Sturmian words: structure, combinatorics, and their arithmetics

Aldo de Luca*

Dipartimento di Matematica, Università di Roma "La Sapienza", Piazzale A. Moro 2, 00185 Roma, and Istituto di Cibernetica del C.N.R., Arco Felice, Italy

Abstract

We prove some new results concerning the structure, the combinatorics and the arithmetics of the set \( \text{PER} \) of all the words \( w \) having two periods \( p \) and \( q \), \( p < q \), which are coprimes and such that \( |w| = pq - 2 \). A basic theorem relating \( \text{PER} \) with the set of finite standard Sturmian words was proved in de Luca and Mignosi (1994). The main result of this paper is the following simple inductive definition of \( \text{PER} \): the empty word belongs to \( \text{PER} \). If \( w \) is an already constructed word of \( \text{PER} \), then also \( (aw)^{(-)} \) and \( (bw)^{(-)} \) belong to \( \text{PER} \), where \( (-) \) denotes the operator of palindrome left-closure, i.e. it associates to each word \( u \) the smallest palindrome word \( u^{(-)} \) having \( u \) as a suffix. We show that, by this result, one can construct in a simple way all finite and infinite standard Sturmian words. We prove also that, up to the automorphism which interchanges the letter \( a \) with the letter \( b \), any element of \( \text{PER} \) can be codified by the irreducible fraction \( p/q \). This allows us to construct for any \( n \geq 0 \) a natural bijection, that we call Farey correspondence, of the set of the Farey series of order \( n + 1 \) and the set of special elements of length \( n \) of the set \( St \) of all finite Sturmian words. Finally, we introduce the concepts of Farey tree and Farey monoid. This latter is obtained by defining a suitable product operation on the developments in continued fractions of the set of all irreducible fractions \( p/q \).

Keywords: Sturmian words; Palindromes; Standard words

1. Introduction

Sturmian words are binary infinite words which are not ultimately periodic and of minimal subword complexity. These words have a long history (cf. [18]), so that they have different names. The term Sturmian is due to Hedlund and Morse [10], whereas the term characteristic was used by Christoffel [4]. Sturmian words have...
several applications in various fields such as Physics, Algebra and Computer Science. For this reason there exists a large literature on this subject (cf. [3]).

The most famous Sturmian word is the Fibonacci word \( f \) which is the limit of the sequence of words \( \{ f_n \}_{n \geq 0} \), inductively defined as: \( f_0 = b \), \( f_1 = a \), \( f_{n+1} = f_n f_{n-1} \), for all \( n > 0 \). The words \( f_n \) of this sequence are called the finite Fibonacci words. The name Fibonacci is due to the fact that for each \( n \), \( |f_n| \) is equal to the \( (n+1) \)th term of the Fibonacci numerical sequence: 1, 1, 2, 3, 5, 8, \ldots . There exist several different but equivalent, definitions of Sturmian words. Some are of ‘combinatorial’ nature and others of ‘geometrical nature’. For instance, a Sturmian word can be defined by considering the sequence of the intersections with a squared-lattice of a semi-line having a slope which is an irrational number. A horizontal intersection is denoted by the letter \( b \), a vertical intersection by \( a \) and an intersection with a corner by \( ab \) or \( ba \). From this point of view the Fibonacci word is obtained by considering a semi-line starting from the origin and having a slope equal to \( g - 1 \), where \( g = \frac{1}{2} (1 + \sqrt{5}) \) is the golden number. Sturmian words represented by a semi-line starting from the origin are usually called standard. They are of great interest from the language point of view since one can prove that the set of all finite subwords of a Sturmian word depends only on the slope of the corresponding semi-line. A finite subword of any Sturmian word is called finite Sturmian word. We shall denote by \( St \) the set of all finite Sturmian words.

Standard Sturmian words can be defined in the following way which is a natural generalization of the definition of the Fibonacci word. Let \( q_0, q_1, \ldots, q_n, \ldots \) be any sequence of natural integers such that \( q_0 > 0 \) and \( q_i > 0 \) \((i = 1, \ldots, n)\). We define, inductively, the sequence of words \( \{ s_n \}_{n \geq 0} \), where \( s_0 = b \), \( s_1 = a \), \( s_{n+1} = s_n^{q_n-1} s_{n-1} \), \( n > 1 \). The sequence \( \{ s_n \}_{n \geq 0} \) has a limit \( s \) which is a standard Sturmian word. Any standard Sturmian word is obtained in this way. The set of all the words \( s_n, n \geq 0 \) of any standard sequence \( \{ s_n \}_{n \geq 0} \) constitutes a language \( Stand \) which has remarkable and surprising properties.

In a previous paper [7] we proved a basic theorem (cf. Theorem 2) which gives three different characterizations of \( Stand \). The first, which generalizes a property of Fibonacci words, is based on palindrome words. More precisely, we proved that \( s \in Stand \) if and only if \( s \in \{ a, b \} \) or \( s = AB = Cxy \), where \( A, B, C \) are palindromes and \( x, y \in \{ a, b \} \), \( x \neq y \). The second is based on the periodicities of words. Let \( PER \) be the set of all words \( w \) having two periods \( p \) and \( q \) which are coprimes and such that \( |w| = p + q - 2 \). We proved that \( Stand = \{ a, b \} \cup PER \{ ab, ba \} \). Finally, the third characterization is of a ‘syntactical’ nature. A word \( s \) belongs to \( PER \) if and only if \( asa, ash, bsa, bsb \in St. \) A word \( w \in St \) with this property is called also a strictly bispecial element.

This theorem has several applications. In particular, one can determine the subword complexity of \( Stand \) and derive in a simple and purely combinatorial way, the subword complexity formula for \( St \) (cf. [7]).

The above results show that the ‘kernel’ of the standard Sturmian words is the set \( PER \). In this paper we present some new results concerning the structure, the combinatorics and the arithmetics of \( PER \). These results can be extended to \( Stand \). Moreover, they are relevant for all finite Sturmian words since \( St \) is equal to the set of all subwords of \( PER \). In Section 5 we prove that a word \( w \) belongs to \( PER \) if and only if either
is a power of a single letter or can be, uniquely, represented as \( w = PxyQ = QyxP \), with \( P, Q \) palindromes, \( |P| < |Q| \) and \( x, y \in \{a, b\}, \ x \neq y \). Moreover, \( p - |P| + 2 \) and \( q = |Q| + 2 \) are periods of \( w \) such that \( \gcd(p, q) = 1 \), \( |w| = p + q - 2 \), \( p \) is the minimal period of \( w \) and \( Q \) the maximal proper palindrome suffix of \( w \). From this we are able to obtain a new characterization of the set \( \text{PER} \) which allows us to construct it in a very simple way. The construction makes use of the operator \((-)\) of palindrome left-closure which associates to any word \( w \) the word \( w(-) \) defined as the smallest palindrome word having \( w \) as a suffix. The set \( \text{PER} \) has the following closure property: if \( w \in \text{PER} \), then the words \( (aw)(-) \) and \( (bw)(-) \) belong to \( \text{PER} \). Moreover, \( \text{PER} \) is the smallest subset of \( \mathcal{A}^* \), \( \mathcal{A} = \{a, b\} \), containing the empty word \( \varepsilon \) and having the above closure property. If we define recursively, the sequence of sets \( \{X_n\}_{n \geq 0} \), where \( X_0 = \{\varepsilon\} \) and \( X_{n+1} = (\mathcal{A}X_n)(-) \), \( n \geq 0 \), then \( \text{PER} = \bigcup_{n \geq 0} X_n \). For each \( n > 0 \), \( X_n \) is a biprefix code having \( 2^n \) elements.

Let \( w = a^{h_1}b^{h_2}a^{h_3} \ldots \) be a finite or infinite word such that the exponents \( h_i, i \geq 0 \) are natural integers and \( h_i > 0 \) for \( i > 1 \). One can associate with \( w \) a finite or infinite sequence \( \{s_n\}_{n \geq 0} \) of elements of \( \text{PER} \) having \( s_0 = \varepsilon \) and for each \( n \geq 0 \), \( s_{n+1} = (w_n s_n)(-) \), where \( w_n \) is the \( n \)th letter of \( w \). We prove that if the sequence \( \{s_n\}_{n \geq 0} \) is infinite, then it converges to a standard Sturmian word. Moreover, the above correspondence is a bijection.

In Section 6 we are concerned with some results of a more arithmetical nature. The starting point is the existence of a natural bijection, up to the automorphism of \( \mathcal{A}^* \) which interchanges the letter \( a \) with the letter \( b \), of \( \text{PER} \) and the set \( \mathcal{F} \) of all irreducible fractions \( p/q \), with \( p \leq q \). The correspondence is obtained by associating with each word \( w \in \text{PER} \) the fraction \( \|w\| = p/q \), where \( p \) is the minimal period of \( w \) and \( q \) the period such that \( |w| = p + q - 2 \) (one sets also \( \|\varepsilon\| = 1/1 \)). For any \( n > 0 \), let \( A_n \) be the set of all the elements \( w \in \text{PER} \) such that \( \|w\| = p/q \) with \( q \leq n + 1 \) and \( p + q - 2 \geq n \). We prove that \( A_n \) is a biprefix code. Moreover, the set of the suffixes of \( A_n \) of length \( n \) coincides with the set \( S_R(n) \) of right special elements of \( St \) of length \( n \). An element \( w \in St \) is right-special if \( wa, wb \in St \). Moreover, \( A_n \) coincides with the left-palindrome closure of \( S_R(n) \).

Let \( \mathcal{F}_n \) be the set \( \mathcal{F}_n = \{p/q \in \mathcal{F} | q \leq n\} \). If the elements of \( \mathcal{F}_n \) are ordered in an increasing way, one obtains the so-called Farey series of order \( n \). By a cardinality argument one knows that for any \( n > 0 \) there exists a bijection of \( S_R(n) \) and \( \mathcal{F}_{n+1} \). By using the previous and further results we are able to construct, for any \( n \), a very natural bijection of \( S_R(n) \) and \( \mathcal{F}_{n+1} \), which we call the Farey correspondence.

In the last section we introduce the concepts of Farey tree and Farey monoid. The first is the usual binary tree representing all binary words beginning with the letter \( a \). To each vertex representing a word \( w \) one can associate the corresponding Farey number \( \|\psi(w)\| = p/q \). The ‘sons’ of \( p/q \) are the fractions \( p/(p+q) \) and \( q/(p+q) \). Some interesting properties of this tree are shown. The Farey monoid is obtained by considering a natural product operation on the developments in continued fractions of the Farey numbers. We prove that there exists an isomorphism of \( a\mathcal{A}^* \cup \{\varepsilon\} \) and \( \mathcal{F} \).
The main results of this paper without complete proofs have been communicated at the Conference "Semigroups, Automata and Languages" held in Porto in June 1994 (cf. [6]).

2. Preliminaries

Let $A$ be a finite non-empty set, or alphabet and $A^*$ the free monoid generated by $A$. The elements of $A$ are called letters, those of $A^*$ words. The identity of $A^*$ is named empty word and denoted by $\varepsilon$.

For any word $w \in A^*$, $|w|$ denotes its length, i.e. the number of letters occurring in $w$. The length of $\varepsilon$ is taken to be equal to 0. For any letter $a \in A$, $|w|_a$ will denote the number of occurrences of the letter $a$ in $w$. One has, of course, that $|w| = \sum_{a \in A} |w|_a$.

For any $w \in A^*$, $\text{alph}(w)$ is the subset of $A$ which is minimal, with respect to the inclusion, and such that $w \in (\text{alph}(w))^*$. The mirror image ($\overline{}$) is the unary operation in $A^*$ recursively defined as $\overline{\varepsilon} = \varepsilon$ and $\overline{(ua)} = a\overline{u}$, for all $u \in A^*$ and $a \in A$. The mirror image is involutory and such that for all $u, v \in A^*$, $\overline{(uv)} = \overline{v}\overline{u}$, i.e. it is an involutory antiautomorphism of $A^*$. For any $L$ subset of $A^*$ we set $\overline{L} = \{ \overline{w} | w \in L \}$.

A word $w$ which coincides with its mirror image is called palindrome. The set of all palindromes over $A$ is denoted by $\text{PAL}(A)$, or simply, by $\text{PAL}$.

When $A = \{a, b\}$ we denote by ($\ast$) the involutory automorphism of $A^*$ defined as: $a^* = b$, $b^* = a$. Thus $\varepsilon = \varepsilon$ and for any $w \in A^*$, $w \neq \varepsilon$, $\hat{w}$ is obtained from $w$ by interchanging the letter $a$ with $b$. For a subset $L$ of $A^*$ we set $\hat{L} = \{ \hat{w} | w \in L \}$.

A word $w$ which coincides with its mirror image is called palindrome. The set of all palindromes over $A$ is denoted by $\text{PAL}(A)$, or simply, by $\text{PAL}$.

A word $w = w_1 \ldots w_n$, $w_i \in A$, $1 \leq i \leq n$, has a period $p$ if the following condition is satisfied:

$$\text{If } i \in [1, n - p], \text{ then } w_i = w_{i + p}. $$

We denote by $\Pi(w)$ the set of all periods of $w$. From the definition one has that any integer $p \geq |w|$ is a period of $w$. We recall the following important theorem due to Fine and Wilf [9]: If $p, q \in \Pi(w)$ and $|w| \geq p + q - \gcd(p, q)$, then $\gcd(p, q) \in \Pi(w)$. Moreover, one can prove (cf. [11]) that the lower bound $p + q - \gcd(p, q)$ to the length of $w$ in order that $w$ admits the period $\gcd(p, q)$ is optimal.

A word $u$ is a factor, or subword, of $w$ if $w \in A^* u A^*$, i.e. there exist $x, y \in A^*$ such that $w = xy$. The factor $u$ is called proper if $u \neq w$. If $x = \varepsilon$ ($y = \varepsilon$), then $u$ is called a prefix (suffix) of $w$. By $F(w)$ we denote the set of all factors of $w$. A subset $L \subseteq A^*$ is called language. For any language $L$ the set $F(L)$ of its factors is defined as $F(L) = \bigcup_{w \in L} F(w)$. A language $L$ is called factorial if it is closed by factors, i.e. $L = F(L)$. For any language $L$ the enumeration function, or subword complexity, $g_L$ of $L$ is the map $g_L : N \rightarrow N$ defined as: for all $n \geq 0$, $g_L(n) = \text{Card}(L \cap A^n)$. If $X, Y$ are languages we denote by $X^{-1}Y$ and $YX^{-1}$ the subsets of $A^*$

$$X^{-1}Y = \{ w \in A^* | Xw \cap Y \neq \emptyset \}, \quad YX^{-1} = \{ w \in A^* | wX \cap Y \neq \emptyset \}. $$
When $X$ is a singleton, i.e. $X = \{v\}$, the sets $\{v\}^{-1}Y$ and $Y\{v\}^{-1}$ will be simply denoted by $v^{-1}Y$ and $Yv^{-1}$.

An infinite word (from left to right) $x$ over $\mathcal{A}$ is any map $x: N \rightarrow \mathcal{A}$. For any $i \geq 0$, we set $x_i = x(i)$ and write:

$$x = x_0x_1 \ldots x_n \ldots .$$

The set of all infinite words over $\mathcal{A}$ is denoted by $\mathcal{A}^\omega$. A word $u \in \mathcal{A}^*$ is a (finite) factor of $x \in \mathcal{A}^\omega$ if $u = \varepsilon$ or there exist integers $i,j$ such that $i \leq j$ and $u = x_i \ldots x_j$. Any pair $(i,j)$ such that the preceding equality is satisfied is called an occurrence of $u$ in $x$. If a factor $u$ of $x$ is empty or has the occurrence $(0,|u| - 1)$, then $u$ is called a prefix of $x$. We denote by $F(x)$ the set of all (finite) factors of $x$ and by $Pref(x)$ the set of prefixes of $x$. An infinite word $x \in \mathcal{A}^\omega$ is called recurrent if for any $u \in F(x)$ there is an infinite number of occurrences of $u$ in $x$. The enumeration function of the language $F(x)$ is also called the enumeration function of $x$ and simply denoted by $g_x$.

