International Journal of Foundations of Computer Science
Vol. 15, No. 2 (2004) 293–306
© World Scientific Publishing Company

ON THE PALINDROMIC COMPLEXITY OF INFINITE WORDS*

S. BRLEK†, S. HAMEL‡, M. NIVAT§ and C. REUTENAUER
†*LACIM, Université du Québec à Montréal*
Montréal (QC) CANADA H3C 3P8
‡*McGill University*
§*LIAFA, Université Denis Diderot*
¶*LaCIM, Université du Québec à Montréal,*
Montréal (QC) CANADA H3C 3P8
†*brlek@lacim.uqam.ca*
‡*hamel@cs.mcgill.ca*
§*mnivat@lacim.uqam.ca*
¶*christo@lacim.uqam.ca*

Received 14 May 2003
Accepted 20 May 2003
Communicated by Juhani Karhumaki

We study the problem of constructing infinite words having a prescribed finite set P of palindromes. We first establish that the language of all words with palindromic factors in P is rational. As a consequence we derive that there exists, with some additional mild condition, infinite words having P as palindromic factors. We prove that there exist periodic words having the maximum number of palindromes as in the case of Sturmian words, by providing a simple and easy to check condition. Asymmetric words, those that are not the product of two palindromes, play a fundamental role and an enumeration is provided.

1. Introduction

The palindrome complexity has received a noticeable attention in various classes of infinite words, and a survey was recently provided by Allouche and al. [3]. The role of palindromic factors in words is ubiquitous. It gives an insight on the intrinsic structure, due to its connection with the usual complexity, the characterization it provides for Sturmian words [10] or the relation with the notion of recurrence [4]. Many other connections exist and we refer to the survey [3] for further reading.

In this paper we consider first the language L_P of words whose palindromic factors belong to a fixed and finite set P of palindromes. The set L_P turns out to be the complement of a local language and therefore is rational. It is easy then to obtain a finite automaton which recognizes L_P. The usual construction can be

*with the support of NSERC and CRC(Canada).

avoided by considering a special case of the de Bruijn graph of factors from which the finite automaton is easily derived. An easy consequence is that if there exists a recurrent infinite word having P for palindromic factors, then there exist a periodic one sharing exactly the same palindromic factors.

In Section 4 we consider infinite periodic words. By using a representation of periodic words by circular words, we obtain a geometric characterization of those having an infinite set of palindromic factors, as well as those having a finite one. More precisely, an infinite periodic word W of primitive period w has an infinite set of palindromes if and only if w is the product of two palindromes. Consequently the periodic words having a finite set of palindromes are the words whose smallest periodic pattern is *asymmetric*.

In a recent paper, Droubay, Justin and Pirillo, showed that finite Sturmian (and even episturmian) words, satisfy the property that their length plus 1 equals the number of its palindromic factors [6], that is to say, they realize the upper bound of the palindrome complexity. We show that with some easy to check condition, a primitive circular word provides a similar result for periodic words.

We also provide an enumeration formula for asymmetric words by using the work of Stockmeyer [12]. Finally, we propose some conjectures about the Lie algebras associated to these words.

2. Definitions and notations

We borrow from M. Lothaire [9] the basic terminology about words. In what follows, Σ is a finite *alphabet* whose elements are called *letters*. By *word* we mean a finite sequence of letters

$$w : [1..n] \longrightarrow \Sigma\ , \quad n \in \mathbb{N},$$

of length n, w_i denotes its i-th letter. The set of n-length words over Σ is denoted Σ^n. By convention the *empty* word is denoted ϵ and its length is 0. The free monoid generated by Σ is defined by $\Sigma^* = \bigcup_{n\geq 0} \Sigma^n$. The set of right infinite words is denoted by Σ^ω and that of bi-infinite words by ${}^\omega\Sigma^\omega$. Also $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$.

Given a word $w \in \Sigma^\infty$, a *factor* f of w is a word $f \in \Sigma^*$ satisfying

$$\exists x \in \Sigma^*, y \in \Sigma^\infty, w = xfy.$$

If $x = \epsilon$ (resp. $y = \epsilon$) then f is called *prefix* (resp. *suffix*). The set of all factors of w is denoted by $F(w)$, and those of length n is $F_n(w) = F(w) \cap \Sigma^n$. Finally Pref($w$) denotes the set of all prefixes of w. The length of a word w is $|w|$, and the number of occurrences of a factor $f \in \Sigma^*$ is $|w|_f$. A word is said to be *primitive* if it is not a power of another word. Two words u and v are *conjugate* when there are words x, y such that $u = xy$ and $v = yx$. The conjugacy class of a word w is denoted by (w); note that the length is invariant under conjugacy. We may represent each conjugacy class by a circular word, written clockwise on a regular polygon. Note that w is primitive if and only if the corresponding circular word is not fixed by any non-trivial rotation of the polygon. This implies that, for w primitive, there is at most one axial symmetry leaving its circular word fixed (since the product of two symmetries is a rotation). Thus, a primitive word which is the product of two palindromes has only one such factorization (see the equivalence (i) $\iff$ (ii) in the proof of Theorem 4).

