Theoretical Computer Science 475 (2013) 120–125

Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Note

Proof of the Brlek–Reutenauer conjecture

L'. Balková[a,*], E. Pelantová[a], Š. Starosta[b]

[a] Department of Mathematics, FNSPE, Czech Technical University in Prague, Trojanova 13, 120 00 Praha 2, Czech Republic

[b] Department of Applied Mathematics, FIT, Czech Technical University in Prague, Thákurova 9, 160 00 Praha 6, Czech Republic

ARTICLE INFO

Article history:
Received 13 August 2012
Accepted 26 December 2012
Communicated by D. Perrin

ABSTRACT

Brlek and Reutenauer conjectured that any infinite word $\mathbf{u}$ with language closed under reversal satisfies the equality $2D(\mathbf{u}) = \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n)$ in which $D(\mathbf{u})$ denotes the defect of $\mathbf{u}$ and $T_{\mathbf{u}}(n)$ denotes $\mathcal{C}_{\mathbf{u}}(n+1)-\mathcal{C}_{\mathbf{u}}(n)+2-\mathcal{P}_{\mathbf{u}}(n+1)-\mathcal{P}_{\mathbf{u}}(n)$, where $\mathcal{C}_{\mathbf{u}}$ and $\mathcal{P}_{\mathbf{u}}$ are the factor and palindromic complexity of $\mathbf{u}$, respectively. This conjecture was verified for periodic words by Brlek and Reutenauer themselves. Using their results for periodic words, we have recently proved the conjecture for uniformly recurrent words. In the present article we prove the conjecture in its general version by a new method without exploiting the result for periodic words.

1. Introduction

Brlek and Reutenauer conjectured in [6] a nice equality which combines together the factor complexity $\mathcal{C}_{\mathbf{u}}$, the palindromic complexity $\mathcal{P}_{\mathbf{u}}$, and the palindromic defect $D(\mathbf{u})$ of an infinite word $\mathbf{u}$. It sounds as follows.

Brlek–Reutenauer Conjecture. *If $\mathbf{u}$ is an infinite word with language closed under reversal, then*

$$2D(\mathbf{u}) = \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n),$$

where $T_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n) + 2 - \mathcal{P}_{\mathbf{u}}(n+1) - \mathcal{P}_{\mathbf{u}}(n)$.

Brlek and Reutenauer proved ibidem that their conjecture holds for periodic infinite words. It is known from [7] that the Brlek–Reutenauer conjecture holds for words with zero defect. In [3], we proved the conjecture for uniformly recurrent words. In our proof, we constructed for any uniformly recurrent word $\mathbf{u}$ whose language is closed under reversal a periodic word $\mathbf{v}$ with language closed under reversal such that $D(\mathbf{u}) = D(\mathbf{v})$ and $T_{\mathbf{u}}(n) = T_{\mathbf{v}}(n)$ for any n. Then we used validity of the conjecture for periodic words.

In this paper, we will prove that the Brlek–Reutenauer conjecture holds in full generality without exploiting the result for periodic words. Since both sides of the equality in the Brlek–Reutenauer conjecture are non-negative, validity of the conjecture will be shown if we prove the following two theorems.

Theorem 1. *If $\mathbf{u}$ is an infinite word with language closed under reversal such that both $D(\mathbf{u})$ and $\sum_{n=0}^{+\infty} T_{\mathbf{u}}(n)$ are finite, then*

$$2D(\mathbf{u}) = \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n). \tag{1}$$

* Corresponding author. Tel.: +420 224358544.
E-mail address: lubomira.balkova@gmail.com (L. Balková).

0304-3975/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2012.12.024

Theorem 2. *If* $\mathbf{u}$ *is an infinite word with language closed under reversal, then*

$$D(\mathbf{u}) < +\infty \quad \textit{if and only if} \quad \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n) < +\infty.$$

In [3], which is devoted mainly to the uniformly recurrent words, we already stated in the section Open problems one part of Theorem 2, namely that $D(\mathbf{u}) < +\infty$ implies $\sum_{n=0}^{+\infty} T_{\mathbf{u}}(n) < +\infty$. As pointed out in [4], there is a gap in our proof, and its corrected version can be found in [2]. In order to make the present paper self-sustained so that the reader understands and checks all steps of the proof without having all previous papers at hand, we recall necessary notations and statements together with the proofs of the essential ones.

