Open and Closed Prefixes of Sturmian Words

Alessandro De Luca[1] Gabriele Fici[2]

[1]Università di Napoli Federico II, Italy

[2]Università di Palermo, Italy

WORDS

Turku, 16–20 September 2013

Open and Closed Words

Definition

A word is closed if it is empty or it contains a factor occurring only as a prefix and as a suffix (with no internal occurrences).
Otherwise, it is open.

Open and Closed Words

Definition

A word is closed if it is empty or it contains a factor occurring only as a prefix and as a suffix (with no internal occurrences).
Otherwise, it is open.

Example

$abab$, $abaabaab$, and a are closed words.

Open and Closed Words

Definition

A word is closed if it is empty or it contains a factor occurring only as a prefix and as a suffix (with no internal occurrences).
Otherwise, it is open.

Example

$abab$, $abaabaab$, and a are closed words.

ab, $abbaba$ and $ababc$ are open words.

Open and Closed Words

Definition

A word is closed if it is empty or it contains a factor occurring only as a prefix and as a suffix (with no internal occurrences).
Otherwise, it is open.

Example

abab, *abaabaab*, and *a* are closed words.

ab, *abbaba* and *ababc* are open words.

Remark

Closed words are also known as periodic-like words, or complete (first) returns.

Open and Closed Words

Let w be a non-empty word. The following are equivalent definitions of closed word:

Open and Closed Words

Let w be a non-empty word. The following are equivalent definitions of closed word:

1. the longest repeated prefix (resp. suffix) of w does not have internal occurrences in w;

Open and Closed Words

Let w be a non-empty word. The following are equivalent definitions of closed word:

1. the longest repeated prefix (resp. suffix) of w does not have internal occurrences in w;
2. the longest repeated prefix (resp. suffix) of w is not right (resp. left) special in w;

Open and Closed Words

Let w be a non-empty word. The following are equivalent definitions of closed word:

1. the longest repeated prefix (resp. suffix) of w does not have internal occurrences in w;
2. the longest repeated prefix (resp. suffix) of w is not right (resp. left) special in w;
3. the longest border of w does not have internal occurrences in w;

Open and Closed Words

Let w be a non-empty word. The following are equivalent definitions of closed word:

1. the longest repeated prefix (resp. suffix) of w does not have internal occurrences in w;
2. the longest repeated prefix (resp. suffix) of w is not right (resp. left) special in w;
3. the longest border of w does not have internal occurrences in w;

Of course, the negations are equivalent definitions of open words.

Open and Closed Words

Remark

Every word that is a power is closed (if $w = z^n$, $n \geq 2$, then z^{n-1} appears only as a prefix and as a suffix in w).

Open and Closed Words

Remark

Every word that is a power is closed (if $w = z^n$, $n \geq 2$, then z^{n-1} appears only as a prefix and as a suffix in w).

Remark

Every binary palindrome of length < 14 is closed. A shortest open binary palindrome is aabbabaababbaa.

Open and Closed Words

Remark

Every word that is a power is closed (if $w = z^n$, $n \geq 2$, then z^{n-1} appears only as a prefix and as a suffix in w).

Remark

Every binary palindrome of length < 14 is closed. A shortest open binary palindrome is aabbabaababbaa.

Actually, for palindromes there is a relation between openness/closedness and richness. Recall that a word of length n is rich (or full) if it contains n non-empty palindromes, that is the maximum possible.

Open and Closed Words

Remark

Every word that is a power is closed (if $w = z^n$, $n \geq 2$, then z^{n-1} appears only as a prefix and as a suffix in w).

Remark

Every binary palindrome of length < 14 is closed. A shortest open binary palindrome is aabbabaababbaa.

Actually, for palindromes there is a relation between openness/closedness and richness. Recall that a word of length n is rich (or full) if it contains n non-empty palindromes, that is the maximum possible.

Theorem (Bucci, de Luca, De Luca, 2009)

A palindrome is rich if and only if all its palindromic factors are closed.

Open and Closed Prefixes

We can separate the factors of a word into open and closed. We focus here on the prefixes.

Open and Closed Prefixes

We can separate the factors of a word into open and closed. We focus here on the prefixes.

Definition

Let w be a finite or infinite word. We define the sequence $oc(w)$ as the sequence whose n-th element is:

- 1 if the prefix of length n of w is closed;
- 0 if it is open.

Open and Closed Prefixes

We can separate the factors of a word into open and closed. We focus here on the prefixes.

Definition

Let w be a finite or infinite word. We define the sequence $oc(w)$ as the sequence whose n-th element is:

- 1 if the prefix of length n of w is closed;
- 0 if it is open.

For example, if $w = abaabab$, then $oc(w) = 1010110$.