The *mirror image* $\widetilde{u}$ of $u = u_1u_2\cdots u_n \in \Sigma^n$ is the word $\widetilde{u} = u_nu_{n-1}\cdots u_1$. A *palindrome* is a word p such that $p = \widetilde{p}$, and for a language $L \subseteq \Sigma^\infty$, we denote by $\mathrm{Pal}(L)$ the set of its palindromic factors.

Finally, an infinite word w is *recurrent* if it satisfies the condition $u \in F(w) \Longrightarrow |w|_u = \infty$. Clearly, every periodic word is recurrent, and there exist recurrent but non-periodic words, the Thue-Morse word M [11], and the Sturmian words being some of these.

3. Words with few palindromes

We intend to construct words whose palindromic factors are contained in a fixed finite set of palindromes P. Evidently we must suppose that P is factorially closed (with respect to palindromes). Define the set Q to be the set of minimal elements of $\mathrm{Pal}(\Sigma^*) \setminus P$, where the minimality is taken with respect to the *factorial* partial order: $u \leq v$ iff u is a factor of v.

Theorem 1. The maximal language whose palindromes are contained in P is rational and is given by $L_P = \Sigma^* \setminus \Sigma^*Q\Sigma^*$.

Proof. It suffices to show that Q is finite. Let $q \in Q$ and p be the factor of q obtained by deleting the first and last letter in q. By minimality we have that $p \in P$. Thus the length of the elements in Q is bounded. ◼

The language L_P may be finite. Take for instance P to be the palindromic factorial closure of $\{aba, bab\}$. Then L_P is the factorial closure of $\{abab, baba\}$.

Example. Let $P = \{\epsilon, a, b, c\}$. Then $Q = \{aa, bb, cc, bab, cac, aba, cbc, aca, bcb\}$ and L_P is the set of all finite prefixes of the six infinite words $(abc)^\omega, \ldots, (cba)^\omega$; equivalently the set of all finite factors of the two bi-infinite words ${}^\omega(abc)^\omega$, ${}^\omega(cba)^\omega$. A deterministic automaton for the language $L_P = \Sigma^* \setminus \Sigma^*Q\Sigma^*$, is given in Figure 1, where the initial state is shaded and all states are terminal.

Corollary 2. *Let P be a finite set of palindromes. Then there exists an infinite recurrent word W with $Pal(W) = P$ if and only if there exists a periodic word U with $Pal(U) = P$.*

Proof. Let $\mathbf{A}$ be a finite and trim automaton recognizing L_P. Since each factor of W is in L_P, there is an infinite path in $\mathbf{A}$ whose label is W. Therefore, there are states

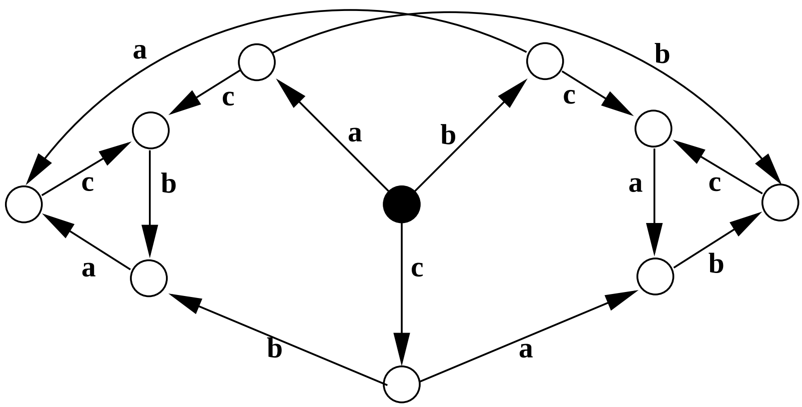


Fig. 1. An automaton for the language L_P.

appearing infinitely often in the path, and consequently, there exist arbitrarily long closed paths whose labels are factors of W. Since W is recurrent and P finite, there exist a finite circuit in $\mathbf{A}$ such that its label x satisfies $P \subseteq F(x)$. Define now $U = x^\omega$. Then we have $P \subseteq \mathrm{Pal}(U)$ and $\mathrm{Pal}(U) \subseteq P$ since U is the label of an infinite path in $\mathbf{A}$ and since $\mathbf{A}$ is trim. Conversely every periodic word is recurrent. ◻

The hypothesis "recurrent" is essential in the corollary above. Indeed, define the word $W = bab \cdot (abbaab)^\omega$ whose set of palindromes is $P = \{\epsilon, a, b, aa, bb, aba, bab, abba, baab, babab\}$. Suppose there exist a periodic word U having the same set of palindromes. By periodicity there would be an infinite number of occurrences of *babab* in it. Then U would contain factors in $\Sigma \cdot babab \cdot \Sigma$, but each of them contains a forbidden palindrome; a contradiction.

