A NEW CHARACTERISTIC PROPERTY OF RICH WORDS

MICHELANGELO BUCCI, ALESSANDRO DE LUCA, AMY GLEN, AND LUCA Q. ZAMBONI

ABSTRACT. Originally introduced and studied by the third and fourth authors together with J. Justin and S. Widmer (2008), *rich words* constitute a new class of finite and infinite words characterized by containing the maximal number of distinct palindromes. Several characterizations of rich words have already been established. A particularly nice characteristic property is that all 'complete returns' to palindromes are palindromes. In this note, we prove that rich words are also characterized by the property that each factor is uniquely determined by its longest palindromic prefix and its longest palindromic suffix.

1. INTRODUCTION

In [5], X. Droubay, J. Justin, and G. Pirillo proved that any finite word w of length $|w|$ contains at most $|w| + 1$ distinct palindromes (including the empty word). Inspired by this result, the third and fourth authors together with J. Justin and S. Widmer recently initiated a unified study of finite and infinite words that are characterized by containing the maximal number of distinct palindromes (see [10]). Such words are called *rich words* in view of their 'palindromic richness'. More precisely, a finite word w is *rich* if and only if it has exactly $|w| + 1$ distinct palindromic factors. For example, the word $abac$ is rich since it is of length 4 and has exactly 5 distinct palindromic factors: ε, a, b, c, aba. However, if we switch the order of the last two letters, then the resulting word $abca$ is not rich since we lose the palindrome aba. An infinite word is *rich* if all of its factors are rich. For example, the infinite words $a^\omega = aaa\cdots$ and $(ab)^\omega = ababab\cdots$ are clearly rich, whereas $(abc)^\omega = abcabcabc\cdots$ is not, since it contains the non-rich word $abca$.

Rich words have appeared in many different contexts; they include episturmian words [5, 9], complementation-symmetric Rote sequences [1], symbolic codings of trajectories of symmetric interval exchange transformations [6, 7], trapezoidal words [4], and a certain class of words associated with β-expansions where β is a Parry number [2]. Another special class of rich words consists of S. Fischler's sequences with "abundant palindromic prefixes", which were introduced and studied in [8] in relation to Diophantine approximation. Some other simple examples of rich words include: non-recurrent infinite words like $abbbb\cdots$ and $abaabaaabaaaab\cdots$; the periodic infinite words: $(aab^kaabab)(aab^kaabab)\cdots$, with $k \geq 0$; the aperiodic recurrent infinite

Date: Submitted: September 17, 2008 / Revised: October 22, 2008.

2000 *Mathematics Subject Classification.* 68R15.

Key words and phrases. combinatorics on words; palindromes; rich words; return words.

Corresponding author: Amy Glen.

word $\psi(\boldsymbol{f})$ where $\boldsymbol{f} = abaababaaba\cdots$ is the *Fibonacci word* and ψ is the morphism: $a \mapsto aab^k aabab$, $b \mapsto bab$; and the recurrent, but not uniformly recurrent, infinite word generated by the morphism: $a \mapsto aba$, $b \mapsto bb$. See [10] for further examples and references.

Let u be a non-empty factor of a finite or infinite word w. We say that u is *unioccurrent* in w if u has exactly one occurrence in w. Otherwise, if u has more than one occurrence in w, then there exists at least one factor r of w having exactly two distinct occurrences of u, one as a prefix and one as a suffix. Such a factor r is called a *complete return* to u in w. For example, $aabcbaa$ is a (palindromic) complete return to the palindrome aa in the rich word $aabcbaaba$. In [10], it was shown that rich words are characterized by the property that all complete returns to palindromes are palindromes. Using this characteristic property, it is easy to see, for instance, that the Finnish palindromes $avattava$ and $ettette$ are rich, whereas the palindrome $saippuakivikauppias$ (meaning soap-stone vendor) is not rich since it contains a non-palindromic complete return to the letter a (namely $aippua$).

The following proposition collects together all of the characteristic properties of rich words that were previously established in [5] and [10].

