# A characterization of Sturmian words by return words

Laurent Vuillon
LIAFA, Université Paris VII,
case 7014, 2 place Jussieu,
75251 Paris Cedex 05, France.
Phone. +33 1 44 27 70 97; Fax. +33 1 44 27 62 86;
e-mail: Laurent.Vuillon@liafa.jussieu.fr

February 18, 2000

*Résumé:* Nous présentons une nouvelle caractérisation des mots de Sturm basée sur les mots de retour. Si l'on considère chaque occurrence d'un mot w dans un mot infini récurrent, on définit l'ensemble des mots de retour de w comme l'ensemble de tous les mots distincts commençant par une occurrence de w et finissant exactement avant l'occurrence suivante de w. Le résultat principal montre qu'un mot est sturmien si et seulement si pour chaque mot w non vide apparaissant dans la suite, la cardinalité de l'ensemble des mots de retour de w est égale à deux.

*Abstract:* We present a new characterization of Sturmian words using return words. Considering each occurrence of a word $w$ in a recurrent word, we define the set of return words over $w$ to be the set of all distinct words beginning with an occurrence of $w$ and ending exactly before the next occurrence of $w$ in the infinite word. It is shown that an infinite word is a Sturmian word if and only if for each non-empty word $w$ appearing in the infinite word, the cardinality of the set of return words over $w$ is equal to two.

# 1 Introduction

Sturmian words are infinite words over a binary alphabet with exactly $n+1$ factors of length $n$ for each $n \geq 0$ (see [2, 6, 12]). In fact, the study of the Sturmian words appears in many areas like combinatorics on words ([6]), symbolic dynamics ([3, 1, 7, 21]), theoretical computer science ([5, 17]) and tilings ([8, 14, 20, 22, 24]). The Sturmian words have many equivalent characterizations (see for a complete presentation of Sturmian words [6]) using complexity function ([3, 8, 24]), balanced words ([12]), cutting sequences ([9]), mechanical words ([12]) and description by morphisms of the Sturmian words ([3, 17, 21]). In this article, the approach is based on the concept of return words introduced for the first time by Durand in order to obtain nice results on the characterization of primitive substitutive sequences (see [10, 11]). This notion is quite natural and can be seen as a symbolic version of the first return map (see for a presentation of symbolic dynamics [4, 13, 15, 18]). In the article of Alessandri and Berthé [1], the notion of first return map also appears in the three gap theorem (this theorem for Sturmian words gives a geometrical proof of Proposition 4.1).

The following construction enlights the structure of the Sturmian words, in particular it focuses on its self-similarity structure. Indeed, whatever the length of a word $w$ appearing in a Sturmian word, we construct two return words ($u$ and $v$) over $w$ such that the Sturmian word is the concatenation of these two return words $u$ and $v$. In terms of tilings, if we associate to $u$ and $v$ two tiles (which are segments of length equal to the number of letters in $u$ and $v$, so that the positions of the tiles are given by the positions of $u$ and $v$ in the Sturmian word) then the line is tiled by this two kinds of tiles. The important point to note here is this new form of self-similarity in the tiling of the line by a Sturmian word. In precedent works, the self-similarity remains from morphism invariance of the tiling (see [14, 19, 20, 23]). Here we show a more general invariance of the tiling, namely the invariance of the number of tiles which give the tiling of the line by Sturmian words.

The last observation is a minimality-like result. In terms of complexity, Sturmian words are aperiodic words with minimal complexity. Furthermore, Proposition 3.1 presents a characterization of ultimately periodic words which is: an infinite word is an ultimately periodic word if and only if there exits $w_0$ such that the set of return words over $w_0$ has one element. In addition to that, the main theorem states that

**Theorem** *A binary infinite word $U$ is Sturmian if and only if the set of return words over $w$ has exactly two elements for every non empty word $w$.*

We find the following minimality-like result as a corollary of the main theorem and of Proposition 3.1: Sturmian words are aperiodic words with minimal cardinality of the set of return words.

The structure of the article is the following. Section 2 contains basic definitions and notations. Section 3 shows a characterization by return words of ultimately periodic words. In Section 4, we prove that if an infinite word is Sturmian then it has two return words over every non-empty factor. In Section 5, we construct two infinite sequences of return words associated to an infinite word with two return words over each factor. Section 6 establishes the relation between return words and standard pairs. In addition to that, we prove that if an infinite word has two return words over each factor then it is a Sturmian word.

## 2 Basic definitions and examples

Let $\mathcal{A} = \{0, 1\}$ be a binary alphabet. We denote by $\mathcal{A}^*$ the set of finite words on $\mathcal{A}$ and by $\mathcal{A}^{\mathbb{N}}$ the set of one-sided infinite words. A word $w$ is a *factor* of a word $x \in \mathcal{A}^*$ if there exist some words $u, v \in \mathcal{A}^*$ such that $x = uwv$. An infinite word $U$ is called *recurrent* if every factor of $U$ appears infinitely many times in $U$. For a finite word $w = w_1 w_2 \cdots w_n$, the length of $w$ is denoted by $|w|$ and is equal to $n$. The set of factors of $U$ with length $n$ is denoted by $L_n(U)$. The language $L(U) = \cup_n L_n(U)$ is the set of factors of $U$. For two finite words $w$ and $u$, the number of occurrences of $w$ on $u$ is denoted by $|u|_w$ and $|u|_w = Card\{i | 0 \leq i \leq |u| - |w| | u_{i+1} u_{i+2} \cdots u_{i+|w|} = w\}$.

The *position set* $i(U, w) = \{i_1, i_2, \cdots, i_k, \cdots\}$ of the word $w$ is a set of integers $i(U, w) = \{i_1, i_2, \cdots, i_k, \cdots\}$ where $i_k$ represents the position of the first letter of the $k$-th occurrence of the word $w$ in the infinite word $U$. In a more formal way, $i_k \in i(U, w)$ if and only if $U_{i_k} U_{i_k+1} \cdots U_{i_k+|w|-1} = w$ and $|U_1 \cdots U_{i_k+|w|-1}|_w = k$. Since the infinite word $U$ is recurrent the set $i(U, w)$ is infinite. For a recurrent word $U$, the set of *return words over* $w$ is the set (denoted by $\mathcal{H}_{U,w}$) of all distinct words with the following form:

$$U_{i_k} U_{i_k+1} \cdots U_{i_{k+1}-1}$$

for all $k \in \mathbb{N}, k > 0$. This definition is best understood on an example. Let $U_1 = (0100100001)^\omega$ be an infinite word on the alphabet $\mathcal{A}$. By definition, the set of return words over 01 is $\mathcal{H}_{U_1,01} = \{010, 01000, 01\}$. Indeed, the infinite word $U_1$ can be written

$$(\underline{0}10\underline{0}1000\underline{0}1\underline{0}10\underline{0}1000\underline{0}1)^\omega$$

where $\underline{0}$ denotes the position of the first letter for each occurrence of the word 01. In the preceding example, the length of each element of $\mathcal{H}_{U_1,01}$ is larger than the word $w = 01$. Let us mention that the length of a return word over $w$ could be smaller than the length of $w$. For example, let the infinite word be $U_2 = (00001)^\omega$, we find $\mathcal{H}_{U_2,000} = \{0, 0001\}$. Indeed, the position of the first occurrence of 000 in $U_2$ is $i_1 = 1$ and the position of the second occurrence of 000 in $U_2$ is $i_2 = 2$, hence 0 is element of the set $\mathcal{H}_{U_2,000}$. From now on, we write $\mathcal{H}_w$ for $\mathcal{H}_{U,w}$.

