there exist words $u_1, u_2, \ldots, u_{n+1}$ such that $w = u_1 a_1 u_2 a_2 \ldots u_n a_n u_{n+1}$.

- A word $v \in \Sigma^*$ is a *suffix* of a word $w \in \Sigma^*$ if there exists a word $u \in \Sigma^*$ such that $w = uv$.

- A word $v \in \Sigma^*$ is a *prefix* of a word $w \in \Sigma^\infty$ if there exists a word $u \in \Sigma^\infty$ such that $w = vu$.

- A word $v \in \Sigma^*$ is a *factor* of a word $w \in \Sigma^\infty$ if there exists words $u_1 \in \Sigma^*$, $u_2 \in \Sigma^\infty$ such that $w = u_1 v u_2$. The set of all factors of a word w is denoted with $\mathrm{Fact}(w)$.

Remark. Some authors by *subword* mean the notion that we here call *factor*, and the herein presented notion of *subword* is then called *scattered* (or *sparse*) *subword*. We proceed with the former convention.

We say that a factor v of a word $w \in \Sigma^\infty$ is *unioccurrent* in w if it occurs in w exactly once, that is, if there exists a unique pair of words $u_1 \in \Sigma^*$, $u_2 \in \Sigma^\infty$ such that $w = u_1 v u_2$.

Definition 1.2. A map $\tilde{\ }\colon \Sigma^* \to \Sigma^*$, called *reversal*, is defined as follows: for $w = a_1 a_2 \ldots a_n$, where $a_1, a_2, \ldots, a_n \in \Sigma$, it holds that $\widetilde{w} = a_n a_{n-1} \ldots a_1$.

We say that the set of factors of w is *closed under reversal* if for any $v \in \mathrm{Fact}(w)$ it holds that $\widetilde{v} \in \mathrm{Fact}(w)$.

Definition 1.3. A word $w \in \Sigma^*$ is a *palindrome* if $w = \widetilde{w}$. The set of all palindromic factors of a word $w \in \Sigma^\infty$ is denoted with $\mathrm{Pal}(w)$.

Definition 1.4. An infinite word w is:

- *periodic* if it is of the form v^∞ for some $v \in \Sigma^*$;

- *aperiodic* if it is not periodic;

- *recurrent* if each of its factors has infinitely many occurrences in w;

- *uniformly recurrent* if it is recurrent and, for each of its factors, the gaps between consecutive occurrences of it in w are bounded (by *gap*, we mean the difference between two positions at which two consecutive occurrences of the considered factor begin).

The following theorems, the proof of which can be found in e.g. [17, Proposition 2.11], [1, Theorem 10.9.4] and [1, Example 10.9.1], respectively, will be useful.

Theorem 1.5. *Given an infinite word w, if* $\mathrm{Fact}(w)$ *is closed under reversal, then w is recurrent.*

Proof. Let $u \in \mathrm{Fact}(w)$. It is enough to prove that, for any copy of u in w, there is another copy of u in w positioned to the right of the previously observed copy. Let there be i letters in w before the observed copy of u. Let v be a prefix of w of length $2i + |u| + 1$. Then there is a copy of $\widetilde{u}$ in $\widetilde{v}$ with $i + 1$ letters before it. Since $\mathrm{Fact}(w)$ is closed under reversal, we have $\widetilde{v} \in \mathrm{Fact}(w)$, and thus there is a copy of $\widetilde{u}$ in w with at least $i + 1$ letters before it; in other words, there is a copy of $\widetilde{u}$ in w positioned to the right of the observed copy of u. We now analogously have that there is a copy of $\widetilde{\widetilde{u}} = u$ in w positioned to the right of the considered copy of $\widetilde{u}$, which finishes the proof. ■

Theorem 1.6. *An infinite word w is uniformly recurrent iff, for each $u \in \mathrm{Fact}(w)$, there exists $n \in \mathbb{N}$ such that $u \in \mathrm{Fact}(v)$ for each $v \in \mathrm{Fact}(w)$ such that $|v| = n$.*

Proof. Assume first that w is uniformly recurrent, and $u \in \mathrm{Fact}(w)$. By the definition of uniformly recurrent word, the gaps between consecutive occurrences of u in w are bounded. Let b be this bound, and take $n = b + |u| - 1$. We claim that, for each $v \in \mathrm{Fact}(w)$ such that $|v| = n$, it must hold that $u \in \mathrm{Fact}(v)$. Indeed: for each occurrence of such a v in w, there exists an occurrence of u in w that begins at one of the first b letters of v. This occurrence of u ends at the position $\leqslant b + |u| - 1$ from the beginning of v, and thus we get $u \in \mathrm{Fact}(v)$.

Let us now prove the other direction. Let $u \in \mathrm{Fact}(w)$, and let $n \in \mathbb{N}$ be such that $u \in \mathrm{Fact}(v)$ for each $v \in \mathrm{Fact}(w)$ such that $|v| = n$. For each occurrence of u in w, the word formed by the n consecutive letters beginning at the second letter of the observed occurrence contains another occurrence of u. Therefore, the gaps between consecutive occurrences of u are bounded by $n - |u| + 1$. ■

Remark. The property from the previous theorem is sometimes used as a definition of uniformly recurrent word.

Theorem 1.7. *If an infinite word is periodic, then it is uniformly recurrent.*

Proof. Each infinite periodic word is of the form w^∞ for some finite word w. Let $u \in \mathrm{Fact}(w^\infty)$. Let $n = |w| + |u| - 1$. We are going to prove that, for

each $v \in \text{Fact}(w^\infty)$ such that $|v| = n$, it holds that $u \in \text{Fact}(v)$. By Theorem 1.6, this is enough to complete the proof.

Let $v \in \text{Fact}(w^\infty)$, $|v| = n$. For any occurrence of v in w^∞, there exists an occurrence of u in w^∞ that begins at the position $\leqslant |w|$ from the beginning of v. Therefore, this occurrence u ends at the position $\leqslant |w| + |u| - 1$ from the beginning of v, and thus we get that $u \in \text{Fact}(v)$. ∎

Definition 1.8. A function $\varphi : \Sigma^* \to \Sigma^*$ is called a *morphism* if, for all $w, v \in \Sigma^*$, it holds that $\varphi(wv) = \varphi(w)\varphi(v)$.

Clearly, a morphism is uniquely determined by images of the letters, and thus it is possible to extend any given morphism to the infinite words in the natural way. We say that a word $w \in \Sigma^\infty$ is a *fixed point* of a morphism φ if $\varphi(w) = w$.

1.2 MP-ratio

In this section we are concerned only with binary words, and therefore we fix the alphabet $\Sigma = \{0, 1\}$.

Clearly, each word $w \in \{0, 1\}^*$ contains a palindromic subword of length at least $\left\lceil \frac{|w|}{2} \right\rceil$: a subword consisting of the dominant letter. This motivates the following definitions:

Definition 1.9. We say that $w \in \{0, 1\}^*$ is *minimal-palindromic* if it does not contain palindromic subwords of length greater than $\left\lceil \frac{|w|}{2} \right\rceil$.

Definition 1.10. For a word $w \in \{0, 1\}^*$, a pair (r, s), where $r, s \in \{0, 1\}^*$, such that rws is minimal-palindromic, is called an *MP-extension* of w. If the length $|r| + |s|$ is minimal possible, then we call the pair (r, s) a *shortest MP-extension* or *SMP-extension*, and the ratio $\frac{|rws|}{|w|}$ is called the *MP-ratio* of w.

We measure the degree of "palindromicity" of w by the MP-ratio. The following theorem [19, Theorem 4] gives an upper bound on the MP-ratio, at the same time settling the question of existence of an (S)MP-extension for a given word w.

Theorem 1.11. *For any word $w \in \{0, 1\}^*$, the MP-ratio of w is less than or equal to* 4.

Proof. For a given $w \in \{0,1\}^*$, let

$$v = 0^{|w|+|w|_1} w 1^{|w|+|w|_0}.$$

We claim that the word v is minimal-palindromic. It holds that $|v| = |w| + |w|_1 + |w| + |w| + |w|_0 = 4|w|$. Let p be a palindromic subword of v. We may assume, w.l.o.g., that p begins and ends with the letter 1. Therefore, p is a subword of $w 1^{|w|+|w|_0}$. We shall distinguish two cases:

Case 1: $p \in 1^*$. In this case, it holds:

$$|p| \leqslant |w|_1 + (|w| + |w|_0) = 2|w| = \frac{|v|}{2}.$$

Case 2: $p = u 0 1^k$. In this case, $u0$ is a subword of w, and thus, since p is a palindrome, we have $k \leqslant |u|_1 \leqslant |w|_1$. Therefore:

$$|p| = |u0| + k \leqslant |w| + |w|_1 \leqslant 2|w| = \frac{|v|}{2}.$$

This shows that the word v is indeed minimal-palindromic, that is, the pair $(0^{|w|+|w|_1}, 1^{|w|+|w|_0})$ is an MP-extension of w. It follows that the MP-ratio of w is at most $\frac{|v|}{|w|} = \frac{4|w|}{|w|} = 4$, which was to be proved. ∎

It turns out that the constant 4 in the previous theorem is the best possible, in the asymptotic sense. Namely, if $R(n)$ denotes the maximal MP-ratio over all the binary words of a given length n, we show the following theorem [19, Theorem 5].

Theorem 1.12. *It holds:*
$$\lim_{n \to \infty} R(n) = 4.$$

We first show the following easy lemma.

Lemma 1.13. *For each minimal-palindromic word w, one of the values $|w|_0$, $|w|_1$ equals $\left\lfloor \frac{|w|}{2} \right\rfloor$, while the other one equals $\left\lceil \frac{|w|}{2} \right\rceil$.*

Proof. Clearly, if, say, $|w|_0 < \left\lfloor \frac{|w|}{2} \right\rfloor$, then $|w|_1 > \left\lceil \frac{|w|}{2} \right\rceil$, which contradicts the definition of a minimal-palindromic word. The other cases are similar. ∎

The main tool in the proof of Theorem 1.12 is the notion of *economic words*.

17

Definition 1.14. A word $w \in \{0,1\}^*$ is called:

- *k-economic*, for a given $k \in \mathbb{N}_0$, if w is a palindrome and the word $w1^k$ has a palindromic subword of length $\geqslant |w|_1 + k + 2$;

- *economic* if w is k-economic for all $0 \leqslant k \leqslant |w|_1$.

The previous definition is somewhat technical, but its importance will be revealed soon.

For a given k-ecomonic word w, each palindromic subword of $w1^k$ of length $\geqslant |w|_1 + k + 2$ is of the form $1^m q 1^m$, where $m \leqslant k$ and q is a palindrome such that $1^m q$ is a subword of w. The pair (q, m) is called a *k-witness* of w (there can be more k-witnesses of an observed word).

Lemma 1.15. *For each MP-extension (r, s) of an economic word w, it holds that $|rs|_1 > |w|_1$.*

Proof. Suppose the opposite: $|rs|_1 \leqslant |w|_1$. Since w is a palindrome, we may assume, w.l.o.g., $|r|_1 \leqslant |s|_1$. We now have $|s|_1 - |r|_1 \leqslant |s|_1 + |r|_1 = |rs|_1 \leqslant |w|_1$, and thus, by Definition 1.14, the word w is $(|s|_1 - |r|_1)$-economic. Denote its $(|s|_1 - |r|_1)$-witness by (q, m). We first have $m \leqslant |s|_1 - |r|_1$, and thus $m + |r|_1 \leqslant |s|_1$. Therefore, since the word $1^m q$ is a subword of w (by the definition of a witness), it holds that $1^{m+|r|_1} q 1^{m+|r|_1}$ is a subword of rws, and it is a palindrome (since q is a palindrome). However, we shall now prove that its length is greater than $\lceil \frac{|rws|}{2} \rceil$, which is a contradiction with the fact that rws is minimal-palindromic.

By Definition 1.14, it holds that $2m + |q| = |1^m q 1^m| \geqslant |w|_1 + (|s|_1 - |r|_1) + 2$. Further, since (r, s) is an MP-extension of w, Lemma 1.13 gives that $|rws|_1 + 1 \geqslant \lceil \frac{|rws|}{2} \rceil$. Therefore,

$$|1^{m+|r|_1} q 1^{m+|r|_1}| = (2m + |q|) + 2|r|_1 \geqslant |w|_1 + (|s|_1 - |r|_1) + 2 + 2|r|_1$$

$$= |w|_1 + |s|_1 + |r|_1 + 2 = |rws|_1 + 2 \geqslant \left\lceil \frac{|rws|}{2} \right\rceil + 1,$$

which finishes the proof. ∎

The previous lemma, together with Lemma 1.13, implies the following inequality for each economic word w:

$$|rws| = |rws|_0 + |rws|_1 \geqslant 2|rws|_1 - 1 = 2|w|_1 + 2|rs|_1 - 1 > 4|w|_1. \quad (1.1)$$

This inequality sheds some light on the role of economic words. Namely, the MP-ratio of economic words that have "a lot of" letters 1 is close to 4. What follows is a construction of a sequence of economic words that have "many" letters 1, which ultimately leads to a proof of Theorem 1.12.

Lemma 1.16. *Let w_0 be an economic word, and define, for $i \in \mathbb{N}_0$,*

$$w_{i+1} = w_i 1^{t_i} w_i, \tag{1.2}$$

where the sequence $t_0, t_1, t_2 \ldots$ of non-negative integers satisfies $t_i < |w_i|_0$ for every $i \in \mathbb{N}_0$. Then, for each $i \in \mathbb{N}_0$, the word w_i is economic.

Proof. Since w_0 is economic and thus a palindrome, it is immediately seen that all the words w_i are palindromes. Therefore, we are left to check whether, for each $i \in \mathbb{N}_0$, the word $w_i 1^k$ has a palindromic subword of length $\geqslant |w_i|_1 + k + 2$, for all $0 \leqslant k \leqslant |w_i|_1$.

We proceed by induction on i. Assume that w_i is economic, and let us prove that w_{i+1} is economic, too. Let $0 \leqslant k \leqslant |w_{i+1}|_1$. We shall distinguish a few cases:

Case 1: $0 \leqslant k \leqslant |w_i|_1$. Since w_i is economic, it is k-economic (because of the assumed bounds on k). Let (q, m) be a k-witness of w_i. We claim that

$$p = 1^m q 1^{t_i + m} q 1^m$$

is a palindromic subword of $w_{i+1} 1^k$, of length $\geqslant |w_{i+1}|_1 + k + 2$.

By the fact that (q, m) is a k-witness of w_i, we see that $1^m q$ is a subword of w_i. Therefore, $p = 1^m q 1^{t_i + m} q 1^m$ is a subword of $w_{i+1} 1^m = w_i 1^{t_i} w_i 1^m$, and thus in turn a subword of $w_{i+1} 1^k$ (because $m \leqslant k$, which follows from the fact that (q, m) is a k-witness of w_i). Finally, since w_i is k-economic and (q, m) is its k-witness, we have:

$$2m + |q| = |1^m q 1^m| \geqslant |w_i|_1 + k + 2,$$

that is,

$$|q| \geqslant |w_i|_1 + k + 2 - 2m,$$

which gives

$$|p| = 2|q| + 3m + t_i \geqslant 2(|w_i|_1 + k + 2 - 2m) + 3m + t_i$$
$$= 2|w_i|_1 + 2k + 4 - m + t_i \geqslant 2|w_i|_1 + 2k + 4 - k + t_i$$
$$= 2|w_i|_1 + k + 4 + t_i > |w_{i+1}|_1 + k + 2.$$

19

Case 2: $|w_i|_1 < k \leqslant |w_i|_1 + t_i$. We claim that

$$p = 1^k w_i 1^k$$

is a palindromic subword of $w_{i+1} 1^k$, of length $\geqslant |w_{i+1}|_1 + k + 2$.

Since $k \leqslant |w_i|_1 + t_i$, we see that $1^k w_i$ is a subword of $w_{i+1} = w_i 1^{t_i} w_i$, and thus p is indeed a subword of $w_{i+1} 1^k$. Further, by the bounds $t_i < |w_i|_0$ and $|w_i|_1 < k$, that is, $|w_i|_0 \geqslant t_i + 1$ and $k \geqslant |w_i|_1 + 1$, it follows that

$$|p| = 2k + |w_i|_1 + |w_i|_0 \geqslant |w_i|_1 + 1 + k + |w_i|_1 + t_i + 1 = |w_{i+1}|_1 + k + 2,$$

which was to be proved.

Case 3: $|w_i|_1 + t_i < k \leqslant |w_{i+1}|_1$. We have $0 < k - |w_i|_1 - t_i \leqslant |w_{i+1}|_1 - |w_i|_1 - t_i = |w_i|_1$. Therefore, the word w_i is $(k - |w_i|_1 - t_i)$-economic. Let (q, m) be a $(k - |w_i|_1 - t_i)$-witness of w_i. We claim that

$$p = 1^{m+|w_i|_1+t_i} q 1^{m+|w_i|_1+t_i}$$

is a palindromic subword of $w_{i+1} 1^k$, of length $\geqslant |w_{i+1}|_1 + k + 2$.

It is clear that $1^{|w_i|_1+t_i}$ is a subword of $w_i 1^{t_i}$. Further, by the choice of (q, m), we see that $1^m q$ is a subword of w_i, and that $m \leqslant k - |w_i|_1 - t_i$, that is, $m + |w_i|_1 + |t_i| \leqslant k$. This shows that p is a subword of $w_{i+1} 1^k$. Finally, by the choice of (q, m), we have:

$$|p| = 2(|w_i|_1 + t_i) + |1^m q 1^m| \geqslant 2(|w_i|_1 + t_i) + |w_i|_1 + (k - |w_i|_1 - t_i) + 2$$

$$= 2|w_i|_1 + t_i + k + 2 = |w_{i+1}| + k + 2.$$

The proof is completed. ∎

Lemma 1.17. *Let $w(t_0, t_1, \ldots, t_{j-1})$ be the word w_j as defined in the statement of Lemma 1.16, where $t_0, t_1, \ldots, t_{j-1}$ are given and satisfy $2^i \leqslant t_i < 2^{i+2}$ for each $0 \leqslant i \leqslant j - 1$, and the initial word is $w_0 = 0000$. Then the word $w(t_0, t_1, \ldots, t_{j-1})$ is economic.*

Proof. It is enough to show that the conditions of Lemma 1.16 are satisfied.

In order to prove that w_0 is economic, we need to prove that it is k-economic for all $0 \leqslant k \leqslant |w_0|_1 = 0$, that is, that w_0 is 0-economic, that is, that $w_0 1^0 = w_0$ has a palindromic subword of length $\leqslant |w_0|_1 + 0 + 2 = 2$. This is, of course, clear.

We now show that, for each $0 \leqslant i \leqslant j$ it holds that

$$|w(t_0, t_1, \ldots, t_{i-1})|_0 = |w_i|_0 = 2^{i+2}. \tag{1.3}$$

Since $t_i < 2^{i+2}$, this is enough to complete the proof. We proceed by induction on i. For $i = 0$ we have

$$|w_0|_0 = |0000|_0 = 4.$$

Therefore, the base holds. Now, assume $|w(t_0, t_1, \ldots, t_{i-1})|_0 = 2^{i+2}$. Using (1.2), we have

$$|w(t_0, t_1, \ldots, t_i)|_0 = |w(t_0, t_1, \ldots, t_{i-1}) \, 1^{t_i} \, w(t_0, t_1, \ldots, t_{i-1})|_0$$
$$= 2|w(t_0, t_1, \ldots, t_{i-1})|_0 = 2 \cdot 2^{i+2} = 2^{i+3}.$$

The proof is completed. ∎

Lemma 1.18. *For every k large enough, there exists a word v_k that satisfies $|v_k| = k$ and that is of the form $w(t_0, t_1, \ldots, t_{j-1})$ for some $t_0, t_1, \ldots, t_{j-1}$ (where the conditions of Lemma 1.17 hold).*

Proof. By induction on j, we first prove:

$$|w(t_0, t_1, \ldots, t_{j-1})| = 2^{j+2} + 2^{j-1}t_0 + 2^{j-2}t_1 + \cdots + 2t_{j-2} + t_{j-1}. \tag{1.4}$$

For $j = 0$, we have $|w_0| = 4 = 2^2$. Assume that (1.4) holds for a given j. We now have:

$$|w(t_0, t_1, \ldots, t_j)| = |w(t_0, t_1, \ldots, t_{j-1}) \, 1^{t_j} \, w(t_0, t_1, \ldots, t_{j-1})|$$
$$= 2|w(t_0, t_1, \ldots, t_{j-1})| + t_j$$
$$= 2(2^{j+2} + 2^{j-1}t_0 + 2^{j-2}t_1 + \cdots + 2t_{j-2} + t_{j-1}) + t_j$$
$$= 2^{j+3} + 2^j t_0 + 2^{j-1}t_1 + \cdots + 4t_{j-2} + 2t_{j-1} + t_j,$$

which was to be proved. In particular, we have:

$$|w(1, 2, 4, \ldots, 2^{j-1})| = 2^{j+2} + 2^{j-1} \cdot 1 + 2^{j-2} \cdot 2 + \cdots + 1 \cdot 2^{j-1}$$
$$= 2^{j+2} + j \cdot 2^{j-1} = 2^{j-1}(j + 8); \tag{1.5}$$

$$|w(3, 7, 15, \dots, 2^{j+1} - 1)| = 2^{j+2} + 2^{j-1} \cdot (2^2 - 1) + 2^{j-2} \cdot (2^3 - 1)$$
$$+ \cdots + 2 \cdot (2^j - 1) + 2^{j+1} - 1$$
$$= 2^{j+2} + j \cdot 2^{j+1} - (2^{j-1} + \cdots + 2 + 1) \qquad (1.6)$$
$$= 2^{j+2} + j \cdot 2^{j+1} - (2^j - 1)$$
$$= 2^j(2j + 3) + 1.$$

For large enough j, we have

$$2^j(j + 9) < 2^j(2j + 3) + 1,$$

that is,

$$|w(1, 2, 4, \dots, 2^{j-1}, 2^j)| < |w(3, 7, 15, \dots, 2^{j+1} - 1)|. \qquad (1.7)$$

Let a large enough k be given. Choose j such that

$$|w(1, 2, 4, \dots, 2^{j-1})| \leqslant k < |w(1, 2, 4, \dots, 2^{j-1}, 2^j)|.$$

By (1.7), we now have

$$|w(1, 2, 4, \dots, 2^{j-1})| \leqslant k < |w(3, 7, 15, \dots, 2^{j+1} - 1)|. \qquad (1.8)$$

Therefore, in order to finish the proof, it is enough to show that, for each large enough j and each k such that (1.8) holds, there exists a word v_k that satisfies $|v_k| = k$ and that is of the form $w(t_0, t_1, \dots, t_{j-1})$ for some $t_0, t_1, \dots, t_{j-1}$. We prove that, if such a word exists for some k from the observed interval, then such a word also exists for $k + 1$ (unless the end of the interval is reached). Therefore, starting from the word $w(1, 2, 4, \dots, 2^{j-1})$, such a procedure will cover the whole interval.

Let

$$k = |w(t_0, t_1, \dots, t_{j-1})| = 2^{j+2} + 2^{j-1} t_0 + 2^{j-2} t_1 + \cdots + 2 t_{j-2} + t_{j-1}.$$

Let i' be the largest index such that $t_{i'} < 2^{i'+2} - 1$ (such an i' exists since $k < |w(3, 7, 15, \dots, 2^{j+1} - 1)|$). Thus, for each $i' + 1 \leqslant i \leqslant j - 1$ we have

$t_i = 2^{i+2} - 1$. By (1.4), it holds:

$$|w(t_0, \ldots, t_{i'-1}, t_{i'} + 1, t_{i'+1} - 1, \ldots, t_{j-1} - 1)|$$

$$= 2^{j+2} + \sum_{l=0}^{i'-1} 2^{j-l-1} t_l + 2^{j-i'-1}(t_{i'} + 1) + \sum_{l=i'+1}^{j-1} 2^{j-l-1}(t_l - 1)$$

$$= 2^{j+2} + \sum_{l=0}^{j-1} 2^{j-l-1} t_l + 2^{j-i'-1} - \sum_{l=i'+1}^{j-1} 2^{j-l-1}$$

$$= 2^{j+2} + \sum_{l=0}^{j-1} 2^{j-l-1} t_l + 2^{j-i'-1} - \sum_{l=0}^{j-i'-2} 2^l$$

$$= 2^{j+2} + \sum_{l=0}^{j-1} 2^{j-l-1} t_l + 2^{j-i'-1} - (2^{j-i'-1} - 1)$$

$$= 2^{j+2} + \sum_{l=0}^{j-1} 2^{j-l-1} t_l + 1 = k + 1.$$

This completes the proof. ∎

We now show that the words v_k obtained in the previous lemma indeed contain "many" letters 1. Namely, we have the following lemma.