As we said in the introduction, infinite Sturmian words are infinite words over the alphabet $\{a,b\}$ which can be defined in several different and equivalent ways. We shall give the following definition:

**Definition 1.** An infinite word $x \in \mathcal{A}^\omega$ is Sturmian if and only if the enumeration function $g_x$ satisfies the following condition: for all $n \geq 0$

$$g_x(n) = n + 1.$$

Let us now define the set $St$ of finite Sturmian words:

**Definition 2.** A word $w \in St$ if and only if there exists an infinite Sturmian word $x$ such that $w \in F(x)$.

By definition $St$ is a factorial language on the alphabet $\{a,b\}$. The following interesting and useful combinatorial characterization of the language $St$ was given by Dulucq and Gouyou-Beauchamps [8]:

**Theorem 1.** The language $St$ is the set of all the words $w \in \{a,b\}^*$ such that for any pair $(u,v)$ of factors of $w$ having the same length one has:

$$||u||_a - ||v||_a|| \leq 1.$$

3. **Standard Sturmian words**

There exist several methods to construct infinite Sturmian words. We shall refer here to the following procedure that we call standard method: Let $(q_0,q_1,q_2,\ldots)$ be an infinite sequence of integers such that $q_0 \geq 0$ and $q_i > 0$ for all $i > 0$. 

We define the sequence \( \{s_n\}_{n \geq 0} \), where \( s_0 = b, \ s_1 = a \) and for all \( n \geq 1 \),
\[
s_{n+1} = s_n^{q_{n-1}} s_{n-1}.
\]
One easily verifies that the sequence \( \{s_n\}_{n \geq 0} \) converges to an infinite sequence \( x \). We call \( \{s_n\}_{n \geq 0} \) the approximation sequence of \( x \) and \( (q_0, q_1, q_2, \ldots) \) the directive sequence of \( x \). It has been proved that \( x \) is an infinite Sturmian word whose representative semi-line starts from the origin. Conversely, any infinite Sturmian word whose representative semi-line starts from the origin can be obtained by the preceding standard method. Moreover, one can also prove that if \( q_0 > 0 \) then \([0, q_0, q_1, q_2, \ldots]\) represents the development in continued fractions of the slope associated with the infinite Sturmian word \( x \).

Let us remark that if \((q_0, q_1, q_2, \ldots)\) is the directive sequence of \( x \) and \( q_0 > 0 \) then, as one easily verifies, \((0, q_0, q_1, q_2, \ldots)\) is the directive sequence of the Sturmian word \( \hat{x} \) which is obtained from \( x \) by interchanging the letter \( a \) with the letter \( b \). An infinite Sturmian word constructed by the standard method will be also called infinite standard Sturmian word. We denote by \( \textbf{Stand} \) the set of all infinite standard Sturmian words.

Let \( x \) be an infinite standard Sturmian word whose approximating sequence is \( \{s_n\}_{n \geq 0} \). For each \( n \geq 0 \) let us set \( \chi(n) = |s_n| \). One has then for \( n > 0 \):
\[
\chi(n + 1) = q_{n-1} \chi(n) + \chi(n - 1).
\]
One easily verifies, by induction, that for all \( n \geq 0 \), \( \gcd(\chi(n), \chi(n + 1)) = 1 \). If the directive sequence of \( x \) is \((1, 1, \ldots, 1, \ldots)\), then \( x \) is the infinite Fibonacci word \( f \), \( \{f_n\}_{n \geq 0} \) is the sequence of finite Fibonacci words and \( \chi(n) \) is the \((n + 1)\)th term of the Fibonacci numerical sequence.

**Definition 3.** A word \( s \in St \) is called standard if there exists an infinite standard Sturmian word \( x \) and an integer \( n \geq 0 \) such that \( s = s_n \) where \( \{s_n\}_{n \geq 0} \) is the approximating sequence of \( x \).

We shall denote by \( \textbf{Stand} \) the set of all finite standard Sturmian words. We say that a standard word \( s \in \textbf{Stand} \) has the directive sequence \((q_0, q_1, \ldots, q_n)\), with \( q_0 \geq 0 \), \( q_i > 0 \), \( 1 \leq i \leq n \), if there exists a sequence of standard words \( s_0, s_1, \ldots, s_n, s_{n+1}, s_{n+2} \) such that
\[
s_0 = b, \quad s_1 = a, \quad s_{i+1} = s_i^{q_i-1} s_{i-1}, \quad 1 \leq i \leq n + 1
\]
and \( s = s_{n+2} \). One can prove that any standard word has a unique directive sequence. This will be proved in Section 5 (cf. Corollary 4) as a consequence of Propositions 8 and 10.

Since the set of factors of an infinite Sturmian word \( y \) depends only on the slope associated to \( y \) and does not depend on its starting point (cf. [12]), then a word \( s \in St \) if and only if there exists an infinite standard Sturmian word \( x \) such that \( s \in F(x) \). Thus it follows that \( St = F(\textbf{Stand}) = F(\textbf{Stand}) \).
Let us remark that infinite, as well as finite, standard Sturmian words can be defined and constructed by other different methods (cf. [18, 16]). In [7] we gave three different characterizations of the set \( \text{Stand.} \) The first is based on the following property, which is expressed by palindrome words, that we called Robinson's property (cf. [17, 7]). This property was considered by the author in [5] in the case of Fibonacci words.

**Definition 4.** A word \( W \in \{a, b\}^* \) has the Robinson property if \( |W| = 1 \) or when \( |W| \geq 2 \) then

\[
W = AB = Cxy,
\]

with \( x, y \in \{a, b\}, x \neq y \) and \( A, B, C \) palindrome words.

We remark that Pedersen et al. [14] proved that there exists, and is unique, a word \( W \) such that

\[
W = AB = Cxy,
\]

if and only if \( \gcd(|A| + 2, |B| - 2) = 1. \)

Let us denote by \( \Sigma \) the set of all the words on the alphabet \( \mathcal{A} \) having Robinson's property. One has that

\[
\Sigma = \mathcal{A} \cup (\text{PAL}^2 \cap \text{PAL}\{ab, ba\}).
\]

In [7] we proved the following noteworthy result:

**Proposition 1.** The set of finite standard Sturmian words coincides with the set of all words having the Robinson property, i.e.

\[
\Sigma = \text{Stand.}
\]

A remarkable application of Proposition 1 to the study of Sturmian words generated by iterated morphisms was recently given by Berstel and Séébold in [2].

A second characterization of finite standard Sturmian words is based on periodicities of words. Let \( w \in \mathcal{A}^* \) and \( \Pi(w) \) be the set of its periods. We define the set \( \text{PER} \) of all words \( w \) having two periods \( p \) and \( q \) which are coprimes and such that \( |w| = p + q - 2. \)

Thus a word \( w \) belongs to \( \text{PER} \) if it is a power of a single letter or is a word of maximal length for which the theorem of Fine and Wilf does not apply. In the sequel we assume that \( v \in \text{PER} \). This is, formally, coherent with the above definition if one takes \( p = q = 1. \) In [7] we proved the following remarkable result:

**Proposition 2.** \( \text{Stand} = \mathcal{A} \cup \text{PER}\{ab, ba\}. \)

The third and last characterization is based on an analysis of some combinatorial properties concerning special, bispecial and strictly bispecial elements of \( St. \)
We recall that any infinite Sturmian word $x$ is recurrent so that for any $s \in F(x)$ there exists always at least one letter $z \in \{a, b\}$ such that $zs \in F(x)$. Let us give the following definitions:

**Definition 5.** A word $s \in St$ is right (left) special if $sa, sb \in St$ ($as, bs \in St$).

**Definition 6.** A word $s \in St$ is bispecial if it is right and left special.

**Definition 7.** A word $s \in St$ is strictly bispecial if $asa, asb, bsa, bsb$ are in $St$.

**Definition 8.** Let $x$ be an infinite Sturmian word. A factor $s$ of $x$ is right special (left special) in $x$ if $sa, sb \in F(x)$ ($as, bs \in F(x)$).

We recall the following proposition (cf. [12]).

**Proposition 3.** If $s \in F(x)$, where $x$ is an infinite Sturmian word, then $s \in F(x)$. Moreover, if $x$ is an infinite standard Sturmian word, then $s$ is right special in $x$ if and only if $s = \bar{p}$, where $p$ is a prefix of $x$.

We shall denote by $SR, SL, BS$ and $SBS$ the sets of right special, left special, bispecial and strictly bispecial elements of $St$, respectively.

The following (cf. [7]) holds:

**Proposition 4.** $PER = SBS$.

We can summarize the previous results in the following basic theorem:

**Theorem 2.**
1. $\text{Stand} = \Sigma = \mathcal{A} \cup PER\{ab, ba\},$
2. $PER = SBS,$
3. $St = F(\Sigma) = F(PER).$

In [7] we proved, as consequences of Theorem 2, the following results concerning the enumeration functions of the previous sets. Let us denote by $s_R, s_L$ and $sbs$ the enumeration functions of the sets $SR, SL$ and $SBS$. If $g_{St}$ and $g_{Stand}$ are the enumeration functions of finite Sturmian and finite standard Sturmian words, one has that the following relations hold for each $n > 0$:

$$
g_{St}(n+1) = g_{St}(n) + s_R(n),$$
$$s_R(n+1) = s_R(n) + sbs(n),$$
$$sbs(n) = \left(1/2\right)g_{Stand}(n+2),$$
$$g_{Stand}(n) = 2\phi(n),$$

where $\phi$ is the totient Euler's function. From the above relations one easily derives (cf. [13, 7]) that

$$s_R(n) = \sum_{i=1}^{n+1} \phi(i),$$
and, moreover
\[ g_{S_L}(n) = 1 + \sum_{i=1}^{n} \phi(i)(n-i+1). \]

4. Combinatorial properties of special elements

We shall give now some lemmas concerning the structure of the sets \( S_R, S_L, BS \) and \( SBS \).

**Lemma 1.** \( S_R = \hat{S}_L, S_R = \hat{S}_R, S_L = \hat{S}_L \).

**Proof.** Let us first prove that \( S_R = \hat{S}_L \). One has:
\[ s \in S_R \iff sa, sb \in St \iff a\tilde{s}, b\tilde{s} \in St \iff \tilde{s} \in S_L. \]

Let us now prove that \( S_R = \hat{S}_R \). One has:
\[ s \in S_R \iff sa, sb \in St \iff \tilde{s}b, \tilde{s}a \in St \iff \tilde{s} \in S_R. \]

In a similar way one proves that \( S_L = \hat{S}_L \). \( \Box \)

From this lemma it follows that \( s_R = s_L \); moreover, the set \( BS \) is invariant under the operators \( (\cdot) \) and \( (\cdot) \).

**Lemma 2.** \( S_R \cap PAL = S_L \cap PAL = SBS \).

**Proof.** Let us prove that \( S_R \cap PAL = SBS \). The inclusion \( SBS \subset S_R \cap PAL \) is trivial since \( SBS \subset S_R \) and from Theorem 2, \( SBS \subset PAL \). In order to prove the inverse inclusion we have to show that a palindrome right-special element of \( St \) is strictly bispecial. Let \( s \in S_R \cap PAL \). One has then \( sa, sb \in St \) and \( s = \tilde{s} \). From Proposition 3 it follows that \( as, bs \in St \). We shall prove now, by Theorem 1, that for \( x, y \in \{a, b\} \) the word \( xsy \) belongs to \( St \). Let \( f, f' \) be two non-overlapping factors of \( xsy \) having the same length. We want to prove that \( |f|_x - |f'|_x | \leq 1 \). If \( f, f' \in F(xs) \) or \( f, f' \in F(sy) \), then the result is obvious since \( xs, sy \in St \). Let us then suppose that \( f = xu \) and \( f' = vy \) with \( |u| = |v| \). Since \( s \) is palindrome one has \( v = \tilde{u} \). Hence
\[ ||f||_x - |f'|_x| = ||xu||_x - |vy||_x | = |1 - |y||_x|. \]

Thus the previous difference is equal to 0 if \( x = y \) and equal to 1 otherwise. This shows that \( xsy \in St \). The proof of the equality \( S_L \cap PAL = SBS \) is perfectly symmetric. \( \Box \)

Let us explicitly observe that a palindrome element of \( St \), in general, is not an element of \( SBS \). For instance, the palindrome word \( baab \in St \) is neither a right-special element nor a left-special element of \( St \) since \( baabb \) and \( bbaab \) do not belong to \( St \).
From Lemma 2 it follows that any palindrome suffix (prefix) of a right (left) special element is strictly bispecial. Moreover, BS ∩ PAL = SBS. Thus a bispecial element of St is not strictly bispecial unless it is a palindrome. For instance for the length n = 4 there are 10 right special elements:

\textit{aaaa, baaa, abaa, aaba, baba, abbb, abbb, babb, bbab, abab.}

The elements \textit{baaa} and \textit{abbb} are not left-special. All the others are bispecial. However the only strictly bispecial elements are \textit{aaaa} and \textit{bbbb}.

Let us now introduce the set:

\textit{SBS \cap A^*S_R.}

An element \(\sigma\) belongs to \(\textit{SBS \cap A^*S_R}\) if and only if there exist \(s \in S_R\) and \(\lambda \in A^*\) such that \(\sigma = \lambda s \in \textit{SBS}\); the word \(\sigma\) is called a left-extension of \(s\) in the set \(\textit{SBS}\). A left-extension \(\sigma\) of \(s\) in \(\textit{SBS}\) is called proper if the word \(\sigma\) has no palindrome suffixes \(\tau\) such that \(|\sigma| > |\tau| \geq |s|\). In a symmetric way one can consider the set \(\textit{SBS \cap S_LA^*}\). If \(s \in S_L\) and \(\sigma = s\lambda \in \textit{SBS}, \lambda \in A^*\), then \(\sigma\) is called a right-extension of \(s\) in \(\textit{SBS}\); \(\sigma\) is called proper if \(\sigma\) has no palindrome prefixes \(\tau\) such that \(|\sigma| > |\tau| \geq |s|\).

**Proposition 5.** Any right-special element of \(St\) has a unique proper left-extension in \(\textit{SBS}\).

**Proof.** Let \(s\) be a right special element of \(St\). If \(s \in \textit{SBS}\) the result is trivial. Let us then suppose that \(s \in S_R \setminus \textit{SBS}\). Let us first prove the 'unicity' and later the 'existence' of a proper left-extension of \(s\) in \(\textit{SBS}\).

