Biinfinite words with maximal recurrent unbordered factors

José Carlos Costa *

March 26, 2002

Abstract

A finite non-empty word z is said to be a *border* of a finite non-empty word w if $w = uz = zv$ for some non-empty words u and v. A finite non-empty word is said to be *bordered* if it admits a border, and it is said to be *unbordered* otherwise. In this paper, we give two characterizations of the biinfinite words of the form ${}^{\omega}uvu^{\omega}$, where u and v are finite words, in terms of its unbordered factors.

The main result of the paper states that the words of the form ${}^{\omega}uvu^{\omega}$ are precisely the biinfinite words $\mathbf{w} = \cdots a_{-2}a_{-1}a_0a_1a_2\cdots$ for which there exists a pair (l_0, r_0) of integers with $l_0 < r_0$ such that, for every integers $l \leq l_0$ and $r \geq r_0$, the factor $a_l \cdots a_{l_0} \cdots a_{r_0} \cdots a_r$ is a bordered word.

The words of the form ${}^{\omega}uvu^{\omega}$ are also characterized as being those biinfinite words $\mathbf{w}$ that admit a left recurrent unbordered factor (i.e., an unbordered factor of $\mathbf{w}$ that has an infinite number of occurrences "to the left" in $\mathbf{w}$) of maximal length that is also a right recurrent unbordered factor of maximal length. This last result is a biinfinite analogue of a result known for infinite words.

1 Introduction

This paper is concerned with a combinatorial problem on biinfinite words which has arisen in the study of a certain class of finite semigroups: the pseudovariety **LSl** of locally idempotent and locally commutative semigroups. The class **LSl** is formed by the finite semigroups S such that $eSe = e$ for each element $e = e^2 \in S$, and is associated via Eilenberg's correspondence with the well known class of locally testable languages, as shown independently by Brzozowski and Simon [2] and McNaughton [7]. Recall that a language L is locally testable if one can decide membership of a given word u in L by considering the factors of a fixed length k of u and its prefix and suffix of length $k-1$. Alternatively, a locally testable language is a language that is a Boolean combination of languages of the form wA^*, A^*w and A^*wA^*, where A is a finite alphabet and w is a word on A. On the other hand, the free pro-**LSl** semigroups, — topological semigroups which play an important role in the study of the pseudovariety **LSl**, — were described by the author [3] in terms of infinite and biinfinite words. It is not surprising therefore that the study of the pseudovariety **LSl** must often use combinatorial properties of words, namely involving infinite and biinfinite words and factors of words.

*Work supported, in part, by FCT through the *Centro de Matemática da Universidade do Minho*, and by the FCT and POCTI approved project POCTI/32817/MAT/2000 which is comparticipated by the European Community Fund FEDER.

The original question, motivated by the study mentioned above, is an interesting property involving the notion of unbordered word. That question is the following: given a biinfinite word $\mathbf{w}$ and a fixed occurrence of a factor of $\mathbf{w}$, is there an occurrence, extending the fixed one, of an unbordered factor of $\mathbf{w}$? Of course the answer to this problem is negative in general, since for the words of the form ${}^{\omega}uvu^{\omega}$, any factor extending uvu is bordered. In this paper, we show that the words of the form ${}^{\omega}uvu^{\omega}$ are the only biinfinite words for which the question above has a negative answer. The ultimately periodic words of the form ${}^{\omega}uvu^{\omega}$ are also shown to be the unique biinfinite words that admit a left recurrent unbordered factor of maximal length that is also a right recurrent unbordered factor of maximal length.

Also known as "primary words" and "mots latéraux", unbordered words have been widely studied by the community. This article deals with the relation between the length of unbordered factors and periodicity in infinite and biinfinite words, and it constitutes an extension of previous works on this subject [1, 4, 5, 6]. In particular, the second characterization of the words of the form ${}^{\omega}uvu^{\omega}$ mentioned above, is the "biinfinite version" of a result established by Ehrenfeucht and Silberger [5] for infinite words.

2 Preliminaries

In this section we recall basic definitions and introduce notations that will be used later. We follow in most part the terminology of Lothaire [6] for finite words and of Perrin and Pin [8] for infinite and biinfinite words.