Corollary 3. *Let P be a finite set of palindromes and factorially closed with respect to palindromes. The following properties are decidable*

- *there exist an infinite word W such that $Pal(W) \subseteq P$;*
- *there exist an infinite word W such that $Pal(W) = P$;*

Proof. The first property amounts to check if L_P is infinite. For the second we proceed as follows: let $\mathbf{A}$ be a finite deterministic automaton recognizing L_P; for each maximal element $p \in P$, choose a path $c(p)$ in $\mathbf{A}$ labeled by p; note that there is a finite number of choices. Now, we have to check if there is a closed path in $\mathbf{A}$ having each $c(p)$ as subpath. This is clearly equivalent to check if a certain rational language is non-empty. ◻

The corollary justifies the following problem.

Problem: Characterize those P's satisfying each of the conditions in Corollary 3.

From the rational expression above, it is straightforward to construct a finite deterministic automaton that recognizes L_P. An alternate construction can be provided by using the De Bruijn graph of factors. Recall that, for a given language L, and a fixed integer m, the graph of m-length factors is defined by

$$xu \xrightarrow{y} uy \in \Gamma_m \iff xu, uy \in F_m(L),$$

where $x, y \in \Sigma$. Observe that the last letter of uy is the label of the edge. Therefore the labelling can be omitted. Let m be the maximal length of words in P, and call a word w *admissible* if $w \in L_P$. For a given m-length palindrome p, choose an admissible left extension v of p having length $m+2$. Then, starting from v construct the graph Γ_{m+2} of factors of length $m+2$ that avoid words in Q as follows.

1. *Initialisation.*: $\Gamma_{m+2} := v$;
2. *Inductive step.* $\Gamma_{m+2} := \Gamma_{m+2} \cup \{xu \xrightarrow{y} uy \mid xu \in \Gamma_{m+2}$ and uy is admissible$\}$.

Example. Let $P = \{\epsilon, a, b, aa, bb, aaa, aba, bab, abba, baab, baaab\}$ be a set of palindromes. Note that P is factorially closed. Then,

$$Q = \{bbb, aaaa, aabaa, babab, ababa, bbabb, aabbaa, \\ \times\, babbab, abaaba, bbaabb, abaaaba, bbaaabb\}.$$

Starting with $abbaaab$ which is an admissible word of length 7, the following graph of admissible factors is obtained. Observe that there are two elementary

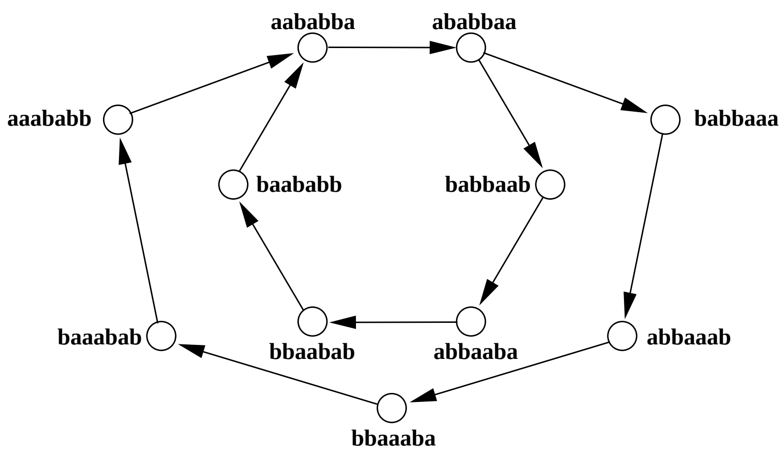


Fig. 2. A strongly connected component containing P.

cycles. None of them contains all factors in P. Therefore a smallest period having all factors in P is $u = abbaababbaaab$. It is unique with respect to conjugation. Note also that there exist non-periodic words having the same set of palindromes.

We have not provided the full example in this case. To be complete, we must check all the admissible words among the 2^7 words of length 7, so that the construction is similar (in complexity) to the determinization, and can contain several strongly connected components: indeed, palindromes are stable by mirror image so that there is a mirror component of the one given above, which does not mean that there is no other in general.