Suppose that \(\sigma\) and \(\sigma'\) are two proper and distinct left-extensions of \(s\) in \(\textit{SBS}\). We can write:

\[ \sigma = \lambda s, \quad \sigma' = \lambda' s \]

with \(\lambda, \lambda' \in \{a, b\}^*\). Since \(\sigma\) and \(\sigma'\) are 'proper', \(\lambda\) cannot be a suffix of \(\lambda'\) and, conversely, \(\lambda'\) cannot be a suffix of \(\lambda\). Hence we can write:

\[ \lambda = \alpha xu, \quad \lambda' = \beta yu, \]

with \(\alpha, \beta, u \in \{a, b\}\) and \(x, y \in \{a, b\}, x \neq y\). One has then:

\[ \sigma = \alpha xus, \quad \sigma' = \beta yus. \]

Since \(xus, yus \in S_R\) it follows:

\[ xusx, xusy, yusx, yusy \in St. \]

i.e. \(us \in \textit{SBS}\), so that \(us \in \textit{PAL}\). If \(u = e\) this contradicts the fact that \(s\) is not a palindrome. If \(u \neq e\), then one contradicts the fact that \(\sigma\), as well as \(\sigma'\), are 'proper' left extensions of \(s\) in \(\textit{SBS}\).
Let us now prove the 'existence' of a proper left-extension of \( s \) in \( SBS \). We suppose first that \( s = xx, \) where \( x \in \{a, b\} \) and \( x \in PAL \). By Lemma 2 one has \( x \in SBS \), so that from Theorem 2, if \( y \in \{a, b\} \) and \( y \neq x \) then \( axy \in \text{Stand} \) so that there exists an infinite standard Sturmian word \( x \) and an integer \( n \geq 0 \) such that \( axy = s_n \), where \( \{s_m\}_{m \geq 0} \) is the approximating sequence of \( x \). One has that \( n \geq 2 \). If \( n = 2 \), then \( x = x|x| \) and \( s \in PAL \). Hence suppose \( n > 2 \). Since \( s_{n-1} = \beta yx \), with \( \beta \in SBS \) and by the fact that \( s_{n-1} \) is still standard, one has
\[
s_{n}s_{n-1} = \beta xy\beta yx
\]
with \( \beta xy\beta yxa = \beta ys \in PER = SBS \). Thus \( s \) admits a left-extension in \( SBS \). If we consider a left-extension of \( s \) in \( SBS \) of minimal length this has to be proper.

Let us now suppose that \( s = \lambda xx, \) where \( \lambda \in \{a, b\}^*, x \in \{a, b\} \) and \( x \) is the maximal palindrome suffix of \( s \). As we have seen above \( xx \) has a unique proper left-extension in \( SBS \). Let us denote by \( \sigma \) this extension:

\[
\sigma = \mu xx.
\]

We want to prove that \( s \) is a suffix of \( \sigma \). Let us observe that \( \mu \) cannot be a suffix of \( \lambda \). Indeed, otherwise, one will contradict the fact that \( x \) is the palindrome suffix of \( s \) of maximal length. Let us suppose, by contradiction, that \( \lambda \) is not a suffix of \( \mu \). This implies since \( \sigma \neq s \), that there exist \( u \in \{a, b\}^* \) and \( x_1, x_2 \in \{a, b\}, \) \( x_1 \neq x_2, \) such that

\[
s = \lambda'x_1uxx, \quad \sigma = \mu'x_2uxx
\]

with \( \lambda', \mu' \in \{a, b\}^* \). Since \( s, \sigma \in S_B \) one derives that \( uxx \in SBS \). If \( u = \epsilon \), then \( xx \in SBS \) so that \( xx \) is a palindrome and this contradicts the fact that \( x \) is the maximal palindrome suffix of \( s \). Let us then suppose \( u \neq \epsilon \). Since \( |uxx| > |x| \) one contradicts the fact that \( \sigma \) is the unique proper left-extension of \( xx \) in \( SBS \).

Let \( s \in S_R \), we denote by \( s^\dagger \) the unique proper left-extension of \( s \) in \( SBS \). We observe that \( s^\dagger \) coincides with the left-extension of \( s \) in \( SBS \) of minimal length. Indeed, this is an obvious consequence of the previous proposition and of the fact that a left-extension of \( s \) in \( SBS \) of minimal length has to be proper.

**Corollary 1.** A right special element of \( S_I \) is a right special factor in an infinite standard Sturmian word.

**Proof.** Let \( s \in S_R \). We consider the proper left-extension \( s^\dagger \) of \( s \) in \( SBS \). One has that:

\[
s^\dagger = \lambda s = \tilde{s}\lambda.
\]

with \( \lambda \in \{a, b\}^* \). From Theorem 2, \( s^\dagger ab \in \text{Stand} \), so that there exists an infinite standard Sturmian word \( x \) and an integer \( n \geq 2 \) such that \( \tilde{s}\lambda ab = s_n \), where \( \{s_m\}_{m \geq 0} \) is
the approximating sequence of \( x \). Since \( s_n \in \text{Pref}(x) \) then \( \tilde{s} \in \text{Pref}(x) \), so that from Proposition 3 it follows that \( s \) is right-special in \( x \). \( \square \)

Let us remark that in a perfect symmetric way one can prove that any left-special element \( s \) of \( S_t \) has a unique proper right-extension in \( SBS \). Moreover, one derives as a corollary that any left-special element of \( S_t \) is a left-special factor in an infinite standard Sturmian word.

**Proposition 6.** Let \( x \) and \( y \) be two infinite Sturmian words having the same right special factor of length \( n \). Then

\[
F(x) \cap \mathcal{A}^i = F(y) \cap \mathcal{A}^i, \quad (i = 1, \ldots, n + 1).
\]

**Proof.** Since for any infinite Sturmian word there exists always an infinite standard Sturmian word having the same set of finite factors, we can assume that \( x \) and \( y \) are standard. From the hypothesis and Proposition 3 one has that \( x \) and \( y \) have the same prefix of length \( n \) and then the same right special factors of lengths \( 1, 2, \ldots, n \). The proof of the proposition is obtained by induction on the integer \( i \).

**Base of the induction.** One has \( F(x) \cap \mathcal{A} = F(y) \cap \mathcal{A} = \{a, b\} \). By hypothesis \( x \) and \( y \) have the same prefix of length 1, say \( a \); hence \( a \) is a right special factor of \( x \) and \( y \). This implies that \( aa, ab \in F(x) \cap F(y) \). Moreover, from Proposition 3, one has \( ba = (ab) \in F(x) \cap F(y) \). Note that \( bb \notin F(x) \cap F(y) \). This completes the base of the induction.

**Induction step.** Suppose that we have proved the property up to \( i - 1, 1 < i < n \). Thus by hypothesis \( F(x) \cap \mathcal{A} = F(y) \cap \mathcal{A} = \{a, b\} \). Let \( \{f_0, f_1, \ldots, f_i\} \) the set of \( i + 1 \) factors of length \( i \) belonging to \( F(x) \cap F(y) \). Let \( f_0 \) be the right special factor of length \( i \). We have then \( f_0a, f_0b \in F(x) \cap F(y) \). Let us now suppose that there exists \( f \in \{f_1, \ldots, f_i\} \) such that \( fa \in F(x) \) and \( fb \in F(y) \). Since \( |f_0| = |f| = i \) and \( f_0 \neq f \) there will exist a word \( u \in \{a, b\}^* \) and \( x, y \in \{a, b\}, x \neq y \) such that

\[
f_0 = f'xu, \quad f = f'yu,
\]

with \( f_0', f' \in \{a, b\}^* \). Hence

\[
f_0'xua, f_0'xub \in F(x) \cap F(y),
\]

and

\[
f'yua \in F(x), \quad f'yub \in F(y)
\]

If \( x = a \), then \( y = b \) and \( ||aua| - |bub|| = 2 \) which is contradiction. If \( x = b \) one reaches a similar contradiction. Hence for every \( f \in \{f_1, \ldots, f_i\} \) there exists a unique letter \( x \) such that \( fx \in F(x) \cap F(y) \). This implies that \( F(x) \cap \mathcal{A}^{i+1} = F(y) \cap \mathcal{A}^{i+1} \). \( \square \)
5. A new characterization of Standard words

Lemma 3. A palindrome word \( w \) has the period \( p < |w| \) if and only if it has a palindrome prefix (suffix) of length \( |w| - p \).

Proof. Let \( w = w_1 \ldots w_n \), \( w_i \in \mathcal{A} \) \((i = 1, \ldots, n)\) be a palindrome word of length \( n \). One has for \( i \in [1, n] \)

\[
w_i = w_{n-i+1}.
\]

Since \( w \in \text{PAL} \) then \( w \) has always the period \( |w| - 1 \). Let \( p \) be any other period such that \( p < |w| \). If we set \( q = n - p > 0 \), then we can write the above relation as

\[
w_i = w_{(q-i+1)+p}.
\]

Now for \( i \in [1, q] \), one has \( q - i + 1 \geq 1 \), so that from the \( p \)-periodicity of \( w \) it follows:

\[
w_i = w_{q-i+1},
\]

for \( i \in [1, q] \) i.e. \( w \) has a palindrome prefix \( Q \) of length \( q \). Since \( w \) is palindrome then \( Q \) is also a suffix of \( w \).

Conversely, suppose that \( w \) is a palindrome word of length \( n \) having the palindrome prefix \( Q \) of length \( q < n \). We can write:

\[
w = Q\tilde{\lambda} = \tilde{\lambda}Q.
\]

From the lemma of Lyndon and Schützenberger (cf. [11]) one derives:

\[
\tilde{\lambda} = \alpha \beta, \quad \lambda = \beta \alpha, \quad Q = (\alpha \beta)^r \alpha, \quad r \geq 0
\]

\[
w = (\alpha \beta)^{r+1} \alpha, \quad \alpha = \tilde{\lambda}, \quad \beta = \tilde{\beta},
\]

so that \( w \) has the period \( |\alpha \beta| = |\tilde{\lambda}| = |w| - |Q| = |w| - q = p \). \( \square \)

Proposition 7. \( \text{PER} = a^* \cup b^* \cup (\text{PAL} \cap (\text{PALab}\text{PAL})) \).

Proof. Let \( w \in \text{PER} \). Thus \( w \) has two periods \( p \) and \( q \) such that \( \gcd(p, q) = 1 \) and \( |w| = p + q - 2 \). This implies (cf. [7, Theorem 4]) that \( w \) is a palindrome word which is either a power of a single letter \((a \text{ or } b)\) or \( w \) has the palindrome prefixes (and suffixes) \( P \) and \( Q \) of lengths \( |P| = p - 2 \) and \( |Q| = q - 2 \). Hence \( w \) can be written as

\[
w = PxyQ = QyxP,
\]

with \( x, y \in \{a, b\} \) and \( x \neq y \) (cf. [7]).

Conversely, if \( w \in a^* \cup b^* \) then \( w \) has the periods \( p = 1 \) and \( q = |w| + 1 \) having \( \gcd(p, q) = 1 \) and \( |w| = p + q - 2 \). If \( w = PxyQ = QyxP \), with \( P, Q \in \text{PAL} \), \( x, y \in \{a, b\} \) and \( x \neq y \), then from the previous lemma, \( w \) has the periods:

\[
p = |w| - |Q|, \quad q = |w| - |P|,
\]
so that \(|w| = p + q - 2\). Since \(w \in PAL\) and \(wyx = PxyQyx \in (PAL)^2\) it follows that \(wyx \subseteq \Sigma\), so that from the theorem of Pedersen et al. [14] one has \(gcd(p, q) = 1\). Hence, \(w \in PER\). □

**Lemma 4.** Let \(w \in PER\) be such that \(|\text{Card}(alph}(w))| > 1\). Then \(w\) satisfies the following properties:

1. \(w\) can be uniquely represented as:
\[
w = PxyQ = QyxP,
\]
with \(x, y\) fixed letters in \([a, b]\), \(x \neq y\) and \(P, Q \in PAL\). Moreover, \(gcd(p, q) = 1\), where \(p = |P| + 2\) and \(q = |Q| + 2\).

2. If \(|P| < |Q|\), then \(Q\) is the maximal proper palindrome suffix (and prefix) of \(w\).

3. \(p = |P| + 2\) is the minimal period of \(w\). Moreover, the standard word \(s = wyx\) will have still the minimal period \(p\).

4. If \(|P| + 1 < |Q|\), then there exist and are unique the integers \(k\) and \(r\) such that \(k > 0, 0 < r < p\) and
\[
Q = (Pxy)^kU,
\]
with \(|U| = r\). Hence \(w = (Pxy)^{k+1}U\).

**Proof.**

1. Let \(w = PxyQ = QyxP\), with \(x, y \in \{a, b\}, x \neq y\) and \(P, Q \in PAL\). Suppose now that there exist \(P', Q' \in PAL\) such that
\[
w = PxyQ = QyxP = P'xyQ' = Q'yxP'.
\]
This implies that:
\[
wyx = P(xyQyx) = P'(xyQ'yx).
\]
If \(P \neq P'\) then \(wyx\) can be factored in two distinct ways in the product of two palindromes. Since \(wyx\) is primitive (cf. [7]) one reaches a contradiction. Thus it follows \(P = P'\) and \(Q = Q'\). Since \(wyx \in \Sigma\) then from [14] it follows that \(gcd(p, q) = 1\).

2. Let \(w = PxyQ = QyxP\) with \(P, Q \in PAL\) and suppose that \(|P| < |Q|\). Let us prove that \(Q\) is the maximal proper palindrome suffix (and prefix) of \(w\). Indeed, suppose by contradiction, that \(Q'\) is a palindrome suffix of \(w\) such that \(|Q'| > |Q|\). From Lemma 3, \(w\) has the period \(p' = |w| - |Q'| < p\). Since \(p' > p' + 1\) it follows:
\[
|w| = p + q - 2 \geq p' + q - 1 \geq p' + q - d,
\]
where \(d = gcd(p', q)\), so that \(w\) has the period \(d\) in view of the theorem of Fine and Wilf. Moreover, since \(q > p\), one has:
\[
|w| = p + q - 2 \geq p + p'.
\]
This implies that \(w\) has also the period \(d' = gcd(p, p')\). Since \(p \geq d'\) and \(p' \geq d\) one derives:
\[
|w| \geq d + d'.
\]
so that \( w \) has the period \( \delta = \gcd(d, d') \). Since \( \gcd(p, q) = 1 \) it follows \( \delta = 1 \). This implies \( \text{Card}(\alpha(w)) = 1 \), which is a contradiction.

3. Let us prove that \( p = |P| + 2 \) is the minimal period of \( w \). Indeed, if \( w \) has a period \( p' < p \), then by Lemma 3, \( w \) will have a palindrome suffix \( Q' \) of length \( |Q'| = |w| - p' > |Q| \). Let us consider now the standard word \( s = wyx = PxyQyx = QyxPyx \). As one easily verifies the word \( s \) saves the period \( p \). This is also the minimal period of \( s \). In fact if \( s \) has a period \( p' < p \), then also \( w \) will have the period \( p' < p \) which is a contradiction.

4. By hypothesis \( |Q| \geq |P| + 2 = p \). This implies that there exist and are unique the integers \( k > 0 \) and \( r \) such that \( 0 \leq r < p \) and \( |Q| = kp + r \). Since \( w = PxyQ = QyxP \) has the period \( p \) then also \( Q \) will have the period \( p \). Moreover, since \( Pxy \) is a prefix of \( Q \) we can write:

\[
Q = (Pxy)^k U,
\]
with \( |U| = r \). Since \( w = PxyQ \) one derives \( w = (Pxy)^{k+1} U \). \( \square \)

From the above lemma one has that if \( w \in \text{PER} \) and \( \text{Card}(\alpha(w)) > 1 \), then \( w \) can be uniquely represented as:

\[
w = PxyQ,
\]
with \( x, y \in \{a, b\} \), \( x \neq y \), \( P, Q \in \text{PAL} \), \( PxyQ = QyxP \) and \( |P| < |Q| \). We call this representation the canonical representation of \( w \). The word \( xy = (P^{-1}w)Q^{-1} \) is uniquely determined and will be called the intermediate word of \( w \).