# 3 Ultimately periodic words

Before proving the main theorem, we establish a simpler result which is a characterization of ultimately periodic words.

**Proposition 3.1** *A recurrent word is ultimately periodic if and only if there exists $w_0$ such that the set of return words over $w_0$ has exactly one element.*

*Proof of the Proposition*

If the infinite word is ultimately periodic, it can be written as $U = pv^\omega$ with $p$ a finite word and $|v|$ the shortest period of the infinite word $U' = v^\omega$. If $|vv|_v = 2$ then $\#\mathcal{H}_{U',v} = 1$. Furthermore, there exists $k$ such that $|v^k| \geq |p| + |v| - 1$. This leads to $\#\mathcal{H}_{U,v^k} = 1$.

Otherwise, if $|vv|_v \geq 3$ then $vv = v_1 v v_2$. Hence the word $v$ can be written as $v = v_1 v_2 = v_2 v_1$. By a classical result on combinatorics on words (see [16]): if $x = yz$ and $x' = zy$ and $x = x'$ then $y = u^p$ and $z = u^q$. The word $v$ is a power of another word, in contradiction with the minimality of $v$.

Conversely, if $\#\mathcal{H}_w = 1$ for a given word $w$, by definition $w$ is a factor of the infinite word $U$. Then $U$ can be written as $U = pwS$ where $pw$ is a prefix of the infinite word with exactly one occurrence of the factor $w$ and $S$ an infinite word. By hypothesis $\mathcal{H}_w = \{v\}$ for a given word $v$, consequently $U = pv^\omega$. It follows that the infinite word $U$ is ultimately periodic. □

# 4 Sturmian implies two return words

This section uses the graph of words associated with the factors of a Sturmian word $U$ (see [3, 6]). In the graph, the vertices are words of length $n$. There is an edge between the vertices $u$ and $v$ if and only if there exist two letters $a$ and $b$ such that $ua$ and $bv$ are factors of $U$ and $ua = bv$ (we label the edge by $[a,b]$, $u \xrightarrow{[a,b]} v$). As $U$ is a Sturmian word, there exists for every $n$ a unique word $R$ (resp. $L$) of length $n$ with two right extensions (resp. with two left extensions). The other words have a unique right extension (resp. left extension).

Consequently, the graph of words for Sturmian words is composed by three paths (see Figures 1 and 2): the first and the second between $R$ and $L$. The first path is

$$R \xrightarrow{[a_1,b_1]} v_1 \xrightarrow{[a_2,b_2]} v_2 \longrightarrow \cdots v_{k-1} \xrightarrow{[a_k,b_k]} L.$$

The second path is

$$R \xrightarrow{[a'_1,b'_1]} v'_1 \xrightarrow{[a'_2,b'_2]} v'_2 \longrightarrow \cdots v'_{k'-1} \xrightarrow{[a'_{k'},b'_{k'}]} L.$$

The third path is between $L$ and $R$ namely

$$L \xrightarrow{[c_1,d_1]} w_1 \xrightarrow{[c_2,d_2]} w_2 \longrightarrow \cdots w_{k''-1} \xrightarrow{[c_{k''},d_{k''}]} R.$$

The third path has length 0 if $R_n = L_n$.

In the proof of the following proposition, we use the balanced characterization of a Sturmian word (see [6, 12]): a non-ultimately periodic word $U$ is Sturmian if and only if $\forall n \in \mathbb{N}^*, \forall u, v \in L_n(U), ||u|_0 - |v|_0| \leq 1$.

**Proposition 4.1** *The set of return words over $w$ of a Sturmian word has exactly two elements for every non-empty word $w$.*

*Proof of the proposition:* Let $G$ be the graph of word appearing in $U$ with length $|w|$. Suppose that $w$ is a factor of the third path $Lc_1c_2\cdots c_{k''}$ then $w$ has two return words. Indeed, the first comes from the first path and the second comes from the second path.

Suppose now that $w$ is factor of length $n$ of the first or the second path. The role of the first and the second paths is symmetrical. Without loss of generality, we suppose that $w$ is a factor of the first path.

In order to study the return words, we precise the symmetry of the graph of words. In a Sturmian word the language is invariant by mirror image, that is if $x = x_1x_2\cdots x_n$ is a factor of $U$ then the mirror image of $x$ namely $\tilde{x} = x_nx_{n-1}\cdots x_1$ is a factor of $U$. Let $\tilde{G}$ be the graph of mirror image words of length $|w|$. The mirror image transformation maps $G$ into $\tilde{G}$. Furthermore, the language of Sturmian word is invariant by mirror image. Thus the two graphs are equals. It follows that the graph of word $G$ has an

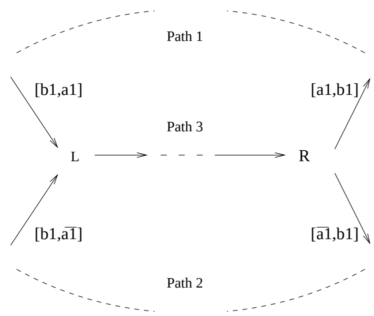


Figure 1: Graph of words.

axial symmetry for the mirror image transformation (see Figure 1 and 2). In consequence, the first path is given by

$$R \xrightarrow{[a_1,b_1]} v_1 \longrightarrow \cdots v_{k-1} \xrightarrow{[b_1,a_1]} L.$$

The second path is given by

$$R \xrightarrow{[\bar{a}_1,b_1]} v'_1 \longrightarrow \cdots v'_{k'-1} \xrightarrow{[b_1,\bar{a}_1]} L,$$

where the complementation operation is $\overline{a} = 1 - a$ for $a = 0$ or 1.

Let $r$ be a return word over $w$. In order to study the construction of the return word $r$, we read the word from left to right and we label the paths in the graph of words. We label by 1 if we use the first path, by 2 if we use the second path and by 3 if we use the third path. For example, we label by 131 if the word begins in the first path, takes the third path and ends in the first path (without taking the second path). We label by $13(23)^m1$ if the word begins in the first path, takes the third path, takes $m$ times the concatenation of the second path and the third path, and ends in the first path.

Consider now the length of each label for a word beginning in the first path. There exists $\ell_0$ such that $13(23)^{\ell_0}1$ is the shortest return word. If we consider two return words, we have $13(23)^{\ell_0}1$ and $13(23)^{\ell_1}1$ with $\ell_0 < \ell_1$. If $\ell_0 + 2 \leq \ell_1$ then both $13(23)^{\ell_0}1$ and $23(23)^{\ell_0}23$ appear in the labels. Let $z$ be the largest common factor associated to the labels $13(23)^{\ell_0}1$ and $23(23)^{\ell_0}23$. The word $z$ is constructed by using the third path and $\ell_0$ times the concatenation of the second path and the third path. By construction $L$ is a prefix of $z$ and $R$ is suffix of $z$. In consequence, if we consider the label

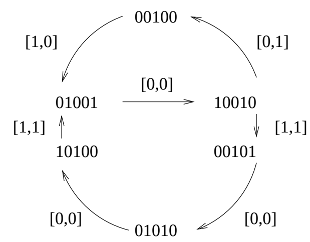


Figure 2: Example of graph of words with length 5.