Proposition 1. *For any finite or infinite word w, the following conditions are equivalent:*

i) *w is rich;*
ii) *every prefix of w has a unioccurrent palindromic suffix (and equivalently, when w is finite, every suffix of w has a unioccurrent palindromic prefix);*
iii) *every factor u of w contains exactly $|u| + 1$ distinct palindromes;*
iv) *for each factor u of w, every prefix (resp. suffix) of u has a unioccurrent palindromic suffix (resp. prefix);*
v) *for each palindromic factor p of w, every complete return to p in w is a palindrome.*

Remark 2. The equivalences: i) $\Leftrightarrow$ ii), i) $\Leftrightarrow$ iii), and i) $\Leftrightarrow$ iv) were proved in [5].

Explicit characterizations of periodic rich infinite words and recurrent *balanced* rich infinite words have also been established in [10]. More recently, we proved the following connection between palindromic richness and complexity.

Proposition 3. [3] *For any infinite word $\mathbf{w}$ whose set of factors is closed under reversal, the following conditions are equivalent:*

- *all complete returns to palindromes are palindromes;*
- *$\mathcal{P}(n) + \mathcal{P}(n+1) = \mathcal{C}(n+1) - \mathcal{C}(n) + 2$ for all $n \in \mathbb{N}$,*

where $\mathcal{P}$ (resp. $\mathcal{C}$) denotes the palindromic complexity *(resp.* factor complexity*) function of $\mathbf{w}$, which counts the number of distinct palindromic factors (resp. factors) of each length in $\mathbf{w}$.*

From the perspective of richness, the above result can be viewed as a characterization of *recurrent* rich infinite words since any rich infinite word is recurrent if and only if its set of factors is closed under reversal (see [10]). Interestingly, the proof of Proposition 3 relied upon another characterization of rich words, stated below.

Let $\tilde{v}$ denote the *reversal* of a given word v.

Proposition 4. [3] *A finite or infinite word w is rich if and only if, for each factor v of w, every factor of w beginning with v and ending with $\tilde{v}$ and containing no other occurrences of v nor of $\tilde{v}$ is a palindrome.*

In this note, we establish yet another interesting characteristic property of rich words. Our main results are the following two theorems.

Theorem 5. *For any finite or infinite word w, the following conditions are equivalent:*

(A) *w is rich;*
(B) *each non-palindromic factor u of w is uniquely determined by a pair (p, q) of palindromes such that p and q are not factors of each other and p (resp. q) is the longest palindromic prefix (resp. suffix) of u.*

Theorem 6. *A finite or infinite word w is rich if and only if each factor of w is uniquely determined by its longest palindromic prefix and its longest palindromic suffix.*

By contrast, a rich word is *not* uniquely determined by its longest palindromic prefix and suffix. For example, consider the words acb and adb where a, b, c, d are mutually distinct letters. These two words are rich with the same longest palindromic prefix (namely a) and the same longest palindromic suffix (namely b).

2. Terminology and notation

In this paper, all words are taken over a finite *alphabet* $\mathcal{A}$, i.e., a finite non-empty set of symbols called *letters*. A *finite word* over $\mathcal{A}$ is a finite sequence of letters from $\mathcal{A}$. A (right) *infinite word* $\boldsymbol{x}$ is a sequence indexed by $\mathbb{N}_+$ with values in $\mathcal{A}$, i.e., $\boldsymbol{x} = x_1x_2x_3\cdots$ with each $x_i \in \mathcal{A}$.

Given a finite word $w = x_1x_2\cdots x_m$ (where each x_i is a letter), the *length* of w, denoted by $|w|$, is equal to m. We denote by $\tilde{w}$ the *reversal* of w, given by $\tilde{w} = x_m\cdots x_2x_1$ (the "mirror image" of w). If $w = \tilde{w}$, then w is called a *palindrome*. By convention, the *empty word* ε (i.e., the unique word of length 0) is assumed to be a palindrome.

A finite word z is a *factor* of a finite or infinite word w if $w = uzv$ for some words u, v. In the special case $u = \varepsilon$ (resp. $v = \varepsilon$), we call z a *prefix* (resp. *suffix*) of w. If $u \neq \varepsilon$ and $v \neq \varepsilon$, then we say that z is an *interior factor* of $w = uzv$. A *proper factor* (resp. *proper prefix*, *proper suffix*) of a word w is a factor (resp. prefix, suffix) of w that is shorter than w. We use the notation $p^{-1}w$ (resp. ws^{-1}) to indicate the removal of a prefix p (resp. suffix s) of the word w.