A finite non-empty set A is called an *alphabet*. The elements of A are called *letters*. A *(finite) word* on A is a finite sequence $w = (a_1, \ldots, a_n)$ of elements of A. We write also $w = a_1 \cdots a_n$. The integer n is called the *length* of w. The empty sequence, called the *empty word*, is denoted by 1 and its length is 0. The length of a word w is denoted by $|w|$. We denote by A^* the set of words on A and by A^+ the set of non-empty words. The *product* of two words $w = a_1a_2 \cdots a_n$ and $z = b_1b_2 \cdots b_m$ is the word $wz = a_1a_2 \cdots a_nb_1b_2 \cdots b_m$.

A word $w \in A^+$ is said to be *primitive* if it is not a power of another word; that is, if $w = u^n$ for some $u \in A^*$ and $n \in \mathbb{N}$ implies $w = u$ (and $n = 1$).

Two words w and z are said to be *conjugate* if there exist words $u, v \in A^*$ such that

$$w = uv, \quad z = vu.$$

A *biinfinite* (resp. *right infinite*, *left infinite*) word on A is a sequence $\mathbf{w} = (a_n)_n$ of letters of A indexed by $\mathbb{Z}$ (resp. $\mathbb{N}$, $-\mathbb{N}$). We denote

$$\mathbf{w} = \cdots a_{-2}a_{-1}a_0a_1a_2 \cdots \quad (\text{resp. } \mathbf{w} = a_1a_2 \cdots, \quad \mathbf{w} = \cdots a_{-2}a_{-1}).$$

The sets of biinfinite, right infinite and left infinite words on A will be denoted, respectively, by $A^{\mathbb{Z}}$, $A^{\mathbb{N}}$ and $A^{-\mathbb{N}}$.

For words $u = a_1a_2 \cdots a_n \in A^+$ and $v = b_1b_2 \cdots b_m \in A^*$, we denote by vu^{ω} the right infinite word

$$vu^{\omega} = vuuu \cdots = b_1b_2 \cdots b_ma_1a_2 \cdots a_na_1a_2 \cdots a_na_1a_2 \cdots a_n \cdots$$

obtained by the infinite repetition (to the right) of the word u after the word v. The word vu^{ω} is said to be *ultimately periodic* and u is said to be a *period* of vu^{ω}. We will use the notation ${}^{\omega}uv$ to represent the (*ultimately periodic of period* u) left infinite word

$${}^{\omega}uv = \cdots uuuv = \cdots a_1a_2 \cdots a_na_1a_2 \cdots a_na_1a_2 \cdots a_nb_1b_2 \cdots b_m$$

obtained by the infinite repetition (to the left) of the word u after the word v.

Let $\mathbf{w} = \cdots a_{-2}a_{-1}a_0a_1a_2\cdots$ be a biinfinite word. For integers i and j such that $i < j$, we denote

$$\mathbf{w}[i,i[=1, \quad \mathbf{w}[i,j[=a_i\cdots a_{j-1}, \quad \mathbf{w}[i,+\infty[=a_ia_{i+1}\cdots .$$

Analogously one would define $\mathbf{w}]i,j]$, $\mathbf{w}[i,j]$, $\mathbf{w}]-\infty,j[$, etc. When they make sense, these notations are used also for finite and infinite words. We say that $\mathbf{w}$ is of the form

$$^{\omega}uvu^{\omega}$$

where $u \in A^+$ and $v \in A^*$, if there exists an integer i such that $\mathbf{w}]-\infty,i[={}^{\omega}u$, and $\mathbf{w}[i,+\infty[= vu^{\omega}$.

A finite word $u \in A^*$ is a *factor* of a (finite, infinite or biinfinite) word w if $u = w[i,j[$ for some integers i and j. In this case $w[i,j[$ is said to be an *occurrence* of the factor u in w. We will say also "the occurrence $u = w[i,j[$ in w" instead of "the occurrence $w[i,j[$ of u in w".

Let w be a (finite, infinite or biinfinite) word. The set of letters that occur in w is denoted by $\mathrm{Alph}(w)$.