For any \( w \in \mathcal{A}^* \) we introduce the set

\[
L_w = \mathcal{A}^* w \cap \text{PAL}.
\]
Any element of \( L_w \) will be called a palindrome left-extension of \( w \).

**Lemma 5.** Let \( w \in \mathcal{A}^* \). There exists in \( L_w \) a unique element \( w^{(-)} \) of minimal length. Moreover, if \( w = Q\delta \), \( \delta \in \mathcal{A}^* \), where \( Q \) is the maximal palindrome prefix of \( w \), then \( w^{(-)} = \delta Q\delta \).

**Proof.** Let \( k \) be the minimal length of the elements of \( L_w \). Suppose now that there exist \( \lambda_1, \lambda_2 \in \mathcal{A}^* \) such that \( \lambda_1 w, \lambda_2 w \in \text{PAL} \) and \( |\lambda_1 w| = |\lambda_2 w| = k \). This implies \( |\lambda_1| = |\lambda_2| = s \), with \( 0 \leq s \leq |w| - 1 \). Moreover, \( \lambda_1 = \lambda_2 = \bar{u} \), where \( u \) is the suffix of \( w \) of length \( s \). Hence there exists in \( L_w \) a unique element \( w^{(-)} \) of minimal length. Let us now write \( w \) as \( w = Q\delta \), where \( Q \) is the maximal palindrome prefix of \( w \). One has then

\[
w^{(-)} = \lambda w = \lambda Q\delta,
\]
with \( |\lambda| \leq |\delta| \). Since \( w^{(-)} \in \text{PAL} \) then \( \delta = \delta' \tilde{\lambda} \), with \( \delta' \in \mathcal{A}^* \). This implies

\[
w^{(-)} = \lambda Q\delta' \tilde{\lambda}.
\]
Since \( w^{(-)} \in PAL \) then \( Q\delta' \in PAL \), so that \( Q\delta' \) is a palindrome prefix of \( w \). If \( \delta' \neq \varepsilon \) we reach a contradiction since \( |Q\delta'| > |Q| \). Hence \( \delta' = \varepsilon \) and \( \lambda = \delta' \). □

It follows from the above lemma that one can introduce the map \((-) : \mathcal{A}^* \rightarrow PAL\) which associates to any word \( w \in \mathcal{A}^* \) the palindrome word \( w^{(-)} \). We call \( w^{(-)} \) the palindrome left-closure of \( w \) and \((-) \) the operator of palindrome left-closure.

Let us remark that one can introduce in a perfect symmetric way an operator \((+) : \mathcal{A}^* \rightarrow PAL\) of palindrome right-closure which associates to any word \( w \in \mathcal{A}^* \) the word \( w^{(+)} \) defined as the (unique) word of minimal length in the set \( R_w = w \mathcal{A}^* \cap PAL \). One easily verifies that if \( w = \delta Q \), where \( Q \) is the palindrome suffix of \( w \) of maximal length, then \( w^{(+)} = \delta Q \delta \). Moreover, one has that for any \( w \in \mathcal{A}^* \).

\[ w^{(-)} = (w^{(+)})^{-1}. \]

Let us now prove the following remarkable:

**Theorem 3.** Let \( s \) be a right special element of \( S \). Then

\[ s^t = s^{(-)}, \]

i.e. the palindrome left-closure of \( s \) coincides with the proper left-extension of \( s \) in \( SBS \).

**Proof.** We prove first the theorem in the case \( s = xQ \) with \( x \in \{a, b\} \) and \( Q \in PAL \). By Theorem 2 and Lemma 2 one has \( Q \in PER = SBS \). If \( s \in SBS \), then there is nothing to prove. Let us then suppose that \( s \) is not a palindrome. If \( Card(alph(Q)) = 1 \), then \( Q = y|Q| \) with \( y \in \mathcal{A} \) and \( x \neq y \). Indeed, \( Q \neq x|Q| \), otherwise \( s \) would be a palindrome. Hence \( s = xy|Q| \) and

\[ s^{(-)} = y|Q|x|Q|. \]

This implies that \( s^{(-)} \in SBS \) since

\[ s^{(-)}yx = y|Q|x|Q|x \in \Sigma. \]

Hence in this case \( s^t = s^{(-)} \). Let us then suppose that \( Card(alph(Q)) = 2 \). By Lemma 4 we can write \( Q \) as:

\[ Q = PxyR = RyxP, \]

with \( P, R \in PAL, x, y \in \{a, b\}, x \neq y \). Hence

\[ s = xPxyR. \]

Let us first suppose that \( |P| > |R| \). Let us prove that in this case the maximal palindrome prefix \( V \) of \( s \) is \( xPx \). In fact, otherwise, there would exist words \( R_1, R_2 \in \{a, b\}^* \) such that \( R = R_1xR_2 \) and

\[ V = xPxyR_1x. \]
This would imply that $P_{xy}R_1$ is a palindrome prefix of $Q$ whose length is greater than $|P|$ which is absurd in view of Lemma 4. It follows then:

$$s(-) = R_yxP_{xy}R.$$

Since by Propositions 7 and 4, $R_yxP_{xy}R \in SBS$, one has $s^1 = s(-)$. 

Let us now suppose that $|P| < |R|$. We want to prove that also in this case the maximal palindrome prefix of $s$ is $xP_x$. If $|R| = |P| + 1$, then since $P_{xy}R = R_yxP$ one has $P_x = R = xP$ so that

$$P = x^{|P|}, \quad R = x^{|P|+1}, \quad Q = x^{|P|+1} y x^{|P|+1}, \quad s = x^{|P|+2} y x^{|P|+1}.$$ 

In this case the maximal palindrome prefix of $s$ is $x^{|P|+2} = xP_x$. Let us then suppose that $|R| \geq |P| + 2 = p$. From Lemma 4 there exist and are unique the integers $k > 0$, $0 \leq r < p$, such that

$$R = (P_{xy})^k U, \quad Q = (P_{xy})^{k+1} U;$$

with $|U| = r$. Hence:

$$s = x(P_{xy})^{k+1} U.$$

Suppose now that $s$ has a palindrome prefix $V$ whose length is greater than $|xP_x|$. We have to consider two cases:

**Case 1:** $|V| < |x(P_{xy})^{k+1}|$. Since $|V| > |xP_x|$ there exist $h \geq 1$ and a prefix $P'$ of $P$ such that:

$$V = x(P_{xy})^h P'.$$

This implies that the word $(P_{xy})^h P'$ is palindrome, i.e.

$$(P_{xy})^h P' = \tilde{P}'(yxP)^h.$$ 

If $P = P'$ we reach a contradiction since one derives $P_{xy} = P_{yx}$. Let us then suppose $|P'| < |P|$. In this case $P'x$ is a prefix of $P$, so that we can write $P = P'xP''$ with $P'' \in S^*$. Thus from the above equation one has

$$(P'xP''y)^h P' = \tilde{P}'(yxP'xP'')^h.$$ 

This implies:

$$P'xP''xyP' = \tilde{P}' y x P' x P''.$$ 

From this it follows $P' = \tilde{P}'$ and then $x = y$ which is a contradiction.

**Case 2:** $|V| > |x(P_{xy})^{k+1}|$. One has

$$|V| > 1 + (k + 1) p = |R| + (p - r) + 1.$$ 

Since $p - r \geq 1$ one has

$$|V| - 2 \geq |R| + 1.$$
Now $V$ is a palindrome prefix of $s=xQ$ so that $V=xV'x$. This implies that $Q$ has the palindrome prefix $V'$ whose length $|V'|=|V|\geq|R|+1$. Moreover, since $|V|<|Q|$ one has $|V'|<|Q|$ which is absurd since $R$ is the maximal proper palindrome prefix of $Q$.

Turning back to our problem we have that also in this case the maximal palindrome prefix of $s$ is $xP'x$ so that by Lemma 5

$$s(-)=RyxPxyR.$$  

Since $RyxPxyR \in SBS$ one has $s^+=s(-)$.

Let us now suppose that $s \in S_R$ is such that $s=\lambda xQ$ where $\lambda \in \mathcal{A}^*$, $\lambda \neq \varepsilon$ and $Q$ is the maximal proper palindrome suffix of $s$. If $s \in \text{PAL}$ there is nothing to prove. Let us then suppose that $s$ is not a palindrome. One has then

$$s(-)=\lambda's = \lambda' \lambda xQ = \lambda' \lambda f,$$

where $\lambda' \in \mathcal{A}^*$ and $f=xQ$. Now, as proved before $f^+=f(-)$. Moreover, since $s(-)$ is a palindrome left-extension of $f$ one has:

$$|s(-)| \geq |f(-)| = |f^+|.$$

Since the left-extension $f^+$ of $f \in S_R$ in $SBS$ is unique one has $s^+=f^+$, so that $|s(-)| \geq |s|$. Moreover, one has obviously that:

$$|s(-)| \leq |s|.$$

Hence $|s(-)|=|s|$. Since by Lemma 5, $s(-)=\tilde{P}P\delta$, where $P$ is the maximal palindrome prefix of $s=P\delta$ and $s^+=\mu P\delta$, it follows that $|\mu|=|\tilde{\delta}|$ so that $\mu=\tilde{\delta}$ and $s^+=s(-)$. □

**Remark 1.** Let us observe that Proposition 5 can be derived as a corollary of the previous theorem since this latter, according to the given proof, can be restated as follows: If $s \in S_R$, then $s(-) \in SBS$. Moreover, from the proof of Theorem 3 one has also that if $w=PxyQ=QyxP \in SBS$ with $P,Q \in \text{PAL}$ and $x,y \in \{a,b\}$, $x \neq y$, then $(xw)(-) = (yw)(-) \in SBS$ and

$$(xw)(-)=QyxPxyQ, \quad (yw)(-)=PxyQyxP.$$  

If $X$ is a subset of $\mathcal{A}^*$ we denote by $X(-)$ the set

$$X(-)=\{w(-) \in \mathcal{A}^* \mid w \in X\}.$$  

Let us define inductively the sequence $\{X_n\}_{n \geq 0}$ of finite subsets of $\mathcal{A}^*$ as

$$X_0 = \{\varepsilon\},$$

$$X_{n+1}=(\mathcal{A}X_n)(-), \quad n \geq 0.$$  

Thus $s \in X_{n+1}$ if and only if there exist $x \in \mathcal{A}$ and $t \in X_n$ such that $s=(xt)(-)$. We set

$$\mathcal{P} = \bigcup_{n \geq 0} X_n.$$
Theorem 4. Let $\mathcal{A} = \{a, b\}$. One has

$$\mathcal{L} = \text{SBS}.$$ 

Proof. Let us first prove the inclusion $\mathcal{L} \subseteq \text{SBS}$. We show by induction on the integer $n$ that for any $n > 0$, $X_n \subseteq \text{SBS}$. The proof of the base of the induction is trivial since $X_0$ and $X_1$ are obviously included in $\text{SBS}$. Suppose now that $X_n \subseteq \text{SBS}$ for $n > 0$; we want to prove that $X_{n+1} \subseteq \text{SBS}$. Let $s \in X_{n+1}$. This implies that there exist $x \in \{a, b\}$ and $t \in X_n$ such that $s = (xt)^\cdot$. Since $t \in \text{SBS}$ then $xt$ is a right special element of $St$ so that by Theorem 3, $s = (xt)^\cdot = (xt)^\dagger \in \text{SBS}$. Thus $X_{n+1} \subseteq \text{SBS}$. Hence $\mathcal{L} \subseteq \text{SBS}$.

Let us now prove the inverse inclusion $\text{SBS} \subseteq \mathcal{L}$. The proof is by induction on the length of the elements of $\text{SBS}$. Let $s \in \text{SBS}$. If $|s| \leq 1$ the result is trivial. Let us then suppose $|s| > 1$. By Theorem 2, $\text{SBS} = \text{PER}$ and by Proposition 7, $s$ is either equal to $x^{|s|}$ with $x \in \{a, b\}$ or $s \in \text{PALabPAL} \cap \text{PAL}$. In the first case, trivially, $x^{|s|} \in X_{|s|} \subseteq \mathcal{L}$. Let us then suppose that

$$s = Px_y Q = QyxP,$$

with $P, Q \in \text{PAL}$, $x, y \in \{a, b\}$, $x \neq y$. We can always suppose that $|P| < |Q|$. If $|Q| = |P| + 1$, then $Q = Px = xP$ so that $Q = x^{|Q|}$ and $s = x^{|Q|}yx^{|Q|}$. Hence, $s = (yx^{|Q|})^\cdot$ where $x^{|Q|} \in X_{|Q|} \subseteq \mathcal{L}$. This implies $s \in X_{|Q|+1}$. Let us now suppose $|Q| \geq |P| + 2$. One has:

$$Q = PxyR,$$

where $R \in \text{PAL}$ by Lemma 3. Hence

$$s = PxyPxyR = PxyRyxP.$$

Now by Proposition 7 and Theorem 2, $RyxP \in \text{PER} = \text{SBS}$. Moreover, $s = (yRyxP)^\cdot$ since, as we have seen in the proof of Theorem 3 (cf. Remark 1) the maximal palindrome prefix of $yRyxP$ is $yRy$. Since $|RyxP| < |s|$ by the inductive hypothesis $RyxP \subseteq X_n$ for a suitable $n > 0$. Hence $(yRyxP)^\cdot \in X_{n+1}$. $\square$

Let $\mathcal{A}$ be any alphabet. A subset $X$ of $\mathcal{A}^*$ is a prefix code if $X \cap X. \mathcal{A}^+ = \emptyset$. In a symmetric way $X$ is a suffix code if $X \cap \mathcal{A}^+X = \emptyset$. The set $X$ is called biprefix code if it is both prefix and suffix (cf. [1]).

Lemma 6. If $X \subseteq \mathcal{A}^*$ is a suffix code, then $Y = (X)^\cdot$ is a biprefix code. Moreover, $Y = (\mathcal{A}X)^\cdot$ is a biprefix code such that $\text{Card}(Y) = \text{Card}(\mathcal{A}) \text{Card}(X)$.

Proof. Let us prove that $Y$ is a suffix code. Suppose, by contradiction, that there exist $\gamma_1, \gamma_2 \in Y$ such that $\gamma_1 = \lambda \gamma_2$, $\lambda \in \mathcal{A}^*$. We can write $\gamma_1 = x_1^\cdot$ and $\gamma_2 = x_2^\cdot$ with $x_1, x_2 \in X$. This implies that $\gamma_1 = \alpha x_1$, $\gamma_2 = \beta x_2$, with $\alpha, \beta \in \mathcal{A}^*$. By the hypothesis one has $\alpha x_1 = \lambda \beta x_2$. Since $X$ is a suffix code it follows that $x_1 = x_2$ and then $\gamma_1 = \gamma_2$. This shows that $Y$ is a suffix code. Since the words of $Y$ are palindromes it follows that $Y$ is also a prefix code and then a biprefix code.
If $X$ is a suffix code, then so will be $\mathcal{A}X$. Thus from the previous result $(\mathcal{A}X)^{-}$ is a biprefix code. Let $x_1, x_2 \subset X$ be such that $x_1 \neq x_2$ and suppose that $(xx_1)^{-} = (yx_2)^{-}$ for $x, y \in \mathcal{A}$. This implies that $\lambda xx_1 = \mu yy_2$, for suitable $\lambda, \mu \in \mathcal{A}^*$. Since $X$ is a suffix code then $x_1 = x_2$ which is absurd. From this trivially follows that $\text{Card}(Y) = \text{Card}(\mathcal{A}) \text{Card}(X)$. 