3. Proofs of Propositions 1 and 4

For the sake of completeness, we will first provide simple proofs of the characteristic properties stated in Propositions 1 and 4.

Proof of Proposition 1. We begin by proving the equivalence of properties i) and ii).

i) $\Leftrightarrow$ ii): Let $P(w)$ denote the number of distinct palindromic factors of w. For any word u and letter x, we have

$$P(ux) = \begin{cases} P(u) & \text{if } ux \text{ does not have a unioccurrent palindromic suffix}, \\ P(u)+1 & \text{if } ux \text{ has a unioccurrent palindromic suffix}. \end{cases}$$

Therefore, by induction (with $P(\varepsilon) = 1$), it follows that $P(w)$ is precisely the number of prefixes of w that have a unioccurrent palindromic suffix. In particular $P(w) \leq |w| + 1$, and moreover we see that $P(w) = |w| + 1$ (i.e., w is rich) if and only if each prefix of w has a unioccurrent palindromic suffix. Similarly, when w is finite, we deduce that w is rich if and only if each suffix of w has a unioccurrent palindromic prefix.

ii) $\Rightarrow$ iii): Suppose w satisfies property ii) (i.e., w is rich) and let u be any factor of w. Then $w = vuv'$ for some words v, v' where v is finite, and v' is finite or infinite depending on w. By property ii), every prefix of vu has a unioccurrent palindromic suffix, and so again by ii), every suffix of u have a unioccurrent palindromic prefix. Thus, by the equivalence of properties i) and ii), u is rich, i.e., u has exactly $|u| + 1$ distinct palindromic factors.

iii) $\Rightarrow$ iv): Suppose w satisfies property iii). Then every factor of w is rich. Hence, for each factor u of w, every prefix (resp. suffix) of u has a unioccurrent palindromic suffix (resp. prefix), by the equivalence of properties i) and ii).

iv) $\Rightarrow$ v): Suppose to the contrary that property v) does not hold for w satisfying property iv). Then w contains a non-palindromic complete return r to a palindrome p. We deduce that $r = pup$ for some non-palindromic word u. Indeed, since r is not a palindrome, $r \neq pp$ and the two occurrences of p in r cannot overlap; otherwise there exists a non-empty word v such that $r = pv^{-1}p$, in which case $p = vf = gv = \tilde{v}\tilde{g} = \tilde{p}$ for some words f, g. Whence $v = \tilde{v}$ and $r = g\tilde{v}\tilde{g} = gv\tilde{g}$, a palindrome. Now, we easily see that p is the longest palindromic suffix of r; otherwise p would occur in the interior of r as a prefix of a longer palindromic suffix of r. But then r does not have a unioccurrent palindromic suffix (as p is also a prefix of r), a contradiction.

v) $\Rightarrow$ i): Suppose not. Let u be a factor of w of minimal length satisfying property v) and not rich. Since all words of length 3 or less are rich (easy to check), we may write $u = xvy$ with x, y letters and v a word of length at least 2. By the minimality of u, xv is rich and by the equivalence of i) and ii), the longest palindromic suffix p of u occurs more than once in u. Hence, by property v), we reach a contradiction to the maximality of p. $\square$

Proof of Proposition 4. ONLY IF: Consider any factor v of w and let u be a factor of w beginning with v and ending with $\tilde{v}$ and containing neither v nor $\tilde{v}$ as an interior factor. If v is a palindrome, then either $u = v = \widetilde{v}$ (in which case u is clearly a palindrome), or u is a complete return to v in w, and hence u is (again) a palindrome by Proposition 1.

Now assume that v is not a palindrome. Suppose by way of contradiction that u is not a palindrome and let p be the longest palindromic suffix of u (which is unioccurrent in u by richness). Then $|p| < |u|$ as u is not a palindrome. If $|p| > |v|$, then $\tilde{v}$ is a proper suffix of p, and hence v is a proper prefix of p. But then v is an interior factor of u, a contradiction. On the other hand, if $|p| \leq |v|$, then $|p| \neq |v|$ and p is a proper suffix of $\tilde{v}$ (as $\tilde{v}$ is not a palindrome), and hence p is a proper prefix of v. Thus p is both a prefix and a suffix of u; in particular p is not unioccurrent in u, a contradiction.