Let $\mathbf{w} \in A^{\mathbb{N}} \cup A^{-\mathbb{N}}$ be an infinite word. A factor of $\mathbf{w}$ that has an infinite number of occurrences in $\mathbf{w}$ is said to be *recurrent* in $\mathbf{w}$. If each factor of $\mathbf{w}$ is recurrent in $\mathbf{w}$, then $\mathbf{w}$ is said to be *recurrent*.

Let $\mathbf{w} \in A^{\mathbb{Z}}$ be a biinfinite word. A factor $u \in A^+$ of $\mathbf{w}$ is said to be *left recurrent* in $\mathbf{w}$ if u is recurrent in a (and so in any) left infinite word of the form $\mathbf{w}]-\infty,i]$. Analogously one can define the notion of a *right recurrent* factor of $\mathbf{w}$. A factor of $\mathbf{w}$ that is simultaneously left recurrent and right recurrent is called *recurrent*.

A word $u \in A^*$ is said to be a *prefix* (resp. a *suffix*) of a word $w \in A^*$, and w is said to be a *right extension* (resp. a *left extension*) of u, if there exists a word v such that $w = uv$ (resp. $w = vu$); if $u \neq w$, then u is said to be a *proper prefix* (resp. a *proper suffix*) of w and w is said to be a *proper right extension* (resp. a *proper left extension*) of u.

Let y be a bordered word and let $\overline{y}$ be the shortest border of y. Then $\overline{y}$ is an unbordered word since, otherwise, $\overline{y}$ would admit a border z and, clearly, this word z would be a border of y shorter than $\overline{y}$. Notice that $|y| \geq 2|\overline{y}|$ so that

$$y = \overline{y}u\overline{y}$$

for some $u \in A^*$ since, otherwise, $\overline{y}$ would have a border.

Let $w \in A^+$ be a word (bordered or not) with $|w| \geq 2$. We will represent by $\overrightarrow{w}$ (resp. $\overleftarrow{w}$) the longest unbordered word that is a proper prefix (resp. suffix) of w. For instance $w = ababbaabb$ is an unbordered word such that $\overrightarrow{w} = ababb$ and $\overleftarrow{w} = aabb$.

3 The characterizations

We begin by presenting a characterization of the ultimately periodic right infinite words. This result was established by Ehrenfeucht and Silberger in [5, Lemma 3.3].

Lemma 3.1 *A right infinite word* $\mathbf{w} \in A^{\mathbb{N}}$ *is ultimately periodic if and only if there exists an unbordered factor of* $\mathbf{w}$ *of maximal length that is recurrent in* $\mathbf{w}$.

Remark that, if $\mathbf{w} \in A^{\mathbb{N}}$ is an ultimately periodic word of period u and x is an unbordered factor of $\mathbf{w}$ as stated in this last result, then x is a conjugate of the (unique) primitive word z such that $u = z^n$ for some $n \in \mathbb{N}$. In particular, x is a period of $\mathbf{w}$.

The following observation will be important in what follows. This result was given as a remark in [1] and established and proved in [4, Corollary 2.8].

Lemma 3.2 *Let $w \in A^+$ be such that $\overrightarrow{w} = az$, where $a \in A$ and $z \in A^+$, and let $b \in A$ be a letter distinct from a. If bz is a factor of w, then w is an unbordered word and bz only occurs in w as a suffix.*

We can now state and prove the main result of the paper.

Theorem 3.3 *Let $\mathbf{w} \in A^{\mathbb{Z}}$ be a biinfinite word. The following conditions are equivalent:*

1) *The word $\mathbf{w}$ is of the form ${}^{\omega}uvu^{\omega}$ for some finite words u and v;*

2) *There exists an unbordered recurrent factor x of $\mathbf{w}$ such that:*

   i) x is of maximal length between the unbordered left recurrent factors of $\mathbf{w}$;

   ii) x is of maximal length between the unbordered right recurrent factors of $\mathbf{w}$;

3) *There exists an occurrence $\mathbf{w}[l_0, r_0]$ in $\mathbf{w}$ such that, for every integers $l \leq l_0$ and $r \geq r_0$, the factor $\mathbf{w}[l, r]$ is a bordered word.*

Proof. 1) $\Leftrightarrow$ 2) This equivalence is an immediate consequence of Lemma 3.1 and its dual for left infinite words.