2. Preliminaries

By $\mathcal{A}$ we denote a finite set of symbols called *letters*; the set $\mathcal{A}$ is therefore called an *alphabet*. A finite string $w = w_0w_1\dots w_{n-1}$ of letters from $\mathcal{A}$ is said to be a *finite word*, its length is denoted by $|w| = n$. Finite words over $\mathcal{A}$ together with the operation of concatenation and the empty word ϵ as the neutral element form a free monoid $\mathcal{A}^*$. The map

$$w = w_0w_1\dots w_{n-1} \quad \mapsto \quad \overline{w} = w_{n-1}w_{n-2}\dots w_0$$

is a bijection on $\mathcal{A}^*$, the word $\overline{w}$ is called the *reversal* or the *mirror image* of w. A word w which coincides with its mirror image is a *palindrome*.

Under an *infinite word* we understand an infinite string $\mathbf{u} = u_0u_1u_2\dots$ of letters from $\mathcal{A}$. A finite word w is a *factor* of a word v (finite or infinite) if there exist words p and s such that $v = pws$. If $p = \epsilon$, then w is said to be a *prefix* of v, if $s = \epsilon$, then w is a *suffix* of v.

The *language* $\mathcal{L}(v)$ of a finite or an infinite word v is the set of all its factors. Factors of v of length n form the set denoted by $\mathcal{L}_n(v)$. We say that the language of an infinite word $\mathbf{u}$ is *closed under reversal* if $\mathcal{L}(\mathbf{u})$ contains with every factor w also its reversal $\overline{w}$.

For any factor $w \in \mathcal{L}(\mathbf{u})$, there exists an index i such that w is a prefix of the infinite word $u_iu_{i+1}u_{i+2}\dots$. Such an index is called an *occurrence* of w in $\mathbf{u}$. If each factor of $\mathbf{u}$ has infinitely many occurrences in $\mathbf{u}$, the infinite word $\mathbf{u}$ is said to be *recurrent*. It is easy to see that if the language of $\mathbf{u}$ is closed under reversal, then $\mathbf{u}$ is recurrent (a proof can be found in [9]). For a recurrent infinite word $\mathbf{u}$, we may define the notion of a *complete return word* of any $w \in \mathcal{L}(\mathbf{u})$. It is a factor $v \in \mathcal{L}(\mathbf{u})$ such that w is a prefix and a suffix of v and w occurs in v exactly twice.

If any factor $w \in \mathcal{L}(\mathbf{u})$ has only finitely many complete return words, then the infinite word $\mathbf{u}$ is called *uniformly recurrent*.

The *factor complexity* of an infinite word $\mathbf{u}$ is the map $\mathcal{C}_{\mathbf{u}} : \mathbb{N} \mapsto \mathbb{N}$ defined by the prescription $\mathcal{C}_{\mathbf{u}}(n) := \#\mathcal{L}_n(\mathbf{u})$. To determine the first difference of the factor complexity, one has to count the possible extensions of factors of length n. A *right extension* of $w \in \mathcal{L}(\mathbf{u})$ is a letter $a \in \mathcal{A}$ such that $wa \in \mathcal{L}(\mathbf{u})$. Of course, any factor of $\mathbf{u}$ has at least one right extension. A factor w is called *right special* if w has at least two right extensions. Similarly, one can define a *left extension* and a *left special* factor. We will deal mainly with recurrent infinite words $\mathbf{u}$. In such a case, any factor of $\mathbf{u}$ has at least one left extension.

In [8] it is shown that any finite word w contains at most $|w| + 1$ distinct palindromes (including the empty word). The *defect* $D(w)$ of a finite word w is the difference between the utmost number of palindromes $|w| + 1$ and the actual number of palindromes contained in w.

In accordance with the terminology introduced in [8], the factor with a unique occurrence in another factor is called *unioccurrent*.

The following corollary gives an insight into the birth of defects.

Corollary 3 ([8]). *The defect $D(w)$ of a finite word w is equal to the number of prefixes w' of w for which the longest palindromic suffix of w' is not unioccurrent in w'. In other words, if b is a letter and w a finite word, then $D(wb) = D(w) + \delta$, where $\delta = 0$ if the longest palindromic suffix of wb occurs exactly once in wb and $\delta = 1$ otherwise.*

Corollary 3 implies that $D(v) \geq D(w)$ whenever w is a factor of v. It enables to give a reasonable definition of the defect of an infinite word (see [5]).