IF: The given conditions tell us that any complete return to a palindromic factor v $(= \tilde{v})$ of w is a palindrome. Hence w is rich by Proposition 1. □

4. PROOF OF THEOREM 5

We will now prove our first main theorem. The following two lemmas establish that (A) implies (B).

Lemma 7. *Suppose w is a finite or infinite rich word and let u be any non-palindromic factor of w with longest palindromic prefix p and longest palindromic suffix q. Then $p \neq q$, and p and q are not proper factors of each other.*

Proof. By Proposition 1, p and q are unioccurrent factors of u. Thus, since u is not a palindrome (and hence $|u| > \max\{|p|, |q|\}$), it follows immediately that $p \neq q$, and p and q are not proper factors of each other. □

Lemma 8. *Suppose w is a finite or infinite rich word. If u and v are factors of w with the same longest palindromic prefix p and the same longest palindromic suffix q, then $u = v$.*

Proof. We first observe that if u or v is a palindrome, then $u = p = q = v$. So let us now assume that neither u nor v is a palindrome.

Suppose to the contrary that $u \neq v$. Then u and v are clearly not factors of each other since neither u nor v is equal to p or q, and p and q are unioccurrent in each of u and v (by Proposition 1). Let z be a factor of w of minimal length containing both u and v. As u and v are not factors of each other, we may assume without loss of generality that z begins with u and ends with v. Then z contains *at least* two distinct occurrences of p (as a prefix of each of u and v). In particular, z begins with a complete return r_1 to p with $|r_1| > |u|$ because p is unioccurrent in u by Proposition 1. Moreover, r_1 is a palindrome by the richness of w, and hence r_1 ends with $\tilde{u}$ since u is a proper prefix of r_1. Similarly, z ends with a complete return r_2 to q with $|r_2| > |v|$ since q is unioccurrent in v by Proposition 1. Hence, since r_2 is a palindrome (by the richness of

w) and v is a proper suffix of r_2, it follows that r_2 begins with $\tilde{v}$. So we have shown that $\tilde{u}$ and $\tilde{v}$ are (distinct) interior factors of z.

Let us first suppose that an occurrence of $\tilde{v}$ is followed by an occurrence of $\tilde{u}$ in z (i.e., z has an interior factor beginning with $\tilde{v}$ and ending with $\tilde{u}$). Then, since q is a unioccurrent prefix of each of the (distinct) factors $\tilde{v}$ and $\tilde{u}$, we deduce that z contains (as an interior factor) a complete return r_3 to q beginning with $\tilde{v}$. In particular, as r_3 is a palindrome (by richness), r_3 ends with v. Thus, z has a proper prefix beginning with u and ending with v, contradicting the minimality of z. On the other hand, if z has an interior factor beginning with $\tilde{u}$ and ending with $\tilde{v}$, then using the same reasoning as above, we deduce that z has a proper suffix beginning with u and ending with v. But again, this contradicts the minimality of z; whence $u = v$. ☐

The proof of "(A) ⇒ (B)" is now complete. The next lemma proves that (B) implies (A).

Lemma 9. *Suppose w is a finite or infinite word with the property that each non-palindromic factor u of w is uniquely determined by a pair (p, q) of distinct palindromes such that p and q are not factors of each other and p (resp. q) is the longest palindromic prefix (resp. suffix) of u. Then w is rich.*

Proof. To prove that w is rich, it suffices to show that each prefix of w has a unioccurrent palindromic suffix (see Proposition 1).

Let u be any prefix of w and let q be the longest palindromic suffix of u. We first observe that if u is a palindrome then $u = q$, and hence q is unioccurrent in u. Now let us suppose that u is not a palindrome and let p be the longest palindromic prefix of u. If q is not unioccurrent in u, then, as p and q are not factors of each other (by the given property of w), we deduce that u has a *proper* factor v beginning with p and ending with q and containing neither p nor q as an interior factor. Moreover, we observe that p is the longest palindromic prefix of v; otherwise p would occur in the interior of v (as a suffix of a longer palindromic prefix of v). Similarly, we deduce that q is the longest palindromic suffix of v. So v has the same longest palindromic prefix and the same longest palindromic suffix as u, a contradiction. Whence q is unioccurrent in u. This completes the proof of the lemma. ☐

5. Proof of Theorem 6

Lemma 8 proves that each factor of a rich word is uniquely determined by its longest palindromic prefix and its longest palindromic suffix.