1) $\Rightarrow$ 3) Suppose that $\mathbf{w}$ is of the form ${}^{\omega}uvu^{\omega}$. Then there exists an integer i such that $\mathbf{w}]-\infty, i[= {}^{\omega}u$ and $\mathbf{w}[i, +\infty[= vu^{\omega}$. Consider the occurrence

$$uvu = \mathbf{w}[i - |u|, i + |vu| - 1] = \mathbf{w}[l_0, r_0].$$

Let $l \leq l_0$ and $r \geq r_0$ be two integers and consider the factor $w = \mathbf{w}[l, r]$ of $\mathbf{w}$. Then, there exist a proper suffix u' of u, a proper prefix u'' of u and positive integers k' and k'' such that

$$w = u'u^{k'}vu^{k''}u''.$$

If $u' = u'' = 1$, then u is a border of w. If $u' \neq 1$ or $u'' \neq 1$, then $u'u''$ is a border of w. Therefore, the factor $w = \mathbf{w}[l, r]$ is a bordered word.

3) $\Rightarrow$ 1) Let $\mathbf{w}[l_0, r_0]$ be an occurrence of a factor in $\mathbf{w}$ and suppose that for every integers $l \leq l_0$ and $r \geq r_0$, the factor $\mathbf{w}[l, r]$ is a bordered word. Denote by $\mathbf{w}_{l_0}$ the left infinite word $\mathbf{w}]-\infty, l_0[$ and by $\mathbf{w}_{r_0}$ the right infinite word $\mathbf{w}]r_0, +\infty[$.

We begin by proving the following crucial lemma.

Lemma 3.4 *The words $\mathbf{w}_{l_0}$ and $\mathbf{w}_{r_0}$ are recurrent and have the same factors.*

Proof. We prove the lemma by induction on the length of the factors of $\mathbf{w}_{l_0}$ and $\mathbf{w}_{r_0}$.

Let $a = \mathbf{w}[i, i]$, with $i < l_0$, be a letter of $\mathrm{Alph}(\mathbf{w}_{l_0})$ and let $q = r_0 - i$. Let $y = \mathbf{w}[i, n]$ where $n > r_0 + q$ (as illustrated in Fig. 1). Since y contains the occurrence $\mathbf{w}[l_0, r_0]$, y is bordered. The shortest border $\overline{y}$ of y is an unbordered word that is a prefix of y. Therefore, the length of $\overline{y}$ is $\leq q$ since otherwise $\overline{y}$ would contain the occurrence $\mathbf{w}[l_0, r_0]$. Hence, the choice of n and the fact that $\overline{y}$ is a suffix of y show that $\overline{y}$ is a factor of $\mathbf{w}_{r_0}$. Since a is the first letter of $\overline{y}$, we deduce that a is a factor of $\mathbf{w}_{r_0}$.

Furthermore, since n is arbitrarily large, this proves that a has an infinite number of occurrences in $\mathbf{w}_{r_0}$. By symmetry, we conclude that

$$\mathrm{Alph}(\mathbf{w}_{l_0}) = \mathrm{Alph}(\mathbf{w}_{r_0})$$

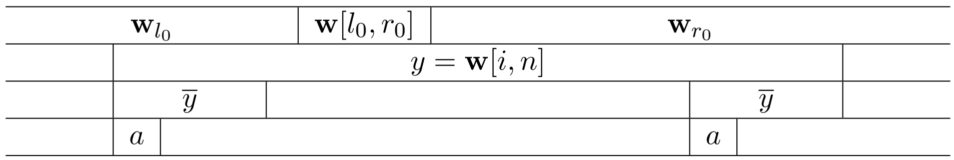

Figure 1:

and that each letter of Alph($\mathbf{w}_{l_0}$) is recurrent in both $\mathbf{w}_{l_0}$ and $\mathbf{w}_{r_0}$.

As a consequence, if Alph($\mathbf{w}_{l_0}$) =Alph($\mathbf{w}_{r_0}$) $= \{a\}$ for some letter a, then $\mathbf{w}_{l_0} = {}^{\omega}a$ and $\mathbf{w}_{r_0} = a^{\omega}$ and the lemma is clearly valid. For the rest of the proof, we assume that Alph($\mathbf{w}_{l_0}$) is not trivial.