Open and Closed Prefixes

Theorem

The sequence oc characterizes every (finite or infinite) Sturmian word, up to exchanging letters.

Open and Closed Prefixes

Theorem

The sequence oc characterizes every (finite or infinite) Sturmian word, up to exchanging letters.

That is, if w_1 and w_2 are Sturmian words, then $oc(w_1) = oc(w_2)$ if and only if $w_1 = w_2$ (up to isomorphisms).

Open and Closed Prefixes

Theorem

The sequence oc characterizes every (finite or infinite) Sturmian word, up to exchanging letters.

That is, if w_1 and w_2 are Sturmian words, then $oc(w_1) = oc(w_2)$ if and only if $w_1 = w_2$ (up to isomorphisms).

The hypothesis that w_1 and w_2 be Sturmian words is necessary. For example, $aaba$ and $aabb$ have the same sequence oc: 1100.

Open and Closed Prefixes

Example

Let $F = abaababaabaababaababaabaab\cdots$ be the Fibonacci word. Then:
$oc(F) = 10101100111000111100000\cdots$

Open and Closed Prefixes

Example

Let $F = abaababaabaababaababaabaab \cdots$ be the Fibonacci word. Then:
$oc(F) = 10101100111000111100000 \cdots$

Remarks:

1. the runs of 1's are followed by equal-length runs of 0's;
2. the lengths of the runs form the Fibonacci sequence.

Open and Closed Prefixes

Example

Let $F = abaababaabaababaababaabaab\cdots$ be the Fibonacci word. Then: $oc(F) = 10101100111000111100000\cdots$

Remarks:

1. the runs of 1's are followed by equal-length runs of 0's;
2. the lengths of the runs form the Fibonacci sequence.

Theorem (Bucci, De Luca, Fici, 2013)

Let w be a prefix of the Fibonacci word F. Then w is open if and only if there exists i such that $F_{i+1} - 1 \leq |w| \leq 2F_i - 2$.

Open and Closed Prefixes

Example

Let $F = abaababaabaababaababaabaab \cdots$ be the Fibonacci word. Then: $oc(F) = 101011001110001111100000 \cdots$

Remarks:

1. the runs of 1's are followed by equal-length runs of 0's;
2. the lengths of the runs form the Fibonacci sequence.

Theorem (Bucci, De Luca, Fici, 2013)

Let w be a prefix of the Fibonacci word F. Then w is open if and only if there exists i such that $F_{i+1} - 1 \leq |w| \leq 2F_i - 2$.

Question: What can we say about other standard Sturmian words?

Standard Sturmian Words

Definition

Let α be an irrational number such that $0 < \alpha < 1$, and $[0; d_0 + 1, d_1, \ldots]$ its continued fraction expansion. The sequence of words defined by:

$$s_{-1} = b,\ s_0 = a \text{ and } s_{n+1} = s_n^{d_n} s_{n-1} \text{ for } n \geq 0$$

converges to the infinite standard Sturmian word w_α.

Standard Sturmian Words

Definition

Let α be an irrational number such that $0 < \alpha < 1$, and $[0; d_0 + 1, d_1, \ldots]$ its continued fraction expansion. The sequence of words defined by:

$$s_{-1} = b,\ s_0 = a \text{ and } s_{n+1} = s_n^{d_n} s_{n-1} \text{ for } n \geq 0$$

converges to the infinite standard Sturmian word w_α.

1. The sequence of words s_n is called the standard sequence of w_α.

Standard Sturmian Words

Definition

Let α be an irrational number such that $0 < \alpha < 1$, and $[0; d_0 + 1, d_1, \ldots]$ its continued fraction expansion. The sequence of words defined by:

$$s_{-1} = b,\ s_0 = a \text{ and } s_{n+1} = s_n^{d_n} s_{n-1} \text{ for } n \geq 0$$

converges to the infinite standard Sturmian word w_α.

1. The sequence of words s_n is called the standard sequence of w_α.
2. A standard word is a finite word belonging to some standard sequence.

Standard Sturmian Words

Definition

Let α be an irrational number such that $0 < \alpha < 1$, and $[0; d_0 + 1, d_1, \ldots]$ its continued fraction expansion. The sequence of words defined by:

$$s_{-1} = b,\ s_0 = a \text{ and } s_{n+1} = s_n^{d_n} s_{n-1} \text{ for } n \geq 0$$

converges to the infinite standard Sturmian word w_α.

1. The sequence of words s_n is called the standard sequence of w_α.
2. A standard word is a finite word belonging to some standard sequence.
3. A central word is a word u such that uxy is a standard word, for letters $x, y \in \Sigma$.