$13(23)^{\ell_0}1$ then we find the word:

$$a_1 z a_1$$

and if we consider the label $23(23)^{\ell_0}23$ then we find the word

$$\overline{a}_1 z \overline{a}_1.$$

We find two factors of the infinite word which are not balanced, thus the infinite word cannot be a Sturmian word. In consequence, the number of return words for a Sturmian word cannot be greater than two. By Proposition 3.1 the number of return words cannot be smaller than two. The only remaining case is two return words labelled by $13(23)^{\ell_0}1$ and $13(23)^{\ell_0+1}1$. We have proved that for a Sturmian word, the number of return words over each non-empty word is exactly two. □

To illustrate the previous proposition, we consider the following graph of words (see Figure 2):

The word 01001 in this example is a factor of the third path. Then the return words over 010010 are the followings, by using the first path we find 01001001 and by using the second path we find 0100101001. Hence $\mathcal{H}_{01001} = \{010, 01001\}$.

The word 00100 is factor of the first path. Then the return words over 00100 are the followings: the label 131 is associated to the word 00100100 and the label 13231 is associated to the word 0010010100100. More generally, the label $13(23)^k1$ is associated to the word $00100(10100)^k100$. It is easy to check that if we consider the labels $13(23)^k1$ and $13(23)^{k+2}1$ then the words are $00100(10100)^k100$ and $0010010100(10100)^k10100100$. Thus the factors

$$00100(10100)^k100$$

and

$$10100(10100)^k101$$

are not balanced. The only remaining case is two return words labelled by $13(23)^k1$ and $13(23)^{k+1}1$.

# 5 Construction of the Sturmian words

In the next three sections, we prove some propositions in order to show the main theorem of this article.

In this section, we suppose that the set of return words over $w$ has two elements for every non-empty word and by induction, we construct an infinite sequence of pairs of return words $(u,v) = (u_n,v_n)_{n\in\mathbb{N}}$ where the sequence $u = (u_n)_{n\in\mathbb{N}}$ tends to a Sturmian word.

In the construction, we use extensively the following property: if $\mathcal{H}_w = \{u,v\}$ then the infinite word $U$ up to a shift can be constructed as the concatenation of words $u$ and $v$ (the shift of the infinite word $U = U_1U_2\cdots U_i\cdots$ is the infinite word $S(U) = U_2\cdots U_i\cdots$; by composition a shift of $U$ is denoted by $S^j(U)$ and is equal to the infinite word $S^j(U) = U_{j+1}\cdots U_i\cdots$). In other words, if $\mathcal{H}_w = \{u,v\}$ then there exists $j$ such that $S^j(U)\in(u+v)^\omega$. In this way, we construct an infinite sequence of pairs of return words $(u_n,v_n)_{n\in\mathbb{N}}$ such that the infinite word $U$ up to a shift can be written as the concatenation of $u_n$ and $v_n$. The idea is to control the sequence $u = (u_n)_{n\in\mathbb{N}}$ in order to tend to a Sturmian word.

## 5.1 First step of the induction

**Proposition 5.1** *Let $U$ be a recurrent infinite word in the alphabet $\{0,1\}$. If the set of return words over $w$ has exactly two elements for every non-empty word $w$, then either*

$$\mathcal{H}_0 = \{0, 01\}, \mathcal{H}_1 = \{10^n, 10^{n+1}\}$$

*with $n > 0$ or*

$$\mathcal{H}_0 = \{01^m, 01^{m+1}\}, \mathcal{H}_1 = \{1, 10\}$$

*with $m > 0$.*

*Proof of the proposition*

As the infinite word is the concatenation of words 0 and 1, then in general form, we have

$$\mathcal{H}_0 = \{01^{m_1}, 01^{m_2}\}, \mathcal{H}_1 = \{10^{n_1}, 10^{n_2}\}$$

with $0 \leq m_1 < m_2$ and $0 \leq n_1 < n_2$.

In the points a), b), c) and d) of this proof, we make a reasoning by contradiction.

a) If $m_1 = n_1 = 0$, the sets of return words yield

$$\mathcal{H}_0 = \{0, 01^{m_2}\}, \mathcal{H}_1 = \{1, 10^{n_2}\}.$$

In other words, a shift of the infinite word $U$ is given by the concatenation of the words 0 and $01^{m_2}$ (resp. 1 and $10^{n_2}$). For this reason, 0 (resp. 1) always occurs in blocks of length $n_2$ ($m_2$). In consequence, the infinite word is a shift of the following periodic word: $\exists i, 0 \leq i \leq m_2 + n_2$ such that $U = S^i(1^{m_2}0^{n_2}1^{m_2}0^{n_2}\cdots)$. By Proposition 3.1, for a periodic word, there exists $w$ such that $\#\mathcal{H}_w = 1$. This is a contradiction. In conclusion, either $m_1$ or $n_1$ is larger than 0.

b) If $m_1 > 0$ and $n_1 > 0$ then 00 and 11 are factors. This implies that 0 is a return word over 0 and 1 is a return word over 1. In other terms

$$\mathcal{H}_1 = \{1, 10^{m_1}, 10^{m_2}\}, \mathcal{H}_0 = \{0, 01^{n_1}, 01^{n_2}\}.$$

There is a contradiction because $\#\mathcal{H}_0 = \#\mathcal{H}_1 = 3$. In conclusion, either $m_1$ or $n_1$ is equal to 0.

c) If $m_1 = 0$, $m_2 \geq 2$ and $n_1 > 0$ then $\mathcal{H}_0 = \{0, 01^{m_2}\}$ and $\mathcal{H}_1 = \{10^{n_1}, 10^{n_2}\}$. But 11 is a factor of $01^{m_2}$ then $\mathcal{H}_1 = \{1, 10^{n_1}, 10^{n_2}\}$. This leads to a contradiction. By the same reasoning, if $n_1 = 0$, $n_2 \geq 2$ and $m_1 > 0$, we have a contradiction. In conclusion, either ($m_1 = 0$ and $m_2 = 1$) or ($n_1 = 0$ and $n_2 = 1$).

d) Suppose that $m_1 = 0$, $m_2 = 1$ and $n_2 = n_1 + l$ with $n_1 > 0$. The sets of return words are $\mathcal{H}_0 = \{0, 01\}$ and $\mathcal{H}_1 = \{10^{n_1}, 10^{n_1+l}\}$.

If $l > 1$ we consider now the return word $0^{n_1+1}$ then

$$\mathcal{H}_{0^{n_1+1}} = \{0, 0^{n_1+1}(10^{n_1})^k 1\}$$

for a given $k > 0$. The infinite word is a shift of the following periodic word: $\exists i, 0 \leq i \leq n_1(k+1)+l+2$ such that $U = S^i(10^{n_1+l}(10^{n_1})^k 10^{n_1+l}(10^{n_1})^k \cdots)$. By Proposition 3.1, there exists $w$ such that $\#\mathcal{H}_w = 1$. By the same reasoning, if $n_1 = 0$, $n_2 = 1$ and $m_2 = m_1 + l$ with $m_1 > 0$ and $l > 1$ then this leads to a contradiction.

