



Return words in fixed points of substitutions

Lubomira Balkova, Edita Pelantová, Wolfgang Steiner

► To cite this version:

Lubomira Balkova, Edita Pelantová, Wolfgang Steiner. Return words in fixed points of substitutions. 2007. hal-00089863v3

HAL Id: hal-00089863

<https://hal.science/hal-00089863v3>

Preprint submitted on 27 Apr 2007 (v3), last revised 27 Sep 2007 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RETURN WORDS IN FIXED POINTS OF SUBSTITUTIONS

LUBOMÍRA BALKOVÁ, EDITA PELANTOVÁ, AND WOLFGANG STEINER

ABSTRACT. We present three simple ideas which enable to determine easily the number of return words in some infinite words. Using these ideas, we give a new and very short proof of the fact that each factor of an Arnoux-Rauzy word of order m has exactly m return words. We describe the structure of return words for the Thue-Morse sequence and focus on infinite words which are associated to simple Parry numbers.

1. INTRODUCTION

Recently, the notion of return words has played an important role in the study of (symbolic) dynamical systems. Roughly speaking, for a given factor w of an infinite word u , a return word of w is a word between two successive occurrences of the factor w . This definition is due to Durand [4] who used it to obtain a nice characterization of primitive substitutive sequences. A slightly different notion of return words was used by Ferenczi, Mauduit and Nogueira [8].

Sturmian words are aperiodic words over a bilateral alphabet with the lowest possible factor complexity; they were defined by Morse and Hedlund [12]. Using return words, Vuillon [16] has found a new equivalent definition of Sturmian words. He has shown that an infinite word u over a bilateral alphabet is Sturmian if and only if any factor of u has exactly two return words.

More generally, an infinite word u is said to satisfy the property \mathcal{R}_m if every factor of u has exactly m return words. For $m \geq 3$, no complete characterization of words satisfying \mathcal{R}_m is known, only partial results for different generalizations of Sturmian words exist. Justin and Vuillon [10] proved that Arnoux-Rauzy words of order m satisfy \mathcal{R}_m . Ferenczi, Holton and Zamboni [7] showed for infinite words coding regular 3-interval exchanges that every bispecial factor has 3 return words. More generally, Vuillon [17] proved that m -interval exchange coding words satisfy \mathcal{R}_m . The factor complexity, i.e. the number of different factors of length n , of all examples in the preceding paragraph is $\mathcal{C}(n) = (m - 1)n + 1$.

In this paper, we present three very simple ideas which allow determining easily the number of return words in some infinite words.

- The first observation is that for the study of return words in a uniformly recurrent word u , it suffices to limit the considerations to bispecial factors of u .
- We explain the role of factor graphs for the determination of the number of return words. This idea has already been used by Vuillon [16]. It is especially powerful when the first observation is taken into account.
- In case of words which are invariant under a substitution, we will make use of the relation between return words of a factor and return words of its image under the substitution.

Using these ideas, we give in Section 3 a new and very short proof for the number of return words for Arnoux-Rauzy sequences. In Section 4, we describe the cardinality of the set of return words for the Thue-Morse sequence. Section 5 is devoted to words u_β which are associated to simple Parry numbers.

For a real number $\beta > 1$, the infinite word u_β codes the sequence of gaps between successive β -integers. If β is a Parry number, then u_β is the fixed point of the canonical substitution associated to β . A simple Parry number is an algebraic integer $\beta > 1$ having a finite Rényi expansion of unity $d_\beta(1) = t_1 t_2 \cdots t_m$ (see Section 5.1 for the definitions). In this case, the alphabet of the infinite word u_β contains m letters. In [9], it has been proved that u_β is Arnoux-Rauzy if and only if $t_1 = t_2 = \cdots = t_{m-1}$ and $t_m = 1$. Consequently, u_β satisfies property \mathcal{R}_m . We will show that this is true more generally for all words u_β with complexity $\mathcal{C}(n) = (m - 1)n + 1$ for every $n \in \mathbb{N}$.

2. BASIC DEFINITIONS

An *alphabet* \mathcal{A} is a finite set of symbols called *letters*. A (possibly empty) concatenation of letters is a *word*. The set \mathcal{A}^* of all finite words provided with the operation of concatenation is a free monoid. The *length* of a word w is denoted by $|w|$. A finite word w is called a *factor* (or *subword*) of the (finite or right infinite) word u if there exist a finite word v and a word v' such that $u = v w v'$. The word w is a *prefix* of u if v is the empty word. Analogously, w is a *suffix* of u if v' is the empty word. A concatenation of k letters a will be denoted by a^k , a concatenation of infinitely many letters a by a^ω .

The *language* $\mathcal{L}(u)$ is the set of all factors of the word u , and $\mathcal{L}_n(u)$ is the set of all factors of u of length n . Let w be a factor of an infinite word u and let $a, b \in \mathcal{A}$. If wa is a factor of u , then we call a a *right extension* of w . Analogously, if bw is a factor of u , we call b a *left extension* of w . We will denote by $\text{Ext}_+(w)$ the set of all right extensions of w and by $\deg_+(w)$ the number of right extensions of w . Analogously, the set of left extensions of w will be denoted by $\text{Ext}_-(w)$ and its cardinality by $\deg_-(w)$. A factor w is *right special* if $\deg_+(w) > 1$, *left special* if $\deg_-(w) > 1$ and *bispecial* if w is right special and left special.