Definition 4. The defect of an infinite word $\mathbf{u}$ is the number (finite or infinite)

$$D(\mathbf{u}) = \sup\{D(w) \colon w \text{ is a prefix of } \mathbf{u}\}.$$

Let us point out two facts.

(1) If we consider all factors of a finite or an infinite word $\mathbf{u}$, we obtain the same defect, i.e.,

$$D(\mathbf{u}) = \sup\{D(w) \colon w \in \mathcal{L}(\mathbf{u})\}.$$

(2) Any infinite word with finite defect contains infinitely many palindromes.

Using Corollary 3 and Definition 4, we obtain immediately the following corollary.

Corollary 5. *Let* $\mathbf{u}$ *be an infinite word with language closed under reversal. The following statements are equivalent.*

(1) *The defect of* $\mathbf{u}$ *is finite.*
(2) *There exists an integer* H *such that the longest palindromic suffix of any prefix* w *of length* $|w| \geq H$ *occurs in* w *exactly once.*

For *the longest palindromic suffix* of a word w we will sometimes use the notation $lps(w)$.

The number of palindromes of a fixed length occurring in an infinite word is measured by the so called *palindromic complexity* $\mathcal{P}_{\mathbf{u}}$, the map which assigns to any non-negative integer n the number

$$\mathcal{P}_{\mathbf{u}}(n) := \#\{w \in \mathcal{L}_n(u) \colon w \text{ is a palindrome}\}.$$

Denote

$$T_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n) + 2 - \mathcal{P}_{\mathbf{u}}(n+1) - \mathcal{P}_{\mathbf{u}}(n).$$

The following proposition is proven in [1] for uniformly recurrent words; however, as also noted in [6], the uniform recurrence is not needed in the proof and it holds for any infinite word with language closed under reversal.

Proposition 6 ([1])**.** *If* $\mathbf{u}$ *is an infinite word with language closed under reversal, then*

$$T_{\mathbf{u}}(n) \geq 0 \quad \text{for all } n \in \mathbb{N}. \tag{2}$$

Let $\mathbf{u}$ be an infinite word with language closed under reversal. Using the proof of Proposition 6, those $n \in \mathbb{N}$ for which $T_{\mathbf{u}}(n) = 0$ can be characterized in the graph language. Before doing that we need to introduce some more notions.

An *n-simple path* e is a factor of $\mathbf{u}$ of length at least $n+1$ such that the only special (right or left) factors of length n occurring in e are its prefix and suffix of length n. If w is the prefix of e of length n and v is the suffix of e of length n, we say that the n-simple path e starts in w and ends in v. We will denote by $G_n(\mathbf{u})$ an undirected graph whose set of vertices is formed by unordered pairs $(w, \overline{w})$ such that $w \in \mathcal{L}_n(\mathbf{u})$ is right or left special. We connect two vertices $(w, \overline{w})$ and $(v, \overline{v})$ by an unordered pair $(e, \overline{e})$ if e or $\overline{e}$ is an n-simple path starting in w or $\overline{w}$ and ending in v or $\overline{v}$. Note that the graph $G_n(\mathbf{u})$ may have multiple edges and loops.

Lemma 7. *If* $\mathbf{u}$ *is an infinite word with language closed under reversal and* $n \in \mathbb{N}$, *then* $T_{\mathbf{u}}(n) = 0$ *if and only if both of the following conditions are met.*

(1) *The graph obtained from* $G_n(\mathbf{u})$ *by removing loops is a tree.*
(2) *Any* n*-simple path forming a loop in the graph* $G_n(\mathbf{u})$ *is a palindrome.*

Proof. It is a direct consequence of the proof of Theorem 1.2 in [1] (recalled in this paper as Proposition 6). □

3. Proof of Theorem 1

The aim of this section is to prove Theorem 1, i.e., to prove the Brlek–Reutenauer conjecture under the additional assumption that the defect $D(\mathbf{u})$ of an infinite word $\mathbf{u}$ and the sum $\sum_{n=0}^{\infty} T_{\mathbf{u}}(n)$ are finite. As observed in [6], it is easy to prove the "finite analogy" of the conjecture, which deals only with finite words. We will also make use of this result.