Lemma 1.19. *For the words v_k obtained in the previous lemma, it holds:*

$$\lim_{k \to \infty} \frac{|v_k|_1}{|v_k|} = 1.$$

Proof. For each $t_0, t_1, \ldots, t_{j-1}$ satisfying the conditions of Lemma 1.17, it clearly holds that

$$|w(t_0, t_1, \ldots, t_{j-1})|_1 \geqslant |w(1, 2, 4, \ldots, 2^{j-1})|_1,$$

while it is easy to see that it also holds that

$$|w(t_0, t_1, \ldots, t_{j-1})|_0 = |w(1, 2, 4, \ldots, 2^{j-1})|_0.$$

23

Therefore:

$$|w(1, 2, 4, \ldots, 2^{j-1})|_0 \cdot |w(t_0, t_1, \ldots, t_{j-1})|_1$$
$$+ |w(1, 2, 4, \ldots, 2^{j-1})|_1 \cdot |w(t_0, t_1, \ldots, t_{j-1})|_1$$
$$\geqslant |w(t_0, t_1, \ldots, t_{j-1})|_0 \cdot |w(1, 2, 4, \ldots, 2^{j-1})|_1$$
$$+ |w(1, 2, 4, \ldots, 2^{j-1})|_1 \cdot |w(t_0, t_1, \ldots, t_{j-1})|_1,$$

and thus

$$|w(1, 2, 4, \ldots, 2^{j-1})| \cdot |w(t_0, t_1, \ldots, t_{j-1})|_1$$
$$\geqslant |w(t_0, t_1, \ldots, t_{j-1})| \cdot |w(1, 2, 4, \ldots, 2^{j-1})|_1,$$

and finally

$$\frac{|w(t_0, t_1, \ldots, t_{j-1})|_1}{|w(t_0, t_1, \ldots, t_{j-1})|} \geqslant \frac{|w(1, 2, 4, \ldots, 2^{j-1})|_1}{|w(1, 2, 4, \ldots, 2^{j-1})|}.$$

By (1.5), we have $|w(1, 2, 4, \ldots, 2^{j-1})| = 2^{j-1}(j+8)$. From this and (1.3), we get $|w(1, 2, 4, \ldots, 2^{j-1})|_1 = 2^{j-1}(j + 8) - 2^{j+2} = 2^{j-1}j$. Therefore,

$$1 \geqslant \frac{|w(t_0, t_1, \ldots, t_{j-1})|_1}{|w(t_0, t_1, \ldots, t_{j-1})|} \geqslant \frac{2^{j-1}j}{2^{j-1}(j+8)} = \frac{j}{j+8}.$$

By the squeeze theorem, we get

$$\lim_{k \to \infty} \frac{|v_k|_1}{|v_k|} = 1,$$

which was to be proved. ∎

We now have enough prerequisites to prove Theorem 1.12.

Proof of Theorem 1.12. It is enough to prove that, for any $\varepsilon > 0$, there exists $k_0 \in \mathbb{N}$ such that for all $k \geqslant k_0$ it holds that $\frac{|rv_k s|}{|v_k|} > 4 - \varepsilon$, where (r, s) is an SMP-extension of v_k. Let $\varepsilon > 0$ be given. The previous lemma shows that there exists $k_0 \in \mathbb{N}$ such that for all $k \geqslant k_0$ it holds that

$$\frac{|v_k|_1}{|v_k|} > 1 - \frac{\varepsilon}{4}.$$

All the words v_k are economic, and thus, by (1.1), we get

$$\frac{|rv_k s|}{|v_k|} > \frac{4|v_k|_1}{|v_k|} > 4 - \varepsilon.$$

The proof is finished. ∎

Holub and Saari asked the following questions about MP-extensions:

Question 1.20. *Consider all the binary words of a given length n. Are those among them which reach the maximal possible MP-ratio necessarily palindromes?*

Question 1.21. *Does every binary word possess an SMP-extension (r, s) with $r, s \in 0^* \cup 1^*$?*

Question 1.22. *Does every binary word possess an SMP-extension (r, s) with $r, s \in 0^*1^* \cup 1^*0^*$?*

To these three questions we append another one of a similar kind.

Question 1.23. *Does every binary word possess an SMP-extension (r, s) such that r and s do not have a letter in common?*

Let us say a few words about the intuition behind these questions.

Clearly, the minimal possible MP-ratio equals 1 and is reached precisely for minimal-palindromic words, which are thought of as the least palindromic words. Question 1.20 deals with the words on the opposite end: since they are thought of as the most palindromic words, it is quite expected, as Question 1.20 predicts, that they must be palindromes. However, in Section 2.1 we show that this is not always the case.

Questions 1.21, 1.22 and 1.23 deal with the possible forms of SMP-extensions. Question 1.21 is based on the following intuition: since we are avoiding palindromic subwords longer than necessary, it seems reasonable to assume that r and s are as simple as possible, that is, powers of a single letter; indeed, other forms of r and s would give rise to more different subwords, thus increasing the chance of a palindrome being among them. Question 1.22 is just a weaker form of Question 1.21. Finally, Question 1.23, arguably the most plausible of all, predicts that it is safe to assume that r and s do not have a letter in common, based on the fact that a common letter to r and s actually increases the length of a longest palindromic subword of a starting

25

word. Nevertheless, in Section 2.1 we disprove all these intuitions. Note that, although any counterexample to Question 1.22 also is a counterexample to Question 1.21, and furthermore, our counterexample to Question 1.23 also is another counterexample to Question 1.21—we still resolve Question 1.21 separately. The reason is that, while Questions 1.22 and 1.23 are resolved by a single counterexample each, we provide an infinite family of counterexamples to Question 1.21.

1.3 Palindromic defect

Definition 1.24. Let an infinite word w be given.

- The *factor complexity* (or only *complexity*) of w is the function $C_w : \mathbb{N}_0 \to \mathbb{N}_0$ defined by

$$C_w(n) = |\{v \in \mathrm{Fact}(w) : |v| = n\}|.$$

- The *palindromic complexity* of w is the function $P_w : \mathbb{N}_0 \to \mathbb{N}_0$ defined by

$$P_w(n) = |\{v \in \mathrm{Pal}(w) : |v| = n\}|.$$

We now recall an inequality due to Droubay, Justin and Pirillo [15, Proposition 2].

Theorem 1.25. *For any finite word w it holds:*

$$|\mathrm{Pal}(w)| \leqslant |w| + 1.$$

Proof. We claim that, for each finite word v and each letter a, the word va has at most one palindromic factor that is not a palindromic factor of v; further, such a factor exists iff there exists a palindromic suffix of va that is unioccurrent in va.

Assume that there does not exist a palindromic suffix of va that is unioccurrent in va. In other words, each palindromic factor of va that contains the final letter a occurs at least once more in va, that is, occurs in v. Therefore, in this case there are no palindromic factors of va that are not palindromic factors of v.

Assume now that p is a palindromic suffix of va that is unioccurrent in va. Then p is clearly a palindromic factor of va that is not a palindromic

factor of v. We now have to show that p is the only such one. In fact, we show that p is the longest palindromic suffix of va. Suppose, on the contrary, that q is palindromic suffix of va longer than p. We then have that p is a suffix of q, and thus, since both p and q are palindromes, $\tilde{p} = p$ is a prefix of $\tilde{q} = q$. Therefore, p is not unioccurrent in va, a contradiction. This proves the claim.

Since $\mathrm{Pal}(\varepsilon) = \{\varepsilon\}$, for any finite word $w = a_1 a_2 \ldots a_n$ we have:

$$|\mathrm{Pal}(w)| \leqslant |\mathrm{Pal}(a_1 a_2 \ldots a_{n-1})| + 1 \leqslant |\mathrm{Pal}(a_1 a_2 \ldots a_{n-2})| + 2$$

$$\leqslant \cdots \leqslant |\mathrm{Pal}(a_1)| + n - 1 \leqslant |\mathrm{Pal}(\varepsilon)| + n = n + 1,$$

which was to be proved. ∎

This inequality motivated Brlek et al. [8] to introduce the following definition:

Definition 1.26. *Palindromic defect* (or only *defect*) of a finite word w is the difference

$$D(w) = |w| + 1 - |\mathrm{Pal}(w)|.$$

The following theorem and its corollary [4, Corollary 2.3] gives an important property of the defect.

Theorem 1.27. *For any $w \in \Sigma^*$ and $a \in \Sigma$, it holds:*

$$D(wa) = \begin{cases} D(w), & \textit{if the longest palindromic suffix of } wa \\ & \textit{is unioccurrent in } wa; \\ D(w) + 1, & \textit{otherwise;} \end{cases}$$

$$D(aw) = \begin{cases} D(w), & \textit{if the longest palindromic prefix of } aw \\ & \textit{is unioccurrent in } aw; \\ D(w) + 1, & \textit{otherwise.} \end{cases}$$

Therefore, $D(w)$ equals the number of prefixes v of w such that the longest palindromic suffix of v occurs in v more than once, and also equals the number of suffixes v of w such that the longest palindromic prefix of v occurs in v more than once.

Proof. The first equality follows by the proof of Theorem 1.25. The second equality is analogous. ∎

Corollary 1.28. *Let $w \in \Sigma^*$ and $v \in \mathrm{Fact}(w)$. Then $D(v) \leqslant D(w)$.*

Proof. Let $w = u_1 v u_2$. By Theorem 1.27, we have:

$$D(v) \leqslant D(vu_2) \leqslant D(u_1 v u_2) = D(w),$$

which was to be proved. ∎

The above corollary motivates the following definition of the defect of an infinite word w:

Definition 1.29. For an infinite word w, we define its defect by:

$$D(w) = \sup_{v \in \mathrm{Fact}(w)} D(v).$$

Clearly, this equality also holds for finite words.

Another important inequality connecting the notions discussed above is proved by Baláži, Masáková and Pelantová [3, Theorem 1.2(ii)]:

Theorem 1.30. *Let w be an infinite word with $\mathrm{Fact}(w)$ being closed under reversal. For each $n \in \mathbb{N}_0$ we have*

$$P_w(n) + P_w(n+1) \leqslant C_w(n+1) - C_w(n) + 2.$$

Proof. Let w be an infinite word with $\mathrm{Fact}(w)$ being closed under reversal. Fix $n \in \mathbb{N}_0$. We define a directed graph G_n, called the *Rauzy graph* of the word w (after [25]), in the following way: the set of vertices of G_n is the set

$$V_n = \{v \in \mathrm{Fact}(w) : |v| = n\},$$

the set of edges of G_n is the set

$$E_n = \{v \in \mathrm{Fact}(w) : |v| = n+1\},$$

and an edge $e \in E_n$ begins at a vertex $x \in V_n$ and ends at a vertex $y \in V_n$ iff x is a prefix of e and y is a suffix of e.

It is easy to see that, if an infinite word is recurrent, then its Rauzy graph is strongly connected. Since $\mathrm{Fact}(w)$ is closed under reversal, Theorem 1.5 gives that w is recurrent, and thus the graph G_n is strongly connected.

As is usual in graph theory, by $\deg_+(x)$, respectively $\deg_-(x)$, we define the outdegree, respectively indegree, of a vertex x (that is: the number of

edges that begin, respectively end, at the vertex x). By the definition of V_n and E_n, we have the following equalities:

$$\deg_+(x) = |\{a \in \Sigma : xa \in E_n\}; \tag{1.9}$$

$$\deg_-(x) = |\{a \in \Sigma : ax \in E_n\}. \tag{1.10}$$

It is well known that the sum of outdegrees of all the vertices of any directed graph equals the number of edges of that graph, and that the same holds for indegrees. Therefore, we have:

$$\sum_{x \in V_n} \deg_+(x) = \sum_{x \in V_n} \deg_-(x) = |E_n| = C_w(n+1).$$

Since $|V_n| = C_w(n)$, we now have:

$$\sum_{x \in V_n} (\deg_+(x) - 1) = \sum_{x \in V_n} (\deg_-(x) - 1) = C_w(n+1) - C_w(n). \tag{1.11}$$

Notice that, in the first of the above sums, the vertices $x \in V_n$ such that $\deg_+(x) = 1$ make no contribution to the sum. An analogous observation can be made about the second sum and the vertices $x \in V_n$ such that $\deg_-(x) = 1$. For this reason, we call the vertex $x \in V_n$ *right special* if $\deg_+(x) \geqslant 2$, and *left special* if $\deg_-(x) \geqslant 2$. In other words, if we think of x as a factor of w, then, by (1.9) and (1.10), we say that x is right special, respectively left special, if there are at least two letters $a_1, a_2 \in \Sigma$ such that $xa_1, xa_2 \in \mathrm{Fact}(w)$, respectively $a_1x, a_2x \in \mathrm{Fact}(w)$. We say that x is *special* if x is right or left special (or both). We can now rewrite 1.11 as:

$$\sum_{\substack{x \in V_n \\ x \text{ is right special}}} (\deg_+(x) - 1) = \sum_{\substack{x \in V_n \\ x \text{ is left special}}} (\deg_-(x) - 1) = C_w(n+1) - C_w(n). \tag{1.12}$$

We define a unary operation ρ on the vertices and edges of G_n in the following way: for a vertex $x \in V_n$, let $\rho(x) = \tilde{x}$, and for an edge $e \in E_n$, let $\rho(e) = \tilde{e}$. The notation $\tilde{\ }$ in this definition denotes, of course, the reversal of x and e, which are here thought of as factors of w. Since $\mathrm{Fact}(w)$ is closed under reversal, the operation ρ is well-defined and maps G_n onto itself. We clearly have:

$$P_w(n) = |\{x \in V_n : \rho(x) = x\}|;$$

$$P_w(n+1) = |\{e \in E_n : \rho(e) = e\}|.$$

We now prove that there do not exist $x \in V_n$ and $e \in E_n$ such that e begins or ends at x, e is not a loop, and $\rho(x) = x$, $\rho(e) = e$. Suppose, on the contrary, that such $x = a_1 a_2 \ldots a_n \in \text{Fact}(w)$ and $e = x a_{n+1} = a_1 a_2 \ldots a_n a_{n+1}$ exist (the case $e = a_{n+1}x$ is analogous). Since $a_n \ldots a_2 a_1 = \tilde{x} = \rho(x) = x$, we have $a_i = a_{n-i+1}$ for each $1 \leqslant i \leqslant n$. Since $a_{n+1} a_n \ldots a_2 a_1 = \tilde{e} = \rho(e) = e$, we have $a_i = a_{n-i+2}$ for each $1 \leqslant i \leqslant n$. From these two observations it follows that $a_1 = a_2 = \cdots = a_{n+1}$, that is, $x = a_1^n$, $e = a_1^{n+1}$. Therefore, e begins and ends at $a_1^n = x$, that is, e is a loop, which is a contradiction.

Let us first treat the case when G_n does not have special vertices. In other words, G_n is a directed cycle, say of length k. The sums in (1.12) are therefore empty, and we thus have $C_w(n+1) - C_w(n) = 0$. It follows that we have to prove $P_w(n) + P_w(n+1) \leqslant 2$. This is clear for $k = 1$. Therefore, assume $k \geqslant 2$. The assertion is also clear if there is no vertex nor edge that is mapped to itself by ρ. Assume now that there is a vertex $x \in V_n$ such that $\rho(x) = x$. There are exactly two edges incident to x, and they cannot be mapped to themselves by ρ. Therefore, they are mapped to each other. It now follows that ρ actually acts like a "mirror symmetry" on G_n, and thus there is exactly one other vertex or edge that is mapped to itself by ρ (if k is even, this will be the vertex, while if k is odd, this will be the edge, halfway through the cycle from x). Finally, assume that there is an edge $e \in E_n$ such that $\rho(e) = e$. In this case, the two vertices incident to e cannot be mapped to themselves by ρ, and thus there are mapped to each other. The further considerations are now analogous as in the previous case, and thus the assertion is proved.

From now on, we assume that G_n has at least one special vertex (that is: right or left special).

Call a directed path $s = (x_1, x_2, \ldots, x_k, x_{k+1})$ in G_n a *simple path* if its initial vertex, x_1, and its final vertex, x_{k+1}, are special, and none of the other vertices on s are special (that is: for each $2 \leqslant i \leqslant k$, it holds that $\deg_+(x_i) = \deg_-(x_i) = 1$). In particular, a special vertex is itself considered as a trivial simple path, of length 0. (The *length* of the path is, as is usual in graph theory, the number of edges on it, that is, one less than the number of vertices on it.)

Since the graph G_n is strongly connected, for each its vertex, and for each its edge, there exists a simple path that contains it. We claim that for vertices and edges that are mapped to themselves by ρ more can be told: for each such a vertex $x \in V_n$, there exists a special path s of an even length such that x is the central vertex of s, while for each such an edge $e \in E_n$, there exists a special path s of an odd length such that e is the central edge of s. Let us prove this claim. Let $x \in V_n$, $\rho(x) = x$. If x is special, then x itself is a trivial special path of length 0, which was to be proven. Therefore, assume that x is not special. There exist a unique edge leading from x, and a unique edge leading to x. These two edges cannot be the same edge (that is, a loop), since x would then be the only vertex in G_n, a contradiction. Therefore, the observed two edges are not mapped to themselves by ρ, and thus they are mapped to each other. Hence, if one of them begins or ends at a special vertex, the same must hold for the other one, and thus x is the central vertex of a simple path of an even length. If neither of these two edges begin nor end at a special vertex, the path constructed so far can again be uniquely extended in both directions, and this procedure is iterated until a special vertex is reached (which must happen at the same time for the both ends). This proves the claim for $x \in V_n$. Let now $e \in E_n$, $\rho(e) = e$. If e is a loop, then it must be a loop at a special vertex, and thus e is a simple path of length 1. If e is not a loop, then the two vertices incident with it cannot be mapped to themselves by ρ, and thus they are mapped to each other. Therefore, they are either both special, or both not special. If they are special, the edge e is the required simple path. If they are not special, we finish the proof in a similar manner as in the case of a vertex.

We note that the path constructed in the previous procedure is unique for each $x \in V_n$ and $e \in E_n$. Further, by the construction it follows that each such path is mapped onto itself by ρ.

Therefore, in order to bound $P_w(n) + P_w(n+1)$ from above, it is enough to give an upper bound on the number of simple paths in G_n that are mapped onto themselves by ρ. To this end, we introduce the notion of the *reduced Rauzy graph*. The set of vertices, say V'_n, of the reduced Rauzy graph G'_n is the set of all special vertices $x \in V_n$, and the set of edges of G'_n is the set of all simple paths in G_n (where each edge in G'_n connects its beginning and ending vertex, when viewed as a simple path in G_n). We allow multiple edges between the two vertices, that is, G'_n is actually a multigraph. Since the graph G_n is strongly connected, the graph G'_n is also strongly connected. We see that the graph G'_n is mapped onto itself by ρ.

Since ρ is an involution (that is, ρ^2 is the identity), for each $x \in V'_n$ it holds that either $\rho(x) = x$ or there exists $y \in V'_n$ such that $x \neq y$, $\rho(x) = y$, $\rho(y) = x$. We thus can divide V'_n into classes in such a way that each vertex $x \in V'_n$ such that $\rho(x) = x$ is alone in its class, while each pair $\{x, y\} \subseteq V'_n$ such that $x \neq y$, $\rho(x) = y$ and $\rho(y) = x$ form a two-element class. Let m_1 be the number of one-element classes (that is, the number of vertices that are mapped to themselves by ρ), and m_2 be the number of two-element classes. We have:

$$|V'_n| = m_1 + 2m_2. \qquad (1.13)$$

Notice that for each edge from x to y, where x and y are not in the same class, there exists another edge between these two classes: namely, the edge from $\rho(y)$ to $\rho(x)$. Therefore, for each two classes that are connected by an edge, there are at least two edges between them. Since the graph G'_n is strongly connected and the number of classes is $m_1 + m_2$, there are at least $m_1 + m_2 - 1$ pairs of classes that are connected by an edge, and hence there are at least $2(m_1 + m_2 - 1)$ edges that connect vertices from different classes. These edges, that correspond to simple paths in the original Rauzy graph, are not mapped onto themselves by ρ.

We now have the following bound:

$$P_w(n) + P_w(n+1) \leqslant \sum_{\substack{x \in V_n \\ x \text{ is special}}} \deg_+(x) - 2(m_1 + m_2 - 1) + m_1. \qquad (1.14)$$

Indeed, the first sum is the number of all simple paths of length greater than zero in G_n, the expression $2(m_1 + m_2 - 1)$ is the lower bound on the number of such paths that are not mapped onto themselves by ρ, while the final m_1 is the number of simple paths of zero length that are mapped onto themselves. Finally, since for any $x \in V_n$ that is not right special it holds that $\deg_+(x) = 1$, it follows:

$$\sum_{\substack{x \in V_n \\ x \text{ is special}}} (\deg_+(x) - 1) = \sum_{\substack{x \in V_n \\ x \text{ is right special}}} (\deg_+(x) - 1). \qquad (1.15)$$

Therefore, by (1.14), (1.13), (1.15) and (1.12), we have:

$$P_w(n) + P_w(n+1) \leqslant \sum_{\substack{x \in V_n \\ x \text{ is special}}} \deg_+(x) - (m_1 + 2m_2) + 2$$

$$= \sum_{\substack{x \in V_n \\ x \text{ is special}}} (\deg_+(x) - 1) + 2$$

$$= \sum_{\substack{x \in V_n \\ x \text{ is right special}}} (\deg_+(x) - 1) + 2$$

$$= C_w(n+1) - C_w(n) + 2,$$

which was to be proved. ∎

Remark. Actually, in [3], the above inequality is formulated only for uniformly recurrent words. However, as can be seen in the above proof, this assumption is never needed, but only the assumption that w is recurrent (which follows by the fact that $\mathrm{Fact}(w)$ is closed under reversal and Theorem 1.5).

Finally, we state the Brlek-Reutenauer conjecture [9], recounted in the Preface. It predicts the following equality dealing with the defect $D(w)$ and the function $T_w : \mathbb{N}_0 \to \mathbb{N}_0$, inspired by Theorem 1.30, defined by

$$T_w(n) = C_w(n+1) - C_w(n) + 2 - P_w(n) - P_w(n+1).$$

Conjecture 1.31. *Let w be an infinite word with $\mathrm{Fact}(w)$ being closed under reversal. It holds:*

$$2D(w) = \sum_{n=0}^{\infty} T_w(n).$$

In the same paper, Brlek and Reutenauer proved that Conjecture 1.31 holds for periodic words [9, Theorem 2].

Theorem 1.32. *Let w be a periodic infinite word with $\mathrm{Fact}(w)$ being closed under reversal. It holds:*

$$2D(w) = \sum_{n=0}^{\infty} T_w(n).$$

They further noted that Conjecture 1.31 holds for words of defect 0, which follows from the following result [17, Theorem 2.14] [10, Theorem 1.1].

Theorem 1.33. *Let w be an infinite word with $\mathrm{Fact}(w)$ being closed under reversal. The following statements are equivalent:*

(a) $D(w) = 0$;

(b) $P_w(n) + P_w(n+1) = C_w(n+1) - C_w(n) + 2$ for all $n \in \mathbb{N}$.

Namely, in [17, Theorem 2.14] it is shown that the statement (a) is equivalent to a property that utilizes the notion of the so-called *complete returns*, while in [10, Theorem 1.1] it is shown that this property is equivalent to the statement (b). Therefore, a direct corollary of Theorem 1.33 is:.

Corollary 1.34. *Let w be an infinite word with $\mathrm{Fact}(w)$ being closed under reversal, such that $D(w) = 0$. Then:*

$$\sum_{n=0}^{\infty} T_w(n) = 0.$$

In other words, Conjecture 1.31 holds for words of defect 0.

Brlek and Reutenauer further tested Conjecture 1.31 for some well-known infinite words and classes of infinite words, namely: the Thue-Morse word, the paperfolding sequences and the generalized Rudin-Shapiro sequences, all of which have infinite defect. It turned out that the conjecture holds for all of them.

Finally, when they tried to make the next logical step, that is, to test the conjecture for aperiodic infinite words of a finite positive defect (at least for some examples), it turned out that examples of aperiodic infinite words of a finite positive defect that have the set of factors closed under reversal are quite hard to find: Brlek and Reutenauer were unable to find even a single example. This is one of the problems we treat in this thesis.