Let w be a factor of an infinite word $u = u_0 u_1 \dots$ (with $u_j \in \mathcal{A}$), $|w| = \ell$. An integer j is an *occurrence* of w in u if $u_j u_{j+1} \dots u_{j+\ell-1} = w$. Let j, k , $j < k$, be successive occurrences of w . Then $u_j u_{j+1} \dots u_{k-1}$ is a *return word* of w . The set of all return words of w is denoted by $\text{Ret}(w)$,

$$\text{Ret}(w) = \{u_j u_{j+1} \dots u_{k-1} \mid j, k \text{ being successive occurrences of } w \text{ in } u\}.$$

An infinite word u is *uniformly recurrent* if, for any $n \in \mathbb{N}$, there exists an integer $R(n)$ such that any factor of u of length $R(n)$ contains all factors of length n . It is not difficult to see that the set of return words of w is finite for any factor w if u is a uniformly recurrent word.

The variability of local configurations in u is expressed by the *factor complexity function* (or simply *complexity*) $\mathcal{C}_u : \mathbb{N} \rightarrow \mathbb{N}$, $\mathcal{C}_u(n) := \#\mathcal{L}_n(u)$.

It is well known that a word u is aperiodic if and only if $\mathcal{C}_u(n) \geq n + 1$ for all $n \in \mathbb{N}$. Infinite aperiodic words with the minimal complexity $\mathcal{C}_u(n) = n + 1$ for all $n \in \mathbb{N}$ are called *Sturmian words*. These words have been studied extensively, and several equivalent definitions of Sturmian words can be found in Berstel [3].

A mapping φ on the free monoid \mathcal{A}^* is called a *morphism* if $\varphi(vw) = \varphi(v)\varphi(w)$ for all $v, w \in \mathcal{A}^*$. Obviously, for defining the morphism, it suffices to give $\varphi(a)$ for all $a \in \mathcal{A}$. The action of a morphism can be naturally extended on right-sided infinite words by

$$\varphi(u_0 u_1 u_2 \dots) := \varphi(u_0) \varphi(u_1) \varphi(u_2) \dots$$

A morphism φ such that $\varphi(b) \neq \varepsilon$ for all $b \in \mathcal{A}$ and for which there exists a letter $a \in \mathcal{A}$ such that $\varphi(a) = aw$ for some non-empty word $w \in \mathcal{A}^*$, is called a *substitution*. An infinite word u such that $\varphi(u) = u$ is called a *fixed point* of the substitution φ . Obviously, every substitution has at least one fixed point, namely $\lim_{n \rightarrow \infty} \varphi^n(a)$.

3. HANDY RULES FOR DETERMINING RETURN WORDS

3.1. Consider only bispecial factors! If a factor w is not right special, i.e., if it has a unique right extension $a \in \mathcal{A}$, then the sets of occurrences of w and wa coincide, and

$$\text{Ret}(w) = \text{Ret}(wa).$$

If a factor w has a unique left extension $b \in \mathcal{A}$, then $j \geq 1$ is an occurrence of w in the infinite word u if and only if $j - 1$ is an occurrence of bw . This statement does not hold for $j = 0$. Nevertheless, if u is a uniformly recurrent infinite word, then the set $\text{Ret}(w)$ of return words of w stays the same no matter whether we include the return word corresponding to the prefix w of u or not. Consequently, we have

$$\text{Ret}(bw) = b\text{Ret}(w)b^{-1} = \{bvb^{-1} \mid v \in \text{Ret}(w)\},$$

where bvb^{-1} means that the word v is prolonged to the left by the letter b and it is shortened from the right by erasing the letter b (which is always the suffix of v for $v \in \text{Ret}(w)$).

For an aperiodic uniformly recurrent infinite word u , each factor w can be extended to the left and to the right to a bispecial factor. To describe the cardinality and the structure of $\text{Ret}(w)$ for arbitrary w , it suffices therefore to consider bispecial factors w .

3.2. Look at factor graphs! Return words and the role of special factors can be well visualized by means of the *factor graphs* (or *de Bruijn graphs* or *Rauzy graphs*). The factor graph Γ_n is an oriented graph with set of vertices $\mathcal{L}_n(u)$ and set of edges $\mathcal{L}_{n+1}(u)$. The edge $w_0w_1 \dots w_n \in \mathcal{L}_{n+1}(u)$ goes from the vertex $w_0 \dots w_{n-1}$ to the vertex $w_1 \dots w_n$. The factor $w = w_0 \dots w_{n-1}$ is left special if and only if at least two edges ends in the vertex w and w is right special if and only if at least two edges begins in w . Each return word of a factor w of the length n is visualized as an oriented walk in Γ_n , which begins and ends in the vertex w , and such that the vertex w is not entered in the course of this walk.

Proposition 3.1. *Let u be a recurrent word and $m \in \mathbb{N}$. Suppose that for every $n \in \mathbb{N}$ at least one of the following conditions is satisfied:*

- *There is a unique left special factor $w \in \mathcal{L}_n(u)$, and $\deg_-(w) = m$.*
- *There is a unique right special factor $w \in \mathcal{L}_n(u)$, and $\deg_+(w) = m$.*

Then u satisfies the property \mathcal{R}_m , i.e., every factor has exactly m return words.

Proof. As we have explained at the beginning of the section, it is enough to prove that every bispecial factor w has exactly m return words. Let $|w| = n$ and suppose first that w is the unique right special factor in $\mathcal{L}_n(u)$. Hence w has outdegree m in Γ_n and all other vertices have outdegree 1. Therefore the choice of the edge starting the walk at w determines the walk uniquely. Consequently, there are exactly m ways to come back to w , i.e., w has exactly m return words. If w is the unique left special factor in $\mathcal{L}_n(u)$, similar arguments can be used. \square

Recall that Arnoux-Rauzy words of order m are defined as infinite words which have for every $n \in \mathbb{N}$ exactly one right special factor w of length n with $\deg_+(w) = m$ and exactly one left special factor w of length n with $\deg_-(w) = m$. Therefore we obtain the following corollary.