Theorem 8 ([6])**.** *For every finite word* w *we have*

$$2D(w) = \sum_{n=0}^{|w|} T_w(n),$$

where $T_w(n) = \mathcal{C}_w(n+1) - \mathcal{C}_w(n) + 2 - \mathcal{P}_w(n+1) - \mathcal{P}_w(n)$ *and the index* w *means that we consider only factors of* w.

It may seem that the Brlek–Reutenauer conjecture for an infinite word $\mathbf{u}$ can be obtained from Theorem 8 by a "limit transition". However, this transition would be far from being kosher. The following lemmas enable us to avoid the incorrectness.

Lemma 9. *Let* $\mathbf{u}$ *be an infinite word with language closed under reversal and finite defect. If* q *is its prefix satisfying* $D(\mathbf{u}) = D(q)$, *then for* $H = |q| + 1$ *one has*

$$\mathcal{C}_{\mathbf{u}}(H) - \mathcal{P}_{\mathbf{u}}(H) = 2\#\{x \in \mathcal{L}(\mathbf{u}) \colon x \text{ is a palindrome shorter than } H \text{ which is not contained in } q\}.$$

Proof. Let us define a mapping $f : S \to T$, where

$$S = \{x \in \mathcal{L}(\mathbf{u}) \colon x \notin \mathcal{L}(q), |x| < H, x = \overline{x}\}$$

and

$$T = \big\{\{w, \overline{w}\} \colon w \in \mathcal{L}_H(\mathbf{u}), w \neq \overline{w}\big\}.$$

Let x be a palindrome from S and i be the first occurrence of x in $\mathbf{u}$. Put $w = u_{i+|x|-H} \cdots u_{i+|x|-1}$. It means that w is a factor of $\mathbf{u}$ of length H and x is a suffix of w. Since $H > |x|$, the factor w is not a palindrome — otherwise it contradicts the fact that i is the first occurrence of the palindrome x. We put $f(x) = \{w, \overline{w}\}$.

To show that f is surjective, we consider $w \in \mathcal{L}_H(\mathbf{u})$ such that $w \neq \overline{w}$. Let p be the prefix of $\mathbf{u}$ which ends in the first occurrence of w or $\overline{w}$ in $\mathbf{u}$. Since $|p| \geq H = |w| > |q|$, we have according to Corollary 3 that $D(q) = D(p)$ and consequently, $lps(p)$ is unioccurrent in p, which implies that $lps(p)$ is not a factor of q. Moreover, $lps(p)$ is shorter than H — otherwise it contradicts the choice of the prefix p. We found $x = lps(p) \in S$ such that $f(x) = \{w, \overline{w}\}$, i.e., f is surjective.

To show that f is injective, we consider two palindromes $y, z \in S$ and we denote $f(y) = \{w_y, \overline{w_y}\}$ and $f(z) = \{w_z, \overline{w_z}\}$. From the definition of w_x we know that the palindrome x occurs as a factor of w_x exactly once, namely as its suffix. It means that x equals $lps(w_x)$. Let us suppose that $f(y) = f(z)$. We have to discuss two cases.

(1) Case $w_y = w_z$. It gives $lps(w_y) = lps(w_z)$ and thus $y = z$.
(2) Case $w_y = \overline{w_z}$. It implies that y is a prefix of w_z and z is a prefix of w_y. The fact that y is a prefix of w_z forces the first occurrence of w_y to be strictly smaller than the first occurrence of w_z. Simultaneously, since z is a prefix of w_y, the first occurrence of w_z is strictly smaller than the first occurrence of w_y - a contradiction.

Consequently, the assumption $f(y) = f(z)$ implies $z = y$ and the mapping f is injective as well.

Existence of the bijection f between the finite sets T and S means $\#T = \#S$. Since from the definition of T it follows that $\mathcal{C}_{\mathbf{u}}(H) - \mathcal{P}_{\mathbf{u}}(H) = 2\#T$, the equality stated in the lemma is proven. $\square$

Remark 10. As it was pointed out by Bojan Bašić, Lemma 9 may be stated in a more general form for $H > |q|$, then the equality changes to

$$\mathcal{C}_{\mathbf{u}}(H) - \mathcal{P}_{\mathbf{u}}(H) = 2\#\{x \in \mathcal{L}(\mathbf{u}) : x \notin \mathcal{L}(q),\ |x| < H,\ x = \overline{x}\} - 2(H - |q| - 1).$$

Thanks to him, we added the assumption $H = |q| + 1$ in Lemma 9 necessary for the validity of the statement.