For the end of this chapter, we recall one of the most interesting claimed results toward the proof of Conjecture 1.31: [4, Corollary 5.10], which states that, for any infinite word w with $\mathrm{Fact}(w)$ being closed under reversal, if $D(w)$ is finite, then $\sum_{n=0}^{\infty} T_w(n)$ is also finite. However, this claimed result relies on technical Theorem 1.35 below [4, Theorem 5.7], which is, as we shall show in Sections 3.1 and 3.4, actually incorrect, and thus the mentioned result remains open.

Theorem 1.35 (**incorrect**). *For any infinite word u with $\mathrm{Fact}(u)$ being closed under reversal and containing infinitely many palindromes, the following statements are equivalent:*

(a) the defect of u is finite;

(b) there exists an integer H such that the longest palindromic suffix of any factor w of u, of length $|w| \geqslant H$, occurs in w exactly once.

<div style="text-align: right;">*2*</div>

MP-ratio

2.1 Answering Questions 1.20, 1.21, 1.22 and 1.23

Let us first prove a useful lemma.

Lemma 2.1. *If (r, s) is an SMP-extension of w and $|r| + |s| > 0$, then $|rws|$ is odd.*

Proof. Suppose the opposite: (r, s) is an SMP-extension of w, $|r| + |s| > 0$ and $|rws|$ is even. Assume, w.l.o.g., $|r| > 0$ (the case $|s| > 0$ is analogous). Let r' be the word obtained by erasing any letter from r. Since rws is minimal-palindromic, it does not contain palindromic subwords of length greater than $\left\lceil \frac{|rws|}{2} \right\rceil$. Since $|rws|$ is even and $|r'ws| = |rws| - 1$, we have $\left\lceil \frac{|r'ws|}{2} \right\rceil = \left\lceil \frac{|rws|}{2} \right\rceil$. Finally, since $r'ws$ clearly cannot contain palindromic subword longer than the palindromic subwords of rws, we have that (r', s) is an MP-extension of w shorter than (r, s), which is impossible. ∎

We are now ready for the main theorems.

Theorem 2.2. *The answer to Question 1.20 is negative.*

Proof. A counterexample will be given for $n = 6$. We claim that the maximal possible MP-ratio of words of length 6 equals $\frac{11}{6}$, and that one of the words achieving it is

$$v = 010110,$$

a non-palindrome.

<div style="text-align: center;">37</div>

w	rws	w	rws	w	rws
000000	00000011111	001011	001011	010110	00001011011
000001	000001111	001100	00011001111	010111	0010111
000010	000010111	001101	0001101	011000	1011000
000011	0000111	001110	000111011	011001	00011001111
000100	000100111	001111	0001111	011010	11011010000
000101	0001011	010000	111010000	011011	000110111
000110	0001101	010001	010001111	011100	101110000
000111	000111	010010	00100101111	011101	000111011
001000	001000111	010011	0100111	011110	10111100000
001001	001001111	010100	110101000	011111	000011111
001010	001010111	010101	001010111		

Table 2.1: MP-extensions of words of length 6.

In the first place, let us prove that the MP-ratio of v is indeed $\frac{11}{6}$. Let (r, s) be an MP-extension of v. Since v contains palindromic subwords of length 5, 01010 and 01110, we have $|rvs| \geqslant 9$. Let us suppose $|rvs| = 9$. In that case, rvs must not have palindromic subwords of length greater than 5. Notice that, if s contains the letter 0, then 001100 is a subword of rvs, a contradiction; if s contains the letter 1, then 101101 is a subword of rvs, and a contradiction again. Therefore, s is an empty word, and $|r| = 3$. Now, if 11 is a subword of r, then 1101011 is a subword of rvs, a contradiction. If $r = 000$, then 000000 is a subword of rvs, a contradiction. Therefore, r contains one letter 1 and two letters 0. If 01 is a subword of r, then 0101010 is a subword of rvs, a contradiction. That leaves only the possibility $r = 100$, but then 100001 is a subword of rvs, a contradiction. Altogether, it must hold that $|rvs| > 9$, and therefore, by Lemma 2.1, $|rvs| \geqslant 11$. Since $\underbrace{000}_{r}\underbrace{010110}_{v}\underbrace{11}_{s}$ is minimal-palindromic, we have that $(000, 11)$ is an SMP-extension of v, and thus the MP-ratio of v is indeed $\frac{11}{6}$.

We now have to prove that all the other words of length 6 have MP-ratio at most $\frac{11}{6}$, that is, that for each word w there exists an MP-extension (r, s) such that $|rws| \leqslant 11$. Such extensions are shown in Table 2.1. (Only the words starting with the letter 0 are considered, since the other half are analogous. Proposed extensions are in fact SMP-extensions, though there is no need to prove that, we only need $|rws| \leqslant 11$.) ∎

Theorem 2.3. *The answer to Question 1.21 is negative.*

Proof. We claim that, for every $k \geqslant 4$, the only SMP-extension of the word

$$v = 010^k 1010$$

is the pair $(\varepsilon, u) = (\varepsilon, 01^{k+2})$, thus providing an infinite family of counterexamples to Question 1.21.

In the first place, let us prove that (ε, u) is an MP-extension of v, that is, that $vu = 010^k 101001^{k+2}$ does not contain palindromic subwords of length greater than $\lceil \frac{|vu|}{2} \rceil = \lceil \frac{2k+9}{2} \rceil = k + 5$. Let p be a palindromic subword of vu. We shall distinguish a few cases:

Case 1: p begins with three or more letters 1. In this case, p is clearly a subword of 111001^{k+2}. It is now obvious that p cannot be longer than 1^{k+5}, that is, $|p| \leqslant k + 5$.

Case 2: p begins with exactly two letters 1. In this case, p is clearly a subword of 11010011, and thus p cannot be longer than 1100011, that is, $|p| \leqslant 7 < k + 5$.

Case 3: p begins with exactly one letter 1. In this case, p is clearly a subword of $10^k 101001$, which has length $k + 7$. Since the considered word is not a palindrome, and since it can be easily checked that erasing any one of its letters does not leave a palindrome (because $k \geqslant 4$), it follows that $|p| \leqslant k + 5$.

Case 4: p begins with the letter 0. In this case, p is clearly a subword of $010^k 10100$, which has length $k + 7$. Since the considered word is not a palindrome, and since it can be easily checked that erasing any one of its letters does not leave a palindrome, it follows that $|p| \leqslant k + 5$.

Therefore, we have proved that (ε, u) is an MP-extension of v. Notice that $010^{k+1}10$ is a palindromic subword of v, of length $k + 5$. It now follows that for any MP-extension (r, s) of v we have $|rvs| \geqslant 2k + 9$, and thus (ε, u) is in fact an SMP-extension of v. We are left to prove that it is unique.

Let (r, s) be an SMP-extension of v. We already know that $|rvs| = 2k+9$, that is, $|r|+|s| = k+3$. Notice that, if r contains the letter 1, then $1010^k 101$ is a palindromic subword of rvs of length $k+6$, a contradiction; if r contains the letter 0, then $0010^k 100$ is a palindromic subword of rvs of length $k + 6$, and a contradiction again. Therefore, $r = \varepsilon$ and $|s| = k + 3$. Since $|v|_1 = 3$ and $|v|_0 = k + 3$, we have either $|s|_1 = k + 1$ and $|s|_0 = 2$, or $|s|_1 = k + 2$ and $|s|_0 = 1$ (because otherwise we would have $|vs|_0 > k + 5$ or $|vs|_1 > k + 5$, which would contradict the fact that vs is minimal-palindromic). If 10 is a

subword of s, then $010^{k+2}10$ is a palindromic subword of vs of length $k+6$, a contradiction. Therefore, $s = 001^{k+1}$ or $s = 01^{k+2}$. Finally, in the former case $10^{k+4}1$ is a palindromic subword of vs of length $k+6$, which is a contradiction, and thus only the latter case remains, which was to be proved. ∎

Theorem 2.4. *The answer to Question 1.22 is negative.*

Proof. We claim that the only SMP-extension of the word

$$v = 0010000010100111$$

is the pair $(\varepsilon, u) = (\varepsilon, 1011111)$, thus providing a counterexample to Question 1.22.

In the first place, let us prove that (ε, u) is an MP-extension of v, that is, that $vu = 00100000101001111011111$ does not contain palindromic subwords of length greater than $\left\lceil \frac{|vu|}{2} \right\rceil = \left\lceil \frac{23}{2} \right\rceil = 12$. Let p be a palindromic subword of vu. We shall distinguish a few cases:

Case 1: p begins with four or more letters 1. In this case, p is clearly a subword of 1111111011111, and since the considered word is not a palindrome, it follows that $|p| \leqslant 12$.

Case 2: p begins with exactly three letters 1. In this case, p is clearly a subword of 1110011110111, and since the considered word is not a palindrome, it follows that $|p| \leqslant 12$.

Case 3: p begins with exactly two letters 1. In this case, p is clearly a subword of 1101001111011, and since the considered word is not a palindrome, it follows that $|p| \leqslant 12$.

Case 4: p begins with exactly one letter 1. In this case, p is clearly a subword of 10000010100111101, and we may write $p = 10p'01$, where p' is a palindromic subword of 0000101001111. Obviously, p' is a palindromic subword of either 000010100 or 101001111, and since these two words are not palindromes, it follows that $|p'| \leqslant 8$ and therefore $|p| = |p'| + 4 \leqslant 12$.

Case 5: p begins with the letter 0 followed by two or more letters 1. In this case, p is clearly a subword of 011010011110, and it follows that $|p| \leqslant 12$.

Case 6: p begins with the letter 0 followed by exactly one letter 1. In this case, p is clearly a subword of 01000001010010. Since the considered word is not a palindrome, and since it can be easily checked that erasing any one of its letters does not leave a palindrome, it follows that $|p| \leqslant 12$.

Case 7: p begins with two or more letters 0. In this case, p is clearly a subword of 00100000101000. Since the considered word is not a palindrome,

and since it can be easily checked that erasing any one of its letters does not leave a palindrome, it follows that $|p| \leqslant 12$.

Therefore, we have proved that (ε, u) is an MP-extension of v. Notice that 001000000100 is a palindromic subword of v, of length 12. It now follows that for any MP-extension (r, s) of v we have $|rvs| \geqslant 23$, and thus (ε, u) is in fact an SMP-extension of v. We are left to prove that it is unique.

Let (r, s) be an SMP-extension of v. We already know that $|rvs| = 23$, that is, $|r| + |s| = 7$. Notice that, if r contains the letter 1, then 10010000001001 is a palindromic subword of rvs of length 14, a contradiction; if r contains the letter 0, then 0001000001000 is a palindromic subword of rvs of length 13, and a contradiction again. Therefore, $r = \varepsilon$ and $|s| = 7$. Since $|v|_1 = 6$ and $|v|_0 = 10$, we have either $|s|_1 = 5$ and $|s|_0 = 2$, or $|s|_1 = 6$ and $|s|_0 = 1$ (because otherwise we would have $|vs|_0 > 12$ or $|vs|_1 > 12$, which would contradict the fact that vs is minimal-palindromic). If 00 is a subword of s, then 00100000000100 is a palindromic subword of vs of length 14, a contradiction. Therefore, $|s|_1 = 6$ and $|s|_0 = 1$. If 110111 is a subword of s, then 1110111110111 is a palindromic subword of vs of length 13, a contradiction. If 111110 is a subword of s, then 0111111111110 is a palindromic subword of vs of length 13, a contradiction. That leaves only the possibilities: $s = 0111111$ or $s = 1011111$ or $s = 1111011$. In the first case 1111110111111 is a palindromic subword of vs of length 13, a contradiction. In the third case 11011111111011 is a palindromic subword of vs of length 14, a contradiction. Thus only the second case remains, which was to be proved. ∎

Theorem 2.5. *The answer to Question 1.23 is negative.*

Proof. We claim that the only SMP-extension of the word

$$v = 01111101001$$

is the pair $(y, u) = (1, 1000000)$, thus providing a counterexample to Question 1.23.

In the first place, let us prove that (y, u) is an MP-extension of v, that is, that $yvu = 1011111010011000000$ does not contain palindromic subwords of length greater than $\lceil \frac{|yvu|}{2} \rceil = \lceil \frac{19}{2} \rceil = 10$. Let p be a palindromic subword of yvu. We shall distinguish a few cases:

Case 1: p begins with two or more letters 1. In this case, p is clearly a subword of 111111010011. Since the considered word is not a palindrome, and since it can be easily checked that erasing any one of its letters does not leave a palindrome, it follows that $|p| \leqslant 10$.

41

Case 2: p begins with exactly one letter 1. In this case, p is clearly a subword of 101111101001. Since the considered word is not a palindrome, and since it can be easily checked that erasing any one of its letters does not leave a palindrome, it follows that $|p| \leqslant 10$.

Case 3: p begins with the letter 0. In this case, p is clearly a subword of 011111010011000000, and we may write $p = 0p'0$, where p' is a palindromic subword of 1111101001100000. Obviously, p' is a palindromic subword of either 11111010011 or 01001100000. It is not hard to see that the longest palindromic subwords of these two words are 11111111, and 00011000 or 00000000, respectively. It follows that $|p'| \leqslant 8$, and therefore $|p| = |p'| + 2 \leqslant 10$.

Therefore, we have proved that (y, u) is an MP-extension of v. Notice that 01111110 is a palindromic subword of v, of length 8. It now follows that for any MP-extension (r, s) of v we have $|rvs| \geqslant 15$, and thus, by Lemma 1, in order to prove that (y, u) is an SMP-extension of v, we have to show that there are no MP-extensions (r, s) of v such that $|rvs| = 15$ or $|rvs| = 17$.

Suppose that there exists an MP-extension (r, s) of v such that $|rvs| = 15$. Since $|v|_1 = 7$ and $|v|_0 = 4$, we have either $|r|_1 + |s|_1 = 1$ and $|r|_0 + |s|_0 = 3$, or $|r|_1 + |s|_1 = 0$ and $|r|_0 + |s|_0 = 4$. In both cases, at least one of r, s contains the letter 0. If r contains the letter 0, then 0011111100 is a palindromic subword of rvs of length 10, a contradiction; if s contains the letter 0, then 011111110 is a palindromic subword of rvs of length 9, and a contradiction again. Therefore, there are no MP-extensions (r, s) such that $|rvs| = 15$.

Suppose that there exists an MP-extension (r, s) of v such that $|rvs| = 17$. Notice that, if r contains the letter 0, then 0011111100 is a palindromic subword of rvs of length 10, a contradiction; if r contains the letter 1, then 1011111101 is a palindromic subword of rvs of length 10, and a contradiction again. Therefore, $r = \varepsilon$ and $|s| = 6$. Since $|v|_1 = 7$ and $|v|_0 = 4$, we have either $|s|_1 = 2$ and $|s|_0 = 4$, or $|s|_1 = 1$ and $|s|_0 = 5$. Notice that, if 10 is a subword of s, then 0111111110 is a palindromic subword of vs of length 10, a contradiction. Therefore, $s = 000011$ or $s = 000001$. In the former case, 11000000011 is a palindromic subword of vs of length 11, a contradiction; in the latter case, 1000000001 is a palindromic subword of vs of length 10, and a contradiction again. Therefore, there are no MP-extensions (r, s) such that $|rvs| = 17$, and thus (y, u) is in fact an SMP-extension of v. We are left to prove that it is unique.

Let (r, s) be an SMP-extension of v. We already know that $|rvs| = 19$, that is, $|r| + |s| = 8$. Notice that, if 00 is a subword of r, then 00011111000

is a palindromic subword of rvs of length 11, a contradiction; if 11 is a subword of r, then 11011111011 is a palindromic subword of rvs of length 11, a contradiction; if 01 is a subword of r, then 01011111010 is a palindromic subword of rvs of length 11, a contradiction; if 10 is a subword of r, then 100111111001 is a palindromic subword of rvs of length 12, and a contradiction again. Therefore, $|r| \leqslant 1$. Since $|v|_1 = 7$ and $|v|_0 = 4$, we have either $|r|_1 + |s|_1 = 3$ and $|r|_0 + |s|_0 = 5$, or $|r|_1 + |s|_1 = 2$ and $|r|_0 + |s|_0 = 6$. We further note that 110 cannot be a subword of s (this shall be needed later), since otherwise 01111111110 would be a palindromic subword of vs of length 11, a contradiction.

Suppose $r = 0$. Since $|r|_0 + |s|_0 \geqslant 5$, it follows that 00 is a subword of s. Therefore, 00111111100 is a palindromic subword of rvs of length 11, a contradiction.

Suppose $r = \varepsilon$. Therefore, we have either $|s|_1 = 3$ and $|s|_0 = 5$, or $|s|_1 = 2$ and $|s|_0 = 6$. Notice that, if 111 is a subword of s, then 11110001111 is a palindromic subword of vs of length 11, a contradiction. That leaves $|s|_1 = 2$ and $|s|_0 = 6$. We know that 110 is not a subword of s, that is, s ends with the letter 1. Therefore, 0000001 is a subword of s. However, 10000000001 is then a palindromic subword of vs of length 11, a contradiction.

Therefore, $r = 1$, and either $|s|_1 = 2$ and $|s|_0 = 5$, or $|s|_1 = 1$ and $|s|_0 = 6$. Notice that, if 01 is a subword of s, then 10111111101 is a palindromic subword of rvs of length 11, a contradiction. It now follows that $s = 1100000$ or $s = 1000000$. Finally, the former case contradicts the earlier observation that 110 is not a subword of s, and thus only the latter case remains, which was to be proved. ∎

2.1. Answering Questions 1.20, 1.21, 1.22 and 1.23

Palindromic defect

3.1 Counterexample to Theorem 1.35

The claimed proof of Theorem 1.35 briefly states that the equivalence follows by the definition of defect. In fact, by the definition of defect and Corollary 1.27, it follows that the statements (a) and

(b$_0$) there exists an integer H such that the longest palindromic suffix of any *prefix* w of u, of length $|w| \geqslant H$, occurs in w exactly once

are equivalent: the direction ($\Longleftarrow$) is clear, while the direction ($\Longrightarrow$) follows from the observation that, if v is a prefix of u such that $D(v) = D(u)$, then each prefix w of u longer than v contains v as a prefix, and thus the longest palindromic suffix of w must occur in w exactly once (since otherwise it would follow $D(w) \geqslant D(v) + 1 = D(u) + 1$, a contradiction). Unfortunately, the same reasoning cannot be applied with factors in place of prefixes, and therefore the mentioned proof is erroneous (only the direction (b) $\Rightarrow$ (a) can be seen to hold, since we have (b) $\Rightarrow$ (b$_0$) $\Rightarrow$ (a)).

We shall now construct an infinite word u for which (a) holds but (b) does not. Let the morphism φ be defined by $\varphi(1) = 1213$, $\varphi(2) = \varepsilon$, $\varphi(3) = 23$, and let $u = \varphi^\infty(1)$.

Claim 3.1. *For each $i \geqslant 1$ we have*

$$\varphi^{i+1}(1) = \varphi^i(1)\,\varphi^i(1)\,23.$$

Proof. Since $\varphi(1) = 1213$ and $\varphi^2(1) = \varphi(1)\varphi(2)\varphi(1)\varphi(3) = 1213\,1213\,23$, the assertion holds for $i = 1$. By induction, we have

$$\varphi^{i+1}(1) = \varphi(\varphi^i(1)) = \varphi(\varphi^{i-1}(1)\,\varphi^{i-1}(1)\,23)$$
$$= \varphi^i(1)\,\varphi^i(1)\,\varphi(2)\,\varphi(3) = \varphi^i(1)\,\varphi^i(1)\,23,$$

which was to be proved. ∎

Claim 3.2. *For each $i \geqslant 1$ we have*

$$\varphi^i(1) = p_i 3(23)^{i-1},$$

where each p_i is a palindrome that begins with 12 (and thus ends with 21). Further, for $i \geqslant 2$, the largest power of 23 that is a factor of p_i is $(23)^{i-2}$, and for $i \geqslant 3$ this factor is unioccurrent in p_i.

Proof. Since $\varphi(1) = 1213$, the assertion holds for $i = 1$ (with $p_1 = 121$). Further, since $\varphi^2(1) = 1213121323$, the second part of the assertion holds for $i = 2$ (with $p_2 = 1213121$). By induction, using Claim 3.1, we have

$$\varphi^{i+1}(1) = p_i 3(23)^{i-1}\,p_i 3(23)^{i-1}\,23 = p_i 3(23)^{i-1} p_i 3(23)^i, \qquad (3.1)$$

and since

$$p_{i+1} = p_i 3(23)^{i-1} p_i \qquad (3.2)$$

is a palindrome, the first part of the claim is proved. Further, since p_i ends with 1 and begins with 1, the largest power of 23 that is a factor of p_{i+1} is $(23)^{i-1}$, which is unioccurrent in p_{i+1} for $i + 1 \geqslant 3$, and thus the proof is finished. ∎

Claim 3.3. Fact(u) *is closed under reversal, and u contains infinitely many palindromes.*

Proof. Each factor w of u is a factor of $\varphi^i(1)$ for i large enough. Since $\varphi^i(1)$ is a factor of p_{i+1} (see Claim 3.2), it follows that w is a factor of p_{i+1}, and thus its reversal is also a factor of p_{i+1} and in turn a factor of u.

The second part is clear by Claim 3.2. ∎

Claim 3.4. *The word u does not satisfy the statement (b).*

Proof. By (3.2), for each $i \geqslant 1$ we have that $(23)^i 12$ is a factor of p_{i+2} and in turn a factor of u. The longest palindromic suffix of this word is clearly only the letter 2, having $i+1$ occurrences in $(23)^i 12$. Thus, there are arbitrarily large factors w of u such that the longest palindromic suffix of w occurs in w more than once. Therefore, (b) fails. ∎

Claim 3.5. *The defect of u is finite.*

Proof. We shall prove that the longest palindromic suffix of any prefix w of u, of length $|w| \geqslant 10$, is unioccurrent in w. Therefore, u satisfies the statement (b$_0$), which is equivalent to (a).

Let w be a prefix of u, $|w| \geqslant 10$. Choose i such that w is not a prefix of $\varphi^i(1)$ (also not equal to it), but is a prefix of $\varphi^{i+1}(1)$. If $|w| = 10$, then $w = 1213121323 = \varphi^2(1)$, and the longest palindromic suffix of w is 323, which is indeed unioccurrent in w. Thus, assume $|w| \geqslant 11$. It now follows that $i \geqslant 2$.

By (3.1), w is a prefix of $p_i 3(23)^{i-1} p_i 3(23)^i$ longer than $p_i 3(23)^{i-1}$. Let us first consider the case when w is a prefix of $p_i 3(23)^{i-1} p_i$. In this case, it holds that $w = p_i 3(23)^{i-1} v$, where v is a prefix of p_i. Therefore, $\tilde{v}$ is a suffix of p_i, and thus $\tilde{v} 3(23)^{i-1} v$ is a palindromic suffix of w. This suffix is also the longest palindromic suffix of w, since if there were a longer one, there would be at least two occurrences of $3(23)^{i-1}$ in it and thus also in $p_{i+1} = p_i 3(23)^{i-1} p_i$, contradicting Claim 3.2. For the same reason, the suffix $\tilde{v} 3(23)^{i-1} v$ is unioccurrent in w, which was to be proved.

Assume now that w is longer than $p_i 3(23)^{i-1} p_i$. Therefore, it holds that either $w = p_i 3(23)^{i-1} p_i 3(23)^j$ for $0 \leqslant j \leqslant i$, or $w = p_i 3(23)^{i-1} p_i (32)^j$ for $1 \leqslant j \leqslant i$.

First, let

$$w = p_i 3(23)^{i-1} p_i 3(23)^j \tag{3.3}$$

for $0 \leqslant j \leqslant i$. If $j = i$, we claim that the longest palindromic suffix of w is $3(23)^i$. Since this suffix is indeed palindromic, it is enough to show that there does not exist a longer one. Suppose that v is a longer palindromic suffix. Since, by Claim 3.2, p_i ends with 1, we see that $v = \ldots 13(23)^i$, and by the fact that v is palindromic we now get $v = 3(23)^i \ldots 13(23)^i$. It follows that $3(23)^i$ is a factor of $p_i 3(23)^{i-1} p_i = p_{i+1}$, while by Claim 3.2 we have that the largest power of 23 that is a factor of p_{i+1} is $(23)^{i-1}$, a contradiction. Therefore, $3(23)^i$ is indeed the longest palindromic suffix of w, and it has to be unioccurrent in w since otherwise it would again follow that $3(23)^i$ is

a factor of $p_i 3(23)^{i-1} p_i$, an already seen contradiction. We shall now treat the case $0 \leqslant j \leqslant i - 1$. In this case, the suffix $3(23)^j p_i 3(23)^j$ of w is clearly palindromic, and we show that there does not exist a longer one. Suppose that v is a longer palindromic suffix. We see that, in the word v, the letter at the position $2j + 2$ from the right is 1 (because p_i ends with 1), and thus, by the fact that v is palindromic, the letter at the position $2j + 2$ from the left also has to be 1. Since v ends with $3(23)^j p_i 3(23)^j$ and is longer than it, it follows that there has to be the letter 1 in v before $3(23)^j p_i 3(23)^j$. Recalling that w is of the form (3.3), we conclude that v encompasses the whole factor $3(23)^{i-1}$, that is, $v = \ldots 13(23)^{i-1} p_i 3(23)^j$. However, in the word v, there are at most $|p_i|$ letters before $3(23)^{i-1}$ (since there are no more letters in w), and there are $|p_i| + 2j + 1 > |p_i|$ letters after it. By this and the fact that v is a palindrome, it follows that $\overbrace{13(23)^{i-1}} = 3(23)^{i-1}1$ must be a factor of $(23)^{i-1} p_i 3(23)^j$, and therefore a factor of $p_i 3(23)^j$. This is a contradiction (by Claim 3.2, the largest power of 23 that is a factor of p_i is $(23)^{i-2}$). Therefore, $3(23)^j p_i 3(23)^j$ is indeed the longest palindromic suffix of w, and it has to be unioccurrent in w since there are only two occurrences of p_i in w and the first one has no letters preceding it.