Corollary 3.2. *Arnoux-Rauzy words of order m satisfy \mathcal{R}_m , in particular Sturmian words satisfy \mathcal{R}_2 .*

3.3. Respect ancestors! For a large class of substitutions, there exists a straightforward relation between return words of a factor w and of its image $\varphi(w)$.

Definition 3.3. *Let w be a factor of a fixed point u of a substitution φ . We say that a word $v_0v_1 \dots v_{n-1}v_n \in \mathcal{L}_{n+1}(u)$ is an ancestor of w if*

- *w is a factor of $\varphi(v_0v_1 \dots v_{n-1}v_n)$,*
- *w is neither a factor of $\varphi(v_1 \dots v_{n-1}v_n)$ nor of $\varphi(v_0v_1 \dots v_{n-1})$.*

The term ancestor has been introduced by Mignosi and Séebold [11]. Clearly, any factor $\varphi(w)$ has at least one ancestor, namely the factor w .

Lemma 3.4. *Let an infinite word u be a fixed point of a substitution φ and w be a factor of u . If the only ancestor of $\varphi(w)$ is the factor w , then*

$$\text{Ret}(\varphi(w)) = \varphi(\text{Ret}(w)).$$

Proof. Let v be a return word of w , i.e., vw is a factor of $u = \varphi(u)$, w is a prefix of vw and w occurs in vw exactly twice. Since w is the only ancestor of $\varphi(w)$, $\varphi(vw)$ contains $\varphi(w)$ only as prefix and suffix, hence $\varphi(v)$ is a return word of $\varphi(w)$ and $\varphi(\text{Ret}(w)) \subseteq \text{Ret}(\varphi(w))$.

If v' is a return word of $\varphi(w)$, then every ancestor of $v'\varphi(w)$ is of the form vw for some return word v of w since it contains w as prefix and suffix and has no other occurrence of w . Furthermore, we have $v' = \varphi(v)$, thus $\text{Ret}(\varphi(w)) \subseteq \varphi(\text{Ret}(w))$. \square

4. RETURN WORDS IN THE THUE-MORSE SEQUENCE

The Thue-Morse sequence is a very well known, intensively studied binary sequence (see e.g. Allouche and Shallit [1]), which is a fixed point of the substitution

$$0 \mapsto \varphi(0) = 01 \quad \text{and} \quad 1 \mapsto \varphi(1) = 10.$$

This substitution has two fixed points. Since the substitution is primitive, both its fixed points have the same language and are uniformly recurrent. We will denote by u_{TM} one of the fixed points, say the one starting with 0. Let us give a prefix of u_{TM} :

$$u_{TM} = 01101001100101101001011001101001 \dots$$

From the form of the substitution, we can deduce some simple facts about the language of u_{TM} .

Lemma 4.1. *We have*

- (1) $000, 111 \notin \mathcal{L}(u_{TM})$,
- (2) $01010, 10101 \notin \mathcal{L}(u_{TM})$,
- (3) $a_1 \dots a_n \in \mathcal{L}(u_{TM})$ if and only if $(1 - a_1) \dots (1 - a_n) \in \mathcal{L}(u_{TM})$,
- (4) any factor $w \in \mathcal{L}(u_{TM})$ with $|w| \geq 4$ has a unique ancestor.

Proof. It is easy to see that (1)–(3) hold. For (4), note first that the unique ancestor of 00 is 10. If v is an ancestor of a factor $w_1 00 w_2$, we have therefore a decomposition $v = v_1 v_2$ such that $w_1 0$ is a suffix of $\varphi(v_1)$ and $0 w_2$ is a prefix of $\varphi(v_2)$. Since the last letters of $\varphi(0)$ and $\varphi(1)$ are different, v_1 is unique. Similarly, v_2 is unique since the first letters of $\varphi(0)$ and $\varphi(1)$ are different. The same is true for a factor $w_1 11 w_2$. By (2), any factor of length ≥ 4 except 1010 and 0101 contains a factor 00 or 11 and has therefore a unique ancestor. Since $000, 111 \notin \mathcal{L}(u_{TM})$, the unique ancestor of 1010 is 11, and the unique ancestor of 0101 is 00, thus (4) holds. \square

We have chosen the Thue-Morse sequence for studying return words, since it is easy to describe the bispecial factors of u_{TM} . Clearly, 0, 01 and 010 and their complements 1, 10 and 101 are bispecial factors. The structure of longer bispecial factors is also simple.

Proposition 4.2. *Let w be a bispecial factor of u_{TM} with length $|w| \geq 4$. Then there exists a bispecial factor v such that $\varphi(v) = w$.*

Proof. By Lemma 4.1 (4), w has a unique ancestor v , and we have $w = \varphi(v)$ since w is bispecial. It is clear that v is bispecial as well. \square

Theorem 4.3. *Every factor of the Thue-Morse sequence has either 3 or 4 return words.*

Proof. In Section 3.1, we have seen that it is sufficient to determine the cardinality of $\text{Ret}(w)$ for bispecial factors w . By Lemma 3.4 and Proposition 4.2, we can restrict to bispecial factors shorter than 4, i.e., to the factors 0, 01 and 010 and their complements. Since the return words of the complement of w are the complements of the return words of w and

$$\begin{aligned} \text{Ret}(0) &= \{0, 01, 011\}, \\ \text{Ret}(01) &= \{01, 010, 011, 0110\}, \\ \text{Ret}(010) &= \{010, 0100110, 01011010, 010110011\}, \end{aligned}$$

the theorem is proved. \square

5. RETURN WORDS OF u_β

In the sequel, we will describe a new class of infinite words with property \mathcal{R}_m . This class consists of words u_β for β being a simple Parry number. We will make use of the description of left special factors obtained in [2]. For two special classes of u_β , we will give a complete description of return words.