4. Infinite periodic words with finitely many palindromes

Periodic infinite words have a nice representation by circular words. Indeed, if $W = w^\omega$, then w is written clockwise on a circle as a necklace. The letters of a

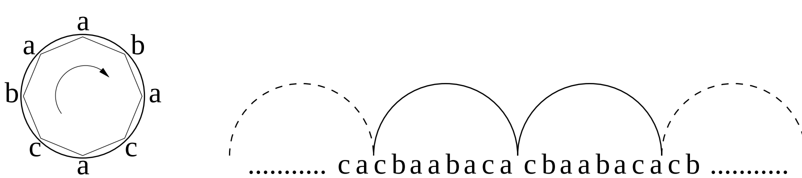


Fig. 3. A circular word and its unfolding.

circular word can be written on the vertices of a regular polygon so that we define an *axial symmetry* as a symmetry of the polygon.

Theorem 4. Let w be a primitive word. Then we have the following equivalent conditions:

(i) w is the product of two palindromes;
(ii) the conjugacy class (w) has an axial symmetry;
(iii) w is conjugate to either a word of the form $a \cdot u$, with $a \in \Sigma$ and $u \in \mathrm{Pal}(\Sigma^*)$, or to a palindrome of even length;
(iv) $\mathrm{Card}(\mathrm{Pal}(w^\omega))$ is infinite;
(v) $\mathrm{Card}(\mathrm{Pal}({}^\omega w^\omega))$ is infinite;
(vi) ${}^\omega w^\omega$ is a bi-infinite palindrome.

We call *symmetric* a word satisfying the equivalent conditions of the theorem, otherwise we call it *asymmetric*.

Proof. (i) $\iff$ (ii) : Figure 4 (a) shows the equivalence. If w is written as a circular word then the centres of the palindromes provide the symmetry. Conversely, if there is a symmetry in (w) with a vertical axis, a symmetrical cut provides a pair of palindromes.
(ii) $\iff$ (iii) : if w is not the conjugate of an even length palindrome, then a in Figure 4 (b) can be chosen to be a letter.

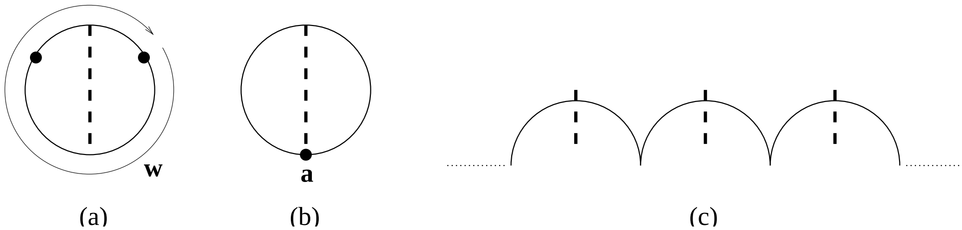
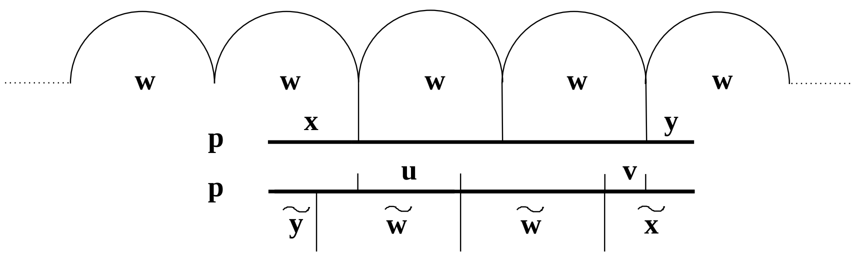


Fig. 4. Symmetric circular words

(ii) $\Longrightarrow$ (vi) $\Longrightarrow$ (v) $\Longrightarrow$ (iv) : is clear (see Figure 4 (c)).
(iv) $\Longrightarrow$ (i) : since arbitrarily long palindromes exist in w^ω, there is one, p say, that contains exactly 2 copies of w. Then $p = xw^2y = \widetilde{y}\widetilde{w}^2\widetilde{x}$ with $|x| < |w|$ and $|y| < |w|$ as shown below Then $ww = u\widetilde{w}v$. Thus $w = uv$ and $\widetilde{w} = vu = \widetilde{v}\,\widetilde{u}$. Therefore u and



v are palindromic. ∎

The statement (iv) $\Longrightarrow$ (i) in Theorem 4 was already noticed by Allouche et al. [3] in a slightly different context and form. Note that the property of being the product of two palindromes is closed under conjugacy. This follows from the theorem, but can be established directly. Observe that in (iii) the two cases are mutually disjoint: indeed, the first case means that (w) has a symmetry with respect to an axis passing through a vertex of the polygon representing (w), and the second means that the axis does not pass through any vertex; moreover, we know by a previous remark that (w) has only one symmetry. In [1], P. Auger calls *dexterpalindrome* a word of the form au as in (iii); he gives enumerative formulas and uses them to compute addition formulas for species of polygons (see also Labelle et al. [2]). For a primitive symmetric conjugacy class, he notes that three cases may occur: the class contains

1) exactly one palindrome and one dexterpalindrome and the length is odd;
2) exactly two palindromes and the length is even;
3) exactly two dexterpalindromes and the length is even.