We now check the case

$$w = p_i 3(23)^{i-1} p_i (32)^j$$

for $1 \leqslant j \leqslant i$. If $j = i$, we claim that the longest palindromic suffix of w is $2(32)^{i-1}$. And indeed, this suffix is indeed palindromic, and in a similar manner as in the previous paragraph we see that there does not exist a longer one (since it would have to be of the form $(23)^i \ldots 1(32)^i$, and a contradiction would be reached). Further, it has to be unioccurrent in w, since otherwise it would follow that $2(32)^{i-1}$ is a factor of either p_i or $3(23)^{i-1}$, a contradiction (the first possibility cannot hold because of Claim 3.2 and $i \geqslant 2$, while the second one clearly is not true). We shall now treat the case $1 \leqslant j \leqslant i - 1$. In this case, the suffix $(23)^j p_i (32)^j$ of w is clearly palindromic, and we show that there does not exist a longer one. Suppose that v is a longer palindromic suffix. In a similar manner as in the previous paragraph, noting that, in the word v, the letter at the position $2j + 1$ from the right is 1, we conclude that v encompasses the whole factor $3(23)^{i-1}$, and get a contradiction as before. Therefore, $(23)^j p_i (32)^j$ is indeed the longest palindromic suffix of w, and it has to be unioccurrent in w since, again, there are only two occurrences of p_i in w and the first one has no letters preceding it. ∎

In conclusion: by Claims 3.3, 3.5 and 3.4, u is a counterexample to the assertion of Theorem 1.35.

In the rest of this section, we check whether the constructed word u perhaps is a counterexample also to the Brlek-Reutenauer conjecture. It turns out that it is not. The techniques used in this proof will be revisited in the following sections, in a more general context.

For the rest of this section, let us denote $|p_i| = l_i$. Since, by (3.2), the sequence $l_1, l_2, l_3 \ldots$ satisfies the recurrent relation $l_{i+1} = 2l_i + 2i - 1$ with $l_1 = |121| = 3$, it is an easy exercise in recurrent relations to show that $l_i = 3 \cdot 2^i - 2i - 1$. Indeed, $l_1 = 3 \cdot 2 - 2 - 1 = 3$, and $2l_i + 2i - 1 = 2 \cdot (3 \cdot 2^i - 2i - 1) + 2i - 1 = 3 \cdot 2^{i+1} - 2(i+1) - 1 = l_{i+1}$.

Claim 3.6. $D(u) = 1$.

Proof. By the proof of Claim 3.5, it is seen that $D(u) = D(121312132)$. Since the word 121312132 is of length 9 and has 9 palindromic factors: ε, 1, 2, 3, 121, 131, 21312, 31213, 1213121, the assertion follows by Definition 1.26. ∎

Claim 3.7. *For each $i \in \mathbb{N}$, the palindromic prefixes (also suffixes) of p_i are precisely: ε, 1 and p_j for $1 \leqslant j \leqslant i$.*

Proof. Since $p_1 = 121$ and $p_2 = 1213121$, the assertion holds for $i = 1$ and $i = 2$. We proceed by induction. Since p_i begins with 1 and ends with 1, each palindromic prefix v of $p_{i+1} = p_i 3(23)^{i-1} p_i$ that is not also a prefix of p_i encompasses the whole factor $3(23)^{i-1}$. However, since for $i \geqslant 2$ the largest power of 23 that is a factor of p_i is $(23)^{i-2}$, there is only one occurrence of $3(23)^{i-1}$ in v, and thus it has to be in the middle of v. Therefore, $v = p_{i+1}$. ∎

Claim 3.8. *For each $i \in \mathbb{N}$ and each copy of p_i in u, there is a copy of p_{i+1} in u such that the considered copy of p_i is positioned either at the beginning or at the end of this copy of p_{i+1}.*

Proof. For an observed copy of p_i in u, choose $j \geqslant i+1$ large enough such that this copy of p_i is contained in a prefix of u equal to $\varphi^j(1) = p_j 3(23)^{j-1}$. Therefore, the considered copy of p_i is contained in a copy of $p_j = p_{j-1}3(23)^{j-2}p_{j-1}$. If $j = i + 1$, the claim is proved; otherwise, we see that the considered copy of p_i is contained in one of the two occurrences of p_{j-1} in p_j, and the result follows by iterating. ∎

Claim 3.9. *Let $n \geqslant 7$ be given. Each $v \in \text{Pal}(u)$ such that $|v| = n$ is of one of the forms: $3(23)^{\frac{n-1}{2}}$, $(23)^{\frac{n-1}{2}}2$ or $w13(23)^k1\overline{w}$ for $2k+3 \leqslant n \leqslant 3 \cdot 2^{k+2}+2k+1$, where, for each k such that $2k + 3 \leqslant n \leqslant 3 \cdot 2^{k+2} + 2k + 1$, v is uniquely determined.*

Proof. Let us first, for a given $i \geqslant 2$, enlist all the palindromic factors of p_i, of length $\geqslant 7$. For $i = 2$, the only palindromic factor of length $\geqslant 7$ of $p_2 = 1213121$ is p_2 itself. Let now $i \geqslant 3$. By (3.2) we have:

$$p_i = p_{i-1}3(23)^{i-2}p_{i-1} = p_{i-2}3(23)^{i-3}p_{i-2}3(23)^{i-2}p_{i-2}3(23)^{i-3}p_{i-2}.$$

Clearly, all the palindromic factors of p_{i-1} also are palindromic factors of p_i. Let v be a palindromic factor of p_i which is not a factor of p_{i-1}, that is, is not contained in $p_{i-2}3(23)^{i-3}p_{i-2}$. Observing that v can not be completely encompassed in $p_{i-2}3(23)^{i-3}p_{i-2}$ (in neither of the two copies), we now distinguish two cases.

Case 1: $3(23)^{i-2}$ is a factor of v. In this case, since there is only one occurrence of $3(23)^{i-2}$ in p_i (by Claim 3.2), it has to be in the middle of v. Therefore,

$$v = 3(23)^{i-2}, \text{ or}$$

$$v \text{ is of the form } w13(23)^{i-2}1\overline{w}, \tag{3.4}$$

where w is an arbitrary suffix of p_{i-1}.

Case 2: $3(23)^{i-2}$ is not a factor of v. In this case, one possibility is $v = (23)^{i-3}2$. Let us check the remaining ones. The word v ends with some prefix of $3(23)^{i-2}$, or begins with a suffix of it, different from ε and $3(23)^{i-2}$ itself. Assume, w.l.o.g., the former occasion. Therefore, $v = v'3(23)^t$ for some $0 \leqslant t \leqslant i - 3$, or $v = v'(32)^t$ for some $1 \leqslant t \leqslant i - 2$, where v' is a suffix of p_{i-1}. In fact, since v is a palindrome, it holds that $v = 3(23)^tv''3(23)^t$ or $v = (23)^tv''(32)^t$, where v'' is a palindromic suffix of p_{i-1} not equal to itself, that is, $v'' = p_j$ for some $1 \leqslant j \leqslant i - 2$ (this follows by Claim 3.7, observing that $v \neq \varepsilon, 1$). By Claim 3.8 we get another constraint: $t \leqslant j-1$. Altogether, the palindromic factors in this case are:

$$(23)^{i-3}2;$$

$$3(23)^tp_j3(23)^t, \text{ for } 0 \leqslant t \leqslant j - 1 \leqslant i - 3; \tag{3.5}$$

$$(23)^tp_j(32)^t, \text{ for } 1 \leqslant t \leqslant j - 1 \leqslant i - 3. \tag{3.6}$$

Some of them are also palindromic factors of p_{i-1}, but it does not matter if a factor is taken into account twice, it was only important not to miss any that are not factors of p_{i-1}.

All the palindromic factors of u are factors of some p_i for i large enough, and thus we now check whether all the obtained factors are of one of the forms given in the statement. This is immediately clear for all the factors but (3.5) and (3.6), where in the case $j = 1$ we do not get the required form. However, $j = 1$ is possible only at (3.5), and it then follows that $t = 1$ and that the considered factor is $3p_1 3 = 31213$; however, this factor is of the length 5, and we assume $n \geqslant 7$.

We now check whether the bound given in the statement hold, and whether the factors are indeed uniquely determined for each n and k given. For k fixed, notice that all the enlisted factors of the form $w13(23)^k 1\overline{w}$ are precisely the middle segments of the word

$$3(23)^{k+1}p_{k+2}3(23)^{k+1} = 3(23)^{k+1}p_{k+1}3(23)^k p_{k+1}3(23)^{k+1}.$$

Indeed, factors from the group (3.4) are those for $i = k + 2$ (and w a suffix of p_{k+1}), while factors from the groups (3.5) and (3.6) are those for $j = k+2$ (and $0 \leqslant t \leqslant k + 1$, respectively $1 \leqslant t \leqslant k + 1$). This guarantees the uniquenees for each n and k given. Further, the shortest of these factors is $13(23)^k 1$, and thus

$$n \geqslant 2k + 3,$$

while the longest of these factors is $3(23)^{k+1}p_{k+2}3(23)^{k+1}$ itself, and thus

$$n \leqslant 2 + 4(k+1) + l_{k+2} = 4k + 6 + 3 \cdot 2^{k+2} - 2(k + 2) - 1 = 3 \cdot 2^{k+2} + 2k + 1.$$

This completes the proof. ∎

Claim 3.10. *It holds:*
- *each occurrence of the letter* 1 *in the word u is contained in a block of the form* 12131213*, and this block is always followed by the letter* 2*;*
- *each occurrence of the letter* 2 *in the word u is followed by the letter* 3*, unless it is contained in a block of the form* 12131213 *(when it is followed by the letter* 1*);*
- *each occurrence of the letter* 3 *in the word u is followed by one of the letters* 1*,* 2*.*

Proof. Follows by induction from Claim 3.1. ∎

51

Claim 3.11. *Let $n \geqslant 2$ be given. For each $v \in \text{Fact}(u)$ such that $|v| = n$, there either exists exactly one letter d such that $vd \in \text{Fact}(u)$, or exist exactly two letters d_1, d_2 such that $vd_1, vd_2 \in \text{Fact}(u)$.*

Further, the latter case occurs if and only if the largest power of 23 that is a suffix of v is $(23)^k$, with $2k \leqslant n \leqslant 3 \cdot 2^{k+1} + 2k - 1$.

Proof. Let $v \in \text{Fact}(u)$, $|v| = n \geqslant 2$. By Claim 3.10, we have:

- if v ends with the letter 1, then exactly one of $v2, v3$ is a factor of u, depending on whether the letter preceding the last 1 in v is 3 or 2, respectively;
- if v ends with the letter 2, then exactly one of $v1, v3$ is a factor of u, depending on whether the letter preceding the last 2 in v is 1 or 3, respectively.

We are thus left to analyze the case when v is of one the forms: $(23)^{\frac{n}{2}}$, $3(23)^{\frac{n-1}{2}}$ or $w13(23)^k$.

For the first two of these forms, the largest power of 23 that is a suffix of v is $(23)^{\frac{n}{2}}$, respectively $(23)^{\frac{n-1}{2}}$, and since $2 \cdot \frac{n}{2} \leqslant n \leqslant 3 \cdot 2^{\frac{n}{2}+1} + 2 \cdot \frac{n}{2} - 1$, we need to prove that then both $(23)^{\frac{n}{2}}1$ and $(23)^{\frac{n}{2}}2$, respectively $3(23)^{\frac{n-1}{2}}1$ and $3(23)^{\frac{n-1}{2}}2$, are factors of u. This is easily seen to hold by (3.2).

Let now
$$v = w13(23)^k.$$

We first note that from Claims 3.1 and 3.2 and the relation (3.2) directly follows that each copy in u of $13(23)^i1$, for any $i \in \mathbb{N}_0$, is contained in a copy of p_{i+2}, which is in turn contained in a copy of $\varphi^{i+2}(1)$. Further, we claim that each copy of $\varphi^{i+2}(1)$ is either positioned at the beginning of u, or preceded by $3(23)^{i+1}$. Indeed, each copy of $\varphi^{i+2}(1)$ coincides with one of the two occurrences of $\varphi^{i+2}(1)$ in a copy of $\varphi^{i+3}(1) = \varphi^{i+2}(1)\,\varphi^{i+2}(1)\,23$. If it coincides with the second occurrence, then it is preceded by $\varphi^{i+2}(1) = p_{i+2}3(23)^{i+1}$, which proves the assertion. Therefore, assume that the considered copy of $\varphi^{i+2}(1)$ is positioned at the beginning of $\varphi^{i+3}(1)$. This copy of $\varphi^{i+3}(1)$ coincides with one of the two occurrences of $\varphi^{i+3}(1)$ in a copy of $\varphi^{i+4}(1) = \varphi^{i+3}(1)\,\varphi^{i+3}(1)\,23$; if it coincides with the second occurrence, then it is preceded by $\varphi^{i+3}(1) = p_{i+3}3(23)^{i+2}$, and thus is indeed preceded by $3(23)^{i+1}$, as claimed; therefore, we again assume that this copy of $\varphi^{i+3}(1)$ is positioned at the beginning of $\varphi^{i+4}(1)$. By repeating this reasoning, we get that, unless the considered copy of $\varphi^{i+2}(1)$ is positioned at the beginning of u, for some $j \geqslant i + 2$ it holds that the considered copy of $\varphi^{i+2}(1)$ is positioned at the

beginning of $\varphi^j(1)$, which in turn coincides with the second occurrence of $\varphi^j(1)$ in $\varphi^{j+1}(1) = \varphi^j(1)\,\varphi^j(1)\,23$. Therefore, the considered copy of $\varphi^{i+2}(1)$ is preceded by $\varphi^j(1) = p_j 3(23)^{j-1}$, and thus is indeed preceded by $3(23)^{i+1}$, as claimed.

Assume $n \leqslant 3 \cdot 2^{k+1} + 2k - 1$. There exists $k_0 \geqslant k$ such that $v(23)^{k_0 - k} 1 = w13(23)^{k_0} 1 \in \text{Fact}(u)$. By the observations from the previous paragraph, we see that each copy of $13(23)^{k_0} 1$ in u is contained in a copy of

$$\underline{3(23)^{k_0+1}}\,\overbrace{\underbrace{p_{k_0+1}3(23)^{k_0}p_{k_0+1}}_{p_{k_0+2}}\,3(23)^{k_0+1}}^{\varphi^{k_0+2}(1)} \tag{3.7}$$

(where the part $\boxed{3(23)^{k_0+1}}$ exists iff the rest is not positioned at the beginning of the word u). Since $3(23)^k p_{k+1}$ is a suffix of $3(23)^{k_0+1} p_{k_0+1}$ (this is clear for $k_0 = k$, while for $k_0 > k$ it holds that $3(23)^k p_{k+1}$ is a suffix of p_{k+2}, which is in turn a suffix of p_{k_0+1}, by Claim 3.7) and $|3(23)^k p_{k+1} 3(23)^k| = 2k+1+l_{k+1}+2k+1 = 4k+2+3\cdot 2^{k+1}-2(k+1)-1 = 3\cdot 2^{k+1}+2k-1 \geqslant n = |v|$, we have that v is a suffix of $3(23)^k p_{k+1} 3(23)^k$. By Claims 3.1 and 3.2, we have

$$\varphi^{k+3}(1) = \varphi^{k+2}(1)\,\varphi^{k+2}(1)\,23 = \varphi^{k+1}(1)\,\varphi^{k+1}(1)\,23\,\varphi^{k+1}(1)\,\varphi^{k+1}(1)\,2323$$

$$= p_{k+1}3(23)^k p_{k+1}3(23)^{k+1} p_{k+1}3(23)^k p_{k+1}3(23)^{k+2}.$$

Therefore, we have that both $3(23)^k p_{k+1} 3(23)^k 1$ and $3(23)^k p_{k+1} 3(23)^k 2$ are factors of $\varphi^{k+3}(1)$, and thus also factors of u. This shows that $v1, v2 \in \text{Fact}(u)$, which was to be proved.

Let now $n > 3 \cdot 2^{k+1} + 2k - 1$. Suppose $v1, v2 \in \text{Fact}(u)$. This means that both $w13(23)^k 1$ and $w13(23)^{k_1} 1$ for some $k_1 > k$ are factors of u. Since each copy of $13(23)^k 1$ in u is contained in a copy of (3.7) for $k_0 = k$, and since $|v| > |3(23)^k p_{k+1} 3(23)^k|$, we have

$$(23)^{k+1} p_{k+1} 3(23)^k \text{ is a suffix of } v. \tag{3.8}$$

Futher, each copy of $13(23)^{k_1} 1$ in u is contained in a copy of (3.7) for $k_0 = k_1$. Since $13(23)^k p_{k+1}$ is a suffix of p_{k+2}, which is in turn a suffix of p_{k_1+1}, we have

$$13(23)^k p_{k+1} 3(23)^k \text{ is a suffix of } v. \tag{3.9}$$

Finally, (3.8) and (3.9) give a contradiction, and thus it is shown that only one of $v1, v2$ is a factor of u. The proof is completed. ∎

Claim 3.12. *For each $n \geqslant 6$ it holds:*

$$T_u(n) = 0.$$

Proof. We first note that there are no palindromic factors of u of even positive length: indeed, in any such factor there would have to be an occurrence of the same two letters next to each other, which is impossible by Claim 3.10.

Let $n \geqslant 6$ be given. By Claim 3.11, we have

$$C_u(n+1) - C_u(n) = |\{k \geqslant 0 : 2k \leqslant n \leqslant 3 \cdot 2^{k+1} + 2k - 1\}|. \qquad (3.10)$$

If n is odd, then $P_u(n+1) = 0$, while Claim 3.9 gives

$$
\begin{aligned}
P_u(n) &= 2 + |\{k \geqslant 0 : 2k + 3 \leqslant n \leqslant 3 \cdot 2^{k+2} + 2k + 1\}| \\
&= 2 + |\{k' \geqslant 1 : 2(k' - 1) + 3 \leqslant n \leqslant 3 \cdot 2^{(k'-1)+2} + 2(k' - 1) + 1\}| \\
&= 2 + |\{k' \geqslant 1 : 2k' + 1 \leqslant n \leqslant 3 \cdot 2^{k'+1} + 2k' - 1\}| \\
&= 2 + C_u(n+1) - C_u(n)
\end{aligned}
$$

(we have $k \neq 0$ in (3.10), because $k = 0$ implies that $n \leqslant 3 \cdot 2^1 - 1 = 5$; since n is odd, we also have $2k < n$ in (3.10), that is, $2k + 1 \leqslant n$). If n is even, then $P_u(n) = 0$, while Claim 3.9 gives

$$
\begin{aligned}
P_u(n+1) &= 2 + |\{k \geqslant 0 : 2k + 3 \leqslant n + 1 \leqslant 3 \cdot 2^{k+2} + 2k + 1\}| \\
&= 2 + |\{k' \geqslant 1 : 2k' + 1 \leqslant n + 1 \leqslant 3 \cdot 2^{k'+1} + 2k' - 1\}| \\
&= 2 + |\{k' \geqslant 1 : 2k' \leqslant n \leqslant 3 \cdot 2^{k'+1} + 2k' - 2\}| \\
&= 2 + C_u(n+1) - C_u(n)
\end{aligned}
$$

(since n is even, we have $n < 3 \cdot 2^{k+1} + 2k - 1$ in (3.10), that is, $n \leqslant 3 \cdot 2^{k+1} + 2k - 2$). Therefore,

$$T_u(n) = C_u(n+1) - C_u(n) + 2 - P_u(n) - P_u(n+1) = 0,$$

which was to be proved. ∎

Claim 3.13. *The word u satisfies the Brlek-Reutenauer conjecture, that is,*

$$2D(u) = \sum_{n=0}^{\infty} T_u(n).$$

n	$v \in \mathrm{Fact}(u), |v| = n$	$v \in \mathrm{Pal}(u), |v| = n$	$T_u(n)$
0	ε	ε	0
1	1 2 3	1 2 3	2
2	12 13 21 23 31 32	none	0
3	121 131 132 213 231 232 312 323	121 131 232 323	0
4	1213 1312 1323 2131 2132 2312 2323 3121 3231 3232	none	0
5	12131 12132 13121 13231 13232 21312 21323 23121 23231 23232 31213 32312 32323	13231 21312 23232 31213 32323	0
6	121312 121323 131213 132312 132323 213121 213231 213232 231213 232312 232323 312131 312132 323121 323231 323232	none	

Table 3.1: Evaluating $T_u(n)$ for $n = 0, 1, \ldots, 5$.

Proof. We first evaluate $T_u(n)$ for $n = 0, 1, \ldots, 5$, this being done in a direct way: we enlist all the factors as well as the palindromic factors of u of length $\leqslant 6$, count them, and evaluate $T_u(n)$ by the definition. All this is presented in Table 3.1. For enlisting the factors systematically, Claim 3.10 is used.

Together with Claim 3.12, this shows:

$$\sum_{n=0}^{\infty} T_u(n) = 2.$$

By Claim 3.6, the proof is finished. ∎

3.2 Highly potential words. Construction and basic properties

Let w be a finite word that is not a palindrome, and let c be a letter that does not occur in w. Define $w_0 = w$ and, for $i \in \mathbb{N}$,

$$w_i = w_{i-1} c^i \widetilde{w_{i-1}}. \tag{3.11}$$

Finally, let

$$\mathrm{hpw}(w) = \lim_{i \to \infty} w_i. \tag{3.12}$$

The meaning of the above limit is clear since each w_i is a prefix of w_{i+1}. We call $\mathrm{hpw}(w)$ the *highly potential word generated by w*.

The following proposition is easy to prove, but is of key importance.

Proposition 3.14. *Let* $\mathrm{hpw}(w)$ *be a highly potential word. It holds:*

a) $\mathrm{Fact}(\mathrm{hpw}(w))$ *is closed under reversal;*

b) $\mathrm{hpw}(w)$ *is recurrent;*

c) $\mathrm{hpw}(w)$ *is not uniformly recurrent;*

d) $\mathrm{hpw}(w)$ *is aperiodic.*

Proof. a) Let $v \in \mathrm{Fact}(\mathrm{hpw}(w))$. Choose $i \in \mathbb{N}_0$ large enough such that $v \in \mathrm{Fact}(w_i)$. Since $w_{i+1} = w_i c^{i+1} \widetilde{w_i}$, we have $\widetilde{v} \in \mathrm{Fact}(w_{i+1})$, and thus $\widetilde{v} \in \mathrm{Fact}(\mathrm{hpw}(w))$.

   b) Follows from a) and Theorem 1.5.

   c) Since $w \in \mathrm{Fact}(\mathrm{hpw}(w))$, and since we can always find two consecutive occurrences of w in $\mathrm{hpw}(w)$ with arbitrarily many letters c in the gap between, the statement follows.

   d) Follows from c) and Theorem 1.7. ∎

The main result of this section is the following theorem.