5.1. Infinite words associated with β -integers. Let $\beta > 1$ be a real number. Then every positive real number x can be represented as

$$x = \sum_{j=-\infty}^k x_j \beta^j = x_k \cdots x_1 x_0 \bullet x_{-1} x_{-2} \cdots \quad \text{with } x_j \in \mathbb{N}.$$

If moreover $x - \sum_{j=\ell}^k x_j \beta^j < \beta^\ell$ for all $\ell \leq k$, then the expression $x_k \cdots x_1 x_0 \bullet x_{-1} x_{-2} \cdots$ is called (greedy) β -*expansion* of x , which we denote by $\langle x \rangle_\beta$. For $x \in [0, 1)$, this expansion can be found using the transformation $T_\beta(x) := \beta x - \lfloor \beta x \rfloor$, where $\lfloor \cdot \rfloor$ denotes the (lower) integer part. Then $\langle x \rangle_\beta = \bullet x_{-1} x_{-2} x_{-3} \cdots$ with $x_{-j} = \lfloor \beta T_\beta^{-j-1}(x) \rfloor$.

For $x = 1$, this algorithm does not provide the β -expansion of x (which is $1 \bullet 00 \cdots$), but we obtain an important sequence, the *Rényi expansion of 1* in base β , which is thus defined as

$$d_\beta(1) = t_1 t_2 t_3 \cdots, \quad \text{where } t_j := \lfloor \beta T_\beta^{j-1}(1) \rfloor.$$

If $d_\beta(1)$ is eventually periodic, then β is called a *Parry number*. If $t_j = 0$ for all $j > m$ and $t_m \neq 0$, then β is a *simple Parry number*, and we write $d_\beta(1) = t_1 \cdots t_m$. From now on, let us limit our considerations to this case. Parry [13] showed that a sequence $t_1 \cdots t_m$ is a Rényi expansion of 1 for some $\beta > 1$ if and only if the sequence satisfies

$$(1) \quad t_j \cdots t_m <_{\text{lex}} t_1 \cdots t_m \quad \text{for all } j, 1 < j \leq m.$$

where $v <_{\text{lex}} w$ means that v is lexicographically less than w . In particular, it follows that

$$(2) \quad t_1 \geq t_j \quad \text{for all } j, 2 \leq j \leq m.$$

Example. The most important example of a simple Parry number β , which is not an integer, is the golden mean, $\beta = (1 + \sqrt{5})/2$, with $\beta^2 = \beta + 1$ and thus $d_\beta(1) = 11$.

In this section, we study infinite words associated with the distances of non-negative β -integers

$$\mathbb{Z}_\beta^+ := \{x \geq 0 \mid \langle x \rangle_\beta = x_k x_{k-1} \cdots x_0 \bullet 00 \cdots \text{ for some } k \geq 0\}.$$

If β is an integer, then clearly $\mathbb{Z}_\beta^+ = \mathbb{N}$, $d_\beta(1) = t_1 = \beta$, and the distance between neighboring elements of \mathbb{Z}_β^+ is always 1. Therefore we are interested only in the case $\beta \notin \mathbb{N}$, thus $d_\beta(1) = t_1 \cdots t_m$ with $m \geq 2$. For these β , Thurston [15] showed that the distances occurring between neighbors of \mathbb{Z}_β^+ form the set $\{\Delta_k \mid 0 \leq k < m\}$, where $\Delta_k := \bullet t_{k+1} \cdots t_m$.

If we assign the letter k to the gap Δ_k and if we write down the order of distances in \mathbb{Z}_β^+ on the real line, we naturally obtain an infinite word u_β . It can be shown that the word u_β is the unique fixed point of the canonical substitution φ (see e.g. Fabre [5]),

$$(3) \quad \begin{array}{rcl} \varphi : & 0 & \mapsto 0^{t_1} 1 \\ & 1 & \mapsto 0^{t_2} 2 \\ & & \vdots \\ & m-2 & \mapsto 0^{t_{m-1}}(m-1) \\ & m-1 & \mapsto 0^{t_m}. \end{array}$$

Obviously, this substitution is primitive, i.e., there exists an integer exponent k such that for each pair of letters $a, b \in \mathcal{A}$, the letter a appears in the word $\varphi^k(b)$. Queffélec [14] showed that a fixed point of a primitive substitution is a uniformly recurrent infinite word. This implies that the infinite word u_β associated to a simple Parry number β is uniformly recurrent and, therefore, the set of return words $\text{Ret}(w)$ of any factor w of u_β is finite. Note that φ is the Fibonacci substitution if β is the golden mean.

5.2. Return words for particular factors of u_β . We will describe return words of all factors of u_β in two different cases which have in common the fact that the study of return words of long factors can be reduced to the study of return words of a small number of short factors.

As we know from Section 3.2, return words of all factors of a uniformly recurrent word can be obtained from return words of bispecial factors. Hence, we focus on bispecial factors in words u_β defined by the substitution (3). In order to describe bispecial factors of u_β , where β is a simple Parry number, we will make use of the description of some special types of factors of u_β from [9].

Definition 5.1. Let $\tau_k := \min\{i \geq 1 \mid t_{k-i} \neq 0\}$ for $2 \leq k \leq m$. ($\tau_k < k$ since $t_1 \geq 1$.)