Lemma 11. *Let $\mathbf{u}$ be an infinite word with language closed under reversal and finite defect. If q is its prefix satisfying $D(\mathbf{u}) = D(q)$, then for any prefix p of $\mathbf{u}$ such that $|p| > |q|$ the number*

$$\#\{x \in \mathcal{L}(p) : x \textit{ is a palindrome of length at most } |q| \textit{ which is not contained in } q\} + \sum_{n=|q|+1}^{|p|} \mathcal{P}_p(n)$$

equals $|p| - |q|$.

Proof. At first we will show the equality

$$|p| - |q| = \#\{x \in \mathcal{L}(p) \setminus \mathcal{L}(q) : x = \overline{x}\}. \tag{3}$$

Let us denote by $u^{(i)}$ the prefix of $\mathbf{u}$ of length i. For any palindrome $x \in \mathcal{L}(p) \setminus \mathcal{L}(q)$ we find the minimal index i such that x occurs in $u^{(i)}$. Since $x \in \mathcal{L}(p) \setminus \mathcal{L}(q)$, we have $|q| < i \leq |p|$. Thus we map any element of $\{x \in \mathcal{L}(p) \setminus \mathcal{L}(q) : x = \overline{x}\}$ to an index $i \in \{|q| + 1, |q| + 2, \ldots, |p|\}$.

Let us look at the details of this mapping. The minimality of i guarantees that x is unioccurrent in $u^{(i)}$. Palindromicity of x gives that $x = lps(u^{(i)})$. It implies that no two different palindromes are mapped to the same index i, i.e., the mapping is injective.

Since $D(q) = D(\mathbf{u})$, according to Corollary 3, $lps(u^{(i)})$ is unioccurrent in $u^{(i)}$ and thus $lps(u^{(i)}) \notin \mathcal{L}(q)$. Thus any index i such that $|q| < i \leq |p|$ has its preimage $x = lps(u^{(i)})$. Therefore the mapping is a bijection and its domain and range have the same cardinality as stated in (3).

To finish the proof, we split elements of $\{x \in \mathcal{L}(p) \setminus \mathcal{L}(q) : x = \overline{x}\}$ into two disjoint parts: elements of length smaller than or equal to $|q|$ and elements of length greater than $|q|$. Since

$$\#\{x \in \mathcal{L}(p) \setminus \mathcal{L}(q) : x = \overline{x},\ |x| > |q|\} = \#\{x \in \mathcal{L}(p) : x = \overline{x},\ |x| > |q|\} = \sum_{n=|q|+1}^{|p|} \mathcal{P}_p(n),$$

the statement of Lemma 11 is proven. $\square$

Now we can complete the proof of Theorem 1.

Proof of Theorem 1. Finiteness of defect means that there exists a constant $L \in \mathbb{N}$ such that $D(\mathbf{u}) = D(q)$ for any prefix q of $\mathbf{u}$ which is longer than or of length equal to L. On the other hand, finiteness of the sum $\sum_{n=0}^{+\infty} T_{\mathbf{u}}(n)$ together with the fact $0 \leq T_{\mathbf{u}}(n) \in \mathbb{Z}$ for any $n \in \mathbb{N}$ implies that there exists a constant $M \in \mathbb{N}$ such that $T_{\mathbf{u}}(n) = 0$ for any $n > M$. Let us fix an integer $H > \max\{L, M\}$ and denote by q the prefix of $\mathbf{u}$ of length $|q| = H - 1$. Consequently,

$$T_{\mathbf{u}}(n) = 0 \quad \text{for any } n \geq H \quad \text{and} \quad D(\mathbf{u}) = D(q).$$

In order to show the equality (1), it thus remains to show $2D(q) = \sum_{n=0}^{H-1} T_{\mathbf{u}}(n)$.