Theorem 3.15. *Let* $\mathrm{hpw}(w)$ *be a highly potential word. Then* $D(\mathrm{hpw}(w)) = D(w) + 1$. *In particular,*

$$0 < D(\mathrm{hpw}(w)) < \infty.$$

Proof. Let $w = w_0 = a_1 a_2 \ldots a_l$, where $a_1, a_2, \ldots, a_l \in \Sigma$. Since

$$w_1 = w_0 c \widetilde{w_0} = a_1 a_2 \ldots a_l c a_l \ldots a_2 a_1,$$

it is easy to see that

$$\mathrm{Pal}(w_1) = \mathrm{Pal}(w_0) \cup \{a_s a_{s+1} \ldots a_l c a_l \ldots a_{s+1} a_s : 1 \leqslant s \leqslant l\} \cup \{c\}.$$

Therefore,

$$D(w_1) = |w_1| + 1 - |\mathrm{Pal}(w_1)| = 2l + 1 + 1 - |\mathrm{Pal}(w_0)| - l - 1$$
$$= l + 1 - |\mathrm{Pal}(w_0)| = |w_0| + 1 - |\mathrm{Pal}(w_0)| = D(w_0) = D(w).$$

Since

$$w_2 = w_1 cc\widetilde{w_1} = a_1 a_2 \ldots a_l ca_l \ldots a_2 a_1 cca_1 a_2 \ldots a_l ca_l \ldots a_2 a_1,$$

having in mind that $w = a_1 a_2 \ldots a_l$ is not a palindrome, it is easy to see that

$$\mathrm{Pal}(w_2) = \mathrm{Pal}(w_1)$$
$$\cup \{a_s a_{s+1} \ldots a_l ca_l \ldots a_2 a_1 cca_1 a_2 \ldots a_l ca_l \ldots a_{s+1} a_s : 1 \leqslant s \leqslant l\}$$
$$\cup \{a_s \ldots a_2 a_1 cca_1 a_2 \ldots a_s : l \geqslant s \geqslant 1\}$$
$$\cup \{ca_l \ldots a_2 a_1 cca_1 a_2 \ldots a_l c, cc\}.$$

Therefore,

$$D(w_2) = |w_2| + 1 - |\mathrm{Pal}(w_2)| = 4l + 4 + 1 - |\mathrm{Pal}(w_1)| - l - l - 2$$
$$= 2l + 3 - |\mathrm{Pal}(w_1)| = |w_1| + 2 - |\mathrm{Pal}(w_1)| = D(w_1) + 1$$
$$= D(w) + 1.$$

We are now going to prove that $D(w_i) = D(w_{i-1})$ for each $i \geqslant 3$. Note that w_i is a palindrome for each $i \geqslant 1$. Let $i \geqslant 3$ be given, and let $w_{i-2} = b_1 b_2 \ldots b_m$, where $b_1, b_2, \ldots, b_m \in \Sigma$. We have

$$w_{i-1} = w_{i-2} c^{i-1} \widetilde{w_{i-2}} = b_1 b_2 \ldots b_m c^{i-1} b_m \ldots b_2 b_1$$

and

$$w_i = w_{i-1} c^i \widetilde{w_{i-1}} = b_1 b_2 \ldots b_m c^{i-1} b_m \ldots b_2 b_1 c^i b_1 b_2 \ldots b_m c^{i-1} b_m \ldots b_2 b_1.$$

We claim that

$$\mathrm{Pal}(w_i) = \mathrm{Pal}(w_{i-1})$$
$$\cup \{b_s b_{s+1} \ldots b_m c^{i-1} b_m \ldots b_2 b_1 c^i b_1 b_2 \ldots b_m c^{i-1} b_m \ldots b_{s+1} b_s$$
$$: 1 \leqslant s \leqslant m\}$$
$$\cup \{c^s b_m \ldots b_2 b_1 c^i b_1 b_2 \ldots b_m c^s : i - 1 \geqslant s \geqslant 1\}$$
$$\cup \{b_s \ldots b_2 b_1 c^i b_1 b_2 \ldots b_s : m \geqslant s \geqslant 1\}$$
$$\cup \{c^i\} \cup \{c^s b_1 b_2 \ldots b_m c^s : 1 \leqslant s \leqslant i - 1\}.$$

57

Indeed: all the palindromes added in the first set are new, because there is no factor c^i in w_{i-1}; all the palindromes added in the second, the third and the fourth set are new for the same reason; finally, all the words added in the fifth set are palindromes because $w_{i-2} = b_1 b_2 \ldots b_m$ is a palindrome, and it can be seen that all of them also are new. Further, it can be easily checked that the list above is complete. Therefore,

$$D(w_i) = |w_i| + 1 - |\mathrm{Pal}(w_i)|$$
$$= 4m + 3i - 2 + 1 - |\mathrm{Pal}(w_{i-1})| - m - (i-1) - m - 1 - (i-1)$$
$$= 2m + i - |\mathrm{Pal}(w_{i-1})| = |w_{i-1}| + 1 - |\mathrm{Pal}(w_{i-1})| = D(w_{i-1}).$$

Altogether, $D(w_0) = D(w_1) = D(w)$ and $D(w_i) = D(w) + 1$ for $i \geqslant 2$. Because of Corollary 1.28 and the equality (3.12), it holds:

$$\sup_{v \in \mathrm{Fact}(\mathrm{hpw}(w))} D(v) = \sup_{i \in \mathbb{N}_0} D(w_i),$$

and thus

$$D(\mathrm{hpw}(w)) = \sup_{i \in \mathbb{N}_0} D(w_i) = D(w) + 1,$$

which was to be proved. ∎

3.3 Conjecture 1.31 for highly potential words

In this section we prove that highly potential words satisfy the Brlek-Reutenauer conjecture.

Theorem 3.16. *For each highly potential word* $\mathrm{hpw}(w)$ *it holds:*

$$2D(\mathrm{hpw}(w)) = \sum_{n=0}^{\infty} T_{\mathrm{hpw}(w)}(n).$$

The proof is preceded by a series of lemmas. For the rest of this section, let $w = w_0 = a_1 a_2 \ldots a_l$, where $a_1, a_2, \ldots, a_l \in \Sigma$, and let $|w_i| = l_i$. Since, by (3.11), the sequence $l_0, l_1, l_2 \ldots$ satisfies the recurrent relation $l_i = 2l_{i-1} + i$ with $l_0 = l$, it is an easy exercise in recurrent relations to show that $l_i = (l+2) \cdot 2^i - i - 2$. Indeed, $l_0 = (l+2) \cdot 2^0 - 0 - 2 = l$, and $2l_{i-1} + i = 2((l+2) \cdot 2^i - i - 2) + i = (l+2) \cdot 2^{i+1} - i + 2 = l_{i+1}$.

Lemma 3.17. *Let $n \geqslant 1$ be given. Each $v \in \mathrm{Pal}(\mathrm{hpw}(w)) \setminus \mathrm{Pal}(w)$ such that $|v| = n$ is uniquely determined by the number of consecutive occurrences of the letter c in the middle of the palindrome v.*

Further, the letter c may consecutively occur exactly $k \geqslant 1$ times in the middle of the palindrome v if and only if $k \leqslant n \leqslant (l + 2) \cdot 2^k + k$ and $k \equiv n \pmod{2}$.

Proof. From (3.11), we see that the letter c occurs exactly k times consecutively only in the word w_k and its further copies in $\mathrm{hpw}(w)$. We have:

$$w_k = w_{k-1}c^k\widetilde{w_{k-1}} = w_{k-1}c^k w_{k-1};$$

$$w_{k+1} = w_k c^{k+1}\widetilde{w_k} = w_k c^{k+1} w_k = w_{k-1}c^k w_{k-1}c^{k+1}w_{k-1}c^k w_{k-1};$$

$$w_{k+2} = w_{k+1}c^{k+2}\widetilde{w_{k+1}} = w_{k+1}c^{k+2}w_{k+1}$$
$$= w_{k-1}c^k w_{k-1}c^{k+1}w_{k-1}c^k w_{k-1}c^{k+2}w_{k-1}c^k w_{k-1}c^{k+1}w_{k-1}c^k w_{k-1}.$$

Therefore, the simultaneous "extending" of both ends of the word c^k can last only till we reach $c^{k+1}w_{k-1}c^k w_{k-1}c^{k+1}$, since at this point the following letter on the right side is c and on the left side is $\neq c$, or vice versa (clearly, the same holds for further copies of w_{k+2}). Thus, the letter c consecutively occurs exactly k times in the middle of a palindrome of a given length n if and only if the palindrome is a middle section of $c^{k+1}w_{k-1}c^k w_{k-1}c^{k+1}$, and therefore it is uniquely determined. Furthermore, we see that such a palindrome exists if and only if $k \equiv n \pmod{2}$ and

$$k \leqslant n \leqslant 2l_{k-1} + 3k + 2 = 2((l+2) \cdot 2^{k-1} - (k-1) - 2) + 3k + 2$$
$$= (l+2) \cdot 2^k - 2k + 2 - 4 + 3k + 2$$
$$= (l+2) \cdot 2^k + k,$$

which was to be proved. ∎

Lemma 3.18. *Let $n \geqslant l + 3$ be given. For each $v \in \mathrm{Fact}(\mathrm{hpw}(w))$ such that $|v| = n$, there either exists exactly one letter d such that $vd \in \mathrm{Fact}(\mathrm{hpw}(w))$, or exist exactly two letters d_1, d_2 such that $vd_1, vd_2 \in \mathrm{Fact}(\mathrm{hpw}(w))$.*

Further, the latter case occurs if and only if v ends with exactly k letters c, with $k \leqslant n \leqslant (l+2) \cdot 2^{k-1} + k - 1$.

Proof. Observe the following easy to see corollary of the definition of $\mathrm{hpw}(w)$: for any occurrence of the letter c in $\mathrm{hpw}(w)$ such that both the letters preceding it and following it are $\neq c$, this letter c is necessarily followed by $\widetilde{w_0}$; for any occurrence of a sequence of two or more consecutive letters c in $\mathrm{hpw}(w)$, this sequence is followed by w_0.

Let $v = b_1 b_2 \ldots b_n$. Assume that v ends with exactly k letters c, where $0 \leqslant k \leqslant n$.

Consider the case $k = 0$, that is, $b_n \neq c$. Since $(l+2) \cdot 2^{k-1} + k - 1 = (l+2) \cdot 2^{-1} + 0 - 1 = \frac{l}{2} < n$, we have to prove that in this case there is uniquely determined letter d such that $vd \in \mathrm{Fact}(\mathrm{hpw}(w))$. Since $n \geqslant l+3$, the letter c must occur in v; in fact, it must occur in $b_3 b_4 \ldots b_n$. Let $b_t = c$ be the last occurrence of c in v, where $3 \leqslant t \leqslant n-1$. We thus have that $b_{t+1} b_{t+2} \ldots b_n$ is a prefix of w_0 or $\widetilde{w_0}$. By the observation given above, we see that if $b_{t-1} = c$, then $b_{t+1} b_{t+2} \ldots b_n$ is a prefix of w_0, while if $b_{t-1} \neq c$, then $b_{t+1} b_{t+2} \ldots b_n$ is a prefix of $\widetilde{w_0}$. Both of these possibilities lead to conclusion that there is only one letter d such that $vd \in \mathrm{Fact}(\mathrm{hpw}(w))$: d is the letter that follows $b_{t+1} b_{t+2} \ldots b_n$ in w_0, respectively $\widetilde{w_0}$, or, if $b_{t+1} b_{t+2} \ldots b_n$ is equal to w_0 or $\widetilde{w_0}$, then $d = c$.

Let now $k = n$, that is, $v = c^n$. Since $(l+2) \cdot 2^{k-1} + k - 1 = (l+2) \cdot 2^{n-1} + n - 1 > n$, we have to prove that in this case there are exactly two letters d_1, d_2 such that $vd_1, vd_2 \in \mathrm{Fact}(\mathrm{hpw}(w))$. And indeed, the only two such letters are $d_1 = c$ and $d_2 = a_1$ (in case $a_1 \neq a_l$, there cannot be $d_2 = a_l$ because of $n > 1$ and the observation from the beginning of the proof).

Finally, let $1 \leqslant k \leqslant n-1$, that is, $b_n = b_{n-1} = \cdots = b_{n-k+1} = c$ and $b_{n-k} \neq c$. It is easy to see that the only two letters d_1, d_2 such that it could possibly hold that $vd_1, vd_2 \in \mathrm{Fact}(\mathrm{hpw}(w))$ are $d_1 = c$ and either $d_2 = a_1$ (in case $k > 1$) or $d_2 = a_l$ (in case $k = 1$).

Assume $n > l_{k-1} + 2k$ and $vc \in \mathrm{Fact}(\mathrm{hpw}(w))$. Wherever the word $b_{n-k+1} b_{n-k+2} \ldots b_n c = c^{k+1}$ is positioned in $\mathrm{hpw}(w)$, it is clearly a part of a middle segment c^s of a copy of $w_s = w_{s-1} c^s \widetilde{w_{s-1}}$ for some $s \geqslant k+1$. We have that $w_k = w_{k-1} c^k \widetilde{w_{k-1}}$ is a prefix of w_{s-1}, and thus $\widetilde{w_k} = w_{k-1} c^k \widetilde{w_{k-1}}$ is a suffix of $\widetilde{w_{s-1}} = w_{s-1}$. Since $n - l_{k-1} - 2k \geqslant 1$, it follows that $b_{n-l_{k-1}-2k} \neq c$ (see Figure 3.1).

Still assuming that $n > l_{k-1} + 2k$, further assume that now $vd_2 \in \mathrm{Fact}(\mathrm{hpw}(w))$ (where $d_2 = a_1$ if $k > 1$, and $d_2 = a_l$ if $k = 1$). Wherever the word $b_{n-k} b_{n-k+1} \ldots b_n d_2 = b_{n-k} c^k d_2$ is positioned in $\mathrm{hpw}(w)$, it is clearly contained in a copy of w_k. We claim that there is a sequence of at least $k+1$ consecutive letters c immediately preceding this copy of w_k. There exists a

$$vc: \quad \overbrace{b_1 \ldots b_{n-l_{k-1}-2k} \mid b_{n-l_{k-1}-2k+1} \ldots b_{n-k-1}b_{n-k} \mid c^k}^{v} \mid c$$

$$\mathrm{hpw}(w): \quad \underbrace{\ldots\ldots w_{k-1} \mid \overbrace{c^k \widetilde{w_{k-1}}}^{} \mid c^k \mid c^{s-k}\widetilde{w_{s-1}} \ldots}_{w_{s-1}}$$

Figure 3.1: $b_{n-l_{k-1}-2k} \neq c$.

copy of $w_{k+1} = w_k c^{k+1}\widetilde{w_k} = w_k c^{k+1} w_k$ such that the considered copy of w_k coincides either with a prefix or with a suffix of this copy of w_{k+1}. In the latter case, it is preceded by c^{k+1}, as claimed. Thus, assume the former case. The considered copy of w_{k+1} now coincides either with a prefix or with a suffix of a copy of $w_{k+2} = w_{k+1}c^{k+2}w_{k+1}$. In the latter case, it is preceded by c^{k+2}, and thus its prefix w_k also is preceded by c^{k+2}, as claimed. Thus, assume again the former case. The copy of $w_k = w_{k-1}c^k\widetilde{w_{k-1}}$ we begin with cannot be positioned in the beginning of $\mathrm{hpw}(w)$, since there should be at least $n - k > l_{k-1} + k > l_{k-1}$ letters before c^k. Therefore, if the procedure above is repeated, it eventually happens that for some $r \geqslant k$ the considered copy of w_k coincides with a prefix of a copy of w_r that in turn coincides with a suffix of a copy of $w_{r+1} = w_r c^{r+1}w_r$. Thus, the considered copy of w_k is preceded by c^{r+1}, which proves the claim. Therefore, $b_{n-l_{k-1}-2k} = c$ (see Figure 3.2).

$$vd_2: \quad \overbrace{b_1 \ldots b_{n-l_{k-1}-2k} \mid b_{n-l_{k-1}-2k+1} \ldots b_{n-k-1}b_{n-k} \mid c^k}^{v} \mid d_2$$

$$\mathrm{hpw}(w): \quad \ldots w_r c^{r+1-k} \mid \underbrace{c^k w_{k-1}}_{w_r} \mid c^k \mid \underbrace{\widetilde{w_{k-1}} \ldots}_{} \ldots$$

Figure 3.2: $b_{n-l_{k-1}-2k} = c$.

Summing the results, we have proved that if

$$n > l_{k-1} + 2k = (l+2) \cdot 2^{k-1} - (k-1) - 2 + 2k = (l+2) \cdot 2^{k-1} + k - 1,$$

then not both vd_1, vd_2 can belong to $\mathrm{Fact}(\mathrm{hpw}(w))$. In order to finish the proof, it is enough to prove the reverse direction for $1 \leqslant k \leqslant n - 1$. Let $n \leqslant l_{k-1} + 2k = (l+2) \cdot 2^{k-1} + k - 1$. There does not exist such n for $k = 1$, since otherwise it would have to hold that $n \leqslant (l+2) \cdot 2^{1-1} + 1 - 1 = l + 2$, contradicting the assumption $n \geqslant l + 3$. Thus, we are left to check the case $2 \leqslant k \leqslant n - 1$. In this case it holds that $\widetilde{w_{k-1}} = w_{k-1}$; therefore, v is a suffix

of $c^k w_{k-1} c^k$, and $vc, va_1 \in \text{Fact}(\text{hpw}(w))$ (Figures 3.1 and 3.2 could again help visualising these conclusions). This completes the proof. $\blacksquare$

Lemma 3.19. *For each $n \geqslant l + 3$ it holds:*

$$T_{\text{hpw}(w)}(n) = 0.$$

Proof. Each $v \in \text{Pal}(\text{hpw}(w))$ such that $|v| \geqslant l + 3$ clearly does not belong to $\text{Pal}(w)$. Let $n \geqslant l + 3$ be given, and let

$$A = \{k \geqslant 1 : k \leqslant n \leqslant (l+2) \cdot 2^k + k \wedge k \equiv n \ (\text{mod } 2)\},$$

$$B = \{k \geqslant 1 : k \leqslant n + 1 \leqslant (l+2) \cdot 2^k + k \wedge k \equiv n + 1 \ (\text{mod } 2)\},$$

$$C = \{k \geqslant 1 : k - 1 \leqslant n \leqslant (l+2) \cdot 2^k + k\}.$$

We claim that $A \cap B = \varnothing$ and $A \cup B = C$. It is easy to see that $A \cap B = \varnothing$ and $A, B \subseteq C$, and thus we are left to prove that $C \subseteq A \cup B$. Let $k \in C$. If $k \equiv n \ (\text{mod } 2)$, then $k - 1 \neq n$, and thus $k - 1 < n$, that is, $k \leqslant n$; therefore, $k \in A$. If $k \equiv n + 1 \ (\text{mod } 2)$, then $n \neq (l+2) \cdot 2^k + k$, and thus $n < (l+2) \cdot 2^k + k$, that is, $n + 1 \leqslant (l+2) \cdot 2^k + k$; therefore, $k \in B$.

By Lemma 3.17, we now have:

$$P_{\text{hpw}(w)}(n) + P_{\text{hpw}(w)}(n+1) = |A| + |B| = |C|.$$

By Lemma 3.18, we have:

$$C_{\text{hpw}(w)}(n+1) - C_{\text{hpw}(w)}(n) = |\{k \geqslant 0 : k \leqslant n \leqslant (l+2) \cdot 2^{k-1} + k - 1\}|.$$

Clearly, the set $\{k \geqslant 0 : k \leqslant n \leqslant (l+2) \cdot 2^{k-1} + k - 1\}$ is an interval, say $[k_{\min}, k_{\max}]$. Actually, it holds that $k_{\max} = n$. We claim that $C = [k_{\min} - 1, n + 1]$. It is easy to see that $n + 1 \in C$ and $n + 2 \notin C$. Let us show the other bound. Since $(l+2) \cdot 2^{1-1} + 1 - 1 = l + 2 < n$, we have $k_{\min} \geqslant 2$. Therefore, $k_{\min} - 1 \geqslant 1$. From $k_{\min} \leqslant n$ and $n \leqslant (l+2) \cdot 2^{k_{\min}-1} + k_{\min} - 1$ we have $k_{\min} - 1 < n$ and $k_{\min} - 1 \in C$. Suppose $k_{\min} - 2 \in C$. We have $k_{\min} - 1 \leqslant n - 1 < n$. Further, from the supposed $k_{\min} - 2 \in C$ it follows that $n \leqslant (l+2) \cdot 2^{k_{\min}-2} + k_{\min} - 2$. Therefore, $k_{\min} - 1 \in [k_{\min}, n]$, which is a clear contradiction, and thus $k_{\min} - 2 \notin C$. Since C is an interval, it holds that $C = [k_{\min} - 1, n + 1]$.

Finally,

$$T_{\mathrm{hpw}(w)}(n) = C_{\mathrm{hpw}(w)}(n+1) - C_{\mathrm{hpw}(w)}(n) + 2 - P_{\mathrm{hpw}(w)}(n) - P_{\mathrm{hpw}(w)}(n+1)$$
$$= |[k_{\min}, n]| + 2 - |[k_{\min} - 1, n+1]|$$
$$= (n - k_{\min} + 1) + 2 - (n + 1 - (k_{\min} - 1) + 1) = 0,$$

which was to be proved. ∎

Lemma 3.20. *It holds:*

$$C_{\mathrm{hpw}(w)}(l+3) = 2|\{v \in \mathrm{Pal}(\mathrm{hpw}(w)) \backslash \mathrm{Pal}(w) : |v| \leqslant l+3\}| - P_{\mathrm{hpw}(w)}(l+3) - 2.$$

Proof. We begin by listing all the factors of $\mathrm{hpw}(w)$ of length $l+3$. However, since $\mathrm{Fact}(\mathrm{hpw}(w))$ is closed under reversal, we shall not include both a factor and its reversal in the list, but choose only one representative for each such pair. We claim that the left column of Table 3.2 presents the described list.

$u \in \mathrm{Fact}(\mathrm{hpw}(w))$, $|u| = l+3$	the longest prefix or suffix of u from $\mathrm{Pal}(\mathrm{hpw}(w)) \backslash (\mathrm{Pal}(w) \cup \{c\})$
wca_la_{l-1}	$a_{l-1}a_lca_la_{l-1}$
$cwca_l$	a_lca_l
$ccwc$	cc
$cccw$	ccc
a_1ccw	a_1cca_1
$a_{l-s}\ldots a_{l-1}a_lca_la_{l-1}\ldots a_s$ $(l-2 \geqslant s \geqslant \lceil\frac{l}{2}\rceil)$	$a_s \ldots a_{l-1}a_lca_la_{l-1}\ldots a_s$
$a_{l+3-t}\ldots a_2a_1c^t$ $(4 \leqslant t \leqslant l+2)$	c^t
$a_{l+2-t}\ldots a_2a_1c^ta_1$ $(3 \leqslant t \leqslant l+1)$	$a_1c^ta_1$
$a_{l+3-s-t}\ldots a_2a_1c^ta_1a_2\ldots a_s$ $(2 \leqslant s \leqslant \lfloor\frac{l+1}{2}\rfloor, \ 2 \leqslant t \leqslant l+3-2s)$	$a_s \ldots a_2a_1c^ta_1a_2\ldots a_s$
c^{l+3}	c^{l+3}

Table 3.2: Factors and its longest palindromic prefixes or suffixes.

The list is compiled by the following approach:

- We first enumerate all $u \in \mathrm{Fact}(\mathrm{hpw}(w))$, $|u| = l + 3$ such that $w \in$ $\mathrm{Fact}(u)$. Depending on whether w begins with the first, the second, the third or the fourth letter of u, we easily see that in all of these cases but the last one the other characters are uniquely determined, while in the last case there are exactly two possibilities. These five factors are shown in the first group. These factors also stand as the representative of factors u such that $\widetilde{w} \in \mathrm{Fact}(u)$.

- We now enumerate all $u \in \mathrm{Fact}(\mathrm{hpw}(w))$, $|u| = l + 3$ such that u ends with a prefix of $\widetilde{w}$, say $a_l a_{l-1} \ldots a_s$. We see that in this case it must hold that $u = a_{l-s} \ldots a_{l-1} a_l c a_l a_{l-1} \ldots a_s$ (the "left end" is calculated so that $(l - (l - s) + 1) + 1 + (l - s + 1) = l + 3$). Since we require $w, \widetilde{w} \notin \mathrm{Fact}(u)$ (in order to avoid repeating a factor already included in the first group), it must hold that $s \geqslant 2$ and $s \leqslant l - 2$. Furthermore, since reversals of factors from this group are of the same form, in order to avoid repeating we require $|a_{l-s} \ldots a_{l-1} a_l| \geqslant |a_l a_{l-1} \ldots a_s|$, that is, $s \geqslant l - s$, that is, $s \geqslant \lceil \frac{l}{2} \rceil$. Altogether: for $l \geqslant 3$ the bounds are $l - 2 \geqslant s \geqslant \lceil \frac{l}{2} \rceil$; for $l = 2$, already $s \leqslant l - 2$ implies that the group is empty, and thus the same bounds as for the case $l \geqslant 3$ can formally stay in this case, too.