The only remaining case is either $m_1 = 0$, $m_2 = 1$ and $n_2 = n_1 + 1$ with $n_1 > 0$ or $n_1 = 0$, $n_2 = 1$ and $m_2 = m_1 + 1$ with $m_1 > 0$. It follows that either

$$\mathcal{H}_0 = \{0, 01\}, \mathcal{H}_1 = \{10^n, 10^{n+1}\}$$

with $n > 0$ or

$$\mathcal{H}_0 = \{01^m, 01^{m+1}\}, \mathcal{H}_1 = \{1, 10\}$$

with $m > 0$. $\square$

### 5.2 Construction of an infinite sequence of return words

In the preceding section, we constructed two return words: either $u_0 = 10^{k_{-1}}$, $v_0 = 10^{k_{-1}+1}$ with $k_{-1} > 0$ or $u_0 = 01^{k_{-1}}$, $v_0 = 01^{k_{-1}+1}$ with $k_{-1} > 0$.

The main difficulty is to construct an infinite sequence of pairs of return words $(u_n, v_n)_{n\in\mathbb{N}}$ such that the infinite word $U$ up to a shift can be written as the concatenation of $u_n$ and $v_n$. In addition to that, we would like to control the sequence $u = (u_n)_{n\in\mathbb{N}}$ so that the sequence tends to a Sturmian word. Furthermore, we would like to construct return words at step $n$ which are the concatenation of return words at step $n-1$. Each pair of return words must encode the repetition of the return words of step $n-1$ (which are $u_n$ and $v_n$) namely at step $n$, we would like to have $u_n = u_{n-1}v_{n-1}^k$ and $v_n = u_{n-1}v_{n-1}^{k+1}$ with $k$ maximal (resp. $u_n = v_{n-1}u_{n-1}^k$ and $v_n = v_{n-1}u_{n-1}^{k+1}$ with $k$ maximal).

In the first part, we have computed two return words $u_0$ and $v_0$, where $u_0 = 10^{k_{-1}}$ is a prefix of $v_0 = 10^{k_{-1}+1}$. If we consider the following example (this is the beginning of the Fibonacci word , given by iteration of the morphism $\sigma(0) = 01$ and $\sigma(1) = 0$):

$$0100101001001010010100100101001001\cdots.$$

We can compute $\mathcal{H}_0 = \{0, 01\}$, $\mathcal{H}_1 = \{10, 100\}$, hence $u_0 = 10$ and $v_0 = 100$. From this, we find $\mathcal{H}_{u_0} = \{10, 100\}, \mathcal{H}_{v_0} = \{100, 10010\}$. Unfortunately, the computation of return words over $u_0$ does not give any information about the number of repetitions of 100 in the infinite word. In particular, the reader could find the same sets of return words over $u_0 = 10$ and $v_0 = 100$ with the following infinite word

$$01001001010010010010100100101001001001010010010 0\cdots.$$

In order to capture the structure of the infinite word, we define the notion of tiling with two tiles $u$ and $v$ which are in the same pair of return words of an infinite word. To encode the structure, we use the vocabulary of tiling theory. A *tiling* (with two tiles $u$ and $v$) denoted by $(A, u, B, v)$ of the infinite word $U$ is defined to be two tiles (namely $u$ and $v$ which are finite factors of $U$) and two sets of integers $A$ and $B$ such that :

- $i$ is an element of $A$, if and only if $u$ is a prefix of $U_iU_{i+1}\cdots$,
- $j$ is an element of $B$, if and only if $v$ is a prefix of $U_jU_{j+1}\cdots$,
- $(\bigcup_{\ell=0\cdots|u|-1}(A+\ell))\bigcap(\bigcup_{m=0\cdots|v|-1}(B+m))=\emptyset$ and
- $(\bigcup_{\ell=0\cdots|u|-1}(A+\ell))\bigcup(\bigcup_{m=0\cdots|v|-1}(B+m))$ is the set of all the integers larger than $\min(A\cup B)$.

In other term, $(A,u,B,v)$ is a tiling of the infinite word $U$ if and only if $(\bigcup_{\ell=0\cdots|u|-1}(A+\ell))$ and $(\bigcup_{m=0\cdots|v|-1}(B+m))$ form a partition of the set of integers larger or equal to $\min(A\cup B)$ (see for general references on this topic [14, 19, 20, 22, 23]).

With this definition and for the Fibonacci word

$$U=0100101001001010010100100101001 00\cdots,$$

$(i(u_0),u_0,i(v_0),v_0))$ is not a tiling of $U$. Indeed, by construction of return words, $i(u_0)\cap i(v_0)$ is non-empty, because $u_0$ is a prefix of $v_0$ and consequently $i(v_0)\subset i(u_0)$. A way to get round this second problem is to extend the return word. If we take $\hat{u}_0=10^{k_{-1}}1$ and $\hat{v}_0=10^{k_{-1}+1}1$ then $(i(\hat{u}_0),u_0,i(\hat{v}_0),v_0)$ is a tiling of $U$. For the following example: $U=010010100100101001010010010100100\cdots$, the position set for the factor 10 is

$$(i(10))=\{2,5,7,10,13,15,18,20,23,26,28,31,\cdots\}.$$

The position set for the factor 100 is

$$(i(100))=\{2,7,10,15,20,23,28,31,\cdots\}.$$

It is clear that the two sets $(i(10))$ and $(i(100))$ overlap. On the other hand, the position set for the factor 101 is $(i(101))=\{5,13,18,26\cdots\}$. The position set for the factor 1001 is $(i(1001))=\{2,7,10,15,20,23,28,31,\cdots\}$. Therefore $(i(101),10,i(1001),100)$ is a tiling of $U$. Indeed, for the Fibonacci word it is easy to check that the two sets $(\bigcup_{\ell=0\cdots|10|-1}(i(101)+\ell))$ and $(\bigcup_{m=0\cdots|100|-1}(i(1001)+m))$ form a partition of the set of integers larger or equal to 2.

In the sequel, we generalize this construction for an infinite sequence of return words.

In order to understand the structure of the infinite word $U$, we say that $u$ is *isolated* in the tiling $(A,u,B,v)$ if for every $i\in A$ $(\min_{j>i}(A\cup B))\in B$ and $(\max_{j<i}(A\cup B))\in B$. In other words, for each occurrence of $u$ in position

$A$, the word $u$ is surrounded by $v$ (i.e., $vuv$ is a factor of the infinite word). $v$ is isolated in the tiling $(A, u, B, v)$ if for every $i \in i(v)$ $(\min_{j>i}(A \cup B)) \in A$ and $(\max_{j<i}(A \cup B)) \in A$.

For example, in the Fibonacci word, the factor 10 is isolated in the tiling $(i(101), 10, i(1001), 100)$. The word 1010 is factor of the Fibonacci word but the factor 10 in position $i(101)$ is always surrounded by 100.