Let us consider a prefix p of $\mathbf{u}$ containing all factors of length H. In this case p is longer than q, thus it holds by Corollary 3 that $D(q) = D(p)$. Using Theorem 8, we have

$$2D(p) = \sum_{n=0}^{|p|} T_p(n) = \sum_{n=0}^{H-1} T_p(n) + \sum_{n=H}^{|p|} T_p(n) = \sum_{n=0}^{H-1} T_{\mathbf{u}}(n) + \sum_{n=H}^{|p|} T_p(n),$$

where the last equality is due to the fact that p contains all factors of length H. It remains to prove that $\sum_{n=H}^{|p|} T_p(n) = 0$. Let us rewrite the sum by definition.

$$\begin{aligned}\sum_{n=H}^{|p|} T_p(n) &= \sum_{n=H}^{|p|} \big(\mathcal{C}_p(n+1) - \mathcal{C}_p(n) + 2 - \mathcal{P}_p(n+1) - \mathcal{P}_p(n)\big) \\ &= -\mathcal{C}_p(H) + 2(|p| - H + 1) - 2\sum_{n=H}^{|p|} \mathcal{P}_p(n) + \mathcal{P}_p(H) \\ &= -\mathcal{C}_{\mathbf{u}}(H) + 2(|p| - H + 1) - 2\sum_{n=H}^{|p|} \mathcal{P}_p(n) + \mathcal{P}_{\mathbf{u}}(H), \end{aligned} \tag{4}$$

where in the last equality we again used the fact that p contains all factors of length H. This fact also allows us to rewrite the set $\{x \in \mathcal{L}(p) \ : \ x \notin \mathcal{L}(q), x = \overline{x}, |x| \le |q|\}$ from Lemma 11 as $\{x \in \mathcal{L}(\mathbf{u}) \ : \ x \notin \mathcal{L}(q), x = \overline{x}, |x| < H\}$. Denote the cardinality of this set by B.

In this notation, Lemmas 9 and 11 say

$$\mathcal{C}_{\mathbf{u}}(H) - \mathcal{P}_{\mathbf{u}}(H) = 2B \quad \text{and} \quad B + \sum_{n=H}^{|p|} \mathcal{P}_p(n) = |p| - H + 1.$$

This implies that the last expression in (4) is zero as desired. $\square$

4. Proof of Theorem 2

If an infinite word $\mathbf{u}$ is periodic with language closed under reversal, then $D(\mathbf{u}) < +\infty$ and $\sum_{n=0}^{+\infty} T_{\mathbf{u}}(n) < +\infty$, as shown in [6]. Consequently, we will limit our considerations in the sequel to aperiodic words.

Proposition 12. *If $\mathbf{u}$ is an aperiodic infinite word with language closed under reversal and N is an integer, then $T_{\mathbf{u}}(n) = 0$ for all $n \ge N$ if and only if for any factor w such that $|w| \ge N$, any factor longer than w beginning in w or $\overline{w}$ and ending in w or $\overline{w}$, with no other occurrences of w or $\overline{w}$, is a palindrome.*

Proof. ($\Leftarrow$): Let us show for any $n \ge N$ that the assumptions of Lemma 7 are satisfied. We have to show two properties of $G_n(\mathbf{u})$ for any $n \ge N$.

(1) Any loop in $G_n(\mathbf{u})$ is a palindrome.
Since any loop e in $G_n(\mathbf{u})$ at a vertex $(w, \overline{w})$ is a word longer than w beginning in a special factor w or $\overline{w}$ and ending in w or $\overline{w}$, with no other occurrences of w or $\overline{w}$, the loop e is a palindrome by the assumption.

(2) The graph obtained from $G_n(\mathbf{u})$ by removing loops is a tree.
Or equivalently, we have to show that in $G_n(\mathbf{u})$ there exists a unique path between any two different vertices $(w', \overline{w'})$ and $(w'', \overline{w''})$. Let p be a factor of $\mathbf{u}$ such that w' or $\overline{w'}$ is its prefix, w'' or $\overline{w''}$ is its suffix and p has no other occurrences of $w', \overline{w'}, w'', \overline{w''}$. Let v be a factor starting in p, ending in w' or $\overline{w'}$ and containing no other occurrences of w' or $\overline{w'}$. By the assumption the factor v is a palindrome, thus $\overline{p}$ is a suffix of v. It is then a direct consequence of the construction of v that the next factor with the same properties as p, i.e., representing a path in the undirected graph $G_n(\mathbf{u})$ between w' and w'', which occurs in $\mathbf{u}$ after p, is $\overline{p}$. This shows that there is only one such path.