5. Maximal palindromicity of periodic words

Recall that Droubay, Justin and Pirillo showed in [6] that for any finite word of length n, its palindromic complexity (the number of its palindromic factors) is bounded by $n+1$, and that Sturmian and also episturmian words realize the bound. Define the *defect* of a finite word u to be

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

This definition is extended to infinite words $W \in \Sigma^\omega$ by setting $D(W)$ to be the maximum of the defect of its factors: it may be finite or infinite. It follows from [6] that if u is a factor of v then $D(u) \leq D(v)$. Hence $D(W)$ is also equal to the maximum of the defect of the finite prefixes of W. We call *full* the words having defect value 0. We already know that there exist periodic infinite words with a finite number of palindromes, and consequently there exist infinite words with infinite defect (equivalently finite words with arbitrarily large defects).

The case of infinite words with small defect values deserves a closer look.

Lemma 5. *Assume that $w = xy = yz$ is a palindrome. Then for some palindromes u, v, and some $i \geq 0$ the following properties hold:*

(i) $x = uv, y = (uv)^i u, z = vu$;
(ii) $xyz = (uv)^{i+2}u$ *is a palindrome.*

Proof. The equation $xy = yz$ in words has for solutions

$$x = uv, y = (uv)^i u, z = vu,$$

(see Lothaire [8], Proposition 1.3.4). Since $w = xy = (uv)^{i+1}u$ is a palindrome u and v are also palindromes. ∎

Theorem 6. Let $w = uv$, with $|u| \geq |v|$ and $u, v \in \mathrm{Pal}(\Sigma^*)$, be a primitive symmetric word. Then the defect of $W = w^\omega$ is bounded by the defect of its prefix of length $|uv| + \lfloor \frac{|u|-|v|}{3} \rfloor$.

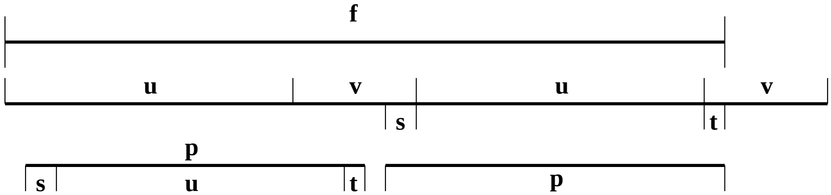
Proof. Recall from [6], that the number of palindromic factors of a word f is equal to the number of prefixes f' of f whose maximal palindromic suffix is unioccurrent in f'. Hence all we have to show is that for each prefix f of W, of length $> |uv| + \lfloor \frac{|u|-|v|}{3} \rfloor$, its maximal palindromic suffix p is unioccurrent. We proceed in three steps according to the length of f.

First case: $|f| > |uvuv|$. Then for some $k \geq 1$, f can be written as

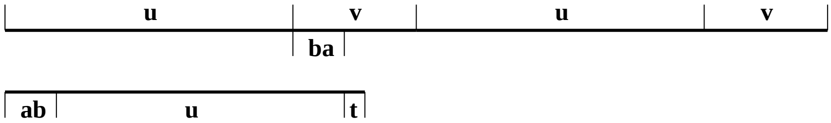
$$f = \begin{cases} u(vu)^k vy & \text{with } y \text{ prefix of } u, \\ \quad \text{or} & \\ uv(uv)^k uy & \text{with } y \text{ prefix of } v, \end{cases}$$

where $(vu)^k v$ and $(uv)^k u$ are palindromic. Let p be the maximal palindromic suffix of f. Then $\widetilde{y}(vu)^k vy$ (resp. $\widetilde{y}(uv)^k uy$) is a suffix of p, so that $f = xp$ where $|x| < |uv|$ and $|p| \geq |uv|$. If we assume, by contradiction, that p has another previous occurrence, then the two occurrences of p must overlap. Using Lemma 5 we obtain a longer palindrome contradicting the maximality of p.

Second case: $|uvu| < |f| \leq |uvuv|$. Hence $f = uvut$ where t is a nontrivial prefix of v. Since $\widetilde{t}$ is a suffix of v (v being a palindrome), $\widetilde{t}ut$ is a suffix of f. Thus the longest palindromic suffix p of f is $p = sut$, where s is a suffix of uv. Assuming by contradiction that p already appears in f, the two occurrences cannot overlap, otherwise f would have by Lemma 5 a palindromic suffix longer than p, contradicting its definition. We have therefore the following situation



where u overlaps itself; using Lemma 5 we obtain the following configuration



where a and b are palindromic, and $u = (ab)^i a$. Note that $|ab| \geq |s|$. Moreover bat is a prefix of v, hence $\widetilde{t}ab$ is a suffix of v. Thus f has the palindromic suffix $\widetilde{t}ab(ab)^i at$ which is strictly longer than $p = sut = s(ab)^i at$, a contradiction.