**Proposition 5.2** *Let $U$ be a recurrent infinite word in the alphabet $\{0, 1\}$. If the set of return words over $w$ has exactly two elements for every non-empty word $w$, then two infinite sequences of return words $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ are constructed as follows $u_{-1} = 0$, $v_{-1} = 1$ and $u_{n+1} = u_n v_n^{k_n}, v_{n+1} = u_n v_n^{k_n+1}$ with $k_n > 0$ or $u_{n+1} = v_n u_n^{k_n}, v_{n+1} = v_n u_n^{k_n+1}$ with $k_n > 0$.*

*Proof of the proposition:*

By induction, we would like to construct $u_{n+1} = u_n v_n^{k_n}, v_{n+1} = u_n v_n^{k_n+1}$ with $k_n > 0$ or $u_{n+1} = v_n u_n^{k_n}, v_{n+1} = v_n u_n^{k_n+1}$ with $k_n > 0$; and the words $\hat{u}_{n+1} = u_n v_n^{k_n} \hat{u}_n, \hat{v}_{n+1} = u_n v_n^{k_n+1} \hat{u}_n$ or $\hat{u}_{n+1} = v_n u_n^{k_n} \hat{v}_n, \hat{v}_{n+1} = v_n u_n^{k_n+1} \hat{v}_n$.

If $u_n$ is isolated in the tiling $(i(\hat{u}_n), u_n, i(\hat{v}_n), v_n)$, then the induction hypothesis is

$$(H_n) \quad \mathcal{H}_{\hat{u}_n} = \{u_{n+1}, v_{n+1}\}$$

$$\text{with } u_{n+1} = u_n v_n^{k_n}, v_{n+1} = u_n v_n^{k_n+1}$$

$$\text{and } \hat{u}_{n+1} = u_n v_n^{k_n} \hat{u}_n \in L(U), \hat{v}_{n+1} = u_n v_n^{k_n+1} \hat{u}_n \in L(U)$$

$$\text{and } (i(\hat{u}_{n+1}), u_{n+1}, i(\hat{v}_{n+1}), v_{n+1}) \text{ is a tiling.}$$

If $v_n$ is isolated in the tiling $(i(\hat{u}_n), u_n, i(\hat{v}_n), v_n)$, then the recurrence hypothesis is the same as before with an exchange of the role of $u_n$ and $v_n$.

First, we prove $H_{-1}$. We have $u_{-1} = \hat{u}_{-1} = 0$ and $v_{-1} = \hat{v}_{-1} = 1$ and $(i(0), 0, i(1), 1)$ is a tiling. By Proposition 5.1, either 0 is isolated and $\mathcal{H}_0 = \{u_0, v_0\}$ with $u_0 = 01^{k_{-1}}, v_0 = 01^{k_{-1}+1}$ or 1 is isolated and $\mathcal{H}_1 = \{u_0, v_0\}$ with $u_0 = 10^{k_{-1}}, v_0 = 10^{k_{-1}+1}$. In the sequel, we suppose that 0 is isolated (if 1 is isolated, the proof is the same by exchanging the role of 0 and 1). We have $\mathcal{H}_0 = \{01^{k_{-1}}, 01^{k_{-1}+1}\}$, thus by definition of return words over 0, $\hat{u}_0 = u_0 0 = 01^{k_{-1}}0$ and $\hat{v}_0 = v_0 0 = 01^{k_{-1}+1}0$ are factors of $U$.

It remains to show that $(i(\hat{u}_0), u_0, i(\hat{v}_0), v_0)$ is a tiling. The infinite word is on the alphabet $\{0, 1\}$, consequently $i(0) \cup i(1) = \mathbb{N}$. In addition to that, 0 is isolated and by definition of return words there exists $j$ such that the infinite word $S^j(U)$ is constructed by concatenation of words $01^{k_{-1}}$ and $01^{k_{-1}+1}$.

It suffices to show that $(i(\hat{u}_0) \cup i(\hat{v}_0)) = i(0)$ and $(i(\hat{u}_0) \cap i(\hat{v}_0)) = \emptyset$.

By contradiction, suppose that $(i(\hat{u}_0) \cup i(\hat{v}_0)) \neq i(0)$ then

- First case: $\exists \ell \in (i(\hat{u}_0) \cup i(\hat{v}_0))$ and $\ell \notin i(0)$. But $\ell \notin i(0)$ implies $U_\ell = 1$ and $\ell \in (i(\hat{u}_0) \cup i(\hat{v}_0))$ implies $U_\ell = 0$. There is a contradiction.

- Second case: $\exists \ell \in i(0)$ and $\ell \notin (i(\hat{u}_0) \cup i(\hat{v}_0))$. But $\ell \in i(0)$ implies that $U_\ell = 0$. Furthermore $U$ is constructed by concatenation of the words $01^{k_{-1}}$ and $01^{k_{-1}+1}$. This implies that after 0 we have either $1^{k_{-1}}0$ or $1^{k_{-1}}10$. There is a contradiction with $\ell \notin (i(\hat{u}_0) \cup i(\hat{v}_0))$.

We have shown that $(i(\hat{u}_0) \cup i(\hat{v}_0)) = i(0)$. Suppose now that $(i(\hat{u}_0) \cap i(\hat{v}_0)) \neq \emptyset$. Take $\ell \in (i(\hat{u}_0) \cap i(\hat{v}_0))$ this leads to $U_\ell U_{\ell+1} \cdots U_{\ell+|01^{k_{-1}}0|-1} = 01^{k_{-1}}0$ and $U_\ell U_{\ell+1} \cdots U_{\ell+|01^{k_{-1}+1}0|-1} = 01^{k_{-1}+1}0$. That is $U_{\ell+|01^{k_{-1}}0|-1}$ is equal to 0 and 1. There is a contradiction.

*Induction step:* Suppose $H_{-1}, H_0, \cdots, H_n$ true. We would like to prove $H_{n+1}$.

By the induction hypothesis $H_n$, $(i(\hat{u}_{n+1}), u_{n+1}, i(\hat{v}_{n+1}), v_{n+1})$ is a tiling. The infinite word is constructed by concatenation of the words $u_{n+1}$ and $v_{n+1}$. We have in general form:

$$\mathcal{H}_{\hat{u}_{n+1}} = \{u_{n+1}v_{n+1}^{m_1}, u_{n+1}v_{n+1}^{m_2}\}, \mathcal{H}_{\hat{v}_{n+1}} = \{v_{n+1}u_{n+1}^{n_1}, v_{n+1}u_{n+1}^{n_2}\}$$

with $0 \leq m_1 < m_2$ and $0 \leq n_1 < n_2$.