Lemma 5.2. [9, Lemma 4.5] All factors of u_β of the form $a0^r b$, where a, b are non-zero letters and $r \in \mathbb{N}$, are the following ones:

$$\begin{aligned} \tau_k 0^{t_k} k & \text{ for } k = 2, \dots, m-1, \\ k 0^{t_1} 1 & \text{ for } k = 1, \dots, m-1, \\ \tau_m 0^{t_1+t_m} 1. \end{aligned}$$

Let us remind that the coefficients t_i , $2 \leq i \leq m$, satisfy $t_i \leq t_1$.

Corollary 5.3. 0^r is a bispecial factor if and only if $r < t_1 + t_m$.

- If $r \leq t_1$, then $\text{Ext}_+(0^r) = \{0, 1\} \cup \{k \mid 2 \leq k < m, r \leq t_k\}$.
- If $t_1 < r < t_1 + t_m$, then $\text{Ext}_+(0^r) = \{0, 1\}$.

Corollary 5.4. Let w be a bispecial factor which contains a non-zero letter, and let $a \neq b$ be non-zero letters satisfying $a, b \in \text{Ext}_+(w)$. Then $t_a = t_b$.

Lemma 5.5. If v is a bispecial factor which contains a non-zero letter, then there exists a unique bispecial factor w and a unique $s \in \{t_1, \dots, t_{m-1}\}$ such that $v = \varphi(w)0^s$.

Proof. Let v' be such that $v = v'0^s$ and v' ends with a non-zero letter. From Lemma 5.2, it is clear that $s \in \{t_1, \dots, t_{m-1}\}$ and v' has the prefix $0^{t_1}1$. Thus, v is either of the form

- $v = 0^{t_1}10^s = \varphi(0)0^s$ or
- $v = 0^{t_1}1v''0^s$.

In the latter case, according to the form of the substitution (3), there exists a unique $w \in \mathcal{L}(u_\beta)$ such that $\varphi(w) = 0^{t_1}1v''$, and w is obviously bispecial. \square

Observation 5.6. If $(m-1)a \in \mathcal{L}(u_\beta)$ for a letter a , then $a = 0$.

Proof. This follows from Lemma 5.2 since $t_1 \geq 1$ and $\tau_k < k$ for every k . \square

Lemma 5.7. If w is a bispecial factor, then there exists $s \in \{t_1, \dots, t_{m-1}\}$ such that $\varphi(w)0^s$ is bispecial. Moreover, if $\deg_+(w) = 2$, then this s is unique and $\deg_+(\varphi(w)0^s) = 2$.

Proof. If w is bispecial, then it has at least two different right extensions, say $a < b$, and:

- If $b < m-1$, then $\varphi(w)0^{t_{a+1}}(a+1)$ and $\varphi(w)0^{t_{b+1}}(b+1)$ are factors of u_β . Hence $\varphi(w)0^s$ with $s = \min\{t_{a+1}, t_{b+1}\}$ is bispecial.
- If $b = m-1$, then Observation 5.6 implies that $\varphi(w(m-1)0) = \varphi(w)0^{t_m}0^{t_1}1 \in \mathcal{L}(u_\beta)$, and $\varphi(w)0^{t_{a+1}}$ is bispecial since $\{0, a+1\} \subset \text{Ext}_+(\varphi(w)0^{t_{a+1}})$.

If the bispecial factor w has only two right extensions, then this choice of s is clearly unique. \square

Corollary 5.8. Let w be a bispecial factor satisfying that both $\varphi(w)0^p$ and $\varphi(w)0^q$ are bispecial factors of u_β for some $p \neq q$, then $\deg_+(w) \geq 3$.

Once, bispecial factors of u_β described, we can investigate their return words. Lemmas 5.5 and 5.7 show that occurrences of w in u_β can be described by means of their “preimages” or “images” and this fact is useful for description of their return words.

Proposition 5.9. Let w be a bispecial factor of u_β such that there exists a unique $s \in \{t_1, \dots, t_{m-1}\}$ with the property that $\varphi(w)0^s$ is a bispecial factor of u_β . Then

$$\text{Ret}(\varphi(w)0^s) = \varphi(\text{Ret}(w)).$$

Proof. It is an obvious consequence of the form of the substitution that if the letter $m-1$ is not a suffix of w , then w is the only ancestor of $\varphi(w)$. According to the Observation 5.6, no bispecial factor ends with $m-1$. Therefore Lemma 3.4 implies $\text{Ret}(\varphi(w)) = \varphi(\text{Ret}(w))$. Let us remind that if a factor v is not right special, then $\text{Ret}(v) = \text{Ret}(va)$, where a is the unique right extension of v . According to our assumption for $k < s$, the factor $\varphi(w)0^k$ is not right special and $\varphi(w)0^{k+1}$ is its unique right extension. Hence we may apply the previous rule s times to obtain the statement. \square

The next lemma will turn out to be useful for determining return words of prefixes of 0^ω .

Lemma 5.10. *For every k , $1 \leq k < m$, there exists a unique word v_k such that $0v_kk \in \mathcal{L}(u_\beta)$ and v_k does not contain zero. Moreover, we have for every $j \in \mathcal{A}$*

$$\{wj \mid w \in \text{Ret}(j)\} = \{j\varphi^j(0)\} \cup \{j\varphi^j(v_kk0) \mid 1 \leq k < m\}.$$

Proof. If $t_k \geq 1$, then k is always preceded by 0 and v_k is the empty word ε . If $t_k = 0$, then Lemma 5.2 shows that k is always preceded by τ_k , thus we have $v_k = v_{\tau_k}\tau_k$. Since $1 \leq \tau_k < k$ and $v_1 = \varepsilon$, the first statement is proved inductively.