Corollary 2. For each $n > 0$ the set $X_n$ is a biprefix code having $2^n$ elements.

Proof. Since $X_1 = \{a, b\}$ is a biprefix code, then by the above lemma it follows that also $X_2$ is biprefix, so that by induction one has that for all $n > 0$, $X_n$ is a biprefix code. Moreover, $\text{Card}(X_n) = 2\text{Card}(X_{n-1})$ that implies by iteration $\text{Card}(X_n) = 2^n$. 

Let $\mathcal{A} = \{a, b\}$. We define the map

$$\psi : \mathcal{A}^* \rightarrow \text{SBS},$$

as

$$\psi(\varepsilon) = \varepsilon, \quad \psi(a) = a, \quad \psi(b) = b,$$

and for all $w \in \mathcal{A}^*$, $x \in \mathcal{A}$,

$$\psi(wx) = (x\psi(w))^{(-)}.$$

Lemma 7. For all $w, u \in \mathcal{A}^*$

$$\psi(wu) \in \mathcal{A}^* \psi(w).$$

Proof. The proof is by induction on the length of $u$. If $u = \varepsilon$ the result is trivial. Let us then suppose $|u| > 0$. We can write $u = vx$ with $v \in \mathcal{A}^*$ and $x \in \mathcal{A}$. One has then:

$$\psi(wvx) = (x\psi(wx))^{(-)} = \lambda x\psi(wx),$$

with $\lambda \in \mathcal{A}^*$. The last equality is due to the fact that any word is a suffix of its left palindrome closure. By the induction hypothesis $\psi(wx) = \lambda' \psi(w)$, $\lambda' \in \mathcal{A}^*$. Thus $\psi(wu) = \lambda x\lambda' \psi(w)$. 

Proposition 8. The map $\psi : \mathcal{A}^* \rightarrow \text{SBS}$ is a bijection.

Proof. Since $\psi(\mathcal{A}^*) = \mathcal{L}$, from Theorem 4 one derives that $\psi$ is a surjection. Let us then prove that $\psi$ is an injection. Let $w_1, w_2 \in \mathcal{A}^*$ be such that $w_1 \neq w_2$ and $\psi(w_1) = \psi(w_2)$. We may always suppose $|w_1| \leq |w_2|$. We have to consider the following two cases:

Case 1: $w_1$ is a proper prefix of $w_2$, i.e. $w_2 = w_1x\zeta$, with $x \in \mathcal{A}$ and $\zeta \in \mathcal{A}^*$. One has then by the previous lemma:

$$\psi(w_2) = \lambda \psi(w_1x) = \lambda (x\psi(w_1))^{(-)} = \lambda \lambda' x\psi(w_1).$$

for suitable $\lambda, \lambda' \in \mathcal{A}^*$. This contradicts the hypothesis that $\psi(w_1) = \psi(w_2)$. 


Case 2: \( w_1 = px \xi , \ w_2 = py \xi ' \) with \( p, \xi , \xi ' \in \mathcal{A}^* \) and \( x, y \in \mathcal{A} , x \neq y \). One has from the previous lemma:
\[
\psi(w_1) = \lambda \psi(px), \quad \psi(w_2) = \lambda' \psi(py),
\]
with \( \lambda, \lambda' \in \mathcal{A}^* \). Since \( \psi(px) = (x \psi(p))(\lambda') = \mu x \psi(p) , \ \mu \in \mathcal{A}^* \) and \( \psi(py) = (y \psi(p))(\lambda') = \mu' y \psi(p) , \ \mu' \in \mathcal{A}^* \), it follows:
\[
\psi(w_1) = \lambda \mu x \psi(p), \quad \psi(w_2) = \lambda' \mu' y \psi(p).
\]
Since \( x \neq y \) it follows \( \psi(w_1) \neq \psi(w_2) \) which is a contradiction. \( \square \)

Lemma 8. Let \( w_1, w_2 \in \mathcal{A}^* \) be such that \( \psi(w_2) = \lambda \psi(w_1) , \ \lambda \in \mathcal{A}^* \), then there exists \( w' \in \mathcal{A}^* \) such that \( w_2 = w_1 w' \).

Proof. If \( \lambda = \varepsilon \) one has \( \psi(w_2) = \psi(w_1) \). Since \( \psi \) is a bijection \( w_2 = w_1 \), so that \( w' = \varepsilon \). Let us then suppose \( |\lambda| > 0 \). We can write \( \lambda = \mu x_1 \), with \( x_1 \in \mathcal{A} \). Thus \( \psi(w_2) = \mu x_1 \psi(w_1) \), so that \( \psi(w_2) \) is a left-palindrome extension of \( x_1 \psi(w_1) \). Let \( \sigma \) be the palindrome suffix of \( \psi(w_2) \) of minimal length such that \( |\sigma| \geq |x_1 \psi(w_1)| \). One has then that \( \sigma = (x_1 \psi(w_1))(\lambda') \). Since \( \lambda' \) is a palindrome one can write
\[
\psi(w_2) = \lambda' \mu x_1 \psi(p), \quad \psi(w_1) = \lambda \mu x \psi(p).
\]
Since \( x \neq y \) it follows \( \psi(w_1) \neq \psi(w_2) \) which is a contradiction. \( \square \)

Corollary 3. If \( X \subseteq \mathcal{A}^* \) is a prefix code, then \( \psi(X) \) is a biprefix code.

Proof. Let \( Y = \psi(X) \) and suppose that there exist \( y_1, y_2 \in Y \) such that \( y_2 = \lambda y_1 \), \( \lambda \in \mathcal{A}^* \). Let us set \( y_1 = \psi(x_1) \) and \( y_2 = \psi(x_2), x_1, x_2 \in X \). One has then \( \psi(x_2) = \lambda \psi(x_1) \). From the above lemma one has \( x_2 = x_1 x' \), \( x' \in \mathcal{A}^* \). Since \( X \) is a prefix code then \( x' = \varepsilon \) so that \( x_1 = x_2 \). This implies \( \psi(x_1) = \psi(x_2) \). Thus \( Y \) is a suffix code. Since the words of \( Y \) are palindromes one has that \( Y \) is biprefix. \( \square \)

Proposition 9. Let \( w \in \mathcal{A}^* \). If \( \psi(w) \) has the canonical representation
\[
\psi(w) = PxyQ,
\]
with \( P, Q, PxyQ \in PAL, \ x, y \in \mathcal{A} , \ |P| < |Q| \), then for any \( k \geq 0 \), \( \psi(wx^k) \) and \( \psi(wy^k) \) have the canonical representations:
\[
\psi(wx^k) = QyxP(xyQ)^k, \quad \psi(wy^k) = PxyQ(yxP)^k.
\]

Proof. The proof is by induction on the integer \( k \). For \( k = 0 \) the result is trivial. Suppose that we have proved the assertion up to \( k - 1 \). Since \( P(xyQ)^{k-1} \) and \( (Pxy)^{k-1}Q \)
are palindromes, one has by the inductive hypothesis and by Theorem 3 (cf. Remark 1):

\[
\psi(wx^k) = (x\psi(wx^{k-1}))^{(-)} = (x(Qyx)^k PxyQ)^{(-)}
\]
\[
= (Qyx)^k PxyQ = QyxP(xyQ)^k,
\]
and

\[
\psi(wy^k) = \phi(y\psi(wy^{k-1}))^{(-)} = (y(Pxy)^k QyxP)^{(-)}
\]
\[
= (Pxy)^k QyxP = PxyQ(yxP)^k. \quad \Box
\]

We can represent any word \( w \in \mathcal{A}^* \) uniquely by a finite sequence \((h_1, h_2, \ldots, h_n)\) of integers, where \( h_1 \geq 0, h_i > 0 \) for \( 1 < i \leq n \) and

\[
w = a^{h_1} b^{h_2} a^{h_3} \ldots.
\]

One has \( |w| = \sum_{i=1}^{n} h_i \). We call such a representation of the words of \( \mathcal{A}^* \) the integral representation.

**Proposition 10.** Let \( w \in \mathcal{A}^* \) and be \((h_1, h_2, \ldots, h_n)\) its integral representation. The standard words have, respectively, the directive sequences

\[
(h_1, \ldots, h_n, 1), \quad (h_1, \ldots, h_{n-1}, h_n + 1),
\]
if \( n \) is even, and, respectively

\[
(h_1, \ldots, h_{n-1}, h_n + 1), \quad (h_1, \ldots, h_n, 1),
\]
if \( n \) is odd.

**Proof.** The proof is by induction on the length \( n \) of the integral representation \((h_1, h_2, \ldots, h_n)\) of \( w \). We shall first suppose that \( h_1 > 0 \).

**Base of the induction.** For \( n = 1 \) we have that \( w = a^{h_1} \) so that \( \psi(a^{h_1}) = a^{h_1} \). The standard words \( a^{h_1}ab \) and \( a^{h_1}ba \) have, as one easily verifies, the directive sequences \((h_1 + 1)\) and \((h_1, 1)\), respectively. We check the basis of the induction also for \( n = 2 \) for reasons which will be clear in the proof of the induction step. For \( n = 2 \)

\[
\psi(a^{h_1}b^{h_2}) = (a^{h_1}b)^{b_2} a^{h_1},
\]
so that the two standard words:

\[
(a^{h_1}b)^{h_2}a^{h_1+1}b, \quad (a^{h_1}b)^{h_2+1}a,
\]
have the directive sequences \((h_1, h_2, 1)\) and \((h_1, h_2 + 1)\), respectively.
The induction step. Suppose now that we have proved the assertion up to \( n - 1 \geq 2 \). Thus consider the sequence \((h_1, h_2, \ldots, h_{n-1})\). This is the integral representation of a word \( w_1 \in \mathcal{A}^* \) having that:

\[
w = w_1 x^{h_n},
\]

where \( x = b \) (resp. \( x = a \)) if \( n \) is even (resp. odd). By the inductive hypothesis the sequence \((h_1, h_2, \ldots, h_{n-1}, 1)\) is the directive sequence of a standard word \( f \) having \( \psi(w_1) \) as a prefix of length \(|f| - 2\). We can consider the sequence of standard words:

\[
f_0, f_1, \ldots, f_{n+1},
\]

where

\[
f_0 = b, \quad f_1 = a, \quad f_s = f_{s-1}^{h_{s-1}} f_{s-2}, \quad 2 \leq s \leq n, \quad f_{n+1} = f_n f_{n-1},
\]

and \( f = f_{n+1} \). Moreover, as one easily derives, one has

\[
f = \psi(w_1)xy,
\]

with \( y \in \mathcal{A} \) and \( y \neq x \). Since \( n \geq 3 \) we can write

\[
f_n = P_{yx}, \quad f_{n-1} = Q_{xy}, \quad f_{n+1} = P_{yx} Q_{xy},
\]

with \( P, Q, P_{yx} Q \in \mathcal{P} \mathcal{A} \). Hence \( \psi(w_1) = P_{yx} Q \). Now by Theorem 3

\[
\psi(w_1 x) = (x P_{yx} Q)^{(-1)} = (x Q_{xy} P)^{(-1)} = P_{yx} Q_{xy} P = P_{yx} P_{yx} Q,
\]

so that \( \psi(w_1 x)xy = f_n^2 f_{n-1} \). By iterating the same argument (cf. Proposition 9) one derives:

\[
\psi(w)xy = \psi(w_1 x^{h_n})xy = f_n^{h_{n+1}} f_{n-1},
\]

so that the standard word \( \psi(w)xy = f_n^{h_{n+1}} f_{n-1} \) has the directive sequence

\[
(h_1, \ldots, h_{n-1}, h_n, 1).
\]

Since the words \( f_n, f_n^{h_{n}}, f_{n-1}, f_n^{h_{n+1}} f_{n-1} \) are standard, one has (cf. [7, Proposition 2]) also

\[
(f_n^{h_{n+1}} f_{n-1})(xy)^{-1} = (f_n^{h_{n}} f_{n-1} f_n)(yx)^{-1}.
\]

Therefore

\[
\psi(w)yx = f_n^{h_n} f_{n-1} f_n
\]

which is a standard word having the directive sequence

\[
(h_1, \ldots, h_{n-1}, h_n, 1).
\]

If \( n \) is even, then \( x = b \) so that \( \psi(w)ba \) has the directive sequence \((h_1, \ldots, h_{n-1}, h_n, 1)\) and \( \psi(w)ab \) has the directive sequence \((h_1, \ldots, h_{n-1}, h_n, 1)\). If \( n \) is odd, then \( x = a \) and the result follows in a similar way.
Let us now suppose that \( h_1 = 0 \), i.e. the integral representation of \( w \) is \((0, h_2, \ldots, h_n)\). We first observe that \( \hat{w} \) has the integral representation \((h_2, \ldots, h_n)\). By the above result the standard words \( \psi(\hat{w})ab = (\psi(w)ba) \) and \( \psi(\hat{w})ba = (\psi(w)ab) \) have, respectively, the directive sequences \((h_2, \ldots, h_n, 1)\) and \((h_2, \ldots, h_{n-1}, h_n + 1)\) if \( n \) is odd and, respectively, \((h_2, \ldots, h_{n-1}, h_n + 1)\) and \((h_2, \ldots, h_n, 1)\) if \( n \) is even. Hence if \( n \) is odd \((\text{resp. even})\) \( \psi(w)ba \) and \( \psi(w)ab \) have the integral representations \((0, h_2, \ldots, h_n, 1)\) \((\text{resp.} (0, h_2, \ldots, h_{n-1}, h_n + 1))\) and \((0, h_2, \ldots, h_n, 1)\) \((\text{resp.} (0, h_2, \ldots, h_n))\). 

**Corollary 4.** If \( s \in \text{Stand} \), then \( s \) has a unique directive sequence.

**Proof.** Let \( s \in \text{Stand} \). Then \( s = \psi(w)xy \) with \( w \in \mathcal{A}^* \) and \( x, y \in \mathcal{A}, x \neq y \). Let \((h_1, h_2, \ldots, h_n)\) be the integral representation of \( w \). We suppose that \( x = a \); the case \( x = b \) is dealt with in a symmetric way. From the previous proposition \( s \) has the directive sequence \((h_1, h_2, \ldots, h_n, 1)\) if \( n \) is even and \((h_1, h_2, \ldots, h_{n-1}, h_n + 1)\) if \( n \) is odd.

Suppose now that \( s \) has also the directive sequence \((k_1, k_2, \ldots, k_m)\). Since \( s \in \mathcal{A}^n ab \) then \( m \) has to be an odd integer. If \( k_m = 1 \), then be \( w' \) the word whose integral representation is \((k_1, k_2, \ldots, k_{m-1})\). From the preceding proposition \( \psi(w')ab \) has the directive sequence \((k_1, k_2, \ldots, k_m)\). Hence \( \psi(w)ab = \psi(w')ab \). This implies \( \psi(w) = \psi(w') \).

Since \( \psi \) is injective one has \( w = w' \). It follows \((k_1, k_2, \ldots, k_{m-1}) = (h_1, h_2, \ldots, h_n)\). Thus \( n = m - 1 \) and \((k_1, k_2, \ldots, k_m) = (h_1, h_2, \ldots, h_n, 1)\). Suppose now that \( k_m > 1 \) and consider the word \( w' \) whose integral representation is \((k_1, k_2, \ldots, k_{m-1}, k_m - 1)\). Since \( m \) is odd from the previous proposition \( \psi(w')ab \) has the directive sequence \((k_1, k_2, \ldots, k_m)\). Thus \( \psi(w)ab = \psi(w')ab \) and \( w = w' \). This implies \((h_1, h_2, \ldots, h_n) = (k_1, k_2, \ldots, k_{m-1}, k_m - 1)\). Hence \( n = m \) and \((k_1, k_2, \ldots, k_m) = (h_1, h_2, \ldots, h_n + 1)\). 