In the points a), b), c) and d) of this proof, we make a reasoning by absurd (the structure of the proof is the same as in the previous proof).

a) If $m_1 = n_1 = 0$, we have

$$\mathcal{H}_{\hat{u}_{n+1}} = \{u_{n+1}, u_{n+1}v_{n+1}^{m_2}\}, \mathcal{H}_{\hat{v}_{n+1}} = \{v_{n+1}, v_{n+1}u_{n+1}^{n_2}\}.$$

The infinite word is constructed by concatenation of the words $u_{n+1}$ and $u_{n+1}v_{n+1}^{m_2}$ (resp. $v_{n+1}$ and $v_{n+1}u_{n+1}^{n_2}$). For this reason, the number of consecutive $u_{n+1}$ (resp. $u_{n+1}$) is $n_2$ (resp. $m_2$). The infinite word is a shift of the following periodic word: $\exists i, 0 \leq i \leq |v_{n+1}| * m_2 + |u_{n+1}| * n_2$ such that $U = S^i(v_{n+1}^{m_2}u_{n+1}^{n_2}v_{n+1}^{m_2}u_{n+1}^{n_2}\cdots)$. By Proposition 3.1 there exists $w_0$ such that $\#\mathcal{H}_{w_0} = 1$. In conclusion, either $m_1$ or $n_1$ is larger than 0.

b) If $m_1 > 0$ and $n_1 > 0$ then $u_{n+1}u_{n+1}$ and $v_{n+1}v_{n+1}$ are factors. We have $(i(\hat{u}_{n+1}), u_{n+1}, i(\hat{v}_{n+1}), v_{n+1})$ is a tiling and the infinite word is constructed by concatenation of the words $u_{n+1}v_{n+1}^{m_1}$ and $u_{n+1}v_{n+1}^{m_2}$ (resp. $v_{n+1}u_{n+1}^{n_1}$ and $v_{n+1}u_{n+1}^{n_2}$). Then $v_{n+1}\hat{v}_{n+1}$ (resp. $u_{n+1}\hat{u}_{n+1}$) is factor of $U$. Consequently, $v_{n+1}$ (resp. $u_{n+1}$) is a return word over $\hat{v}_{n+1}$ (resp. $\hat{u}_{n+1}$). In other terms

$\mathcal{H}_{\hat{u}_{n+1}} = \{u_{n+1}, u_{n+1}v_{n+1}^{m_1}, u_{n+1}v_{n+1}^{m_2}\}, \mathcal{H}_{\hat{v}_{n+1}} = \{v_{n+1}, v_{n+1}u_{n+1}^{n_1}, v_{n+1}u_{n+1}^{n_2}\}.$

There is a contradiction because $\#\mathcal{H}_{\hat{u}_{n+1}} = \#\mathcal{H}_{\hat{v}_{n+1}} = 3$. In conclusion, either $m_1$ or $n_1$ is equal to 0.

c) If $m_1 = 0$, $m_2 \geq 2$ and $n_1 > 0$ then $\mathcal{H}_{\hat{u}_{n+1}} = \{u_{n+1}, u_{n+1}v_{n+1}^{m_2}\}$ and $\mathcal{H}_{\hat{v}_{n+1}} = \{v_{n+1}u_{n+1}^{n_1}, v_{n+1}u_{n+1}^{n_2}\}$. By the same argumentation as in b), $v_{n+1}\hat{v}_{n+1}$ is a factor of $U$. Then $\mathcal{H}_{\hat{v}_{n+1}} = \{v_{n+1}u_{n+1}^{n_1}, v_{n+1}, v_{n+1}u_{n+1}^{n_2}\}$. This leads to a contradiction. By the same reasoning, if $n_1 = 0$, $n_2 \geq 2$ and $m_1 > 0$, we have a contradiction. In conclusion, either ($m_1 = 0$ and $m_2 = 1$) or ($n_1 = 0$ and $n_2 = 1$).

d) Suppose that $m_1 = 0$, $m_2 = 1$ and $n_2 = n_1 + l$ with $n_1 > 0$. We have $\mathcal{H}_{\hat{u}_{n+1}} = \{u_{n+1}, u_{n+1}v_{n+1}\}$ and $\mathcal{H}_{\hat{v}_{n+1}} = \{v_{n+1}u_{n+1}^{n_1}, v_{n+1}u_{n+1}^{n_1+l}\}$. If $l > 1$ we consider now the return word over $u_{n+1}^{n_1}\hat{u}_{n+1}$ then

$$\mathcal{H}_{u_{n+1}^{n_1}\hat{u}_{n+1}} = \{u_{n+1}, u_{n+1}^{n_1+1}(v_{n+1}u_{n+1}^{n_1})^k v_{n+1}\}$$

for a given $k$. The infinite word is a shift of the following periodic word: $\exists i, 0 \leq i \leq |v_{n+1}| + |u_{n+1}| * (n_1 + \ell) + (|v_{n+1}| + |u_{n+1}| * (n_1)) * k$ such that $U = S^i(v_{n+1}u_{n+1}^{n_1+l}(v_{n+1}u_{n+1}^{n_1})^k v_{n+1}u_{n+1}^{n_1+l}(v_{n+1}u_{n+1}^{n_1})^k \cdots)$. By Proposition 3.1 there exists $w$ such that $\#\mathcal{H}_w = 1$. By the same reasoning, $n_1 = 0$, $n_2 = 1$ and $m_2 = m_1 + l$ with $m_1 > 0$ and $l > 0$, we have a contradiction.

The only remaining case is either $m_1 = 0$, $m_2 = 1$ and $n_2 = n_1 + 1$ with $n_1 > 0$ or $n_1 = 0$, $n_2 = 1$ and $m_2 = m_1 + 1$ with $m_1 > 0$. It follows that either

$$\mathcal{H}_{\hat{u}_{n+1}} = \{u_{n+1}, u_{n+1}v_{n+1}\}, \mathcal{H}_{\hat{v}_{n+1}} = \{v_{n+1}u_{n+1}^{n_1}, v_{n+1}u_{n+1}^{n_1+1}\}$$

or

$$\mathcal{H}_{\hat{v}_{n+1}} = \{v_{n+1}, v_{n+1}u_{n+1}\}, \mathcal{H}_{\hat{u}_{n+1}} = \{u_{n+1}v_{n+1}^{n_1}, u_{n+1}v_{n+1}^{n_1+1}\}.$$

We have constructed either

($v_{n+1}$ isolated) $u_{n+2} = v_{n+1}u_{n+1}^{k_{n+1}}$, $v_{n+2} = v_{n+1}u_{n+1}^{k_{n+1}+1}$ and by definition of return words, $\hat{u}_{n+2} = v_{n+1}u_{n+1}^{k_{n+1}}\hat{v}_{n+1}$ and $\hat{v}_{n+2} = v_{n+1}u_{n+1}^{k_{n+1}+1}\hat{v}_{n+1}$ are factors of $U$

or

($u_{n+1}$ isolated) $u_{n+2} = u_{n+1}v_{n+1}^{k_{n+1}}$, $v_{n+2} = u_{n+1}v_{n+1}^{k_{n+1}+1}$ and by definition of return words, $\hat{u}_{n+2} = u_{n+1}v_{n+1}^{k_{n+1}}\hat{u}_{n+1}$ and $\hat{v}_{n+2} = u_{n+1}v_{n+1}^{k_{n+1}+1}\hat{u}_{n+1}$ are factors of $U$.

We now show that $(i(\hat{u}_{n+2}), u_{n+2}, i(\hat{v}_{n+2}), v_{n+2})$ is a tiling. The infinite word $U$ can be written as the concatenation of the words $u_{n+1}v_{n+1}^{k_{n+1}}$ and $u_{n+1}v_{n+1}^{k_{n+1}+1}$.