For the second statement, we proceed by induction on j . For $j = 0$, we have to determine the factors v without zeros such that $0v0 \in \mathcal{L}(u_\beta)$. Since 0 can be preceded by every letter $k \in \mathcal{A}$, these factors are $\{v_kk \mid 1 \leq k < m\} \cup \{\varepsilon\}$, and the statement is true for $j = 0$. For a return word w of $j+1$, it follows from the form of the substitution that $w(j+1) = (j+1)\varphi(w'j)$, and jw' is a return word of j . The induction assumption implies that $w(j+1) = (j+1)\varphi(w'j) = (j+1)\varphi^{j+1}(v0)$ with $v \in \{v_kk \mid 1 \leq k < m\} \cup \{\varepsilon\}$. \square

5.3. Complete description of return words for two classes of β . Now we can describe return words in the so-called confluent case and in the case of mutually different numbers t_1, \dots, t_{m-1} .

Theorem 5.11. *Let u_β be the infinite word associated with a simple Parry number β satisfying $d_\beta(1) = t_1 \cdots t_{m-1}t_m$, where $t_1 = t_2 = \cdots = t_{m-1} = t$ (the so-called confluent case).*

- If $t_m = 1$, then every factor $w \in \mathcal{L}(u_\beta)$ has exactly m return words.
- If $t_m \geq 2$, then every factor $w \in \mathcal{L}(u_\beta)$ has either m or $m+1$ return words (and both values occur).

Proof. The condition $t_1 = \cdots = t_{m-1} = t$ and Lemma 5.7 imply that for each bispecial factor w there exists a unique s such that $\varphi(w)0^s$ is a bispecial factor ($s = t$). Using Proposition 5.9, we have $\#\text{Ret}(\varphi(w)0^s) = \#\text{Ret}(w)$. Lemma 5.5 implies that to determine the cardinality of the set of all return words, it suffices to study the cardinality of the set $\text{Ret}(0^r)$, where 0^r is a bispecial factor of u_β , i.e., $1 \leq r < t + t_m$ by Corollary 5.3. In the next considerations, we use mainly Lemma 5.2. For $1 \leq r \leq t$, the block 0^r is followed either by

- 0, then the return word of 0^r is 0, or by
- $k0^t$, $1 \leq k < m$, hence we obtain the return word 0^rk .

Obviously, there are no other cases, thus one gets

$$\text{Ret}(0^r) = \{0, 0^r1, 0^r2, \dots, 0^r(m-1)\} \quad \text{and} \quad \#\text{Ret}(0^r) = m.$$

For $t < r < t + t_m$ (this case appears only for $t_m \geq 2$), the block 0^r is followed either by

- 0, then the return word of 0^r is 0, or by
- 1, then 0^r1 is suffix of $0^{t_m+t}1 = \varphi((m-1)0)$. If we denote the return word of 0^r by w , then, obviously, w is followed by the very next occurrence of $0^{t_m+t}1$. Thus, the factor $0^{t_m+t-r}w0^{t_m}$ can be written as $\varphi((m-1)v(m-1))$, where v does not contain $m-1$. Therefore $(m-1)v$ is a return word of $m-1$, and by Lemma 5.10, either $v(m-1) = \varphi^{m-1}(0)$ or $v(m-1) = \varphi^{m-1}(k0)$ for some $k \geq 1$.

Consequently, since there are no other cases, we have

$$\text{Ret}(0^r) = \{0, 0^{r-t}\varphi^m(0)0^{-t_m}\} \cup \{0^{r-t}\varphi^m(k0)0^{-t_m} \mid 1 \leq k < m\}$$

and $\#\text{Ret}(0^r) = m+1$. \square

Theorem 5.12. *Let u_β be the infinite word associated with a simple Parry number β satisfying $d_\beta(1) = t_1 \dots t_{m-1}t_m$ such that $t_i \neq t_j$ for all i, j , $1 \leq i < j < m$.*

- If $t_m = 1$, then every factor $w \in \mathcal{L}(u_\beta)$ has exactly m return words.
- If $t_m \geq 2$, then every factor $w \in \mathcal{L}(u_\beta)$ has either m or $m + 1$ return words (and both values occur).

Proof. By Lemma 5.5, every bispecial factor is either of the form 0^r or $\varphi(w)0^s$. If s is unique, then the problem is reduced to the determination of return words of w , by Proposition 5.9. Suppose that there exist $p > q$ such that $\varphi(w)0^p$ and $\varphi(w)0^q$ are bispecial. Then w has at least two non-zero right extensions, by Corollary 5.8. Thus, by Corollary 5.4 and the assumption of the theorem, w does not contain a non-zero letter. Consequently, it suffices to describe the number of return words for bispecial factors of the form 0^r , $1 \leq r < t_1 + t_m$, and for $\varphi(0^r)0^s$. In the latter case, it suffices to consider $r < t_1$, since 0^r , $t_1 \leq r < t_1 + t_m$, has just the right extensions $0, 1$ and, according to Lemma 5.7, there exists a unique s such that $\varphi(0^r)0^s$ is bispecial.

Return words of 0^r , $1 \leq r < t_1 + t_m$.

If $1 \leq r \leq t_1$, then $0^{r+1} \in \mathcal{L}(u_\beta)$, thus $0 \in \text{Ret}(0^r)$. As an obvious consequence of the form of the substitution, it follows that $k0^r$, $k \geq 1$, is always suffix of $\varphi^k(0)0^r$. Since $r \leq t_1$, the factor $\varphi^k(0)$ contains 0^r (particularly as a prefix) for each $k \geq 1$. Therefore, there is exactly one return word of 0^r ending by k for every $k \geq 1$.