Consequently, Lemma 7 implies that $T_n(\mathbf{u}) = 0$ for any $n \ge N$.
($\Rightarrow$): First we prove an auxiliary claim.

Claim. *If $\mathbf{u}$ is an aperiodic infinite word with language closed under reversal and N is an integer such that $T_{\mathbf{u}}(n) = 0$ for all $n \ge N$, then for any w such that $|w| \ge N$ and any factor v longer than w beginning in w and ending in w or $\overline{w}$, with no other occurrences of w or $\overline{w}$, there exists a letter $a \in \mathcal{A}$ such that v has prefix wa and suffix $a\overline{w}$.*

It is clear that repeated application of the previous claim to factors w of length gradually increased by one gives the proof of implication ($\Rightarrow$) of Proposition 12.

We split the proof of the auxiliary claim into two cases.

- Case 1: Assume that w is a special factor.
  If v does not contain any other special factor of length $n = |w|$ except for w and $\overline{w}$, then v is a loop in the graph $G_n(\mathbf{u})$ and according to Lemma 7, the factor v is a palindrome. Necessarily, v begins in wa for some letter a and ends in $a\overline{w}$.
  Suppose now that $v = v_0 v_1 \cdots v_m$ contains a special factor $z \ne w, \overline{w}$ of length n at the position i, i.e., $z = v_i v_{i+1} \cdots v_{n+i-1}$. Without loss of generality, we consider the smallest index i with this property. The pair $(z, \overline{z})$ is a vertex in the graph $G_n(\mathbf{u})$ and a prefix of v, say e, corresponds to an edge in $G_n(\mathbf{u})$ starting in $(w, \overline{w})$ and ending in $(z, \overline{z})$. Since the graph $G_n(\mathbf{u})$ is a tree, the word v which corresponds to a walk from $(w, \overline{w})$ to the same vertex $(w, \overline{w})$ has a suffix f representing an edge in $G_n(\mathbf{u})$ connecting again vertices $(z, \overline{z})$ and $(w, \overline{w})$. It means that the suffix f starts in z or $\overline{z}$ and ends in w or $\overline{w}$. Since $G_n(\mathbf{u})$ has no multiple edges connecting distinct vertices, necessarily $f = \overline{e}$, which already gives the claim.

- Case 2: Assume w is not a special factor.

  It means that there exists a unique letter a such that wa belongs to the language of $\mathbf{u}$. As the language is closed under reversal, the factor $\overline{w}$ has a unique left extension, namely a. If v starts in w and ends in $\overline{w}$, then the claim is proven.

  It remains to exclude that v begins and ends in a non-palindromic factor w. Suppose this situation happens. In this case, there exists a unique q such that wq is a right special factor and it is the shortest right special factor having the prefix w. The factor wq has only one occurrence of the factor w — otherwise we can find a shorter prolongation of w which is right special. Since w is a suffix of v, we deduce that $|wq| < |v|$. Because wq is the shortest right special factor with prefix w, the factor vq belongs to the language and its prefix and suffix wq is a special factor. According to already proven Case 1, we have $wq = \overline{wq} = \overline{q}\,\overline{w}$. It means together with the inequality $|wq| < |v|$ that $\overline{w}$ is contained in v as well — a contradiction. □

The proof of the implication ($\Rightarrow$) of Proposition 12 is taken from [10], where we showed a more general statement for an infinite word whose language is closed under a larger group of symmetries.

Corollary 13. *Let $\mathbf{u}$ be an aperiodic infinite word with language closed under reversal and let N be an integer. If $T_{\mathbf{u}}(n) = 0$ for all $n \geq N$, then the occurrences of w and $\overline{w}$ in $\mathbf{u}$ alternate for any factor w of $\mathbf{u}$ of length at least N.*

The following lemma builds a bridge between Corollary 5 and Proposition 12.