It suffices to show that

$$(i(\hat{u}_{n+2}) \cup i(\hat{v}_{n+2})) = i(\hat{u}_{n+1})$$

and

$$(i(\hat{u}_{n+2}) \cap (i(\hat{v}_{n+2})) = \emptyset.$$

By contradiction, suppose that $(i(\hat{u}_{n+2}) \cup i(\hat{v}_{n+2})) \neq i(\hat{u}_{n+1})$ then

-First case: $\exists \ell \in (i(\hat{u}_{n+2}) \cup i(\hat{v}_{n+2}))$ and $\ell \notin i(\hat{u}_{n+1})$. But $\ell \notin i(\hat{u}_{n+1})$ implies $U_\ell U_{\ell+1} \cdots U_{\ell+|u_{n+1}|-1} \neq u_{n+1}$ and $\ell \in (i(\hat{u}_{n+1}) \cup i(\hat{v}_{n+1}))$ implies $U_\ell U_{\ell+1} \cdots U_{\ell+|u_{n+1}|-1} = u_{n+1}$. There is a contradiction.

-Second case: $\exists \ell \in i(\hat{u}_{n+1})$ and $\ell \notin (i(\hat{u}_{n+2}) \cup i(\hat{v}_{n+2}))$. But $\ell \in i(u_{n+1})$ implies $U_\ell U_{\ell+1} \cdots U_{\ell+|u_{n+1}|-1} = u_{n+1}$. Furthermore, the infinite word is constructed by the concatenation of words $u_{n+1}v_{n+1}^{k_{n+1}}$ and $u_{n+1}v_{n+1}^{k_{n+1}+1}$. This fact implies that after $u_{n+1}$, we have either $v_{n+1}^{k_{n+1}}$ or $v_{n+1}^{k_{n+1}+1}$. There is a contradiction with $\ell \notin (i(\hat{u}_{n+1}) \cup i(\hat{v}_{n+1}))$.

We have shown that $(i(\hat{u}_{n+2}) \cup (i(\hat{v}_{n+2})) = i(\hat{u}_{n+1})$. Suppose now that $(i(\hat{u}_{n+2}) \cap i(\hat{v}_{n+2})) \neq \emptyset$. Take $\ell \in (i(\hat{u}_{n+2}) \cap i(\hat{v}_{n+2}))$ we have

$$U_\ell U_{\ell+1} \cdots U_{\ell+|\hat{u}_{n+1}|-1} = u_{n+1}v_{n+1}^{k_{n+1}}\hat{u}_{n+1}$$

and

$$U_\ell U_{\ell+1} \cdots U_{\ell+|\hat{v}_{n+1}|-1} = u_{n+1}v_{n+1}^{k_{n+1}}v_{n+1}\hat{u}_{n+1}.$$

That is $U_{\ell+|u_{n+1}v_{n+1}^{k_{n+1}}v_{n+1}|}$ is equal to 0 and 1. There is a contradiction. □

## 6 Construction of the Sturmian word.

This section gives a link between the two infinite sequences of return words and an infinite sequence of standard pairs.

We have constructed the following sequence of pairs of return words : $u_{-1} = 0, v_{-1} = 1$ and $u_{n+1} = u_n v_n^{k_n}, v_{n+1} = u_n v_n^{k_n+1}$ or $u_{n+1} = v_n u_n^{k_n}, v_{n+1} = v_n u_n^{k_n+1}$. We call the sequence

$$k = k_{-1}k_0k_1 \cdots k_i \cdots$$

the directive sequence of the return words.

This construction is closely related to the construction of characteristic Sturmian words by using Rauzy rules (see for general references on Rauzy's rules [5, 17, 21]).

Two sequences of words $(A_n)_{n\in\mathbb{N}}$ and $(B_n)_{n\in\mathbb{N}}$ are constructed as follows: $A_{-1}=0$ and $B_{-1}=1$ and

$$(R_1)\ \begin{array}{ll} A_{n+1} & = A_n \\ B_{n+1} & = A_n B_n \end{array} \quad \text{or} \quad (R_2)\ \begin{array}{ll} A_{n+1} & = B_n A_n \\ B_{n+1} & = B_n \end{array}$$

The pairs $(A_n, B_n)$ are called standard pairs. A Sturmian word $x$ is characteristic if both $0x$ and $1x$ are Sturmian (see [6]).

**Proposition 6.1 (Rauzy [21])** *Both sequences $(A_n)_{n\in\mathbb{N}}$ and $(B_n)_{n\in\mathbb{N}}$ have the same limit which is a characteristic word; conversely, any characteristic word is the limit of two such sequences.*

By composition of the Rauzy rules, we construct

$$S_1^{(k)}(u,v) = R_2 \circ (R_1)^k(u,v) = (u^k vu, u^k v)$$

and

$$S_2^{(k)}(u,v) = R_1 \circ (R_2)^k(u,v) = (v^k u, v^k uv).$$

The difference between our pair of return words and Rauzy's construction (standard pair) is the following: we have the word $uv^k$ instead of $v^k u$ and $uv^{k+1}$ instead of $v^k uv$. The next proposition gives a way to find standard pairs by a cyclic permutation of our return words.

**Proposition 6.2** *Let $U$ be a recurrent infinite word on the alphabet $\{0,1\}$. If the set of return words over $w$ has exactly two elements for every non-empty word $w$. Then two infinite sequences $(\vec{u}_n)_{n\in\mathbb{N}}$ and $(\vec{v}_n)_{n\in\mathbb{N}}$ factors of $U$ and elements of a standard pair are constructed as follows: $\vec{u}_{-1}=0$, $\vec{v}_{-1}=1$ and $\vec{u}_{n+1}=\vec{v}_n^{k_n}\vec{u}_n, \vec{v}_{n+1}=\vec{v}_n^{k_n}\vec{u}_n\vec{v}_n$ with $k_n>0$ or $\vec{u}_{n+1}=\vec{u}_n^{k_n}\vec{v}_n, \vec{v}_{n+1}=\vec{u}_n^{k_n}\vec{v}_n\vec{u}_n$ with $k_n>0$. Where $k=k_{-1}k_0k_1\cdots k_i\cdots$ is the directive sequence of the return words.*

*Proof of the proposition:*

A shift of the infinite word $U$ can be written as $(u_0+v_0)^\omega$ with $u_0 = 10^{k_{-1}}, v_0 = 10^{k_{-1}+1}$ (resp. $u_0 = 01^{k_{-1}}, v_0 = 01^{k_{-1}+1}$). In the sequel, we focus on the case $u_0 = 10^{k_{-1}}, v_0 = 10^{k_{-1}+1}$. By definition, there exists $s$ such that $S^s(U) \in (10^{k_{-1}} + 10^{k_{-1}+1})^\omega$. But $L(0^{k_{-1}}S^s(U)) = L(S^s(U))$ and

$0^{k-1}S^s(U) \in (0^{k-1}1 + 0^{k-1}10)^\omega$. Hence, we construct $\vec{u}_0 = 0^{k-1}1$ and $\vec{v}_0 = 0^{k-1}10$. Therefore, a shift of the infinite word can be written as $(\vec{u}_0 \cup \vec{v}_0)^\omega$. Furthermore, these words are constructed by composition of Rauzy rules $S_1^{(k-1)}(0,1) = (\vec{v}_0, \vec{u}_0)$ (resp. $S_2^{(k-1)}(0,1) = (\vec{u}_0, \vec{v}_0)$. In conclusion, $\vec{u}_0$ and $\vec{v}_0$ are factors of $U$ and elements of a standard pair.