Let now $k > 1$ be an integer and assume by induction hypothesis that $\mathbf{w}_{l_0}$ and $\mathbf{w}_{r_0}$ have the same factors of length $k-1$ and that these factors are recurrent in both $\mathbf{w}_{l_0}$ and $\mathbf{w}_{r_0}$. Let

$$w = \mathbf{w}[i, j]$$

be a factor of $\mathbf{w}_{l_0}$ of length k.

Suppose first that w is of the form $w = a^k$ for some letter a. By assumption, Alph($\mathbf{w}_{r_0}$) contains a letter $b \neq a$. Let $b = \mathbf{w}[n, n]$ be an occurrence in $\mathbf{w}_{r_0}$ and let $y = \mathbf{w}[i, n]$. Then y is bordered and, as b is recurrent in $\mathbf{w}_{r_0}$, we may choose the occurrence of b in such a way that the occurrence of $\overline{y}$ as a suffix of y is contained in $\mathbf{w}_{r_0}$. On the other hand b is a suffix of $\overline{y}$. Since b is not a factor of w and $\overline{y}$ is a prefix of y, we deduce that w is a prefix of $\overline{y}$. Therefore w is a factor of $\mathbf{w}_{r_0}$. Moreover since n is arbitrarily large and the length of $\overline{y}$ is upper bounded, we conclude that w is recurrent in $\mathbf{w}_{r_0}$.

Suppose now that w is not of the form $w = a^k$ with $a \in A$. Let w' be the suffix of w of length $k-1$. By induction hypothesis, w' is recurrent in $\mathbf{w}_{r_0}$. Let

$$w' = \mathbf{w}[m, n]$$

be an occurrence in $\mathbf{w}_{r_0}$. We will consider two cases.

First case Suppose first that w is an unbordered word. Let $y = \mathbf{w}[i, n]$. The word y is bordered and, since n can be chosen arbitrarily large, we may assume that the occurrence of $\overline{y}$ as a suffix of y is contained in $\mathbf{w}_{r_0}$. Since w' is a suffix of w and w is unbordered, w' does not have any suffix that is a prefix of w. Therefore $|\overline{y}| > |w'| = k-1$ and so w is a prefix of $\overline{y}$. Thus, we deduce that w is a factor of $\mathbf{w}_{r_0}$. Moreover, as above, we conclude that also in this case w is recurrent in $\mathbf{w}_{r_0}$.

Second case Suppose now that w is a bordered word. Let

$$\overrightarrow{w} = az$$

be the longest unbordered proper prefix of w, where $a \in A$. Recall that we are assuming that w is not of the form $w = a^k$ with $a \in A$. Therefore z is not the empty word since Alph(w) contains at least one letter $c \neq a$ and w is not of the form $w = a^{k-1}c$ because we are assuming w to be bordered. Let b be the letter $b = \mathbf{w}[m-1, m-1]$ and consider the occurrence

$$bw' = \mathbf{w}[m-1, n].$$

If b is equal to a, then $bw' = w$ and so w is a factor of $\mathbf{w}_{r_0}$ since we may assume that $m-1 > r_0$.

Suppose that $b \neq a$. As z is a prefix of w', the word $bz = \mathbf{w}[m-1, h]$, where $m-1 < h \leq n$, is a prefix of bw'. On the other hand, by Lemma 3.2, the word bz is not a factor of w since w is bordered. Consider the bordered word $y = \mathbf{w}[i, h]$. Since y is a word of the form

$$y = azy'bz \ (y' \in A^*)$$

and az is an unbordered word, we have $|\overline{y}| \geq |az| = |bz|$ whence bz is a suffix of $\overline{y}$. On the other hand y is of the form

$$y = wy''bz \ (y'' \in A^*)$$

and so, since bz is not a factor of w, $|\overline{y}| > |w|$. This proves that w is a prefix of $\overline{y}$. Since, as above, we may assume that $\overline{y}$ is a factor of $\mathbf{w}_{r_0}$, we deduce that w is a factor of $\mathbf{w}_{r_0}$. Moreover, as above, w is recurrent in $\mathbf{w}_{r_0}$.