- In the third group we enumerate all the considered factors u that end with c^t, but $u \neq c^{l+3}$. It cannot be $t = 1$, since u would contain w or $\widetilde{w}$. Therefore, $u = a_{l+3-t} \ldots a_2 a_1 c^t$. The bounds are $2 \leqslant t - 2 \leqslant l$, that is, $4 \leqslant t \leqslant l + 2$.

- We now check what are the possibilities when u ends with a prefix of w, say $a_1 a_2 \ldots a_s$. In fact, it would be helpful to distinguish cases $s = 1$ and $s > 1$. Thus, in this group we let $u = a_{l+2-t} a_{l+1-t} \ldots a_1 c^t a_1$. The bounds are $l - 1 \geqslant l + 2 - t \geqslant 1$, that is, $3 \leqslant t \leqslant l + 1$.

- Let now $u = a_{l+3-s-t} \ldots a_2 a_1 c^t a_1 a_2 \ldots a_s$, $s \geqslant 2$. The bound $t \geqslant 2$ is clear. In order to avoid including both a factor and its reversal, we require $|a_{l+3-s-t} \ldots a_2 a_1| \geqslant |a_1 a_2 \ldots a_s|$, that is, $l + 3 - s - t \geqslant s$, that is, $t \leqslant l + 3 - 2s$. For a fixed s, we have the bounds $l - 1 \geqslant l + 3 - s - t \geqslant 1$, that is, $4 - s \leqslant t \leqslant l + 2 - s$. Since $4 - s \leqslant 2$ and $l + 3 - 2s \leqslant l + 2 - s$, the bounds for t are $2 \leqslant t \leqslant l + 3 - 2s$. Considering the bounds for s, we already have $s \geqslant 2$, and an upper bound follows from the requirement that t exists: $2 \leqslant l + 3 - 2s$, that is, $s \leqslant \lfloor \frac{l+1}{2} \rfloor$.

- Finally, there is one more factor not included so far: c^{l+3}.

For each of the enumerated factors, we find out that either its longest palindromic prefix or longest palindromic suffix, but not both, belongs to

$\operatorname{Pal}(\operatorname{hpw}(w)) \setminus (\operatorname{Pal}(w) \cup \{c\})$. These prefixes and suffixes are shown in the right column of Table 3.2. We claim that such a correspondence is in fact a bijection between the left column and the set $\{v \in \operatorname{Pal}(\operatorname{hpw}(w)) \setminus \operatorname{Pal}(w) : 2 \leqslant |v| \leqslant l+3\}$. Therefore, it is enough to check whether each v from this set appears exactly once in the right column.

We shall enumerate these palindromes by ordering them with respect to the number of consecutive occurrences of the letter c in the middle (by Lemma 3.17, this parameter and the length uniquely determine palindrome). If there is one letter c in the middle, the palindromes of length 3 and 5 are in the first group, while the palindromes of length 7 and more are in the second group. If there are two letters c in the middle, the palindromes of length 2 and 4 are in the first group, while the palindromes of length 6 and more are in the fifth group (for $t = 2$, s ranges from 2 to $\lfloor \frac{l+1}{2} \rfloor$, and thus the length of observed polynomials takes all the even values from 6 to $l+3$ or $l+2$, depending on the parity of l). If there are three letters c in the middle, the palindrome of length 3 is in the first group, the palindrome of length 5 is in the fourth group, while the palindromes of length 7 and more are in the fifth group (for $t = 3$, s ranges from 2 to the largest value meeting the requirement $3 \leqslant l + 3 - 2s$, which is $\lfloor \frac{l}{2} \rfloor$, and thus the length of observed polynomials takes all the odd values from 7 to $l+3$ or $l+2$). Continuing in this manner, we enumerate all the considered polynomials, and prove the claim.

Therefore, there are $|\{v \in \operatorname{Pal}(\operatorname{hpw}(w)) \setminus \operatorname{Pal}(w) : 2 \leqslant |v| \leqslant l+3\}|$ factors in the left column. Since for each pair $\{u, \tilde{u}\}$ of factors of $\operatorname{hpw}(w)$ of length $l+3$ only one representative is included in the left column, we have that the number of factors of $\operatorname{hpw}(w)$ of length $l+3$ equals the number of palindromic factors of $\operatorname{hpw}(w)$ of length $l + 3$ plus twice the number of non-palindromic factors in the left column. In short:

$$
\begin{aligned}
C_{\operatorname{hpw}(w)}&(l+3) \\
&= P_{\operatorname{hpw}(w)}(l+3) \\
&\quad + 2\big(|\{v \in \operatorname{Pal}(\operatorname{hpw}(w)) \setminus \operatorname{Pal}(w) : 2 \leqslant |v| \leqslant l+3\}| - P_{\operatorname{hpw}(w)}(l+3)\big) \\
&= 2|\{v \in \operatorname{Pal}(\operatorname{hpw}(w)) \setminus \operatorname{Pal}(w) : 2 \leqslant |v| \leqslant l+3\}| - P_{\operatorname{hpw}(w)}(l+3) \\
&= 2|\{v \in \operatorname{Pal}(\operatorname{hpw}(w)) \setminus \operatorname{Pal}(w) : |v| \leqslant l+3\} \setminus \{c\}| - P_{\operatorname{hpw}(w)}(l+3) \\
&= 2|\{v \in \operatorname{Pal}(\operatorname{hpw}(w)) \setminus \operatorname{Pal}(w) : |v| \leqslant l+3\}| - P_{\operatorname{hpw}(w)}(l+3) - 2,
\end{aligned}
$$

which was to be proved. ∎

Proof of Theorem 3.16. We have:

$$
\sum_{n=0}^{\infty} T_{\mathrm{hpw}(w)}(n) \overset{\mathrm{L3.19}}{=} \sum_{n=0}^{l+2} T_{\mathrm{hpw}(w)}(n)
$$

$$
= \sum_{n=0}^{l+2} \left(C_{\mathrm{hpw}(w)}(n+1) - C_{\mathrm{hpw}(w)}(n) + 2 - P_{\mathrm{hpw}(w)}(n) - P_{\mathrm{hpw}(w)}(n+1) \right)
$$

$$
= \sum_{n=0}^{l+2} C_{\mathrm{hpw}(w)}(n+1) - \sum_{n=0}^{l+2} C_{\mathrm{hpw}(w)}(n) + 2(l+3)
$$

$$
- \sum_{n=0}^{l+2} P_{\mathrm{hpw}(w)}(n) - \sum_{n=0}^{l+2} P_{\mathrm{hpw}(w)}(n+1)
$$

$$
= C_{\mathrm{hpw}(w)}(l+3) - C_{\mathrm{hpw}(w)}(0) + 2(l+3)
$$

$$
- 2 \sum_{n=0}^{l+3} P_{\mathrm{hpw}(w)}(n) + P_{\mathrm{hpw}(w)}(0) + P_{\mathrm{hpw}(w)}(l+3)
$$

$$
= C_{\mathrm{hpw}(w)}(l+3) - 1 + 2(l+3) - 2 \sum_{n=0}^{l+3} P_{\mathrm{hpw}(w)}(n) + 1 + P_{\mathrm{hpw}(w)}(l+3)
$$

$$
\overset{\mathrm{L3.20}}{=} 2|\{v \in \mathrm{Pal}(\mathrm{hpw}(w)) \setminus \mathrm{Pal}(w) : |v| \leqslant l+3\}| - P_{\mathrm{hpw}(w)}(l+3) - 2
$$

$$
+ 2l + 6 - 2 \sum_{n=0}^{l+3} P_{\mathrm{hpw}(w)}(n) + P_{\mathrm{hpw}(w)}(l+3)
$$

$$
= 2|\{v \in \mathrm{Pal}(\mathrm{hpw}(w)) \setminus \mathrm{Pal}(w) : |v| \leqslant l+3\}| + 2l + 4 - 2 \sum_{n=0}^{l+3} P_{\mathrm{hpw}(w)}(n)
$$

$$
= 2|\{v \in \mathrm{Pal}(\mathrm{hpw}(w)) \setminus \mathrm{Pal}(w) : |v| \leqslant l+3\}| + 2l + 4
$$

$$
- 2|\{v \in \mathrm{Pal}(\mathrm{hpw}(w)) : |v| \leqslant l+3\}|
$$

$$
= 2l + 4 - 2|\{v \in \mathrm{Pal}(w) : |v| \leqslant l+3\}| = 2l + 4 - 2|\mathrm{Pal}(w)|
$$

$$
= 2(D(w) + 1) \overset{\mathrm{T3.15}}{=} 2D(\mathrm{hpw}(w)),
$$

which was to be proved. ∎

3.4 Longest palindromic suffixes of factors of a highly potential word

We hereby show that actually all the highly potential words are counterexamples to Theorem 1.35. Since each highly potential word has the set of factors closed under reversal, contains infinitely many palindromes and is of a finite defect, it is enough to prove:

Theorem 3.21. *Each highly potential word* $\mathrm{hpw}(w)$ *contains arbitrarily long factors* v *such that the longest palindromic suffix of* v *occurs in* v *more than once.*

Proof. For each $i \geqslant 2$ the word $wc = w_0 c$ is a prefix of the word $w_{i-1} = \widetilde{w_{i-1}}$. Therefore, $c^i wc \in \mathrm{Fact}(w_i) \subseteq \mathrm{Fact}(\mathrm{hpw}(w))$. Since the letter c does not occur in the word w, and $w \neq \widetilde{w}$, the longest palindromic suffix of the word $c^i wc$ is clearly only the letter c, having $i + 1$ occurrences in $c^i wc$. ∎

3.5 Highly potential word fixed by a morphism

As mentioned in the Preface, Brlek and Reutenauer showed that, under the conjecture that there does not exists an aperiodic word of a finite positive defect that is a fixed point of a non-identical morphism, Conjecture 1.31 holds for all fixed points of non-identical morphisms. However, in this section we construct a highly potential word that is a fixed point of a non-indentical morphism, thus showing that the conjecture assumed by Brlek and Reutenauer is false.

Theorem 3.22. *Let* $\Sigma = \{a, b, c\}$, *let the morphism* φ *be defined by* $\varphi(a) = abcbac$, $\varphi(b) = \varepsilon$, $\varphi(c) = c$, *and let* $w = ab$. *It then holds:*

$$\varphi(\mathrm{hpw}(w)) = \mathrm{hpw}(w).$$

Proof. It is enough to prove that for each $i \in \mathbb{N}_0$ the word $\varphi(w_i)$ is a prefix of $\mathrm{hpw}(w)$. By induction on i, we shall prove that for each $i \in \mathbb{N}_0$ it holds that $\varphi(w_i) = w_{i+1}c$ (and since this is a prefix of w_{i+2} and therefore also a prefix of $\mathrm{hpw}(w)$, the proof would thus be completed). We have:

$$w_0 = ab;$$

67

$$w_1 = abcba;$$

$$w_2 = abcbaccabcba.$$

For $i = 0$ it holds:

$$\varphi(w_0) = \varphi(ab) = abcbac = w_1c.$$

For $i = 1$ it holds:

$$\varphi(w_1) = \varphi(abcba) = abcbac\,c\,abcbac = w_2c.$$

For $i \geqslant 2$ it holds:

$$\varphi(w_i) = \varphi(w_{i-1}c^i\widetilde{w_{i-1}}) = \varphi(w_{i-1}c^iw_{i-1})$$
$$= \varphi(w_{i-1})\varphi(c)^i\varphi(w_{i-1})$$
$$= w_ic\,c^i\,w_ic = w_ic^{i+1}w_ic = w_{i+1}c,$$

which was to be proved. ∎

Remark. The highly potential word considered in the previous theorem (which is the essentially unique highly potential word generated by a word of length 2) is also fixed by a *non-erasing* morphism φ (that is: a morphism which maps none of the letters to ε) defined by $\varphi(a) = \varphi(b) = abcbacc$ and $\varphi(c) = c$. Let us prove this.

We again prove that for each $i \in \mathbb{N}_0$ the word $\varphi(w_i)$ is a prefix of $\mathrm{hpw}(w)$, in particular, $\varphi(w_i) = w_{i+2}cc$ (which is a prefix of w_{i+3} and therefore also a prefix of $\mathrm{hpw}(w)$). We have:

$$w_0 = ab;$$

$$w_1 = abcba;$$

$$w_2 = abcbaccabcba;$$

$$w_3 = abcbaccabcbacccabcbaccabcba;\,.$$

For $i = 0$ it holds:

$$\varphi(w_0) = \varphi(ab) = \varphi(a)\varphi(b)abcbacc\,abcbacc = w_2cc.$$

For $i = 1$ it holds:

$$\varphi(w_1) = \varphi(abcba) = abcbacc\,abcbacc\,c\,abcbacc\,abcbacc = w_3c.$$

For $i \geqslant 2$ it holds:

$$\varphi(w_i) = \varphi(w_{i-1}c^i\widetilde{w_{i-1}}) = \varphi(w_{i-1}c^i w_{i-1})$$
$$= \varphi(w_{i-1})\varphi(c)^i\varphi(w_{i-1})$$
$$= w_{i+1}cc\,c^i\,w_{i+1}cc = w_{i+1}c^{i+2}w_{i+1}cc = w_{i+2}cc,$$

which was to be proved.

4

Conclusion

We believe that the results of this thesis offer a better insight into many notions related to problems on palindromes in finite and infinite words, especially the MP-ratio and the palindromic defect.

Concerning the MP-ratio, our negative answers to the very plausible-looking questions show that many things around here are probably much deeper than they seem to be, and that, perhaps, there are many other surprises just waiting to be discovered.

Concerning the palindromic defect and related notions, our main result here is the construction of the so-called highly potential words. The fact that they are all aperiodic infinite words of a finite positive defect, having the set of factors closed under reversal, is already quite interesting—since in some recent works the construction of even a single word having these properties turned out to be quite hard. In fact, our construction provides a method to obtain such a word from *any* non-palindromic finite word. However, it turns out that this is just a tip of the iceberg, and that highly potential words seem to be a very fruitful supply of counterexamples regarding various problems on words (though, of course, they cannot be counterexamples to just about any assertion on words—one of our results shows that highly potential words do satisfy the Brlek-Reutenauer conjecture; this result, however, has its own significance, both in terms of the techniques used in the proof, as well as in terms of the fact that the Brlek-Reutenauer conjecture resisted another challenge, not a benign one). We sincerely hope that highly potential words will be established as an often visited "playground" regarding various problems

on words (possibly even those that have no connections to the palindromic defect, possibly even no connections to palindromes at all), that is, that many researches will, before stating a conjecture, find it useful to check it for highly potential words.

Prošireni izvod

Kombinatorika na rečima je grana matematike s vrlo širokim spektrom primena. Nešto slično može se reći i za palindrome, tj. reči koje se „isto čitaju" sleva nadesno i zdesna nalevo. Naime, oni imaju veoma važnu ulogu u izučavanju tzv. *Šturmovih nizova* [22, 16], a koji dalje nalaze primenu u teoriji brojeva, optimizaciji putnih mreža, kompjuterskoj grafici i obradi slika, prepoznavanju obrazaca i nadalje [1, Chapter 9]. Palindromi nalaze primenu i u mnogim oblastima s kojima naizgled nemaju nikakve veze, kao što su kvantna fizika [18, 2, 14], molekularna biologija [21, 20] [23, Chapter 4] i odnedavno čak i teorija muzike [24, 13, 11].

Dakle, bolji uvid u ponašanje palindroma od sve je većeg značaja. Jedan od pravaca aktuelnih istraživanja tiče se upostavljanja kriterijuma koja od dve date reči (ne obavezno palindroma) jeste „palindromičnija" od druge, tj. određivanje stepena „palindromičnosti" date reči. Jasno, mogu se zamisliti različiti pristupi, u zavisnosti od interpretacije pojma „biti palindromičniji". Istraživanje koje je tema disertacije izdvaja dva aktuelna pristupa ovom problemu, i odgovara na više otvorenih pitanja u vezi s njima.

Holub and Saari [19] razmatrali su sledeći pristup. Ograničavajući se na binarne reči, primetili su da svaka reč w sadrži palindromsku podreč dužine bar $\left\lceil \frac{|w|}{2} \right\rceil$: podreč koja se sastoji od zastupljenijeg slova. Na osnovu te konstatacije, reč w koje ne sadrži palindromske podreči duže od $\left\lceil \frac{|w|}{2} \right\rceil$ nazvali su *minimalno palindromična*: intuitivno, ove reči su najmanje palindromične. Stepen „palindromičnosti" potom se određuje putem tzv. *MP-razmere*, koja

je definisana sa sledećom koncepcijom na umu: reč je utoliko palindromičnija ukoliko je teže proširiti je do minimalo palindromične reči (stroga definicija biće data u nastavku). Na kraju rada, Holub i Saari postavili su nekoliko pitanja, intuitivno uverljivih, koja bi, u slučaju pozitivnog razrešenja, znatno pojednostavila izračunavanje MP-razmere.

Drugi pristup određivanju stepena „palindromičnosti" rečî (ne nužno binarnih) zasniva se na pojmu tzv. *palindromskog defekta*. Naim, prema rezultatu Droubaya, Justina i Pirilla [15], broj palindromskih faktora date reči najviše je za jedan veći od dužine te reči. Na osnovu ove nejednakosti uvodi se pojam (palindromskog) defekta kao razlika ove dve vrednosti (dakle, defekt date reči uvek je nenegativan). U skladu s ovakvim pristupom, najpalindromičnijima se smatraju one reči čiji je defekt jednak 0. Definicija defekta prirodno se proširuje i na beskonačne reči. Na beskonačnim rečima čiji je skup faktora zatvoren za preokretanje (engl. 'reversal'), Brlek i Reutenauer [9] definisali su funkciju unekoliko srodnu palindromskom defektu, zasnivajući je na nejednakosti koja povezuje tzv. *složenost* i tzv. *palindromsku složenost* reči, dokazanoj u [3]. Tom prilikom postavili su hipotezu koja predviđa određenu jednakost koja povezuje defekt, palindromsku složenost i (faktorsku) složenost beskonačne reči w, pod pretpostavkom da je skup faktora reči w zatvoren za preokretanje.

Brlek i Reutenauer dokazali su svoju hipotezu za periodične reči, i konstatovali su da, na osnovu nekih ranijih rezultata, hipoteza takođe važi i za reči defekta 0. Dalje su ispitali hipotezu za neke reči beskonačnog defekta, i ispostavilo se da hipoteza u ovim slučajevima važi. Prilikom pokušaja provere hipoteze za aperiodične beskonačne reči konačnog pozitivnog defekta uočen je sledeći problem: ispostavilo se da je prilično teško konstruisati beskonačne aperiodične reči čiji je skup faktora zatvoren za preokretanje, i koje imaju konačan pozitivan defekt – Brlek i Reutenauer navode da nisu uspeli konstruisati niti jedan primer.

Beskonačne aperiodične reči konačnog pozitivnog defekta predmet su i ranije hipoteze Blondin-Masséa i koautorâ [7]: hipoteza predviđa da ne postoji takva reč koja je fiksna tačka nekog primitivnog morfizma. Pod nešto jačom pretpostavkom – da ne postoji beskonačna aperiodična reč koja ima konačan pozitivan defekt i koja je fiksna tačka ma kakvog neidentičkog morfizma – Brlek i Reutenauer, u već pomenutom radu, pokazali su da njihova hipoteza važi i za fiksne tačke neidentičkih morfizama. Pretpostavljeno pojačanje hipoteze Blondin-Masséa i koautorâ ostalo je otvoreno pitanje.

Balková, Pelantová i Starosta [4] dokazali su hipotezu Brleka i Reutena-

uera za uniformno rekurentne reči. Osim ovog dokaza, izneli su i nekoliko srodnih teorema, za jednu od kojih se ispostavlja da je nekorektna.

Cilj ove disertacije je pružanje boljeg uvida u ove teme. Teza je organizovana na sledeći način.

U glavi 1 navodimo notaciju i neophodne definicije, kao i sve prethodne rezultate na koje se istraživanje iz disertacije neposredno nadovezuje. Za svaki od njih navedena je referenca, a za najveći deo ponuđen je i dokaz. Glava je podeljena na tri odeljka, pri čemu je prvi od njih uopštenog tipa, drugi uvodi MP-razmeru i srodne pojmove, a treći uvodi palindromski defekt i srodne pojmove.

Glave 2 i 3 predstavljaju u potpunosti originalan doprinos.

Glava 2 posvećena je MP-razmeri i srodnim konceptima. Glavni rezultat ove glave čine odgovori na tri pitanja koja su postavili Holub i Saari, kao i još jedno pitanje sličnog tipa. Pomalo iznenađujuće, odgovori na sva četiri pitanja su negativni. Rezultati ove glave objavljeni su u radu [6].

U glavi 3 izučavaju se palindromski defekt i srodni koncepti. Glava je podeljena na pet odeljaka. U odeljku 3.1 iznosimo konstrukciju kontraprimera za ranije pomenutu teoremu Balkove, Pelantove i Staroste. Rezultati ovog odeljka objavljeni su u radu [5].

Dalje, uvodimo konstrukciju klase reči povezanih sa svim problemima pominjanim iznad. Sama konstrukcija definisana je u odeljku 3.2. Kako deluje da ove reči imaju visok potencijal da predstavljaju primere i kontraprimere u vezi s raznim problemima na rečima, nazvali smo ih *visokopotencijalne reči*. Ustanovljavamo da svaka visokopotencijalna reč ima skup faktora zatvoren za preokretanje, da je aperiodična, rekurentna, ali nije uniformno rekurentna. Dokazujemo da svaka visokopotencijalna reč ima konačan pozitivan defekt.

U odeljku 3.3 dokazujemo da hipoteza Brleka i Reutenauera zaista važi za visokopotencijalne reči. Primetimo da, s obzirom na to što visokopotencijalne reči nisu uniformno rekurentne, ovaj rezultat ne sledi iz rezultata Balkove, Pelantove i Staroste.

U odeljku 3.4 pokazujemo da su visokopotencijalne reči kontraprimer za teoremu Balkove, Pelantove i Staroste. Kako je u odeljku 3.1 prezentovan samo jedan primer, koji ostavlja pomalo patološki utisak, doprinos ovog odeljka je činjenica da postoji još kontraprimera, koji čine familiju manje veštačkog izgleda.

U odeljku 3.5 konstruišemo visokopotencijalnu reč koja je fiksna tačka neidentičkog morfizma. Kako su visokopotencijalne reči aperiodične reči

konačnog pozitivnog defekta, ova konstrukcija obara Brlekovo i Reutenauerovo pojačanje hipoteze Blondin-Masséa i koautorâ.

Prikažimo sada detaljnije sadržaj svake glave rada. Navodimo većinu dokazanih tvrđenja, a za najbitnija među njima dajemo i glavnu ideju dokaza.

1 Uvod

Započinjemo notacijom i neophodnim definicijama.

Elemente zadatog skupa Σ, nazvanog *alfabet*, zovemo *slova*, a konačne, odnosno beskonačne, nizove slova nazivamo *reči*, odnosno *beskonačne reči*, respektivno. Neka Σ^* označava skup svih konačnih reči nad alfabetom Σ, i neka Σ^∞ označava skup svih konačnih ili beskonačnih reči nad alfabetom Σ. Za reči $w = a_1 a_2 \ldots a_n$ i $u = b_1 b_2 \ldots b_m$ (gde su $a_1, \ldots, a_n, b_1, \ldots, b_m \in \Sigma$), zapis wu označava konkatenaciju (nadovezivanje) reči w i u, to jest, $wu = a_1 a_2 \ldots a_n b_1 b_2 \ldots b_m$. Za zadatu reč w i $k \in \mathbb{N}_0$ (gde $\mathbb{N}_0$ označava skup nenegativnih celih brojeva), sa w^k označavamo reč $\underbrace{ww \ldots w}_{k \text{ puta}}$ (koju zovemo k-*ti stepen* reči w), a sa w^∞ označavamo beskonačnu reč $www \ldots$.

Sa a^*, gde je a slovo, označavamo skup $\{a^k : k \geqslant 0\}$, a sa $a^* b^*$, gde je i b slovo, označavamo skup $\{a^k b^l : k, l \geqslant 0\}$.

Dužinu reči w označavamo sa $|w|$. Zapisom $|w|_a$, gde je a slovo, označavamo ukupan broj pojavljivanja slova a u reči w. Jedinstvenu reč dužine jednake 0, koju nazivamo *prazna reč*, označavamo sa ε.

Definicija 1.1. Definišemo sledeće osnovne odnose među rečima:

- Reč $v = a_1 a_2 \ldots a_n$ je *podreč* reči w ako postoje reči $u_1, u_2, \ldots, u_{n+1}$ takve da važi $w = u_1 a_1 u_2 a_2 \ldots u_n a_n u_{n+1}$.