Lemma 14. *Let $\mathbf{u}$ be an aperiodic infinite word with language closed under reversal. There exists $H \in \mathbb{N}$ such that the longest palindromic suffix of any prefix w of $\mathbf{u}$ of length $|w| \geq H$ occurs in w exactly once if and only if there exists $N \in \mathbb{N}$ such that for any factor w with $|w| \geq N$, any factor longer than w beginning in w or $\overline{w}$ and ending in w or $\overline{w}$, with no other occurrences of w or $\overline{w}$, is a palindrome.*

Proof. ($\Rightarrow$): We will show that N may be set equal to H. Let us proceed by contradiction. Suppose there exists a factor $w \in \mathcal{L}(\mathbf{u})$ such that $|w| \geq H$ and there exists a non-palindromic factor of $\mathbf{u}$ longer than w beginning in w or $\overline{w}$ and ending in w or $\overline{w}$, with no other occurrences of w or $\overline{w}$. Let us find the first non-palindromic factor of the above form in $\mathbf{u}$ and let us denote it as r. Let p be the prefix of $\mathbf{u}$ ending in the first occurrence of r in $\mathbf{u}$, i.e., $p = tr$ for some word t and r is unioccurrent in p. Denote by s the longest palindromic suffix of p. By the assumption, s is unioccurrent in p. No matter how long the suffix s is, we will obtain a contradiction.

(1) If $|s| \leq |w|$, then we have a contradiction to the unioccurrence of s.
(2) If $|r| > |s| > |w|$, then we can find at least 3 occurrences of w or $\overline{w}$ in r which is a contradiction to the form of r.
(3) The equality $|r| = |s|$ contradicts the fact that we supposed r to be non-palindromic.
(4) Finally, if $|r| < |s|$, then there is an occurrence of the mirror image of r which is a non-palindromic factor having the same properties as r which occurs before r and contradicts the choice of p.

($\Leftarrow$): Take a prefix containing all factors of length N. Set H equal to its length. Let us show that any prefix p of length greater than or equal to H has $lps(p)$ of length greater than or equal to N. Consider a suffix of p of length N, say w. Either w is a palindrome, then $lps(p)$ is of length greater than or equal to N. Or w is not a palindrome, then we find a suffix of p beginning in $\overline{w}$ and containing exactly two occurrences of w or $\overline{w}$. Such a suffix exists since all factors of length N are contained in p. By assumptions, such a suffix is a palindrome, hence $lps(p)$ is longer than N.

Any prefix p of $\mathbf{u}$ of length greater than or equal to H has $lps(p)$ unioccurrent. Assume there are more occurrences of $lps(p)$ in p and consider its suffix v starting in the last-but-one occurrence of $lps(p)$. Since the length of $lps(p)$ is greater than or equal to N, the factor v is a palindrome by assumptions, which contradicts the choice of $lps(p)$. □

Proof of Theorem 2. For periodic words, the statement was shown in [6]. If $\mathbf{u}$ is aperiodic, then the statement is a direct consequence of Lemma 14, Corollary 5, and Proposition 12. □

Acknowledgments

This work was supported by the Czech Science Foundation grant GAČR 201/09/0584 and by the grant MSM6840770039 of the Ministry of Education, Youth, and Sports of the Czech Republic.

We would like to thank Bojan Bašić for his careful reading, useful remarks and in particular for his correction of a lemma.

References

[1] P. Baláži, Z. Masáková, E. Pelantová, Factor versus palindromic complexity of uniformly recurrent infinite words, Theoret. Comput. Sci. 380 (2007) 266–275.
[2] Ľ. Balková, E. Pelantová, Š. Starosta, Corrigendum: "On Brlek–Reutenauer conjecture", Theoret. Comput. Sci. 465 (2012) 73–74.
[3] Ľ. Balková, E. Pelantová, Š. Starosta, On Brlek-Reutenauer conjecture, Theoret. Comput. Sci. 412 (2011) 5649–5655.
[4] B. Bašć, A note on the paper On Brlek-Reutenauer conjecture, Theoret. Comput. Sci. 448 (2012) 94–96.
[5] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer, On the palindromic complexity of infinite words, Internat. J. Found. Comput. 15 (2004) 293–306.
[6] S. Brlek, C. Reutenauer, Complexity and palindromic defect of infinite words, Theoret. Comput. Sci. 412 (2011) 493–497.
[7] 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.
[8] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, Theoret. Comput. Sci. 255 (2001) 539–553.
[9] A. Glen, J. Justin, S. Widmer, L.Q. Zamboni, Palindromic richness, European J. Combin. 30 (2009) 510–531.
[10] E. Pelantová, Š Starosta, Palindromic richness and Coxeter groups, preprint available at http://arxiv.org/abs/1108.3042.