Therefore, we have proved in all cases that w is a recurrent factor of $\mathbf{w}_{r_0}$. By symmetry, we deduce that $\mathbf{w}_{l_0}$ and $\mathbf{w}_{r_0}$ have the same factors of length k and that these factors are recurrent in both $\mathbf{w}_{l_0}$ and $\mathbf{w}_{r_0}$.

The result follows by induction. □

Let us now return to the proof of Theorem 3.3. Let $r_1 < r_0$ be the maximal integer such that there exists an unbordered factor x of $\mathbf{w}_{l_0}$ with an occurrence of the form

$$x = \mathbf{w}[i, r_1] \tag{1}$$

where $i \leq l_0$. Now, let

$$x_1 = \mathbf{w}[l_1, r_1]$$

be the unbordered word of the form (1) with minimal length; that is, such that $l_1 \leq l_0$ is maximal.

Remark. Notice that $r_1 \geq l_0$ since, for instance, the first letter $\mathbf{w}[l_0, l_0]$ of the factor $\mathbf{w}[l_0, r_0]$ is surely an unbordered factor of $\mathbf{w}_{r_0}$ and $l_0 < r_0$ since $\mathbf{w}[l_0, r_0]$ is bordered.

Notice furthermore that the word $\mathbf{w}]-\infty, r_1]$ has the same factors as $\mathbf{w}_{r_0}$. That is, every word of the form $\mathbf{w}[i, r_1]$ is a factor of $\mathbf{w}_{r_0}$. This is clear when $i \geq l_0$ since in this case $\mathbf{w}[i, r_1]$ is a factor of x_1 and x_1 is a factor of $\mathbf{w}_{r_0}$. That $\mathbf{w}[i, r_1]$, with $i \leq l_0$, is a factor of $\mathbf{w}_{r_0}$ can be shown as in Lemma 3.4.

Notice at last that $\mathrm{Alph}(\mathbf{w}_{l_0}) = \mathrm{Alph}(x_1)$. In fact, if $a \in \mathrm{Alph}(\mathbf{w}_{l_0})$ and $a = \mathbf{w}[n, n]$ is an occurrence in $\mathbf{w}_{r_0}$, then $y = \mathbf{w}[l_0, n]$ is bordered and $a \in \mathrm{Alph}(\overline{y})$. Now, $\overline{y}$ is of the form (1) whence $\overline{y}$ is a factor of x_1 and so $a \in \mathrm{Alph}(x_1)$. ■

The next lemma will permit us to obtain the word u stated in condition 1) of Theorem 3.3.

Lemma 3.5 *For every integer $k > |x_1|$, there exists exactly one left extension w of x_1 of length k such that w is a factor of $\mathbf{w}_{l_0}$.*

Proof. Let $k > |x_1|$ be an integer. The factor

$$w = \mathbf{w}[r_1 - k - 1, r_1] = \mathbf{w}[r_1 - k - 1, l_1[x_1$$

of $\mathbf{w}_{l_0}$ is a left extension of x_1 of length k.

Now, suppose that there exists a factor $w' \neq w$ of $\mathbf{w}_{l_0}$ of length k such that x_1 is a suffix of w'. Assume that k is the minimal integer for which this happens. Then

$$w = aw'' \quad \text{and} \quad w' = bw''$$

where a and b are the (distinct) first letters of w and w', respectively, and w'' is the suffix of length $k-1$ of both w and w'. In particular $\mathrm{Alph}(\mathbf{w}_{l_0})$ is not trivial and so, we deduce from the remark above that $\mathrm{Alph}(x_1)$ is also not trivial and that $\overrightarrow{w}$ is of the form

$$\overrightarrow{w} = az$$

for some prefix $z \neq 1$ of w''. From Lemma 3.2, we deduce that either bz is not a factor of w or bz only occurs in w as a suffix. Let us consider these two cases.

First case Suppose that bz is not a factor of w. Since $\mathbf{w}_{l_0}$ and $\mathbf{w}_{r_0}$ have the same factors, we can choose an occurrence $bz = \mathbf{w}[m,n]$ in $\mathbf{w}_{r_0}$ and consider the bordered word $y = \mathbf{w}[r_1 - k - 1, n]$. Now, as in the second case of Lemma 3.4, one can deduce that $|\overline{y}| > |w|$. This is absurd because in that case $\overline{y}$ is an unbordered factor of $\mathbf{w}_{l_0}$ with an occurrence of the form

$$\overline{y} = \mathbf{w}[r_1 - k - 1, i]$$

with $r_1 < i < r_0$, contradicting the choice of r_1.