- Reč $v \in \Sigma^*$ je *sufiks* reči $w \in \Sigma^*$ ako postoji reč $u \in \Sigma^*$ takva da važi $w = uv$.

- Reč $v \in \Sigma^*$ je *prefiks* reči $w \in \Sigma^\infty$ ako postoji reč $u \in \Sigma^\infty$ takva da važi $w = vu$.

- Reč $v \in \Sigma^*$ je *faktor* reči $w \in \Sigma^\infty$ ako postoje reči $u_1 \in \Sigma^*$, $u_2 \in \Sigma^\infty$ takve da važi $w = u_1 v u_2$. Skup svih faktora reči w označavamo sa $\text{Fact}(w)$.

Napomena. Neki autori pod pojmom *podreč* podrazumevaju ono što mi ovde zovemo faktor *faktor*, a ovdašnji pojam podreči nazivaju *raštrkana podreč*. Mi ćemo se držati gornje konvencije.

Definicija 1.2. Preslikavanje $\tilde{}: \Sigma^* \to \Sigma^*$, koje zovemo *preokretanje*, definisano je na sledeći način: za $w = a_1 a_2 \ldots a_n$, gde su $a_1, a_2, \ldots, a_n \in \Sigma$, važi $\widetilde{w} = a_n a_{n-1} \ldots a_1$.

Kažemo da je skup faktora reči w *zatvoren za preokretanje* ako za sve $v \in \mathrm{Fact}(w)$ važi $\tilde{v} \in \mathrm{Fact}(w)$.

Definicija 1.3. Reč $w \in \Sigma^*$ je *palindrom* ako je $w = \widetilde{w}$. Skup svih palindromskih faktora reči $w \in \Sigma^\infty$ označavamo sa $\mathrm{Pal}(w)$.

Definicija 1.4. Beskonačna reč w je:

- *periodična* ako je oblika v^∞ za neko $v \in \Sigma^*$;

- *aperiodična* ako nije periodična;

- *rekurentna* ako se svaki faktor reči w pojavljuje beskonačno mnogo puta u reči w;

- *uniformno rekurentna* ako je rekurentna i, za svaki njen faktor, skokovi između uzastopnih pojavljivanja tog faktora u w su ograničeni (pod *skokom* podrazumevamo razliku između dve pozicije na kojima počinju dva uzastopna pojavljivanja posmatranog faktora).

Naredne poznate teoreme (videti npr. [17, Proposition 2.11], [1, Theorem 10.9.4] i [1, Example 10.9.1], redom) pokazaće se korisnima.

Teorema 1.5. *Za datu beskonačnu reč w, ako je* $\mathrm{Fact}(w)$ *zatvoreno za preokretanje, tada je reč w rekurentna.*

Teorema 1.6. *Beskonačna reč w je uniformno rekurentna akko, za sve $u \in \mathrm{Fact}(w)$, postoji $n \in \mathbb{N}$ takvo da $u \in \mathrm{Fact}(v)$ za sve $v \in \mathrm{Fact}(w)$ za koje važi $|v| = n$.*

Napomena. Osobina iz prethodne teoreme ponekad se koristi kao definicija uniformno rekurentne reči.

Teorema 1.7. *Ako je beskonačna reč periodična, tada je uniformno rekurentna.*

Definicija 1.8. Funkcija $\varphi : \Sigma^* \to \Sigma^*$ naziva se *morfizam* ako, za sve $w, v \in \Sigma^*$, važi $\varphi(wv) = \varphi(w)\varphi(v)$.

Jasno, morfizam je jednoznačno određen slikama slova, pa sledi da je bilo koji uočen morfizam moguće proširiti i na beskonačne reči, na prirodan način. Kažemo da je reč $w \in \Sigma^\infty$ *fiksna tačka* morfizma φ ako je $\varphi(w) = w$.

1.1 MP-razmera

U ovom odeljku razmatramo samo binarne reči, pa zato fiksiramo alfabet $\Sigma = \{0, 1\}$.

Jasno, svaka reč $w \in \{0, 1\}^*$ sadrži palindromsku podreč dužine bar $\left\lceil \frac{|w|}{2} \right\rceil$: podreč koja se sastoji od zastupljenijeg slova. Ovo je motivacija za sledeće definicije.

Definicija 1.9. Kažemo da je reč $w \in \{0, 1\}^*$ *minimalno palindromična* ako ne sadrži palindromske podreči dužine veće od $\left\lceil \frac{|w|}{2} \right\rceil$.

Definicija 1.10. Za reč $w \in \{0, 1\}^*$, uređen par (r, s), gde su $r, s \in \{0, 1\}^*$, takav da je reč rws minimalno palindromična, nazivamo *MP-proširenje* reči w. Ako je dužina $|r| + |s|$ najmanja moguća, tada par (r, s) nazivamo *najkraće MP-proširenje* ili *SMP-proširenje* (od engl. 'shortest MP-extension'), a razmeru $\frac{|rws|}{|w|}$ nazivamo *MP-razmera* reči w.

Stepen „palindromičnosti" reči w određujemo MP-razmerom. Naredna teorema [19, Theorem 4] postavlja gornje ograničenje MP-razmere, ujedno rešavajući pitanje egzistencije (S)MP-proširenja date reči w.

Teorema 1.11. *Za svaku reč $w \in \{0, 1\}^*$, MP-razmera reči w manja je od ili jednaka sa 4.*

Ideja dokaza. Za datu reč $w \in \{0, 1\}^*$, pokazujemo da je reč

$$v = 0^{|w| + |w|_1} w 1^{|w| + |w|_0}$$

minimalno palindromična. ∎

Ispostavlja se da je konstanta 4 u prethodnoj teoremi najbolja moguća, u asimptotskom smislu. Naime, ako $R(n)$ označava maksimalnu MP-razmeru među svim binarnim rečima zadate dužine n, pokazujemo sledeću teoremu [19, Theorem 5].

Teorema 1.12. *Važi:*
$$\lim_{n\to\infty} R(n) = 4.$$

Dokazu prethodi niz lema. Glavno sredstvo u dokazu je pojam tzv. *ekonomičnih reči.*

Lema 1.13. *Za svaku minimalno palindromičnu reč w, jedna od vrednosti $|w|_0$, $|w|_1$ iznosi $\left\lfloor \frac{|w|}{2} \right\rfloor$, a druga iznosi $\left\lceil \frac{|w|}{2} \right\rceil$.*

Definicija 1.14. Za reč $w \in \{0,1\}^*$ kažemo da je:

- *k-ekonomična*, za dato $k \in \mathbb{N}_0$, ako je w palindrom i reč $w1^k$ ima palindromsku podreč dužine $\geqslant |w|_1 + k + 2$;

- *ekonomična* ako je k-ekonomična za sve $0 \leqslant k \leqslant |w|_1$.

Lema 1.15. *Za svako MP-proširenje (r, s) ekonomične reči w, važi $|rs|_1 > |w|_1$.*

Prethodna lema, uz lemu 1.13, implicira sledeću nejednakost:

$$|rws| = |rws|_0 + |rws|_1 \geqslant 2|rws|_1 - 1 = 2|w|_1 + 2|rs|_1 - 1 > 4|w|_1.$$

Ova nejednakost donekle razjašnjava ulogu ekonomičnih reči. Naime, MP-razmera ekonomičnih reči koje imaju „mnogo" slova 1 bliska je broju 4. U nastavku sledi konstrukcija niza ekonomičnih reči koje imaju „mnogo" slova 1, što na kraju vodi do dokaza teoreme 1.12.

Lema 1.16. *Neka je w_0 ekonomična reč, i definišimo, za $i \in \mathbb{N}_0$,*

$$w_{i+1} = w_i 1^{t_i} w_i,$$

gde niz $t_0, t_1, t_2 \ldots$ nenegativnih celih brojeva zadovoljava $t_i < |w_i|_0$ za sve $i \in \mathbb{N}_0$. Tada je, za sve $i \in \mathbb{N}_0$, reč w_i ekonomična.

Lema 1.17. *Neka je $w(t_0, t_1, \ldots, t_{j-1})$ reč w_j definisana u formulaciji leme 1.16, gde su brojevi $t_0, t_1, \ldots, t_{j-1}$ zadati i zadovoljavaju $2^i \leqslant t_i < 2^{i+2}$ za sve $0 \leqslant i \leqslant j-1$, a početna reč je $w_0 = 0000$. Tada je reč $w(t_0, t_1, \ldots, t_{j-1})$ ekonomična.*

Lema 1.18. *Za svako dovoljno veliko k, postoji reč v_k koja ispunjava $|v_k| = k$ i koja je oblika $w(t_0, t_1, \ldots, t_{j-1})$ za neke $t_0, t_1, \ldots, t_{j-1}$ (gde važe uslovi leme 1.17).*

Sada pokazujemo da reči v_k dobijene u prethodnoj lemi zaista sadrže „mnogo" slova 1. Naime, imamo sledeću lemu.

Lema 1.19. *Za reči v_k dobijene u prethodnoj lemi važi:*

$$\lim_{k \to \infty} \frac{|v_k|_1}{|v_k|} = 1.$$

Najzad, u mogućnosti smo da dokažemo teoremu 1.12.

Ideja dokaza teoreme 1.12. Direktno pokazujemo da, za svako $\varepsilon > 0$, postoji $k_0 \in \mathbb{N}$ takvo da za sve $k \geqslant k_0$ važi $\frac{|rv_ks|}{|v_k|} > 4 - \varepsilon$, gde je (r, s) SMP-proširenje reči v_k. ∎

Holub i Saari postavili su sledeća pitanja o MP-proširenjima:

Pitanje 1.20. *Posmatrajmo sve binarne reči date dužine n. Da li one među njima koje dostižu najveću moguću MP-razmeru moraju biti palindromi?*

Pitanje 1.21. *Da li za svaku binarnu reč postoji njeno SMP-proširenje (r, s) za koje je $r, s \in 0^* \cup 1^*$?*

Pitanje 1.22. *Da li za svaku binarnu reč postoji njeno SMP-proširenje (r, s) za koje je $r, s \in 0^*1^* \cup 1^*0^*$?*

Ovim pitanjima pridružujemo još jedno srodno.

Pitanje 1.23. *Da li za svaku binarnu reč postoji njeno SMP-proširenje (r, s) takvo da r i s nemaju zajedničkih slova?*

Kažimo nekoliko reči o intuiciji iza ovih pitanja.

Jasno, najmanja moguća MP-razmera iznosi 1 i dostiže se upravo za minimalno palindromične reči, koje smatramo najmanje palindromičnima. Pitanje 1.20 razmatra reči na suprotnom kraju: kako njih smatramo najpalindromičnijima, očekivano je, kao što pitanje 1.20 predviđa, da one moraju biti palindromi. Međutim, u odeljku 2.1 pokazujemo da ovo nije slučaj.

Pitanja 1.21, 1.22 i 1.23 bave se mogućim oblicima SMP-proširenja. Pitanje 1.21 bazirano je na sledećem razmišljanju: kako pokušavamo izbeći palindromske podreči duže nego što je neophodno, deluje razumno pretpostaviti da su reči r i s što je moguće prostije, to jest, stepeni jednog slova; zaista, drugi oblici reči r i s proizveli bi više različitih podreči, i na taj način povećali šansu da se među njima javi palindrom. Pitanje 1.22 samo je slabija

verzija pitanja 1.21. Najzad, pitanje 1.23, možda i najuverljivije od svih, predviđa da možemo pretpostaviti da r i s nemaju zajedničkih slova, oslanjajući se na činjenicu da zajedničko slovo za r i s zapravo povećava dužinu najduže palindromske podreči polazne reči. Uprkos svemu, u odeljku 2.1 pokazujemo da ovi intutitivni zaključci nisu tačni. Primetimo da, iako je svaki kontraprimer za pitanje 1.22 ujedno i kontraprimer za pitanje 1.21, i štaviše, naš kontraprimer za pitanje 1.23 takođe je još jedan kontraprimer za pitanje 1.21 – ipak pitanje 1.21 razrešavamo zasebno. Razlog je u tome što, dok pitanja 1.22 i 1.23 razrešavamo samo sa po jednim kontraprimerom, za pitanje 1.21 navodimo beskonačnu familiju kontraprimera.

1.2 Palindromski defekt

Definicija 1.24. Neka je data beskonačna reč w.

- *Faktorska složenost* (ili samo *složenost*) reči w jeste funkcija $C_w : \mathbb{N}_0 \to \mathbb{N}_0$ definisana sa

$$C_w(n) = |\{v \in \text{Fact}(w) : |v| = n\}|.$$

- *Palindromska složenost* reči w jeste funkcija $P_w : \mathbb{N}_0 \to \mathbb{N}_0$ definisana sa

$$P_w(n) = |\{v \in \text{Pal}(w) : |v| = n\}|.$$

Sada navodimo nejednakost Droubaya, Justina i Pirilla [15, Proposition 2].

Teorema 1.25. *Za sve konačne reči w važi:*

$$|\text{Pal}(w)| \leqslant |w| + 1.$$

Ideja dokaza. Dokazujemo da za sve konačne reči $w = a_1 a_2 \ldots a_n$ važi:

$$|\text{Pal}(w)| \leqslant |\text{Pal}(a_1 a_2 \ldots a_{n-1})| + 1 \leqslant |\text{Pal}(a_1 a_2 \ldots a_{n-2})| + 2$$
$$\leqslant \cdots \leqslant |\text{Pal}(a_1)| + n - 1 \leqslant |\text{Pal}(\varepsilon)| + n = n + 1,$$

što je i trebalo dokazati. ∎

Ova nejednakost motivisala je Brleka i koautore [8] da uvedu sledeću definiciju:

Definicija 1.26. *Palindromski defekt* (ili samo *defekt*) konačne reči w jeste razlika

$$D(w) = |w| + 1 - |\mathrm{Pal}(w)|.$$

Naredna teorema i njena posledica [4, Corollary 2.3] ustanovljavaju važno svojstvo defekta.

Teorema 1.27. *Za sve $w \in \Sigma^*$ i $a \in \Sigma$, važi:*

$$D(wa) = \begin{cases} D(w), & \text{ako se najduži palindromski sufiks reči} \\ & \text{wa pojavljuje tačno jednom u wa;} \\ D(w)+1, & \text{inače;} \end{cases}$$

$$D(aw) = \begin{cases} D(w), & \text{ako se najduži palindromski prefiks reči} \\ & \text{aw pojavljuje tačno jednom u aw;} \\ D(w)+1, & \text{inače.} \end{cases}$$

Dakle, defekt $D(w)$ jednak je broju prefiksâ v reči w takvih da se najduži palindromski sufiks reči v pojavljuje u v više od jednom, i ujedno je jednak broju sufiksâ v reči w takvih da se najduži palindromski prefiks reči v pojavljuje u v više od jednom.

Posledica 1.28. *Neka je $w \in \Sigma^*$ i $v \in \mathrm{Fact}(w)$. Tada je $D(v) \leqslant D(w)$.*

Prethodna posledica daje motivaciju za sledeću definiciju defekta beskonačne reči w:

Definicija 1.29. Za beskonačnu reč w, definišemo njen defekt sa:

$$D(w) = \sup_{v \in \mathrm{Fact}(w)} D(v).$$

Jasno, ova jednakost važi i za konačne reči.

Još jednu bitnu nejednakost koja povezuje razmatrane pojmove dokazali su Baláži, Masáková i Pelantová [3, Theorem 1.2(ii)]:

Teorema 1.30. *Neka je w beskonačna reč za koju je $\mathrm{Fact}(w)$ zatvoreno za preokretanje. Za sve $n \in \mathbb{N}_0$ imamo*

$$P_w(n) + P_w(n+1) \leqslant C_w(n+1) - C_w(n) + 2.$$

Ideja dokaza. Definišemo usmeren graf G_n, pod nazivom *Rauzyjev graf* reči w (prema [25]), na sledeći način: skup čvorova grafa G_n je skup

$$V_n = \{v \in \text{Fact}(w) : |v| = n\},$$

skup grana grafa G_n je skup

$$E_n = \{v \in \text{Fact}(w) : |v| = n + 1\},$$

gde grana $e \in E_n$ počinje u čvoru $x \in V_n$ i završava se u čvoru $y \in V_n$ akko je x prefiks reči e a y sufiks reči e. Čvorove ovog grafa čiji je izlazni stepen bar 2 nazivamo *desno specijalnim*, a one čiji je ulazni stepen bar 2 nazivamo *levo specijalnim*. Kažemo da je čvor *specijalan* ako je desno ili levo specijalan (ili i jedno i drugo).

Dokaz razdvajamo na dva slučaja: kada G_n nema nijedan specijalan čvor, i kada ima. Ispostavlja se da je u prvom slučaju graf zapravo orijentisana kontura, i u tom slučaju dokazujemo da važi $C_w(n+1) - C_w(n) = 0$ i $P_w(n) + P_w(n+1) \leqslant 2$. U drugom slučaju svođenjem grafa na tzv. *redukovan Rauzyjev graf* dokazujemo sledeći niz (ne)jednakosti:

$$P_w(n) + P_w(n+1) \leqslant \sum_{\substack{x \in V_n \\ x \text{ je specijalan}}} \deg_+(x) - |\{x \in V_n : x \text{ je specijalan}\}| + 2$$

$$= \sum_{\substack{x \in V_n \\ x \text{ je specijalan}}} (\deg_+(x) - 1) + 2$$

$$= C_w(n+1) - C_w(n) + 2,$$

odakle imamo željeno tvrđenje. ∎

Najzad, formulišemo hipotezu Brleka i Reutenauera, komentarisanu na početku. Ona predviđa sledeću jednakost u kojoj se pojavljuje defekt $D(w)$ i funkcija $T_w : \mathbb{N}_0 \to \mathbb{N}_0$, inspirisana teoremom 1.30, definisana sa

$$T_w(n) = C_w(n+1) - C_w(n) + 2 - P_w(n) - P_w(n+1).$$

Hipoteza 1.31. *Neka je w beskonačna reč za koju je* $\text{Fact}(w)$ *zatvoreno za preokretanje. Važi:*

$$2D(w) = \sum_{n=0}^{\infty} T_w(n).$$

Brlek i Reutenauer pokazali su da hipoteza 1.31 važi za periodične reči, a na osnovu ranijih rezultata [17, Theorem 2.14] i [10, Theorem 1.1] utvrdili su da njihova hipoteza važi i za reči defekta 0. Dalje su ispitivali hipotezu za neke poznate beskonačne reči i klase beskonačnih reči koje imaju beskonačan defekt, naime: Thue-Morseovu reč, tzv. nizove u vezi sa savijanjem papira (engl. 'paperfolding sequences') i uopštene Rudin-Šapirove nizove. Ispostavilo se da hipoteza u ovim slučajevima važi.

Najzad, prilikom pokušaja da preduzmu sledeći logičan korak, tj. da testiraju hipotezu za aperiodične beskonačen reči konačnog pozitivnog defekta (bar za neke primere), ispostavilo se da je prilično teško pronaći primere aperiodičnih beskonačnih reči konačnog pozitivnog defekta takvih da im je skup faktora zatvoren za preokretanje: Brlek i Reutenauere nisu uspeli da pronađu niti jedan takav. Ovo je jedan od problema koje razmatramo u ovoj disertaciji.

Za kraj pominjemo jedno od najzanimljivijih tvrđenja koja bi mogla predstavljati bitne korake na putu ka dokazu hipoteze 1.31: [4, Corollary 5.10], gde se tvrdi da, za svaku beskonačnu reč w za koju je $\mathrm{Fact}(w)$ zatvoreno za preokretanje, ako je defekt $D(w)$ konačan, tada je suma $\sum_{n=0}^{\infty} T_w(n)$ takođe konačna. Međutim, ovo tvrđenje oslanja se na tehničku teoremu 1.35 niže [4, Theorem 5.7], koja je, kao što ćemo pokazati u odeljcima 3.1 i 3.4, zapravo netačna, pa dakle pomenuto tvrđenje ostaje otvoreno.

Teorema 1.35 (netačna). *Za svaku beskonačnu reč u za koju je* $\mathrm{Fact}(u)$ *zatvoreno za preokretanje i sadrži beskonačno mnogo palindroma, naredna tvrđenja su ekvivalentna:*

(a) *defekt reči u je konačan;*

(b) *postoji prirodan broj H takav da se najduži palindromski sufiks svakog faktora w reči u, dužine $|w| > H$, pojavljuje u reči w tačno jednom.*

2 MP-razmera

2.1 Odgovori na pitanja 1.20, 1.21, 1.22 i 1.23

Odeljak započinje lemom veoma korisnom za dalji rad, a korisnom i u opštijem kontekstu.

Lema 2.1. *Ako je (r, s) SMP-ekstenzija reči w i važi $|r| + |s| > 0$, tada je dužina $|rws|$ neparna.*

Potom slede glavni rezultati odeljka.

Teorema 2.2. *Odgovor na pitanje 1.20 je negativan.*

Ideja dokaza. Daje se kontraprimer za $n = 6$. Pokazujemo da maksimalna moguća MP-razmera reči dužine 6 iznosi $\frac{11}{6}$, i da je jedna od reči za koje se ona postiže

$$v = 010110,$$

što nije palindrom. ∎

Teorema 2.3. *Odgovor na pitanje 1.21 je negativan.*

Ideja dokaza. Pokazujemo da, za svako $k \geqslant 4$, jedino SMP-proširenje reči

$$v = 010^k 1010$$

jeste uređen par $(\varepsilon, u) = (\varepsilon, 01^{k+2})$. Na taj način dobijamo beskonačnu familiju kontraprimera za pitanje 1.21. ∎

Teorema 2.4. *Odgovor na pitanje 1.22 je negativan.*

Ideja dokaza. Pokazujemo da je jedino SMP-proširenje reči

$$v = 0010000010100111$$

uređen par $(\varepsilon, u) = (\varepsilon, 1011111)$. Na taj način dobijamo kontraprimer za pitanje 1.22. ∎

Teorema 2.5. *Odgovor na pitanje 1.23 je negativan.*

Ideja dokaza. Pokazujemo da je jedino SMP-proširenje reči

$$v = 01111101001$$

uređen par $(y, u) = (1, 1000000)$. Na taj način dobijamo kontraprimer za pitanje 1.23. ∎

3   Palindromski defekt

3.1   Kontraprimer za teoremu 1.35

U predloženom dokazu teoreme 1.35 samo se kratko tvrdi da navedena ekvivalencija sledi na osnovu definicije defekta. Zapravo, na osnovu definicije defekta i posledice 1.27, sledi da su tvrđenja (a) i

(b$_0$) postoji prirodan broj H takav da se najduži palindromski sufiks svakog
    prefiksa w reči u, dužine $|w| \geqslant H$, pojavljuje u reči w tačno jednom

ekvivalentna: smer ($\Leftarrow$) je jasan, dok smer ($\Rightarrow$) sledi iz konstatacije da, ako je v prefiks reči u takav da je $D(v) = D(u)$, tada svaki prefiks w reči u duži od v sadrži v kao prefiks, pa odatle sledi da se najduži palindromski sufiks reči w mora pojavljivati u reči w tačno jednom (budući da bi u suprotnom sledilo $D(w) \geqslant D(v) + 1 = D(u) + 1$, kontradikcija). Nažalost, slično rezonovanje ne može se primeniti za faktore na mestu prefiksâ, pa je predloženi dokaz nekorektan (važi samo smer (b) $\Rightarrow$ (a), jer imamo (b) $\Rightarrow$ (b$_0$) $\Rightarrow$ (a)).

U ovom odeljku konstruišemo beskonačnu reč u za koju (a) važi ali (b) ne. Neka je morfizam φ definisan sa: $\varphi(1) = 1213$, $\varphi(2) = \varepsilon$, $\varphi(3) = 23$, i neka je $u = \varphi^\infty(1)$.

Preko nekoliko pomoćnih tvrđenja dokazujemo da je reč u kontraprimer za teoremu 1.35. Ovde navodimo najznačajnija.

Tvrđenje 3.1. *Za sve $i \geqslant 1$ imamo*

$$\varphi^{i+1}(1) = \varphi^i(1)\,\varphi^i(1)\,23.$$

Tvrđenje 3.3. Fact(u) *je zatvoreno za preokretanje, i u sadrži beskonačno mnogo palindroma.*

Tvrđenje 3.4. *Reč u ne zadovoljava uslov (b).*

Tvrđenje 3.5. *Defekt reči u je konačan.*

Dakle, na osnovu tvrđenja 3.3, 3.5 i 3.4, reč u je kontraprimer za tvrđenje teoreme 1.35.