Third case: $|uv| + \lfloor \frac{|u|-|v|}{3} \rfloor < |f| \leq |uvu|$. That is $f = uvs$, where s is a prefix of u and $|s| > \lfloor \frac{|u|-|v|}{3} \rfloor$. It follows that the longest palindromic suffix p of f satisfies

$$|p| \geq |v| + 2|s|. \tag{1}$$

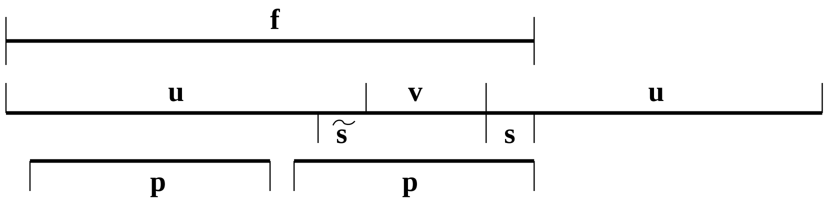
If we assume that there is a previous non-overlapping occurrence of p we must have

$$|u| - |s| \geq |p| \tag{2}$$

as shown below

Putting (1) and (2) together we obtain that

$$|u| - |s| \geq |v| + 2|s| \iff 3|s| \leq |u| - |v|. \tag{3}$$



Hence, $|s| \leq \frac{|u|-|v|}{3} \Longrightarrow |s| \leq \lfloor \frac{|u|-|v|}{3} \rfloor$, a contradiction. ∎

Corollary 7. *Let $w = uv$ be a primitive symmetric word such that $u, v \in Pal(\Sigma^*)$, with $|v| \leq |u| \leq |v| + 2$. Then $D((uv)^\omega) = D(uv)$.*

We also have the following more general formulation.

Corollary 8. *Let $w = uv$ be a primitive symmetric word. Then for some conjugate w' of w we have $D(w^\omega) = D(w')$.*

Proof. We may take $w' = uv$ with $|v| \leq |u| \leq |v| + 2$. Indeed, consider the circular word (w): it has a symmetry axis; take the perpendicular axis through the middle; cutting the circular word along the latter, one obains the desired factorization $w' = uv$ of a conjugate w' of w satisfying $\lfloor \frac{|u|-|v|}{3} \rfloor$ =0 (there are several cases to consider according to the length parity).

Now we can apply the theorem in order to obtain $D(w^\omega) = D((w')^\omega) = D(w')$. ∎

Note that the proof shows how to obtain w'. Moreover, w' has the greatest defect among the conjugates of w. In order to illustrate Corollary 8, take $w = ab^k ab^{k-1} aab^{k-1} ab^k a$, where $k > 1$. Then it is easy to see that $D(w) = 0$, and

$$D(w^\omega) = D(w') = D(b^k aab^k \cdot ab^{k-1} aab^{k-1} a) = k.$$

We will exhibit now a few examples showing that the bound is not optimal, but is not far from being sharp. For instance, let $w = u \cdot v = aabaa \cdot bab$. Then, $|uv| + \lfloor \frac{|u|-|v|}{3} \rfloor = |uv|$. It is easy now to check that $D(aabaa \cdot bab) = 0$. More generally, for any integer $k > 1$,

$$D((aab^k aa \cdot bab)^\omega) = D((aab^k aa \cdot bab) \cdot aab^k) = 0,$$

showing that there exist an infinite family of full infinite periodic words, a result in the spirit of that of Droubay, Justin and Pirillo on Sturmian words [6].

Consider now the word $w = u \cdot v = a^{k+1}ba^kca^kba^{k+1} \cdot c$, then $\lfloor \frac{|u|-|v|}{3} \rfloor = \lfloor \frac{4k+4}{3} \rfloor$, and we have

$$D(w) = 1 \quad \text{and} \quad D(w^\omega) = D(w \cdot a^{k+1}ba^k) = D(w \cdot a^k) = k + 1.$$

This example also solves the problem of constructing periodic sequences with a fixed finite defect value.