Let \( \mathcal{A} = \{a, b\} \) and be \( \mathcal{A}^\omega \) the set of all infinite words on \( \mathcal{A} \). We consider the subset \( \mathcal{A}_0^\omega \) defined as

\[
\mathcal{A}_0^\omega = \{ y \in \mathcal{A}^\omega \mid y \notin \mathcal{A}^* x^\omega, x \in \mathcal{A} \}.
\]

In other words \( y \notin \mathcal{A}^\omega \) if and only if there exists a word \( u \in \mathcal{A}^* \) and a letter \( x \in \mathcal{A} \) such that

\[
y = ux^\omega = uxxx \ldots x \ldots.
\]

Hence any infinite word \( x \in \mathcal{A}_1^\omega \) can be uniquely expressed as

\[
x = a^{h_1} b^{h_2} a^{h_3} \ldots,
\]

with \( h_i > 0 \) and \( h_i > 0 \) for \( i > 1 \). We call the infinite sequence \((h_1, h_2, \ldots, h_n, \ldots)\) the **integral representation** of \( x \).

Let us now associate to each \( x \in \mathcal{A}_0^\omega \) the sequence of words \( \{s_n\}_{n \geq 0} \) defined as:

\[
s_0 = e, \quad s_{n+1} = (x_n s_n)(-), \quad n \geq 0.
\]
Since \( s_n \) is a proper suffix and prefix of \( s_{n+1} \) for any \( n \geq 0 \), the above sequence \( \{s_n\}_{n \geq 0} \) converges to an infinite sequence \( s = \lim s_n \). Thus one can introduce a map \( \psi : \mathcal{A}_0^\omega \to \mathcal{A}_0^\omega \) which associates to \( x \in \mathcal{A}_0^\omega \) the infinite word \( \psi(x) = s \).

**Theorem 5.** Let \( x \in \mathcal{A}_0^\omega \) and be \( (h_1, h_2, \ldots, h_n, \ldots) \) its integral representation. Then \( \psi(x) \) is the infinite standard Sturmian word having the directive sequence \( (h_1, h_2, \ldots, h_n, \ldots) \).

**Proof.** Let \( x \in \mathcal{A}_0^\omega \) and be \( (h_1, h_2, \ldots, h_n, \ldots) \) its integral representation. Let \( \psi(x) = s = \lim s_n \), where \( s_0 = e \) and \( s_{n+1} = (x_n s_n)(-), n \geq 0 \). We consider the subsequence \( \{s_n\}_{n \geq 0} \) defined for all \( n > 0 \) as

\[
\sigma_n = s_{h_1} + h_2 + \ldots + h_n.
\]

Since for all \( i > 0 \), \( \sigma_i \) is a proper prefix of \( \sigma_{i+1} \) and \( \sigma_i \in \text{Pref}(s) \) one has \( \lim \sigma_n - \lim s_n = s \).

Let us now consider the sequence of standard words \( \{t_n\}_{n \geq 0} \) defined for any \( n \geq 0 \) as

\[
t_{2n+1} = \sigma_{2n+1} ba, \quad t_{2n} = \sigma_{2n} ab.
\]

One derives from Proposition 10 that for each \( n > 0, t_n \) has the directive sequence:

\[
(h_1, h_2, \ldots, h_n, 1).
\]

Let us consider now the infinite standard Sturmian word \( y \) whose directive sequence is \( (h_1, h_2, \ldots, h_n, \ldots) \). Thus there exists an infinite sequence of standard words:

\[
f_0, f_1, \ldots, f_n, \ldots
\]

such that

\[
f_0 = b, \quad f_1 = a, \quad f_{n+1} = f_n^{h_n} f_{n-1}, \quad n > 0.
\]

It follows that for any \( n > 0, t_n = f_{n+1} f_n \).

Since \( t_{n+1} = f_{n+2} f_{n+1} = f_n^{h_{n+1}} f_n f_{n+1} \) and \( f_{n+1} = f_n^{h_n} f_n f_{n-1} \) one derives that for any \( n > 0, t_n \) is a proper prefix of \( t_{n+1} \). This implies that there exists the \( \lim t_n \), and, moreover,

\[
t = \lim t_n = \lim \sigma_n = \lim s_n = s.
\]

Since for any \( n > 0, f_n \in \text{Pref}(t) \) one has also \( f_n = y = t \). Hence \( s = y \) that concludes our proof. \( \square \)

**Proposition 11.** The map \( \psi \) is a bijection \( \psi : \mathcal{A}_0^\omega \to \text{Stand} \).
Proof. From the preceding theorem one has $\psi(\mathcal{A}^0) = \text{Stand}$ so that $\psi$ is a surjection. Let us now prove that $\psi$ is injective. Let $x, x' \in \mathcal{A}^0$ be such that $x \neq x'$. We denote by $n$ the minimal integer such that $x_n \neq x'_n$. Let $\psi(x) = s$ and $\psi(x') = s'$. One has then

$$s_n - (x_n s_{n-1})^(-), \quad s'_n - (x'_n s'_{n-1})^(-).$$

Thus $s_n$ has the suffix $x_n s_{n-1}$ and, since $s_n$ and $s_{n-1}$ are palindromes, the prefix $s_{n-1} x_n$. Similarly, $s'_n$ has the prefix $s_{n-1} x'_n$. Hence $s$ will have the prefix $s_{n-1} x_n$ and $s'$ the prefix $s_{n-1} x'_n$. This shows that $s \neq s'$. \qed

In conclusion of this section we remark that a formalism for some respects similar to our has been considered by Raney [15] for the study of some problems on continued fractions in relation with automata theory.

6. The Farey correspondence

We denote by $SBS(a)$ the set of all elements of $SBS$ whose first letter is $a$, i.e. $SBS(a) = SBS \cap \alpha \mathcal{A}^*$. Similarly, $SBS(b)$ will be the set $SBS(b) = SBS \cap b \mathcal{A}^*$. Hence $SBS = \{ \epsilon \} \cup SBS(a) \cup SBS(b)$. One easily verifies that $s \in SBS(a)$ if and only if $s \in SBS(b)$, so that the operation $(-)$ determines a bijection of $SBS(a)$ in $SBS(b)$.

In the following we denote by $\mathcal{F}$ the set of all fractions $p/q$ such that $0 < p \leq q$ and $\gcd(p, q) = 1$. We call $\mathcal{F}$ the set of Farey numbers.

Lemma 9. For any $s \in SBS$ there exists a unique fraction $p/q \in \mathcal{F}$ such that $p, q \in \Pi(s)$, $p$ is the minimal period of $s$ and $|s| = p + q - 2$. The map $\eta : SBS \rightarrow \mathcal{F}$ defined as

$$\eta(s) = p/q,$$

is a surjection. Moreover, for $s \neq \epsilon$

$$\eta(s) = (|s| - |Q|)(|Q| + 2),$$

where $Q$ is the maximal proper palindrome suffix of $s$.

The restrictions $\eta_a$ and $\eta_b$ of $\eta$, respectively, to $SBS(a) \cup \{ \epsilon \}$ and to $SBS(b) \cup \{ \epsilon \}$, are bijections.

Proof. Let $s \in SBS$. If $s = \epsilon$, then the unique fraction in $\mathcal{F}$ such that $|\epsilon| = 0 = p + q - 2$ is $p/q = 1/1$. Let us now suppose that $s = x^{|s|}$ with $x \in \{a, b\}$. The word $s$ has the minimal period $p = 1$. The unique period $q$ of $s$ for which $|s| = q - 1$ is then $q = |s| + 1$. In this case the maximal proper palindrome suffix of $s$ is $Q = x^{|s| - 1}$ so that

$$p/q = 1/(|s| + 1) = (|s| - |Q|)(|Q| + 2).$$

Let us now suppose that $\text{Card}(\text{alph}(s)) = 2$. By Lemma 4, $s$ can be uniquely represented as $s = PxyQ = QyxP$, with $P, Q \in \text{PAL}$, $|P| < |Q|$ and $x, y \in \{a, b\}$, $x \neq y$. 
Moreover, $s$ has the periods $p = |P| + 2$, $q = |Q| + 2$ such that $\gcd(p, q) = 1$ and $|s| = p + q - 2$. Since $|P| < |Q|$, then $p$ is the minimal period of $s$ and $Q$ is the maximal proper palindrome suffix of $s$. We can write the ratio $p/q$, uniquely determined by $s$, as

$$p/q = (|P| + 2)/(|Q| + 2) = (|s| - |Q|)/(|Q| + 2).$$

Let now $p/q \in \mathcal{F}$. We want show that there exist and are only two the words $s$, $\hat{s} \in SBS$ such that $\eta(s) = \eta(\hat{s}) = p/q$. If $p = 1$ then the words $s = a^q$ and $\hat{s} = b^q$ are, trivially, the only two words of $SBS$ such that $\eta(s) = \eta(\hat{s}) = 1/q$. Let us then suppose $p > 1$. Since $\gcd(p, q) = 1$ then from the theorem of Pedersen et al. [14] there exists and is uniquely determined, the word $W \in \Sigma$ such that

$$W = AB = Cab,$$

with $A, B, C \in PAL$ and $|A| = p - 2$, $|B| = q + 2$. Thus $C$ is uniquely determined, $C \in PER$ and has the periods $p$ and $q$ such that $|C| = p + q - 2$ (cf. [7]). By Theorem 2 and Lemma 4, $C$ can be uniquely expressed as

$$C = PxyQ = QyxP,$$

with $P, Q \in PAL$, $x, y \in \{a, b\}$, $x \neq y$, $q = |Q| + 2$, $p = |P| + 2$. Moreover, from the above equation one derives $A = P$, $xy = ba$ and $B = baQab$. One has also, by Lemma 4, that $p$ is the minimal period of $C$. From the equation $W = AB = Cab$ one has that

$$\hat{W} = \hat{A}\hat{B} = \hat{C}ba,$$

so that one derives that $C$ and $\hat{C}$ are the only two words in $SB$ such that $\eta(C) = \eta(\hat{C}) = p/q$. From this it follows, trivially, that the restrictions $\eta_a$ and $\eta_b$ of $\eta$ to $SBS(a) \cup \{e\}$ and to $SBS(b) \cup \{e\}$, respectively, are bijections. □

In the following we denote for $s \in SBS$, $\eta(s) = \|s\|$.

Lemma 10. Let $s \in SBS$ such that $\|s\| = p/q$. There exists a letter $x \in \{a, b\}$ such that

$$\|(xs)^{(-1)}\| = p/(p + q) \quad \text{and} \quad \|(ys)^{(-1)}\| = q/(p + q),$$

with $y \in \{a, b\}$ and $x \neq y$.

Proof. Let $s \in SBS$ such that $\|s\| = p/q$. If $s = e$, then $\|e\| = 1/1$, so that $(ae)^{(-1)} = a$ and $(be)^{(-1)} = b$. Since $\|(ae)^{(-1)}\| = \|(be)^{(-1)}\| = 1/2$.

If $\text{Card(alph}(s)) = 1$, then $s = x^{|s|}$ with $x \in \{a, b\}$ and $\|s\| = p/q = 1/(|s| + 1)$, i.e. $p = 1$ and $q = |s| + 1$. Thus $xs = x^{|s| + 1}$ and $\|(xs)^{(-1)}\| = 1/(|s| + 2) = p/(p + q)$. Moreover, in this case $(ys)^{(-1)} = x^{|s|}yxs^{|s|}$. Hence $\|(ys)^{(-1)}\| = (|s| + 1)/(|s| + 2) = q/(p + q)$.

Let us now suppose that $\text{Card(alph}(s)) = 2$. By Lemma 4, $s$ can be canonically expressed as

$$s = PxyQ = QyxP,$$
with $P, Q \in PAL$, $|P| < |Q|$, $x, y \in \{a, b\}$, $x \neq y$. Moreover, $p = |P| + 2$ and $q = |Q| + 2$. One has then

$$xs = xPxyQ, \quad ys = yQyxP,$$

so that (cf. Proposition 9)

$$(xs)^{-1} = QyxPxyQ = sxyQ, \quad (ys)^{-1} = PxyQyxP = syxP.$$

Hence, recalling that $|s| = p + q - 2$, it follows:

$$\|xs\|^{-1} = q/(p + q), \quad \|ys\|^{-1} = p/(p + q). \quad \square$$

**Corollary 5.** If $s \in S_R$, then $\|s\|^{-1} = p/q$ with $q \leq |s| + 1$.

**Proof.** If $s \in SBS = PER$ then the result is trivial since $s^{-1} = s$ so that $q - 2$ is the length of the maximal proper palindrome suffix of $s$. This length is $\leq |s| - 1$. Thus $q \leq |s| + 1$.

Let us then suppose that $s \notin SBS$. We can write $s$ as

$$s = \lambda xt, \quad \lambda \in A^*, \quad x \in A,$$

where $t \in SBS$ is the maximal proper palindrome suffix of $s$. Hence from Theorem 3, $s^{-1} = (xt)^{-1}$. Let $\|s^{-1}\| = p/q$ and $\|t^{-1}\| = p'/q'$. By the preceding lemma $q = p' + q'$.

Since $p' + q' - 2 = |t|$ it follows that

$$q = p' + q' = |t| + 2 = |xt| + 1 \leq |s| + 1. \quad \square$$

Let us now introduce for $x \in \{a, b\}$ the following maps:

$$\pi_x, \rho_x : SBS(x) \rightarrow SBS(x),$$

defined as: for $s \in SBS(x)$ such that $\|s\| = p/q$ then

$$\pi_x(s) = \eta_x^{-1}(p/(p + q)), \quad \rho_x(s) = \eta_x^{-1}(q/(p + q)).$$

Hence $\|\pi_x(s)\| = p/(p + q)$ and $\|\rho_x(s)\| = q/(p + q)$. By the above lemma if $\|(xs)^{-1}\| = q/(p + q)$, then $\pi_x(s) = (ys)^{-1}$ and $\rho_x(s) = (xs)^{-1}$ with $y \in \{a, b\}$, $x \neq y$. Conversely, if $\|(xs)^{-1}\| = p/(p + q)$, then $\pi_x(s) = (xs)^{-1}$ and $\rho_x(s) = (ys)^{-1}$. Hence $\pi_x(s)$ is a left palindrome extension of $s$ which ‘saves’ the minimal period $p$, whereas $\rho_x(s)$ ‘saves’ the period $q$.

For $k > 0$ let us define the map $\pi_x^{(k)}$ inductively as $\pi_x^{(0)}$ is the identity map and for $k > 0$, $\pi_x^{(k)} = \pi_x \circ \pi_x^{(k-1)}$, where $\circ$ denotes the composition of maps. For $s \in SBS(x)$ such that $\|s\| = p/q$ one has $\|\pi_x^{(k)}(s)\| = p/(kp + q)$, $k > 0$. Let us define for $k > 0$, $\delta_x^{(k)} = \rho_x \circ \pi_x^{(k-1)}$. One easily verifies that $\|\delta_x^{(k)}(s)\| = q/(kq + p)$.

**Lemma 11.** If $s, t \in SBS(x)$, $x \in \{a, b\}$ and there exist $h, k > 0$ such that $\delta_x^{(h)}(s) = \delta_x^{(k)}(t)$, then $s = t$. 