Second case Suppose that bz is a suffix of w. Then w'' is of the form $w'' = ebz$ for some $e \in A^*$. Let $w' = bw'' = \mathbf{w}[m,n]$ be an occurrence of w' in $\mathbf{w}_{r_0}$ and consider the bordered word $y = \mathbf{w}[r_1 - k - 1, n]$. As in the proof of the second case in Lemma 3.4, one deduces that bz is a suffix of $\overline{y}$. Therefore, since bz only occurs in w as a suffix, $|\overline{y}| \geq |w|$. But, as we saw in the first case, $|\overline{y}| \leq |w|$. Therefore $|\overline{y}| = |w|$ and so $w = \overline{y} = w'$. This contradicts the assumption that $w \neq w'$.

Therefore w' does not exist, concluding the proof of the lemma. □

Let us return again to the proof of Theorem 3.3. Let $x_1 = \mathbf{w}[i_1, j_1]$ and $x_1 = \mathbf{w}[i_2, j_2]$ be, respectively, the first and the second occurrences of x_1 in $\mathbf{w}_{r_0}$. Let $w = \mathbf{w}[i_1, j_2]$. Then w has exactly two occurrences of x_1 and it is of the form

$$w = x_1 u = y x_1$$

for some words $u, y \in A^+$. Moreover, since x_1 is unbordered, $|u| \geq |x_1|$ whence u is a left extension of x_1.

Let w' be any factor of $\mathbf{w}_{r_0}$ with exactly two occurrences of x_1, being those occurrences of x_1 as prefix and as suffix of w'. Then $w' = w$ since otherwise, by Lemma 3.5, one of w or w' would be a proper left extension of the other, which is impossible since they have both exactly two occurrences of x_1.

Since x_1 is recurrent in $\mathbf{w}_{r_0}$, this means that $\mathbf{w}_{r_0}$ is of the form $\mathbf{w}_{r_0} = zu^\omega$ for some $z \in A^*$. Moreover, since $\mathbf{w}_{r_0}$ is recurrent, the factor z of $\mathbf{w}_{r_0}$ is of the form $z = u'u^k$ for some suffix u' of u and $k \in \mathbb{N}_0$, so that

$$\mathbf{w}_{r_0} = u'u^\omega.$$

On the other hand, since $\mathbf{w}_{l_0}$ has the same factors of $\mathbf{w}_{r_0}$, $\mathbf{w}_{l_0}$ is of the form

$$\mathbf{w}_{l_0} = {}^{\omega}u u''$$

for some prefix u'' of u. We conclude that $\mathbf{w}$ is of the form

$$^{\omega}uvu^{\omega}$$

where $v = u''\mathbf{w}[l_0, r_0]u'$, which establishes condition 1) of Theorem 3.3.

This concludes the proof of Theorem 3.3. □

4 Acknowledgements

I would like to thank Assis Azevedo and Lurdes Teixeira for useful suggestions on a first version of this paper.

References

[1] R. Assous, M. Pouzet, Une caractérisation des mots périodiques, Discrete Mathematics 25 (1979) 1-5.

[2] J. Brzozowski, I. Simon, Characterization of locally testable events, Discrete Mathematics 4 (1973) 243-271.

[3] J. C. Costa, Free profinite locally idempotent and locally commutative semigroups, Journal of Pure and Applied Algebra 163 (2001) 19-47.

[4] J. P. Duval, Relationship between the period of a finite word and the length of its unbordered segments, Discrete Mathematics 40 (1982) 31-44.

[5] A. Ehrenfeucht, D. M. Silberger, Periodicity and unbordered segments of words, Discrete Mathematics 26 (1979) 101-109.

[6] M. Lothaire, Combinatorics on words, Cambridge University Press, 1997.

[7] R. McNaughton, Algebraic decision procedures for local testability, Math. Systems and Theory 8 (1974) 60-76.

[8] D. Perrin, J.-E. Pin, Mots infinis, Tech. Report LITP 97/04, 1997.