U nastavku odeljka ispitujemo da li reč u zadovoljava hipotezu Brleka i Reutenauera. Ispostavlja se da zadovoljava. Značaj ovog ispitivanja je dvojak: prvo, konstruisana reč u je prva aperiodična reč konačnog pozitivnog defekta, sa skupom faktora zatvorenim za preokretanje, što je upravo primer

koji je nedostajao Brleku i Reutenaueru prilikom ispitivanja validnosti njihove hipoteze; drugo, na tehnike koje se koriste u ovom ispitivanju vraćamo se i u narednim odeljcima, gde će one biti stavljene u znatno opštiji kontekst.

Tvrđenje 3.6. $D(u) = 1$.

Tvrđenje 3.12. *Za sve $n \geqslant 6$ važi:*

$$T_u(n) = 0.$$

Tvrđenje 3.13. *Reč u zadovoljava hipotezu Brleka i Reutenauera, to jest,*

$$2D(u) = \sum_{n=0}^{\infty} T_u(n).$$

3.2 Visokopotencijalne reči: konstrukcija i osnovne osobine

Neka je w konačna reč koja nije palindrom, i neka je c slovo koje se ne pojavljuje u w. Definišimo $w_0 = w$ i, za $i \in \mathbb{N}$,

$$w_i = w_{i-1} c^i \widetilde{w_{i-1}}.$$

Najzad, neka je

$$\mathrm{hpw}(w) = \lim_{i \to \infty} w_i.$$

Opravdanost gornjeg limesa je jasna budući da je svaka reč w_i prefiks reči w_{i+1}. Reč $\mathrm{hpw}(w)$ nazivamo *visokopotencijalna reč generisana sa w*.

Naredna propozicija ima jednostavan dokaz, ali je od ključne važnosti.

Propozicija 3.14. *Neka je $\mathrm{hpw}(w)$ visokopotencijalna reč. Važi:*

 a) $\mathrm{Fact}(\mathrm{hpw}(w))$ *je zatvoreno za preokretanje;*

 b) $\mathrm{hpw}(w)$ *je rekurentna;*

 c) $\mathrm{hpw}(w)$ *nije uniformno rekurentna;*

 d) $\mathrm{hpw}(w)$ *je aperiodična.*

Glavni rezultat ovog odeljka je sledeća teorema.

Teorema 3.15. *Za svaku visokopotencijalnu reč $\mathrm{hpw}(w)$ važi $D(\mathrm{hpw}(w)) = D(w) + 1$. Specijalno,*
$$0 < D(\mathrm{hpw}(w)) < \infty.$$

Ideja dokaza. Neka je $|w| = l$. Nabrajanjem svih palindromskih faktora reči w_1 koji nisu faktori reči w_0, izračunavamo

$$|\mathrm{Pal}(w_1)| = |\mathrm{Pal}(w_0)| + l + 1,$$

a odatle se onda može dobiti

$$D(w_1) = D(w_0) = D(w).$$

Slično, nabrajanjem svih palindromskih faktora reči w_2 koji nisu faktori reči w_1, izračunavamo

$$|\mathrm{Pal}(w_2)| = |\mathrm{Pal}(w_1)| + 2l + 2,$$

a odatle se onda može dobiti

$$D(w_2) = D(w_1) + 1 = D(w) + 1.$$

Nadalje se na sličan način pokazuje $D(w_i) = D(w_{i-1})$ za sve $i \geqslant 3$. Naime, ako je $|w_{i-2}| = m$, nabrajanjem palindromskih faktora reči w_i izračunavamo

$$|\mathrm{Pal}(w_i)| = |\mathrm{Pal}(w_{i-1})| + 2m + 2i - 1,$$

a odatle se onda može dobiti

$$D(w_i) = D(w_{i-1}) = D(w) + 1.$$

Najzad, na osnovu posledice 1.28, dobija se

$$\sup_{v \in \mathrm{Fact}(\mathrm{hpw}(w))} D(v) = \sup_{i \in \mathbb{N}_0} D(w_i),$$

to jest,

$$D(\mathrm{hpw}(w)) = \sup_{i \in \mathbb{N}_0} D(w_i) = D(w) + 1,$$

što je i trebalo dokazati. ∎

3.3 Hipoteza 1.31 za visokopotencijalne reči

U ovom odeljku pokazujemo da visokopotencijalne reči zadovoljavaju hipotezu Brleka i Reutenauera.

Teorema 3.16. *Za sve visokopotencijalne reči* $\mathrm{hpw}(w)$ *važi:*

$$2D(\mathrm{hpw}(w)) = \sum_{n=0}^{\infty} T_{\mathrm{hpw}(w)}(n).$$

Dokazu prethodi niz lema. Neka je $|w| = l$.

Lema 3.17. *Neka je dato $n \geqslant 1$. Svako $v \in \mathrm{Pal}(\mathrm{hpw}(w)) \setminus \mathrm{Pal}(w)$ takvo da je $|v| = n$ jednoznačno je određeno brojem uzastopnih pojavljivanja slova c u sredini palindroma v.*

Štaviše, slovo c može se uzastopno pojaviti tačno $k \geqslant 1$ puta u sredini palindroma v ako i samo ako je $k \leqslant n \leqslant (l+2) \cdot 2^k + k$ i $k \equiv n \pmod 2$.

Lema 3.18. *Neka je dato $n \geqslant l + 3$. Za sve $v \in \mathrm{Fact}(\mathrm{hpw}(w))$ takve da je $|v| = n$, ili postoji tačno jedno slovo d takvo da je $vd \in \mathrm{Fact}(\mathrm{hpw}(w))$, ili postoje tačno dva slova d_1, d_2 takva da $vd_1, vd_2 \in \mathrm{Fact}(\mathrm{hpw}(w))$.*

Štaviše, drugonavedeni slučaj važi ako i samo ako se v završava sa tačno k slova c, gde je $k \leqslant n \leqslant (l+2) \cdot 2^{k-1} + k - 1$.

Lema 3.19. *Za sve $n \geqslant l + 3$ važi:*

$$T_{\mathrm{hpw}(w)}(n) = 0.$$

Ideja dokaza. Svaka reč $v \in \mathrm{Pal}(\mathrm{hpw}(w))$ takva da je $|v| \geqslant l+3$ očito nije u skupu $\mathrm{Pal}(w)$. Neka je dato $n \geqslant l+3$, i neka je

$$A = \{k \geqslant 1 : k \leqslant n \leqslant (l+2) \cdot 2^k + k \wedge k \equiv n \pmod 2\},$$

$$B = \{k \geqslant 1 : k \leqslant n+1 \leqslant (l+2) \cdot 2^k + k \wedge k \equiv n+1 \pmod 2\},$$

$$C = \{k \geqslant 1 : k - 1 \leqslant n \leqslant (l+2) \cdot 2^k + k\}.$$

Može se pokazati da važi $A \cap B = \varnothing$ i $A \cup B = C$.

Prema lemi 3.17, imamo:

$$P_{\mathrm{hpw}(w)}(n) + P_{\mathrm{hpw}(w)}(n+1) = |A| + |B| = |C|.$$

Prema lemi 3.18, imamo:

$$C_{\mathrm{hpw}(w)}(n+1) - C_{\mathrm{hpw}(w)}(n) = |\{k \geqslant 0 : k \leqslant n \leqslant (l+2) \cdot 2^{k-1} + k - 1\}|.$$

Iz ovih jednakosti može se izračunati

$$T_{\mathrm{hpw}(w)}(n) = C_{\mathrm{hpw}(w)}(n+1) - C_{\mathrm{hpw}(w)}(n) + 2 - P_{\mathrm{hpw}(w)}(n) - P_{\mathrm{hpw}(w)}(n+1)$$
$$= 0,$$

što je i trebalo dokazati. ∎

Lema 3.20. *Važi:*

$$C_{\mathrm{hpw}(w)}(l+3) = 2|\{v \in \mathrm{Pal}(\mathrm{hpw}(w))\backslash\mathrm{Pal}(w) : |v| \leqslant l+3\}| - P_{\mathrm{hpw}(w)}(l+3) - 2.$$

Ideja dokaza. Uspostavljamo bijekciju između skupa faktora reči $\mathrm{hpw}(w)$ dužine $l + 3$, takvih da za svaki par faktora oblika $\{u, \tilde{u}\}$ uzimamo samo po jednog predstavnika, i skupa $\mathrm{Pal}(\mathrm{hpw}(w)) \setminus (\mathrm{Pal}(w) \cup \{c\})$. Bijekciju uspostavljamo na sledeći način: ispostavlja se da, za svaki od posmatranih faktora, recimo u, najduži palindromski prefiks ili sufiks, ali ne i jedno i drugo, pripada skupu $\mathrm{Pal}(\mathrm{hpw}(w))\backslash(\mathrm{Pal}(w)\cup\{c\})$. Na osnovu toga, možemo izračunati:

$$
\begin{aligned}
&C_{\mathrm{hpw}(w)}(l + 3) \\
&\quad = P_{\mathrm{hpw}(w)}(l + 3) \\
&\qquad + 2\big(|\{v \in \mathrm{Pal}(\mathrm{hpw}(w)) \setminus \mathrm{Pal}(w) : 2 \leqslant |v| \leqslant l + 3\}| - P_{\mathrm{hpw}(w)}(l + 3)\big) \\
&\quad = 2|\{v \in \mathrm{Pal}(\mathrm{hpw}(w)) \setminus \mathrm{Pal}(w) : |v| \leqslant l + 3\}| - P_{\mathrm{hpw}(w)}(l + 3) - 2,
\end{aligned}
$$

što je i trebalo dokazati. ∎

Ideja dokaza teoreme 3.16. Teorema se sada može pokazati direktnim računom, uz korišćenje dokazanih tvrđenja. ∎

3.4   Najduži palindromski sufiksi faktorâ visokopotencijalne reči

Ovde pokazujemo da su zapravo sve visokopotencijalne reči kontraprimeri za teoremu 1.35. Kako svaka visokopotencijalna reč ima skup faktora zatvoren za preokretanje, sadrži beskonačno mnogo palindroma i konačnog je defekta, dovoljno je pokazati:

Teorema 3.21. *Svaka visokopotencijalna reč* $\mathrm{hpw}(w)$ *sadrži proizvoljno dugačke faktore v takve da se najduži palindromski sufiks reči v pojavljuje u reči v više od jednom.*

Ideja dokaza. Za sve $i \geqslant 2$ reč $c^i wc$ faktor je reči $\mathrm{hpw}(w)$, a njen najduži palindromski sufiks je samo slovo c. ∎

3.5   Visokopotencijalna reč koja je fiksna tačka morfizma

Kao što je ranije pomenuto, Brlek i Reutenauer pokazali su da, pod pretpostavkom da ne postoji aperiodična reč konačnog pozitivnog defekta koja je fiksna tačka neidentičkog morfizma (pojačanje hipoteze Blondin-Masséa i koautorâ), hipoteza 1.31 važi za sve fiksne tačke neidentičkih morfizama. Međutim, u ovom odeljku konstruišemo visokopotencijalnu reč koja je fiksna tačka neidentičkog morfizma, na taj način pokazujući da je Brlekovo i Reutenauerovo pojačanje hipoteze Blondin-Masséa i koautorâ netačno.

Teorema 3.22. *Neka je* $\Sigma = \{a, b, c\}$, *neka je morfizam* φ *definisan sa* $\varphi(a) = abcbac$, $\varphi(b) = \varepsilon$, $\varphi(c) = c$, *i neka je* $w = ab$. *Tada važi:*

$$\varphi(\mathrm{hpw}(w)) = \mathrm{hpw}(w).$$

Napomena. Visokopotencijalna reč posmatrana u prethodnoj teoremi (koja je zapravo jedina visokopotencijalna reč generisana rečju dužine 2) takođe je fiksirana i *nebrišućim* morfizmom φ (tj.: morfizmom koji nijedno slovo ne preslikava u ε) definisanim sa $\varphi(a) = \varphi(b) = abcbacc$ i $\varphi(c) = c$.

3.6   Zaključak

Verujemo da rezultati ove disertacije pružaju bolji uvid u mnoge pojmove povezane sa problemima o palindromima u konačnim i beskonačnim rečima, naročito u vezi sa MP-razmerom i palindromskim defektom.

U vezi sa MP-razmerom, naši negativni odgovori na intuitivno vrlo uverljiva pitanja pokazuju da su mnoge stvari ovde verovatno mnogo dublje nego što izgledaju, i da, možda, postoje još mnoga iznenađenja koja će se tek otkriti.

U vazi s palindromskim defektom i srodnim pojmovima, naš glavni rezultat ovde je konstrukcija tzv. visokopotencijalnih reči. Činjenica da su sve one aperiodične beskonačne reči konačnog pozitivnog defekta, i da im je skup faktora zatvoren za preokretanje, već je prilično zanimljiva – budući da se u nekim skorašnjim radovima konstrukcija makar jedne reči s ovim osobinama pokazala kao prilično teška. Štaviše, naša konstrukcija daje metod kojim se takva reč dobija od *bilo koje* konačne reči koja nije palindrom. No, ispostavlja se da je ovo samo vrh ledenog brega, i da visokopotencijalne reči deluju kao vrlo korisna zaliha kontraprimera za razne probleme o rečima (iako, naravno, ne mogu biti kontraprimeri za baš sva tvrđenja o rečima –

jedan od naših rezultata pokazuje da visokopotencijalne reči zaista zadovoljavaju hipotezu Brleka i Reutenauera; ovaj rezultat, međutim, takođe je od značaja, kako što se tiče tehnika korišćenih u dokazu, tako i zbog činjenice da se hipoteza Brleka i Reutenauera oduprla još jednom izazovu, ne baš bezazlenom). Iskreno se nadamo da će se visokopotencijalne reči ustaliti kao često posećivano „igralište" povodom raznih problema na rečima (moguće čak i onih koji se ne bave palindromskim defektom, možda čak ni uopšte palindromima), tj. da će mnogi istraživači, pre nego što formulišu hipotezu, proveriti tu hipotezu za visokopotencijalne reči.

Bibliography

[1] J.-P. Allouche & J. Shallit, *Automatic Sequences. Theory, Applications, Generalizations,* Cambridge University Press, Cambridge, 2003.

[2] M. Baake, A note on palindromicity, *Lett. Math. Phys.* **49** (1999), 217–227.

[3] P. Baláži & Z. Masáková & E. Pelantová, Factor versus palindromic complexity of uniformly recurrent infinite words, *Theoret. Comput. Sci.* **380** (2007), 266–275.

[4] L. Balková & E. Pelantová & Š. Starosta, On Brlek-Reutenauer conjecture, *Theoret. Comput. Sci.* **412** (2011), 5649–5655.

[5] B. Bašić, A note on the paper "On Brlek-Reutenauer conjecture", *Theoret. Comput. Sci.*, doi:10.1016/j.tcs.2012.04.039.

[6] B. Bašić, Counter-intuitive answers to some questions concerning minimal-palindromic extensions of binary words, *Discrete Appl. Math.* **160** (2012), 181–186.

[7] A. Blondin-Massé & S. Brlek & A. Garon & S. Labbé, Combinatorial properties of f-palindromes in the Thue-Morse sequence, *Pure Math. Appl.* **19 (2-3)** (2008), 39–52.

[8] S. Brlek & S. Hamel & M. Nivat & C. Reutenauer, On the palindromic complexity of infinite words, *Internat. J. Found. Comput. Sci.* **15** (2004), 293–306.

[9] S. Brlek & C. Reutenauer, Complexity and palindromic defect of infinite words, *Theoret. Comput. Sci.* **412** (2011), 493–497.

[10] M. Bucci & A. de Luca & A. Glen & L.Q. Zamboni, A connection between palindromic and factor complexity using return words, *Adv. in Appl. Math.* **42** (2009), 60–74.

[11] N. Carey, On a class of locally symmetric sequences: The right infinite word Λ_θ, in C. Agon & M. Andreatta & G. Assayag & E. Amiot & J. Bresson & J. Mandereau (eds.), *Mathematics and Computation in Music. Proceedings of the Third International Conference (MCM 2011)*, Springer, Berlin, 2011, pp. 42–55.

[12] C. Choffrut & J. Karhumäki, Combinatorics of words, in G. Rozenberg & A. Salomaa (eds.), *Handbook of Formal Languages, Vol. 1. Word, Language, Grammar,* Springer, Berlin, 1997, pp. 329–438.

[13] D. Clampitt & T. Noll, Regions and standard modes, in E. Chew & A. Childs & C.-H. Chuan (eds.), *Mathematics and Computation in Music. Proceedings of the Second International Conference (MCM 2009)*, Springer, Berlin, 2009, pp. 81–92.

[14] D. Damanik & J.-M. Ghez & L. Raymond, A palindromic half-line criterion for absence of eigenvalues and applications to substitution Hamiltonians, *Ann. Henri Poincaré* **2** (2001), 927–939.

[15] X. Droubay & J. Justin & G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* **255** (2001), 539–553.

[16] X. Droubay & G. Pirillo, Palindromes and Sturmian words, *Theoret. Comput. Sci.* **223** (1999), 73–85.

[17] A. Glen & J. Justin & S. Widmer & L. Q. Zamboni, Palindromic richness, *European J. Combin.* **30** (2009), 510–531.

[18] A. Hof & O. Knill & B. Simon, Singular continuous spectrum for palindromic Schrödinger operators, *Comm. Math. Phys.* **174** (1995), 149–159.

[19] Š. Holub & K. Saari, On highly palindromic words, *Discrete Appl. Math.* **157** (2009), 953–959.

[20] L. Kari & K. Mahalingam, Watson-Crick palindromes in DNA computing, *Nat. Comput.* **9** (2010), 297–316.

[21] M.-Y. Leung & K. P. Choi & A. Xia & L. H. Y. Chen, Nonrandom clusters of palindromes in herpesvirus genomes, *J. Computational Biology* **12** (2005), 331–354.

[22] A. de Luca, Sturmian words: structure, combinatorics, and their arithmetics, *Theoret. Comput. Sci.* **183** (1997), 45–82.

[23] D. Nolan & T. Speed, *Stat Labs. Mathematical Statistics Through Applications,* Springer, Berlin, 2000.

[24] T. Noll, Sturmian sequences and morphisms. A music-theoretical application, in A. Yves (ed.), *Mathématique et musique, Journée annuelle de la Société Mathématique de France*, 2008, pp. 79–102.

[25] G. Rauzy, Suites á termes dans un alphabet fini, *Séminaire de théorie des nombres. 1982–1983,* Université de Bordeaux I, U. E. R. de Mathématiques et d'Informatique, Laboratoire de Théorie des Nombres, Talence, 1983, Exp. No. 25, 16 pp.

Bibliography

Kratka biografija

Bojan Bašić je rođen 2. 9. 1986. u Odžacima. Završio je Gimnaziju „Jovan Jovanović Zmaj" u Novom Sadu, kao nosilac Vukove diplome i đak generacije matematičkog smera. Tokom osnovne i srednje škole učestvovao je na brojnim takmičenjima u zemlji i inostranstvu iz matematike, informatike i fizike, a kao najveće uspehe s tih takmičenja izdvaja bronzanu medalju na Internacionalnoj olimpijadi iz matematike u Meksiku, 2005. godine, i bronzanu medalju na Internacionalnoj olimpijadi iz informatike u Poljskoj, 2005. godine.

Diplomirao je matematiku na Prirodno-matematičkom fakultetu u Novom Sadu 2009. godine, s prosečnom ocenom 10,00. Naziv diplomskog rada bio je *Aksioma izbora i matematika*. Izabran je za najboljeg studenta Prirodno-matematičkog fakulteta 2009. godine. Iste godine upisao se na doktorske studije, gde je položio sve ispite s prosečnom ocenom 10,00. Savetnik tokom doktorskih studija bio mu je prof. dr Miloš Kurilić.

Od 2005. godine angažovan je u nastavi na Departmanu za matematiku i informatiku Prirodno-matematičkog fakulteta u Novom Sadu, najpre kao istraživač-saradnik, a potom kao asistent. Držao je ili drži vežbe iz predmeta:

Osnovi geometrije 1, Osnovi geometrije 2, Teorija brojeva i drugih.

Član je naučnog projekta *Teorija skupova, teorija modela i skup-teoretska topologija* pod pokroviteljstvom Ministarstva prosvete i nauke Republike Srbije. Ima tri naučna rada objavljena ili prihvaćena za štampu u časopisima sa ISI liste. Njegovi rezultati (što samostalni, što sa koautorima) u tri navrata su izlagani na međunarodnim konferencijama, od toga jedanput po pozivu.

Novi Sad, maj 2012   Bojan Bašić

Izvod: U tezi razmatramo aktuelne probleme u vezi s palindromskim podrečima i palindromskim faktorima konačnih i beskonačnih reči. Glavni pravac istraživanja jesu kriterijumi za određivanje koja od dve date reči je „palindromičnija" od druge, tj. određivanje stepena „palindromičnosti" date reči. Akcenat stavljamo na dva aktuelna pristupa: tzv. *MP-razmeru* i tzv. *palindromski defekt*, i odgovaramo na više otvorenih pitanja u vezi s njima.

Naime, u vezi sa MP-razmerom u literaturi je postavljeno više pitanja, intuitivno uverljivih, koja bi, u slučaju pozitivnog razrešenja, znatno pojednostavila izračunavanje MP-razmere. Ovim pitanjima dodajemo još jedno srodno, a zatim pokazujemo da, prilično neočekivano, sva ova pitanja imaju negativan odgovor.

U vezi s palindromskim defektom, glavni rezultat rada je konstrukcija beskonačne klase beskonačnih reči koje imaju više osobina za kojima je iskazana potreba u skorašnjim radovima iz ove oblasti. Među najzanimljivije spada činjenica da su sve aperiodične reči konačnog pozitivnog defekta, i da im je skup faktora zatvoren za preokretanje – u nekim skorašnjim radovima konstrukcija makar jedne reči s ovim osobinama pokazala se kao prilično teška. Pomoću ovih reči, koje nazivamo *visokopotencijalne reči*, ispitujemo validnost više otvorenih hipoteza, i za više njih ustanovljavamo da nisu validne.

IZ
Datum prihvatanja teme od strane
NN veća: 8. mart 2012.
DP
Datum odbrane:
DO
Članovi komisije: Predsednik: Dr Petar Marković, vanredni profesor Priro-
dno-matematičkog fakulteta u Novom Sadu
Član: Dr Igor Dolinka, redovni profesor Prirodno-mate-
matičkog fakulteta u Novom Sadu
Član: Dr Ivica Bošnjak, vanredni profesor Prirodno-mate-
matičkog fakulteta u Novom Sadu
Član: Dr Rade Doroslovački, redovni profesor Fakulteta
tehničkih nauka u Novom Sadu
Član: Dr Vojislav Petrović, redovni profesor Prirodno-ma-
tematičkog fakulteta u Novom Sadu

KO

PD

Scientific field: Mathematics

SF

Scientific discipline: Discrete Mathematics

SD

Subject, key words: palindrome, subword, factor,
                                            factor complexity
                                            palindromic complexity, defect

SKW

UC

Holding data: Library of Department of Mathematics
                                            and Informatics, Novi Sad

HD

Note:

N

Abstract: In the thesis we are concerned with actual problems on palindromic subwords and palindromic factors of finite and infinite words. The main course of the research are the ways of determining which of two given words is "more palindromic" than the other one, that is, defining a measure for the degree of "palindromicity" of a word. Particularly, we pay attention to two actual approaches: the so-called MP-ratio and the so-callem *palindromic defect*, and answer several open questions about them.

Namely, concerning the MP-ratio, a few plausible-looking question have been asked in the literature, which would have, if answered positively, made computations of MP-ratios significantly simpler. We add one more related question to these ones, and then show that, rather unexpectedly, all these questions have negative answer. Concerning the palindromic defect, the main result of this work is a construction of an infinite class of infinite words that have several properties that were sought after in some recent works in this area. Among the most interesting facts is that that all these words are aperiodic words of a finite positive defect, having the set of factors closed under reversal—in some recent works, the construction of even a single word having these properties turned out to be quite hard. Using these words, which we are calling *highly potential words*, we check the validity of several open conjectures, and for several of them we find out that they are false.

AB

Accepted on Scientific board on: March 8th 2012

AS

Defended:

DE

Thesis Defend board: President: Petar Marković, Ph.D., Associate Professor,
Faculty of Science, Novi Sad
Member: Igor Dolinka, Ph.D., Full Professor,
Faculty of Science, Novi Sad
Member: Ivica Bošnjak, Ph.D., Associate Professor,
Faculty of Science, Novi Sad
Member: Rade Doroslovački, Ph.D., Full Professor,
Faculty of Engineering, Novi Sad
Member: Vojislav Petrović, Ph.D., Full Professor,
Faculty of Science, Novi Sad

DB