Proof. Suppose that $||s|| = p/q$ and $||t|| = p'/q'$. If there exist $h, k > 0$ such that $\delta^{(k)}_s = \delta^{(h)}_t$, then one has

$$||\delta^{(k)}_s|| = q/(kq + p) = ||\delta^{(h)}_t|| = q'/(hq' + p').$$

Hence $q = q'$ and

$$p + kq = p' + hq'.$$

or

$$p = p' + (h - k)q.$$ If $h > k$ then $p > q$ which is absurd. If $h < k$, then since $p' < q$ it follows $p < 0$ which is absurd. Thus the only possibility is $h = k$, i.e. $p = p'$. This implies $||s|| = ||t||$ and from Lemma 9, $s = t$. \qed

Let $n$ be a positive integer. We define the set $\mathcal{F}_n$ as

$$\mathcal{F}_n = \{ p/q \in \mathcal{F} \mid q \leq n \}.$$ One has $\text{Card}(\mathcal{F}_n) = F(n)$, where

$$F(n) = \sum_{i=1}^{n} \phi(i).$$

If we order the elements of $\mathcal{F}_n$ in an increasing way we obtain a sequence of irreducible fractions called the Farey series of order $n$ and length $F(n)$ (cf. [18]).

In the following table we report the Farey series for $n \leq 5$:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/2</td>
<td>2/3</td>
<td>3/4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/3</td>
<td>1/2</td>
<td>2/3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/4</td>
<td>1/3</td>
<td>2/3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/5</td>
<td>1/4</td>
<td>2/3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/6</td>
<td>1/5</td>
<td>2/5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/7</td>
<td>1/6</td>
<td>2/5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/8</td>
<td>1/7</td>
<td>2/6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1/9</td>
<td>1/8</td>
<td>2/7</td>
</tr>
</tbody>
</table>

We have seen in Section 1 that $s_R(n) = F(n + 1)$. Hence, for any $n > 0$ there exists a bijection of the set $\mathcal{F}_{n+1}$ in $s_R(n) = S_R \cap \mathcal{A}^n$. We shall determine a natural bijection:

$$\Phi_n : \mathcal{F}_{n+1} \rightarrow S_R(n),$$

that we call the Farey correspondence.

Lemma 12. For each $n > 0$, let $\mathcal{G}_n$ be the set

$$\mathcal{G}_n = \{ p/q \in \mathcal{F}_{n+1} \mid p + q - 2 \geq n \}.$$ One has that

$$\text{Card}(\mathcal{G}_n) = \frac{1}{2} F(n + 1) = \frac{1}{2} s_R(n).$$
Proof. An element $p/q \in \mathcal{T}_{n+1}$ does not belong to $\mathcal{G}_n$ if and only if $p + q < n + 2$. Let us recall that the number of all pairs $(p, q)$ of positive integers such that $\text{gcd}(p, q) = 1$ and

$$p + q = i,$$

for a fixed $i > 0$ is given by $\phi(i)$. Let us now add the further condition that $p \leq q$. If $2 < i \leq n + 1$, the number of solutions of the previous equation is then $\frac{1}{2} \phi(i)$. If $i = 2$, then the only solution is $p = q = 1$. Since $1 = \frac{1}{2}(\phi(1) + \phi(2))$ we can state that in any case the number of solutions of the equation $p + q < n + 2$ under the above conditions is given by $\frac{1}{2} \sum_{i=1}^{n+1} \phi(i)$. Hence, $\text{Card}(\mathcal{T}_{n+1} \setminus \mathcal{G}_n) = \frac{1}{2} \sum_{i=1}^{n+1} \phi(i)$ and $\text{Card}(\mathcal{G}_n) = \text{Card}(\mathcal{T}_{n+1}) - \text{Card}(\mathcal{T}_{n+1} \setminus \mathcal{G}_n) = \frac{1}{2} s_R(n)$.

For each $n > 0$ let us define the set $Z_n = \eta_{a}^{-1}(\mathcal{G}_n)$. Since $\eta_a$ is a bijection one has from Lemma 12 that $\text{Card}(Z_n) = \text{Card}(\mathcal{G}_n) = \frac{1}{2} s_R(n)$. We introduce also the sets $\eta_{b}^{-1}(\mathcal{G}_n) = \hat{Z}_n$ and

$$A_n = Z_n \cup \hat{Z}_n = \eta_{a}^{-1}(\mathcal{G}_n) = \{s \in \mathcal{SBS} \mid ||s|| \in \mathcal{G}_n\}.$$

Thus $s \in A_n$ if and only if $s$ is a strictly bispecial element of $\mathcal{S}$ whose Farey number $p/q$ satisfies the condition $|s| = p + q - 2 \geq n$ with $q \leq n + 1$.

Theorem 6. For each $n > 0$ let $S_R(n) = S_R \cap \mathcal{A}^n$. The map defined for $s \in A_n$ as

$$f_n(s) = (\mathcal{A}^*)^{-1} s \cap \mathcal{A}^n$$

is a bijection $f_n : A_n \rightarrow S_R(n)$. Thus one has:

$$(\mathcal{A}^*)^{-1} A_n \cap \mathcal{A}^n = S_R(n),$$

and, moreover,

$$A_n = (S_R(n))^{(-)}.$$

Proof. For each $s \in A_n$, $f_n(s)$ gives the suffix of $s$ of length $n$. Thus $f_n(s) \in S_R(n)$. We want to prove that $f_n$ is a bijection. Since $\text{Card}(A_n) = \text{Card}(Z_n) + \text{Card}(\hat{Z}_n) = s_R(n)$, it is sufficient to prove that $f_n$ is a surjection.

Indeed, for any $s \in S_R(n)$ one has from Corollary 5, $||s^{(-)}|| = p/q$ with $q \leq n + 1$. Since $|s^{(-)}| = p + q - 2 \geq n$ one has $s^{(-)} \in A_n$ and $f_n(s^{(-)}) = s$. This proves that $f_n$ is a bijection and then $(\mathcal{A}^*)^{-1} A_n \cap \mathcal{A}^n = S_R(n)$. As we have seen above $(S_R(n))^{(-)} \subseteq A_n$. Let now $t \in A_n$. One has $f_n(t) = s \in S_R(n)$. Since $f_n(s^{(-)}) = s$ it follows, in view of the fact that $f_n$ is a bijection, $t = s^{(-)}$. This shows $A_n \subseteq (S_R(n))^{(-)}$ and then $A_n = (S_R(n))^{(-)}$. $\square$

Corollary 6. For each $n > 0$ one has:

$$A_n(\mathcal{A}^*)^{-1} \cap \mathcal{A}^n = S_L(n),$$
and, moreover,

\[ A_n = (S_L(n))^{(+)} \).

**Proof.** From the previous theorem \((\mathcal{A}^*)^{-1} A_n \cap \mathcal{A}^n = S_R(n)\). From Lemma 1, \(S_R = S_L\). Moreover, \(A_n = A_n\) since the words in \(A_n\) are palindromes. Hence by taking the mirror images of both sides of the previous relation it follows \(A_n(\mathcal{A}^*)^{-1} \cap \mathcal{A}^n = S_L(n)\). Since for any \(w \in \mathcal{A}^*\) one has \(w^{(-)} = (\bar{w})^{(+)}\) it follows:

\[ A_n = (S_R(n))^{(-)} = (S_L(n))^{(-)} = (S_L(n))^{(+)} \). □

Let \(i, j\) be two positive integers such that \(i \leq j\). We introduce the following number-theoretic function \(\phi_{[i,j]}\), that we call **generalized Euler’s function** defined as:

\[ \phi_{[i,j]}(n) = \text{Card}\{x \in [i, j] | \gcd(x, n) = 1\} \].

In other words \(\phi_{[i,j]}(n)\) gives the number of integers in the interval \([i, j]\) which are primes with the integer \(n\). One has, of course, that \(\phi_{[1,n]}(n) = \phi_{[1,n-1]}(n) = \phi(n)\).

**Lemma 13.** Let \(n > 1\) and \(k\) be such that \(k < n < 2k\). The number of pairs \((p, q)\) of positive integers such that

\[ p + q = n, \quad 1 \leq p < q < k, \quad \gcd(p, q) = 1, \]

is given by

\[ \frac{1}{2} \phi_{[n-k,k]}(n). \]

**Proof.** Let us first count the number of pairs \((p, q)\) of positive integers such that \(\gcd(p, q) = 1\), \(p + q = n\) and \(p, q \in [1, k]\). Let us observe that \(\gcd(p, q) = 1\) if and only if \(\gcd(q, n) = 1\). Since \(p + q = n\) and \(q \leq k\), \(q\) can run in the interval \([n - k, k]\) (note that \(n - k < k\)). Hence the number of the above pairs \((p, q)\) is given by \(\phi_{[n-k,k]}(n)\). If we add the constraint \(1 \leq p \leq q\), then the number of the pairs \((p, q)\) becomes \(\frac{1}{2} \phi_{[n-k,k]}(n)\). □

**Proposition 12.** For each \(n > 0\), \(A_n\) is a biprefix code. The minimal length \(l_{\text{min}}\) and the maximal length \(l_{\text{max}}\) of the words of \(A_n\) are \(l_{\text{min}} = n\), \(l_{\text{max}} = 2n - 1\). Moreover, for each \(h \in [0, n - 1]\)

\[ \text{Card}(A_n \cap \mathcal{A}^{n+h}) = \phi_{[n-1,n-1]}(n + h + 2). \]

**Proof.** From Theorem 6, for each \(n > 0\), \(A_n = (S_R(n))^{(-)}\). Since \(S_R(n)\) is, trivially, a biprefix code then from Lemma 6 it follows that for each \(n > 0\), \(A_n\) is a biprefix code.

Let \(s \in A_n\) and \(|s| = p/q\) its Farey number. One has \(|s| = p + q - 2 \geq n\) and \(q \leq n + 1\). Since \(p \leq q - 1\) it follows \(|s| = p + q - 2 \leq 2q - 3 \leq 2n - 1\). Hence

\[ n \leq |s| \leq 2n - 1. \]
To achieve the result it is sufficient to observe that $a^n$, $a^{n-1}ba^{n-1} \in \Delta_n$; indeed, $a^n$, $a^{n-1}ba^{n-1} \in SBS$, $|a^n| = 1/(n+1)$, $|a^{n-1}ba^{n-1}| = n/(n+1)$ and $|a^n| = n$, $|a^{n-1}ba^{n-1}| = 2n - 1$.

Let $s \in \Delta_n \cap \mathcal{A}_{n+h}$ with $h \in [0, n-1]$. One has $|s| = p + q - 2 = n + h$. Hence

$$p + q = n + h + 2, \quad 1 \leq p \leq q \leq n + 1, \quad gcd(p, q) = 1.$$ 

By the previous lemma the number of pairs $(p, q)$ of positive integers for which the above condition is satisfied is given by $\frac{1}{2} \phi[h+1,n+1](n + h + 2)$. Since to each Farey number $p/q \in \mathcal{G}_n$ correspond two words, namely $s$ and $s^*$ having the Farey number $p/q$ one derives: $Card(\Delta_n \cap \mathcal{A}_{n+h}) = \phi[h+1,n+1](n + h + 2)$. 

Let us observe that $Card(\Delta_n) = F(n+1)$ so that the following identity holds for any $n \geq 0$

$$\sum_{r=1}^{n} \phi_{[r,n+1]}(n + r + 1) = \sum_{i=1}^{n+1} \phi(i).$$

For each $n \geq 0$ we shall call $\Delta_n$ the $n$-Farey code. Let us consider now the subset $S'_R(n)$ of $S_R(n)$ defined as

$$S'_R(n) = (\mathcal{A}^*)^{-1}Z_n \cap \mathcal{A}^n.$$ 

One has that $Card(S'_R(n)) = \frac{1}{2} S_R(n)$. Moreover, if $S''_{R}(n) = (\mathcal{A}^*)^{-1}Z_n \cap \mathcal{A}^n$ one has $S''_{R}(n) = S'_R(n)$. Let us introduce now the map $\Phi'_n : \mathcal{G}_n \rightarrow S'_R(n)$ defined as

$$\Phi'_n = \eta_a^{-1} \circ f_n.$$ 

The map $\Phi'_n$ is a bijection that we call the first restricted Farey correspondence.

**Example.** Let us consider the Farey series of order $n = 7$:

\[
\begin{array}{cccccccc}
1 & 1 & 1 & 1 & 2 & 1 & 4 & 3 \\
7 & 6 & 5 & 4 & 3 & 2 & 1 & 7
\end{array}
\]

The length of the Farey series is 18. We construct the first restricted Farey correspondence $\Phi'_6 : \mathcal{G}_6 \rightarrow S'_R(6)$. The set $\mathcal{G}_6$ is formed by the following 9 elements ordered in an increasing way:

\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 9 \\
7 & 6 & 5 & 4 & 3 & 2 & 1 & 7
\end{array}
\]

In Table 1 we report in the first column the elements of $\mathcal{G}_6$, in the second column the values of $\eta_a^{-1}$, in the third column their lengths and in the fourth column the values of the first restricted Farey correspondence.

Let us set for each $n \geq 0$, $\mathcal{G}'_n = \mathcal{G}_{n+1} \setminus \mathcal{G}_n$. Let $Y_n$ be the set:

$$Y_n = \eta_a^{-1}(\mathcal{G}'_n).$$
\begin{table}
\caption{Farey number s.bisp.element Length Special element}
\begin{tabular}{llll}
\hline
Farey number & s.bisp.element & Length & Special element \\
\hline
1/2 & aaaaaa & 6 & aaaaaa \\
2/3 & abababa & 7 & baba \\
3/5 & aabaabaa & 8 & baaa \\
4/7 & aabaaabaa & 9 & aaaa \\
5/8 & ababa & 6 & aaba \\
3/5 & ababaababa & 10 & aaba \\
4/5 & aaabaaa & 7 & aaaa \\
6/5 & aabaabaa & 9 & aaaa \\
6/5 & aaaaaba & 11 & aaaa \\
\hline
\end{tabular}
\end{table}

We introduce the map $g_n$ defined in $Y_n$ as follows: for any $s \in Y_n$

$$g_n(s) = \delta_b^{k(s)}(s),$$

where $k(s) = \min\{k \in N_+ \mid |\delta_b^{(k)}(s)| \geq n\}$.

\textbf{Theorem 7.} For each $n > 0$, $g_n$ is a bijection $g_n : Y_n \to \hat{Z}_n$.

\textbf{Proof.} From Lemma 12 one has that for each $n > 0$, $\text{Card}(\mathcal{F}_n') = \text{Card}(\mathcal{F}_n) = \frac{1}{2}s_R(n)$, so that $\text{Card}(Y_n) = \frac{1}{2}s_R(n)$. Moreover, for any $k > 0$, $\delta_b^{(k)} = \rho_b \circ \pi_b^{(k-1)}$. Let $s \in Y_n$ be such that $\|s\| = p/q$. One has then $\|\delta_b^{(k)}(s)\| = q/(p + kq)$. Let $k = k(s)$; from the definition of $k(s)$ it follows that $p + (k+1)q \geq n$ and $p + kq - 2 < n$. Hence $p + kq \leq n + 1$. This implies that $g_n(s) \in A_n$. Since $g_n(s)$ is a palindrome left-extension of $s$ and $s$ terminates with the letter $b$ then $g_n(s)$ will begin with the letter $b$. Thus $g_n(s) \in \hat{Z}_n$. From Lemma 11 it follows that $g_n$ is injective. Since $\text{Card}(\hat{Z}_n) = \text{Card}(Y_n)$ one has that $g_n$ is a bijection. \hfill $\square$

Let us introduce the map $\Phi'' : \mathcal{F}_n' \to S''_R(n)$ defined as

$$\Phi'' = \eta_b^{-1} \circ g_n \circ f_n.$$

The map $\Phi''$ is a bijection that we call the second restricted Farey correspondence. Hence the Farey correspondence $\Phi_n : \mathcal{F}_{n+1} \to S_R(n)$, is completely determined by its restrictions $\Phi'_n$ and $\Phi''_n$.