If $t_1 < r < t_1 + t_m$ (this case appears only for $t_m \geq 2$), then it suffices again to apply Lemma 5.10 and an analogous method as in the second part of the proof of Theorem 5.11. (Note that $t_k = 0$ at most for one k , and then, by Definition 5.1, $\tau_k = 1$.) One obtains

$$\begin{aligned} \text{Ret}(0^r) = & \{0\} \cup \{0^{r-t_1}\varphi^m(0)0^{-t_m}\} \cup \{0^{r-t_1}\varphi^m(k0)0^{-t_m} \mid 1 \leq k < m, t_k \neq 0\} \\ & \cup \{0^{r-t_1}\varphi^m(1k0)0^{-t_m} \mid 1 \leq k < m, t_k = 0\}. \end{aligned}$$

Consequently, $\#\text{Ret}(0^r) = m$ for $1 \leq r \leq t_1$ and $\#\text{Ret}(0^r) = m + 1$ for $t_1 < r < t_1 + t_m$.

Return words of $\varphi(0^r)0^s$, $1 \leq r < t_1$, $s \in \{t_1, \dots, t_{m-1}\}$.

We will show that, for every letter k , there exists exactly one return word ending with k and thus $\#\text{Ret}(\varphi(0^r)0^s) = m$. We distinguish the cases $k \geq 1$ and $k = 0$.

- For $k \geq 1$, it is an immediate consequence of the form of the substitution that each factor of u_β having suffix $k\varphi(0^r)0^s$ has as well the factor $\varphi^k(0)\varphi(0^r)0^s$ as suffix. Moreover, the factor $\varphi^k(0)\varphi(0^r)0^s$ has $\varphi(0^r)0^s$ as prefix. The only return word ending with k is therefore the shortest suffix v of $\varphi^k(0)$ such that $v\varphi(0^r)0^s$ has $\varphi(0^r)0^s$ as prefix.
- For $k = 0$, the factor $0\varphi(0^r)0^s$ is always suffix of $\varphi^m(0)\varphi(0^r)0^s$. The only return word ending with 0 is therefore the shortest suffix of $\varphi^m(0)$ starting with $\varphi(0^r)0^s$.

□

5.4. Property \mathcal{R}_m for u_β . In [2], necessary and sufficient conditions for affine complexity of infinite words u_β associated with simple Parry numbers were given.

Theorem 5.13. [2] *Let $d_\beta(1) = t_1 \dots t_m$. Then the following four statements are equivalent.*

- (1) *The infinite word u_β has an affine complexity.*
- (2) *For every $n \in \mathbb{N}$, there exists exactly one left special factor w of length n and its degree is $\deg_-(w) = m$.*
- (3) *For every left special factor w there exists a letter k such that wk is left special as well.*
- (4) *For every right special factor w there exists a letter k such that kw is right special as well.*

If one of the conditions is satisfied, then the complexity $\mathcal{C}(n)$ is $(m - 1)n + 1$ for every $n \in \mathbb{N}$.

The necessary and sufficient condition on affine complexity can be reformulated as a condition on parameters t_1, \dots, t_m .

Theorem 5.14. [2] *Let $d_\beta(1) = t_1 \dots t_m$. The infinite word u_β has an affine complexity if and only if the two following conditions are satisfied*

- (1) $t_m = 1$
- (2) *either $t_1 \dots t_{m-1}$ has no common proper prefix and suffix or there exists a word v and $k \in \mathbb{N}$, $k \geq 2$, such that $t_1 \dots t_{m-1} = v^k$.*

The following result is a direct consequence of the form of Rauzy graphs of factors of u_β having affine complexity (see Section 3.2 and apply Theorem 5.13 (2)).

Theorem 5.15. *Let u_β have an affine complexity. Then u_β satisfies property \mathcal{R}_m .*

It can be shown that the inverse implication is also true, i.e., if at least one of the conditions of Theorem 5.14 does not hold, then there exists a factor with more than m return words.

Proposition 5.16. *Let $d_\beta(1) = t_1 \dots t_m$ and $t_m \geq 2$. Then u_β does not satisfy property \mathcal{R}_m .*

Proof. We will show that the factor $0^{t_m+t_1-1}$ has $m+1$ return words. Clearly, $0^{t_m+t_1-1}$ has a unique ancestor $(m-1)0$, therefore it occurs only as factor of $\varphi((m-1)0) = 0^{t_m+t_1}1$. If it occurs as prefix of $0^{t_m+t_1}1$, then the corresponding return word is 0. If the factor $0^{t_m+t_1-1}$ is followed by 1, then the set of the corresponding return words is equal to $0^{-1}\text{Ret}(\varphi(m-1)) = 0^{-1}\varphi(\text{Ret}(m-1))$. By Lemmas 3.4 and 5.10 we have $0^{-1}\text{Ret}(\varphi(m-1)) = m$, and consequently $\#\text{Ret}(0^{t_m+t_1-1}) = m+1$. \square

If the second condition of Theorem 5.14 is not satisfied, then there exist left special factors which are not prefixes of u_β . The construction of such factors is described in [2]. It can be proved that these factors have more than m return words. Since the description of the construction is technical in general, we illustrate it only on a concrete example.

Example. If $d_\beta(1) = 11011$, i.e. $\varphi(0) = 01$, $\varphi(1) = 02$, $\varphi(2) = 3$, $\varphi(3) = 04$, $\varphi(4) = 0$, then $w = 0102010 = \varphi^2(0)\varphi(0)0$ is a bispecial factor which is not a prefix of u_β since $u_\beta = \varphi^\infty(0)$ starts with

0 1 02 013 010204 0102013010 0102013010204010201 0102013010204010201301001020130102 0102...