Conversely, suppose w is a finite or infinite word with the property that each factor of w is uniquely determined by its longest palindromic prefix and its longest palindromic suffix. To prove that w is rich, we could use very similar reasoning as in the proof of Lemma 9. But for the sake of interest, we give a slightly different proof. Specifically, we show that all complete returns to any palindromic factor of w are palindromes; whence w is rich by Proposition 1.

Let us suppose to the contrary that w contains a non-palindromic complete return r to a palindromic factor p. Then $r = pvp$ for some non-palindromic word v (as already observed in the proof vi) ⇒ v) in Proposition 1). We easily see that p is both the longest palindromic prefix and the longest palindromic suffix of r; otherwise p would occur in the interior of r as a suffix of a longer palindromic prefix of r, or as a prefix of a longer palindromic suffix of r. As $r \neq p$, we have reached a contradiction to the fact that p is the *only* factor of w having itself as both its longest palindromic prefix and its longest palindromic suffix. Thus, all complete returns to p in w are palindromes. This completes the proof of Theorem 6. □

REFERENCES

[1] J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik, Palindrome complexity, *Theoret. Comput. Sci.* 292 (2003) 9–31.

[2] P. Ambrož, C. Frougny, Z. Masáková, E. Pelantová, Palindromic complexity of infinite words associated with simple Parry numbers, *Ann. Inst. Fourier (Grenoble)* 56 (2006) 2131–2160.

[3] M. Bucci, A. De Luca, A. Glen, L.Q. Zamboni, A connection between palindromic and factor complexity using return words, *Adv. in Appl. Math.* (in press), doi:10.1016/j.aam.2008.03.005.

[4] A. de Luca, A. Glen, L.Q. Zamboni, Rich, Sturmian, and trapezoidal words, *Theoret. Comput. Sci.* (in press), doi:10.1016/j.tcs.2008.06.009.

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

[6] S. Ferenczi, L.Q. Zamboni, Language of k-interval exchange transformations, *Bull. London Math. Soc.* 40 (2008) 705–714.

[7] S. Ferenczi, L.Q. Zamboni, Structure of k-interval exchange transformations: induction, trajectories, and distance theorems, Preprint, 2008.

[8] S. Fischler, Palindromic prefixes and episturmian words, *J. Combin. Theory Ser. A* 113 (2006) 1281–1304.

[9] A. Glen, J. Justin, Episturmian words: a survey, Preprint, 2007, arXiv:0801.1655.

[10] A. Glen, J. Justin, S. Widmer, L.Q. Zamboni, Palindromic richness, *European J. Combin.* (in press), doi:10.1016/j.ejc.2008.04.006.

DIPARTIMENTO DI MATEMATICA E APPLICAZIONI "R. CACCIOPPOLI", UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II, VIA CINTIA, MONTE S. ANGELO, I-80126 NAPOLI, ITALY
E-mail address: micbucci@unina.it

DIPARTIMENTO DI MATEMATICA E APPLICAZIONI "R. CACCIOPPOLI", UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II, VIA CINTIA, MONTE S. ANGELO, I-80126 NAPOLI, ITALY
E-mail address: alessandro.deluca@unina.it

LACIM, UNIVERSITÉ DU QUÉBEC À MONTRÉAL, C.P. 8888, SUCCURSALE CENTRE-VILLE, MONTRÉAL, QUÉBEC, H3C 3P8, CANADA / THE MATHEMATICS INSTITUTE, REYKJAVIK UNIVERSITY, KRINGLAN 1, IS-103 REYKJAVIK, ICELAND
E-mail address: amy.glen@gmail.com

UNIVERSITÉ DE LYON, UNIVERSITÉ LYON 1, CNRS UMR 5208 INSTITUT CAMILLE JORDAN, BÂTIMENT DU DOYEN JEAN BRACONNIER, 43, BLVD DU 11 NOVEMBRE 1918, F-69622 VILLEURBANNE CEDEX, FRANCE / SCHOOL OF COMPUTER SCIENCE, REYKJAVIK UNIVERSITY, KRINGLAN 1, IS-103 REYKJAVIK, ICELAND
E-mail address: luca.zamboni@wanadoo.fr; lqz@ru.is