\textbf{Example.} We report in Table 2 the second restricted Farey correspondence $\Phi''_6 : \mathcal{F}_6' \to S''_R(6)$ in the case $n = 6$. The elements of $\mathcal{F}_6'$ ordered in an increasing way are reported in the first column. In the second column there are the values of the function $\eta_b^{-1}$, in
Table 2

<table>
<thead>
<tr>
<th>Farey number</th>
<th>s.bispec.element</th>
<th>$g_6$</th>
<th>Farey number</th>
<th>Special element</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1}{6}$</td>
<td>bbbbb</td>
<td>bbbbbabbbbb</td>
<td>$\frac{5}{7}$</td>
<td>abbbbb</td>
</tr>
<tr>
<td>$\frac{1}{5}$</td>
<td>bbb</td>
<td>bbbabbbbb</td>
<td>$\frac{5}{6}$</td>
<td>babbbb</td>
</tr>
<tr>
<td>$\frac{1}{4}$</td>
<td>bb</td>
<td>bbbabbb</td>
<td>$\frac{4}{5}$</td>
<td>babbbb</td>
</tr>
<tr>
<td>$\frac{1}{3}$</td>
<td>bb</td>
<td>bbabbb</td>
<td>$\frac{3}{4}$</td>
<td>abbb</td>
</tr>
<tr>
<td>$\frac{2}{5}$</td>
<td>babab</td>
<td>babababab</td>
<td>$\frac{5}{7}$</td>
<td>babab</td>
</tr>
<tr>
<td>$\frac{1}{2}$</td>
<td>b</td>
<td>bababab</td>
<td>$\frac{2}{3}$</td>
<td>abab</td>
</tr>
<tr>
<td>$\frac{2}{3}$</td>
<td>bab</td>
<td>babab</td>
<td>$\frac{3}{4}$</td>
<td>babab</td>
</tr>
<tr>
<td>$\frac{1}{4}$</td>
<td>bbabb</td>
<td>bbabbbabb</td>
<td>$\frac{4}{7}$</td>
<td>bbbab</td>
</tr>
<tr>
<td>$\frac{1}{1}$</td>
<td>e</td>
<td>bbbbb</td>
<td>$\frac{1}{2}$</td>
<td>bbbbb</td>
</tr>
</tbody>
</table>

the third column the values of $g_6$, in the fourth column their Farey numbers. In the fifth column the values of $\Phi^*_6$.

7. Farey numbers and standard words

We have seen in the previous sections that there exists a bijection $\psi: \mathcal{A}^* \rightarrow SBS$ and a surjection $\eta: SBS \rightarrow \mathcal{F}$. This latter becomes a bijection if $\eta$ is restricted to $\{e\} \cup SBS_{(a)}$ (or to $\{e\} \cup SBS_{(b)}$). Hence any strictly bispecial element of $St$ can be 'codified', up to the automorphism ($^*$), by a Farey number. Moreover, any binary word faithfully represents a strictly bispecial element. If we consider the restriction $\psi_a$ of $\psi$ to $a\mathcal{A}^* \cup \{e\}$, then we obtain a bijection $\psi_a : a\mathcal{A}^* \cup \{e\} \rightarrow SBS_{(a)} \cup \{e\}$. Hence the composition

$$\zeta = \psi_a \circ \eta_a,$$

is a bijection $\zeta : a\mathcal{A}^* \cup \{e\} \rightarrow \mathcal{F}$. We shall see in this section some remarkable consequences of the previous correspondences.

**Proposition 13.** The set $\mathcal{F}$ of Farey numbers is the smallest subset $Y$ of the set $\mathbb{Q}$ of rational numbers, which contains $\frac{1}{1} \text{ and such that}$

$$\frac{p}{q} \in Y \Rightarrow \frac{p}{p+q}, \frac{q}{p+q} \in Y.$$

**Proof.** It is trivial by Lemma 10 that $\mathcal{F} \supseteq Y$. The proof that $\mathcal{F} \subseteq Y$ is then obtained by induction on the length $n = |\zeta^{-1}(f)|$ with $f \in \mathcal{F}$. If $n = 0$ the result is trivial since $\psi(e) = \varepsilon$ and $\eta(e) = \varepsilon$. Let us then suppose that the assertion is true up to $n - 1$ and consider a word $w \in a\mathcal{A}^*$ of length $n$ and such that $|\psi(w)| = f \in \mathcal{F}$. We can write
A. De Luca / Theoretical Computer Science 183 (1997) 45–82 79

Fig. 1.

$w = ux$, $x \in \mathcal{A}$. Thus $|u| = n - 1$ so that by the inductive hypothesis $|\psi(u)| = p/q \in Y$. Now $\psi(ux) = (x\psi(u))^{(-1)}$. By Lemma 10 it follows that $f = |\psi(w)|$ is either $p/(p+q)$ or $q/(p+q)$. But these fractions belong to $Y$ because of the induction hypothesis. □

The words of the set $\{\varepsilon\} \cup a.\mathcal{A}^*$ can be represented by the vertices of a binary tree, where words are ordered lexicographically. Words of smaller length are to the left of words of greater length. The root represents the empty word $\varepsilon$. The subtree having the root $a$ is a complete binary tree. The edges represent the ‘covering’ relation relative to the prefixial ordering, i.e. there is an edge from $u$ to $v$ if and only if there is a letter $a \in \mathcal{A}$ such that $v = ua$. In view of the previous bijections one can associate with each vertex a strictly bispecial element of $\mathcal{S}t$ and a Farey number (see Fig. 1).

If a vertex denotes a word $w$, then its corresponding Farey number is $p/q = |\psi(w)|$. If $w_1$ and $w_2$ are ‘son’ vertices of $w$, then their Farey numbers will be $p/(p+q)$ and $q/(p+q)$. If $s = \psi(w)$, $s_1 = \psi(w_1)$, $s_2 = \psi(w_2)$, then $s_1$ and $s_2$ are obtained from $s$ by the left-palindrome closures of $as$ and $bs$. We call the above tree the Farey tree. Let us consider a set of vertices on the Farey tree representing a prefix code $X \subseteq \mathcal{A}^*$. As we have seen in Section 5 the set $\psi(X)$ of the corresponding strictly bispecial elements is a biprefix code. This result is easily interpreted by the Farey tree. In fact, any element $y_1 \in \psi(X)$ cannot be derived from any other element $y_2 \in \psi(X)$, $y_2 \neq y_1$, by left palindrome closures so that $\psi(X)$ is a prefix code and then a biprefix code.

For a finite sequence $(a_1, \ldots, a_n)$ of integers such that $a_i > 0$, $1 \leq i < n$ and $a_n \geq 0$, we set $(a_1, \ldots, a_n)$ equal to the continued fraction $[0, a_1, \ldots, a_{n-1}, a_n + 1]$, i.e.:

\[
(a_1, \ldots, a_n) = \frac{1}{\frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\ddots \frac{1}{a_{n-1} + \frac{1}{a_n + 1}}}}}}
\]
Theorem 8. Let $w$ be a word on the alphabet $\mathcal{A} = \{a,b\}$ and $\psi(w)$ be the corresponding strictly bispecial element. If $(h_1, \ldots, h_n)$ is the integral representation of $w$, then the Farey number $\|\psi(w)\|$ has a development in continued fractions given by $(h_n, \ldots, h_1)$.

Proof. The proof is by induction on the length $n$ of the integral representation $(h_1, \ldots, h_n)$ of $w$. Let us first suppose that $h_1 > 0$.

Base of the induction. In the case $n = 1$ the word $w = a^{h_1}$ and $\psi(a^{h_1}) = a^{h_1}$ whose Farey number is $1/(h_1 + 1)$. In this case the result is trivial. Suppose $n = 2$ so that $w = a^{h_1}b^{h_2}$. One has $\psi(a^{h_1}b^{h_2}) = (a^{h_1}b^{h_2})^{h_1}a^{h_1}$. Now

$$\|\psi(a^{h_1}b^{h_2})\| = \frac{h_1 + 1}{h_2(h_1 + 1) + 1} = \frac{1}{h_2 + \frac{1}{h_1 + 1}}.$$

Thus the base of the induction is proved.

Induction step. Suppose that we have proved our assertion up to $n - 1$. We shall write

$$w = w'x^{h_n},$$

where $(h_1, \ldots, h_{n-1})$ is the integral representation of $w'$. Thus by the inductive hypothesis $\|\psi(w')\| = (h_{n-1}, \ldots, h_1)$. Let $P'xyQ'$ with $P', Q' \in PAL$, $|P'| < |Q'|$, $x, y \in \mathcal{A}$, $x \neq y$, be the canonical representation of $\psi(w)$.

We first prove that if $n$ is even, then the intermediate word $xy = ab$ and if $n$ is an odd integer $> 1$, then $xy = ba$. Indeed, for $n = 2$, $\psi(w) = (a^{h_1}b^{h_2})^{h_1}a^{h_1}$, so that $P' = a^{h_1-1}$, $Q' = (a^{h_1}b^{h_2-1})a^{h_1}$ and $xy = ab$. Suppose the assertion is true up to $n - 1$. We want to prove it for $n$. Since $w = w'x^{h_n}$ if $n$ is odd then $x = a$ and by induction $\psi(w') = PabQ$ with $P, Q \in PAL$, $|P| < |Q|$. On the contrary, if $n$ is even then $x = b$ and by induction $\psi(w') = PbaQ$. By Proposition 9, one has that in the first case the intermediate word of $\psi(w)$ is $ba$ and in the second case is $ab$.

Let $n$ be any integer greater than 2. By the above result one has that the intermediate word of $\psi(w')$ is $xy$ so that $\psi(w') = PxyQ$. By Proposition 9 one has that

$$\psi(w) = \psi(w'x^{h_n}) = (Qyx)^{h_n}PxyQ,$$

so that setting $p = |P| + 2$, $q = |Q| + 2$ one derives:

$$\|\psi(w)\| = \frac{q}{p + h_nq} = \frac{1}{h_n + \frac{p}{q}}.$$

By the inductive hypothesis $p/q = \|\psi(w')\| = (h_{n-1}, \ldots, h_1)$ so that $\|\psi(w)\| = (h_n, \ldots, h_1)$.

Let us now suppose $h_1 = 0$. In this case the word $\tilde{w}$ has the integral representation $(h_2, \ldots, h_n)$, so that by the above result $\|\psi(\tilde{w})\| = (h_n, \ldots, h_2)$. Since $(h_n, \ldots, h_2) = (h_n, \ldots, h_2, 0)$ and $\|\psi(w)\| = \|\psi(\tilde{w})\|$, the result follows. □
Let $FC$ be the set of all continued fractions $\langle h_1, \ldots, h_n \rangle$ with $n > 0$ and $h_1 > 0$, $i \in [1, n]$. We can introduce in $FC$ a product operation $\circ$ defined as follows. If $x = \langle h_1, \ldots, h_n \rangle$ and $y = \langle k_1, \ldots, k_m \rangle$ are in $FC$, then:

$$x \circ y = \begin{cases} 
\langle k_1, \ldots, k_m, h_1, \ldots, h_n \rangle & \text{if } n \text{ is even}, \\
\langle k_1, \ldots, k_{m-1}, k_m + h_1, h_2, \ldots, h_n \rangle & \text{if } n \text{ is odd}.
\end{cases}$$

One can easily verify that the above product is associative so that $FC$ is a semigroup. Let $FC^1$ be the monoid obtained from $FC$ by adding to $FC$ an identity element $1$. From the theory of continued fractions one has that any Farey number can be faithfully represented by one element of $FC^1$. We shall denote by $\sigma: \mathcal{F} \to FC^1$ this bijection (to the fraction $\frac{1}{n}$ corresponds the identity of $FC^1$). Thus the monoid operation in $FC^1$ can be naturally transmitted to $\mathcal{F}$ by defining for $x, y \in \mathcal{F}$ the product $x \circ y$, as

$$x \circ y = \sigma^{-1}(\sigma(x) \circ \sigma(y)).$$

Hence $\mathcal{F}$ is a monoid that we call the Farey monoid.

**Proposition 14.** The map $\zeta: a \mathcal{A}^* \cup \{e\} \to \mathcal{F}$ is a monoid isomorphism.

**Proof.** Let $w_1, w_2 \in a \mathcal{A}^* \cup \{e\}$. We want to prove that $\zeta(w_1 w_2) = \zeta(w_1) \circ \zeta(w_2)$. The result is trivial if $w_1$ or $w_2$ is the empty word. Let us then suppose $w_1, w_2 \in a \mathcal{A}^*$. Let us denote, respectively, by $(h_1, \ldots, h_n)$ and $(k_1, \ldots, k_m)$ the integral representations of $w_1$ and $w_2$. By Theorem 8,

$$||\psi(w_1)|| = \langle h_n, \ldots, h_1 \rangle, \quad ||\psi(w_2)|| = \langle k_m, \ldots, k_1 \rangle,$$

where we identify a continued fraction with its value. The word $w_1 w_2$ has the integral representation $(h_1, \ldots, h_n, k_1, \ldots, k_m)$ if $n$ is even and $(h_1, \ldots, h_{n-1}, h_n + k_1, k_2, \ldots, k_m)$ if $n$ is odd. Thus if $n$ is even

$$\zeta(w_1 w_2) = ||\psi(w_1 w_2)|| = \langle k_m, \ldots, k_1, h_n, \ldots, h_1 \rangle = \langle h_n, \ldots, h_1 \rangle \circ \langle k_m, \ldots, k_1 \rangle = ||\psi(w_1)|| \circ ||\psi(w_2)|| = \zeta(w_1) \circ \zeta(w_2).$$

If $n$ is odd, then

$$\zeta(w_1 w_2) = ||\psi(w_1 w_2)|| = \langle k_m, \ldots, k_2, k_1 + h_n, h_{n+1}, \ldots, h_1 \rangle = \langle h_n, \ldots, h_1 \rangle \circ \langle k_m, \ldots, k_1 \rangle = ||\psi(w_1)|| \circ ||\psi(w_2)|| = \zeta(w_1) \circ \zeta(w_2).$$

Since $\zeta$ is a bijection, the result follows. \(\square\)

We have seen in Section 5 that any infinite standard Sturmian word $s$ is the limit of a sequence $\{s_n\}_{n \geq 0}$ of words defined for any $x \in \mathcal{A}_0^\omega$ as

$$s_0 = e, \quad s_{n+1} = (x_n s_n)^{-1}, \quad n \geq 0.$$
Let us remark that if $s_1 = x_1 = x \in \{a, b\}$, then for all $n > 0$, $s_n \in SBS(x)$. By Lemma 9 the above sequence is completely determined, up to the automorphism ($\,\,\,$), by the infinite sequence of Farey numbers:

$$f_0, f_1, \ldots, f_n, \ldots,$$

where for any $n \geq 0$, $f_n = \|s_n\|$. Let us set for any $n \geq 0$, $f_n = p_n/q_n$, with $p_n \leq q_n$ and $gcd(p_n, q_n) = 1$. One has then $f_0 = \frac{1}{1}$ and by Lemma 10, for any $n$,

$$f_{n+1} = p_n/(p_n + q_n) \quad \text{or} \quad f_{n+1} = q_n/(p_n + q_n).$$

In this way we obtain a complete arithmetical description of standard words.

References