6. Counting asymmetric words

There is a natural action of the dihedral group D_n on the words of length n. The orbit of a word w consists in the conjugates of both w and its mirror image $\widetilde{w}$. The words of length n whose orbit contains $2n$ elements are the asymmetric words. For instance, *abc* is asymmetric while *abac* is not. The number A_n of orbits under D_n of asymmetric words can be counted easily by using the results of Stockmeyer ([12], p. 37 and p. 40). The enumeration is obtained by substituting in the subgroup index $S(\overline{C_1}, D_n)$ each variable z_i by q, where q is the cardinality of the alphabet. One obtains the following formulas. If n is odd then

$$A_n = \frac{1}{2n} \sum_{d|n} \mu(d) \left(q^{\frac{n}{d}} - nq^{\frac{n+d}{2d}} \right)$$

and when n is even

$$A_n = \frac{1}{2n} \sum_{d|n} \mu(d) q^{\frac{n}{d}} - \frac{1}{2} \sum_{d \,|\, n \,;\, d \nmid \frac{n}{2}} \mu(d) q^{\frac{n+d}{2d}} - \frac{1}{4} \sum_{d \,|\, \frac{n}{2}} \mu(d) \left(q^{\frac{n+2d}{2d}} + q^{\frac{n}{2d}} \right)$$

This yields the following examples.

- $n = 1 : A(1) = 0$ (a letter is not asymmetric).
- $n = 2 : A(2) = \frac{1}{4}(q^2 - q) - \frac{1}{2}(-q) - \frac{1}{4}(q^2 + q) = 0$, and there is no asymmetric word of length 2.
- $n = 3 : A(3) = \frac{1}{6}(q^3 - 3q^2 - q + 3q) = \dfrac{1}{6}(q^3 - 3q^2 + 2q)$.
- $n = 4 : A(4) = \frac{1}{8}(q^4 - q^2) - \frac{1}{2} \cdot 0 - \frac{1}{4}(q^3 + q^2 - q^2 - q) = \frac{1}{8}(q^4 - 2q^3 - q^2 + 2q)$.
- $n = 5 : A(5) = \frac{1}{10}(q^5 - 5q^3 - q + 5q) = \frac{1}{10}(q^5 - 5q^3 + 4q)$.
- $n = 6 : A(6) = \frac{1}{12}(q^6 - 3q^4 - 4q^3 + 8q^2 - 2q)$.

Moreover, when n is an odd prime number,

$$A(n) = \frac{1}{2n} \left(q^n - nq^{\frac{n+1}{2}} + (n-1)q \right).$$

One concludes that over a 2-letter alphabet, there is no asymmetric word of length $n \leq 5$; and for $n = 6$, there is only one orbit, the orbit of the word $w = aababb$.

7. A conjecture on asymmetric words

Let w be any word of length n and associate to it the following square matrices of order n. For each letter a, M_a is the sum of the elementary matrices $E_{i,i+1}$, for i such that letter a appears in the i-th position in w, with subscripts taken modulo n. In other words, these matrices are the matrices of the linear representation of the minimal automaton recognizing the language w^*.

Now consider the Lie algebra $L(w)$ generated by these matrices, under the natural Lie bracket $[M, N] = MN - NM$.

Theorem 9. Let w be a symmetric word of length $n = 2l$.

(i) If w is conjugate to a palindrome then $\dim(L(w)) \le 2l^2 - l$.
(ii) If w is conjugate to a dexterpalindrome then $\dim(L(w)) \le 2l^2 + l$.

Note that, by a previous remark, these two cases are mutually exclusive.

Proof. Let $B : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ be a bilinear form. The set of matrices $M \in \mathbb{C}^{n\times n}$ such that

$$\forall u, v \in \mathbb{C}^n, B(uM, v) + B(u, vM) = 0 \tag{4}$$

is a sub-Lie algebra of $\mathbb{C}^{n\times n}$, as is well-known and easily verified. We show that in the two cases of the theorem there exists a special bilinear form such that (4) holds for the matrices M_a, and then apply results on classical simple Lie algebras.

If w is conjugate to w' (resp. if $w = \widetilde{w}'$) then $L(w)$ is isomorphic to $L(w')$. Thus we may assume in (i) and (ii) that w is palindromic or $w = u\sigma$, for some palindrome u and letter σ.

(i) Assume that w is palindromic. The row and column indices are taken in $\{0, 1, \ldots, n-1\}$, identified with $\mathbb{Z}/n\mathbb{Z}$. Write $w = w_0 \ldots w_{n-1}$ and note that $w_i = w_j$ if $i + j = -1$ in $\mathbb{Z}/n\mathbb{Z}$. Define the bilinear form B on the canonical basis (e_i) of $\mathbb{C}^n$ by

$$B(e_i, e_{n-i}) = (-1)^i \quad \text{and} \quad B(e_i, e_j) = 0 \quad \text{otherwise.}$$

Then B is symmetric and we have to show that

$$B(e_r M_a, e_s) + B(e_r, e_s M_a) = 0 \quad \text{for any} \quad r, s.$$

Note that the first term is nonzero if and only if $w_r = a$ (hence $e_r M_a = e_{r+1}$) and $r + 1 + s = 0$ in $\mathbb{Z}/n\mathbb{Z}$, and in this case its value is $(-1)^{r+1}$; the second term is nonzero if and only if $w_s = a$ and $r + s + 1 = 0$, and in this case its value is $(-1)^r$. We conclude since $r + s + 1 = 0$ implies $w_r = w_s$.