Since the factors 100 and 2040 are in $\mathcal{L}(u_\beta)$, it follows, by applying φ twice, that $3w2$ and $4w1$ are in $\mathcal{L}(u_\beta)$. Thus, w is a bispecial factor. The ancestors of w are 0101 and 0100, and the ancestors of these words are 00 and 040. Since 4 is always preceded and followed by 0, the number of return words of w equals thus the number of factors of the form xvy , where $x, y \in \{00, 4\}$ and v does not contain 00 and 4. Of the $m = 5$ factors in $\text{Ret}(4)4$ given by Lemma 5.10, the 3 factors $4\varphi^4(0)$, $4\varphi^4(20)$ and $4\varphi^4(40)$ do not contain 00 and satisfy therefore these conditions. For the factors $4\varphi^4(10)$ and $4\varphi^4(130)$, the (common) prefix $4\varphi^4(1)0$ and the suffixes $0\varphi^4(0)$ and $0\varphi^4(3)0$ satisfy these conditions. Therefore w has $m+1 = 6$ return words, even though $t_m = 1$.

6. CONCLUSION

We have shown in the case of infinite words u_β associated with simple Parry numbers that if the complexity of u_β is $\mathcal{C}(n) = (m-1)n + 1$ for all $n \in \mathbb{N}$, then u_β satisfies the property \mathcal{R}_m . We conjecture more generally that this condition is necessary for all words to satisfy \mathcal{R}_m . Unlike the case $m = 2$, this condition is not sufficient for $m \geq 3$. E.g., the fixed point of a certain recoding of the Chacon substitution on 3 letters has complexity $2n + 1$ (Ferenczi [6]) and it is easy to see that it contains factors with 4 return words (Vuillon [17]).

A difference between u_β with affine complexity and the fixed point v of the Chacon substitution is that the language of v contains maximal right and maximal left special factors. (We call a factor w maximal right special if aw is not right special for any $a \in \mathcal{A}$, and define maximal left special factors analogously.) We conjecture that an infinite word u has property \mathcal{R}_m if and only if u has complexity $\mathcal{C}(n) = (m-1)n + 1$ for every $n \in \mathbb{N}$ and the language of u contains neither maximal right special nor maximal left special factors.

ACKNOWLEDGEMENTS

The authors acknowledge financial support by the Czech Science Foundation GA ĆR 201/05/0169 and by the grant LC06002 of the Ministry of Education, Youth and Sports of the Czech Republic.

REFERENCES

- [1] J.-P. Allouche, J. Shallit, *The ubiquitous Prouhet-Thue-Morse sequence*, Sequences and their applications (Singapore, 1998), 1–16, Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London (1999)
- [2] J. Bernat, Z. Masáková, E. Pelantová, *On a class of infinite words with affine factor complexity*, preprint arXiv:math/0612466
- [3] J. Berstel, *Recent results on extensions of Sturmian words*, Int. J. Algebra Comput. **12** (2002), 371–385
- [4] F. Durand, *A characterization of substitutive sequences using return words*, Discrete Math. **179** (1998), 89–101
- [5] S. Fabre, *Substitutions et β -systèmes de numération*, Theor. Comput. Sci. **137** (1995), 219–236
- [6] S. Ferenczi, *Les transformations de Chacon: combinatoire, structure géométrique, lien avec les systèmes de complexité $2n + 1$* , Bull. Soc. Math. Fr. **123** (1995), 271–292
- [7] S. Ferenczi, Ch. Holton, L. Q. Zamboni, *Structure of three interval exchange transformations. II. A combinatorial description of the trajectories.*, J. Anal. Math. **89** (2003), 239–276
- [8] S. Ferenczi, C. Mauduit, A. Nogueira, *Substitution dynamical systems: algebraic characterization of eigenvalues*, Ann. Sci. Éc. Norm. Supér. **29** (1996), 519–533
- [9] Ch. Frougny, Z. Masáková, E. Pelantová, *Complexity of infinite words associated with beta-expansions*, Theor. Inform. Appl. **38** (2004), 163–185; Corrigendum, Theor. Inform. Appl. **38** (2004), 269–271
- [10] J. Justin, L. Vuillon, *Return words in Sturmian and episturmian words*, Theor. Inform. Appl. **34** (2000), 343–356
- [11] F. Mignosi, P. Séebold, *If a DOL-language is k -power-free then it is circular*, in: A. Lingas, R. Karlsson, S. Carlsson (editors), ICALP’93, Lect. Notes Comp. Sci. **700**, 507–518, Springer
- [12] M. Morse, G. A. Hedlund, *Symbolic dynamics II. Sturmian trajectories*, Amer. J. Math. **62** (1940), 1–42
- [13] W. Parry, *On the β -expansions of real numbers*, Acta Math. Acad. Sci. Hung. **11** (1960), 401–416
- [14] M. Queffélec, *Substitution Dynamical Systems – Spectral Analysis*, Lecture Notes in Math. **1294**, Springer, Berlin, 1987
- [15] W. P. Thurston, *Groups, tilings, and finite state automata*, AMS Colloquium lectures, 1989
- [16] L. Vuillon, *A characterization of Sturmian words by return words*, Eur. J. Comb. **22** (2001), 263–275
- [17] L. Vuillon, *On the number of return words in infinite words with complexity $2n + 1$* , LIAFA Research Report 2000/15

DOPPLER INSTITUTE FOR MATHEMATICAL PHYSICS AND APPLIED MATHEMATICS, AND DEPARTMENT OF MATHEMATICS, FNSPE, CZECH TECHNICAL UNIVERSITY, TROJANOVA 13, 120 00 PRAHA 2, CZECH REPUBLIC

E-mail address: l.balkova@centrum.cz, Pelantova@km1.fjfi.cvut.cz

LIAFA, CNRS, UNIVERSITÉ PARIS DIDEROT, 2 PLACE JUSSIEU, CASE 7014, 75251 PARIS CEDEX 05, FRANCE

E-mail address: steiner@liafa.jussieu.fr