Suppose that the words $\vec{u}_0, \cdots \vec{u}_n$ and $\vec{v}_0, \cdots \vec{v}_n$ are factors of $U$ and elements of standard pairs, we would like to prove that $\vec{u}_{n+1}$ and $\vec{v}_{n+1}$ are factors of $U$ and elements of a standard pair.

A shift of the infinite word $U$ can be written as $(u_{n+1} \cup v_{n+1})^\omega$ with $u_{n+1} = u_n v_n^{k_n}$ and $v_{n+1} = u_n v_n^{k_n+1}$. There exists $s$ such that $S^s(U) \in (u_n v_n^{k_n} + u_n v_n^{k_n+1})^\omega$. By hypothesis, $\vec{u}_n$ and $\vec{v}_n$ are elements of a standard pair and there exists $s'$ such that $S^{s'}(U) \in (\vec{u}_n \vec{v}_n^{k_n} + \vec{u}_n \vec{v}_n^{k_n+1})^\omega$.

But $L(\vec{v}_n^{k_n} S^{s'}(U)) = L(S^{s'}(U))$ and $\vec{v}_n^{k_n} S^{s'}(U) \in (\vec{v}_n^{k_n}\vec{u}_n + \vec{v}_n^{k_n}\vec{u}_n\vec{v}_n)^\omega$. Hence, we construct $\vec{u}_{n+1} = \vec{v}_n^{k_n}\vec{u}_n$ and $\vec{v}_{n+1} = \vec{v}_n^{k_n}\vec{u}_n\vec{v}_n$. Therefore, a shift of the infinite word can be written as $(\vec{u}_{n+1} + \vec{v}_{n+1})^\omega$. Furthermore, these words can be constructed by composition of Rauzy rules $S_1^{(k_n)}(\vec{u}_n, \vec{v}_n) = (\vec{v}_{n+1}, \vec{u}_{n+1})$ (resp. $S_2^{(k_n)}(\vec{u}_n, \vec{v}_n) = (\vec{u}_{n+1}, \vec{v}_{n+1})$).

In conclusion, $\vec{u}_{n+1}$ and $\vec{v}_{n+1}$ are factors of $U$ and elements of a standard pair. □

The sequence $(\vec{u}_n, \vec{v}_n)_{n\in\mathbb{N}}$ is a sequence of standard pairs. By Rauzy's proposition the limit when $n$ goes to infinity of $\vec{u}_n$ and $\vec{v}_n$ is a Sturmian word denoted by $\vec{u}_\infty$.

Since $U$ is recurrent, the number of return words over $w$ is finite and $\vec{u}_n$ is a factor of $U$ for all $n$, this gives $L(U) = L(\vec{u}_\infty)$. Consequently, $U$ is a Sturmian word.

With the three last propositions, we have shown that a binary recurrent infinite word $U$ whose set of return words over $w$ has two elements for every non empty word $w$ is a Sturmian word. The converse is given by Proposition 4.1. In conclusion, we have a proof of the main theorem.

# 7 Acknowledgments

The author is greatly indebted to J. Berstel for suggesting the problem and for discussions. The author also wishes to express his thanks to J.-P. Allouche and J. Cassaigne for many useful comments.

## References


[1] P. Alessandri and V. Berthé, *Three distance theorems and combinatorics on words,* l'Enseignement Mathématique **44** (1998), 103–132.

[2] J.-P. Allouche, *Sur la complexité des suites infinies,* Bull. Belg. Math. Soc. **1** (1994), 133–143.

[3] P. Arnoux et G. Rauzy, *Représentation géométrique de suites de complexité 2n+1,* Bull. Soc. Math. France. **119** (1991), 199–215.

[4] M.-P. Béal and D. Perrin, *Symbolic Dynamics and Finite Automata,* In A. Salomaa and G. Rosenberg (Eds.), Handbook of Formal Languages, Vol. 2, pp. 463–503 (1997), Springer Verlag.

[5] J. Berstel and A. De Luca , *Sturmian words, Lyndon words and trees,* Theoret. Comput. Sci. **178** (1997), 171–203.

[6] J. Berstel and P. Séébold , *Sturmian words,* In M. Lothaire, Algebraic combinatorics on Words. (1999). To appear.

[7] V. Berthé, *Fréquences des facteurs des suites sturmiennes*, Theoret. Comput. Sci. **165** (1996), 295–309.

[8] V. Berthé and L. Vuillon, *A two-dimensional generalization of Sturmian sequences: tilings and rotations*, prétirage 97–19, IML.

[9] D. Crisp, W. Moran, A. Pollington and P. Shiue, *Substitution invariant cutting sequences,* J. Théorie des Nombres de Bordeaux, **5** (1993), 123–138.

[10] F. Durand, *A characterization of substitutive sequences using return words,* Discrete Math. **179** (1998), 89–101.

[11] F. Durand, *Contributions à l'étude des suites et systèmes dynamiques substitutifs,* Ph.D. Thesis, Université de la Méditerranée (Aix-Marseille II) 1996.

[12] G. A. Hedlundand M. Morse *Symbolic dynamics II: Sturmian trajectories*, Amer. J. Math. **62** (1940), 1–42.

[13] C. Holton, L. Q. Zamboni *Geometric realizations of substitutions*, Bull. Soc. Math. France **126** (1998), no. 2, 149–179.

[14] R. Kenyon, *The construction of self-similar tilings,* Geometric and Funct. Anal. **6** (1996), 471–488.

[15] D. Lind and B. Marcus, *An introduction to Symbolic Dynamics and Coding,* 1996, Cambridge University Press.

[16] M. Lothaire, *Combinatorics on Words,* Encyclopedia of Mathematics and its applications, Vol. 17, Addison-Wesley, Reading, MA (1983).

[17] P. Mignosi et P. Séébold, *Morphismes sturmiens et règles de Rauzy,* J. Théorie des Nombres de Bordeaux **5** (1993), 221–233.

[18] K. Petersen, *Ergodic Theory,* 1995, Cambridge University Press.

[19] B. Pragatis, *Markov partitions for hyperbolic toral automorphisms,* Ph. D. Thesis, University of Washington, 1992.

[20] C. Radin, *Space tilings and substitutions,* Geom. Dedicata, **55** (1995), 257–264.

[21] G. Rauzy, *Mots infinis en arithmétique,* in M. Nivat and D. Perrin (Eds.), Automata on Infinite Words, No. 192 in Lect. Notes Comp. Sci., pp. 165–171 (1985), Springer Verlag.

[22] M. Senechal, *Quasicrystals and geometry*, 1995, Cambridge University Press.

[23] B. Solomyak, *Dynamics of self-similar tilings,* Ergod. Theory Dynam. Syst. **17** (1997), 695–738.

[24] L. Vuillon *Combinatoire des motifs d'une suite sturmienne bidimensionnelle,* Theoret. Comput. Sci. **209** (1998), 261–285.