Since B is a nondegenerate symmetric bilinear form, the set of M satisfying (4) is a Lie algebra, called the *orthogonal algebra*, of dimension $2l^2 - l$: it is the classical Lie algebra of type D_l (see Humphreys [7] page 3). It contains $L(w)$, so that (i) is proved.

(ii) We take here $\mathbb{Z}/n\mathbb{Z} = \{1, 2, \ldots, n\}$ as indexing set for rows and columns and set $w = w_1 \ldots w_n$ with $w_1 \ldots w_{n-1}$ palindromic; hence for $i \in \{1, \ldots, n-1\}, w_i = w_{n-i}$; note that this holds also for $i = n$ (with indices taken modulo n). Define the bilinear form B by

$$B(e_i, e_{n+1-i}) = (-1)^i, \quad \text{and} \quad B(e_i, e_j) = 0 \quad \text{otherwise.}$$

Then B is skew-symmetric and we have to verify that

$$B(e_r M_a, e_s) + B(e_r, e_s M_a) = 0 \quad \text{for any} \quad r, s.$$

The first term is nonzero if and only if $w_r = a$ and $r+1+s = n+1$, and in this case its value is $(-1)^{r+1}$; the second is nonzero if and only if $w_s = a$ and $r+s+1 = n+1$, and in this case its value is $(-1)^r$. We conclude since $r + s = n$ implies $w_r = w_s$.

Since B is a nondegenerate skew-symmetric bilinear form, the set of M satisfying (4) is a Lie algebra, called the *symplectic Lie algebra*, of dimension $2l^2 + l$: it is the classical Lie algebra of type C_l (see Humphreys [7] page 3). It contains $L(w)$, so that (ii) is proved. ∎

The previous proof and computer calculations suggest the following conjectures.

Conjecture 10 Let w be a primitive word of length n.

1. If n is odd, or if n is even and w is asymmetric, $L(w)$ is the Lie algebra $sl_n(\mathbb{C})$ of all $n \times n$ matrices of trace 0, of dimension $n^2 - 1$ (it is the classical Lie algebra of type A_{n-1}).
2. If n is even, and w is conjugate to a palindrome, then $L(w)$ is the orthogonal Lie algebra.
3. If n is even, and w is conjugate to a dexterpalindrome, then $L(w)$ is the symplectic Lie algebra.

Note that these conjectures are equivalent to their dimensional counterparts, as implies the proof of Theorem 9 and the fact that the matrices M_a have trace 0. In other words one has only to show that the dimensions in the three cases are: $n^2 - 1, 2l^2 - l, 2l^2 + l$, with $n = 2l$. In order to illustrate these conjectures we provide the following table of values for $\dim(L(w))$ obtained by computer calculations.

w	ab	aab	aaab	aabb	aabc	aaaab	aaaaab	abccba	aaaabc
$\dim(L(w))$	3	8	10	6	15	24	21	15	35

Acknowledgements. The last author is grateful to Gilbert Labelle for putting him on the track of the thesis of P.K. Stockmeyer.

References

1. P. Auger, Analyse d'équations combinatoires en théorie des espèces, Ph.D., Université du Québec à Montréal, Publications du LaCIM, Vol. 26 (2000)
2. P. Auger, G. Labelle, P. Leroux, Combinatorial additive formulas and applications, *Advances in Applied Maths*, **28** (2002) 302–342.
3. J.P. Allouche, M. Baake, J. Cassaigne, D. Damanik, Palindrome complexity, *Theoret. Comput. Sci.*, **292** (2003) 9–31.
4. S. Brlek, A. Ladouceur, A note on differentiable palindromes, *Theoret. Comput. Sci.*, **302**/1-3 (2003) 167–178.
5. X. Droubay, G. Pirillo, Palindromes and Sturmian words, *Theoret. Comput. Sci.*, **223** (1999) 73–85.
6. X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.*, **255** (2001) 539–553.
7. J.E. Humphreys, Introduction to Lie algebras and representation theory, Springer Verlag (1980).
8. M. Lothaire, Combinatorics on words, Addison-Wesley, 1983.
9. M. Lothaire, Algebraic Combinatorics on words, Cambridge University Press, 2002.
10. A. de Luca, Sturmian words: structure, combinatorics, and their arithmetics, *Theoret. Comput. Sci.*, **183** (1997) 45–82.
11. M. Morse and G. Hedlund, Symbolic Dynamics, *Amer. J. Math.*, **60** (1938) 815–866.
12. P. K. Stockmeyer, Enumeration of graphs with prescribed automorphism group, *Ph. D. thesis*, University of Michigan (1971).