Standard Sturmian Words

Example

The Fibonacci word F is the standard Sturmian word of slope

$$\alpha = (3 - \sqrt{5})/2 = [0; 2, 1, 1, 1, \dots]$$

so that $d_n = 1$ for every $n \geq 0$. Therefore, the standard sequence of F is:

$$f_{-1} = b,\ f_0 = a,\ f_{n+1} = f_n f_{n-1} \text{ for } n \geq 0.$$

This sequence is also called the sequence of Fibonacci finite words.

Central Words

Theorem (Aldo de Luca, 1997)

A word is central if and only if w is the power of a single letter or there exist palindromes w_1, w_2 such that $w = w_1xyw_2 = w_2yxw_1$, for different letters x, y.
Moreover, if $|w_1| < |w_2|$, then w_2 is the longest palindromic suffix of w.

Central Words

Theorem (Aldo de Luca, 1997)

A word is central if and only if w is the power of a single letter or there exist palindromes w_1, w_2 such that $w = w_1xyw_2 = w_2yxw_1$, for different letters x, y.
Moreover, if $|w_1| < |w_2|$, then w_2 is the longest palindromic suffix of w.

1. Every central word is closed.

Central Words

Theorem (Aldo de Luca, 1997)

A word is central if and only if w is the power of a single letter or there exist palindromes w_1, w_2 such that $w = w_1 xy w_2 = w_2 yx w_1$, for different letters x, y.
Moreover, if $|w_1| < |w_2|$, then w_2 is the longest palindromic suffix of w.

1. Every central word is closed.
2. Every central word is a palindrome.

Semicentral Words

Definition

A semicentral word is a word in which the longest repeated prefix, repeated suffix, left special factor and right special factor all coincide.

Definition

A semicentral word is a word in which the longest repeated prefix, repeated suffix, left special factor and right special factor all coincide.

Theorem (Bucci, De Luca, Fici, 2013)

A word w is semicentral if and only if $w = w_1 xy w_1$ for a central word w_1 and different letters x, y.

Semicentral Words

Definition

A semicentral word is a word in which the longest repeated prefix, repeated suffix, left special factor and right special factor all coincide.

Theorem (Bucci, De Luca, Fici, 2013)

A word w is semicentral if and only if $w = w_1 x y w_1$ for a central word w_1 and different letters x, y.

In a semicentral word $w_1 x y w_1$, the central word w_1 is the longest repeated prefix, repeated suffix, left special factor and right special factor.

Semicentral Words

Definition

A semicentral word is a word in which the longest repeated prefix, repeated suffix, left special factor and right special factor all coincide.

Theorem (Bucci, De Luca, Fici, 2013)

A word w is semicentral if and only if $w = w_1xyw_1$ for a central word w_1 and different letters x, y.

In a semicentral word w_1xyw_1, the central word w_1 is the longest repeated prefix, repeated suffix, left special factor and right special factor.

Semicentral words are therefore open words.

For every $n \geq -1$, one has

$$s_n = u_n xy, \tag{1}$$

for x, y letters such that $xy = ab$ if n is odd or ba if n is even. Indeed, the sequence $(u_n)_{n \geq -1}$ can be defined by: $u_{-1} = a^{-1}$, $u_0 = b^{-1}$, and, for every $n \geq 1$,

$$u_{n+1} = (u_n xy)^{d_n} u_{n-1}\,, \tag{2}$$

where x, y are as in (1).

For every $n \geq -1$, one has

$$s_n = u_n xy, \tag{1}$$

for x, y letters such that $xy = ab$ if n is odd or ba if n is even. Indeed, the sequence $(u_n)_{n \geq -1}$ can be defined by: $u_{-1} = a^{-1}$, $u_0 = b^{-1}$, and, for every $n \geq 1$,

$$u_{n+1} = (u_n xy)^{d_n} u_{n-1}\,, \tag{2}$$

where x, y are as in (1).

Theorem

Let $(u_n xy)$ be the standard sequence of a standard Sturmian word w. Let vx, $x \in \{a, b\}$, be a prefix of w. Then:

1. *v is closed and vx is open if and only if there exists $n \geq 0$ such that $v = u_n xy u_{n+1} = u_{n+1} yx u_n$ (central prefixes);*
2. *v is open and vx is closed if and only if there exists $n \geq 1$ such that $v = u_n xy u_n$ (semicentral prefixes).*

Standard Sturmian Words

prefix of w	open/closed	example
$u_n xyu_n$	open	$aaba$
$u_n xyu_n x$	closed	$aabaa$
$u_n xyu_n xy$	closed	$aabaab$
...	...	...
$u_n xyu_{n+1} = u_{n+1} yxu_n$	closed	$aabaabaa$
$u_{n+1} yxu_n y$	open	$aabaabaaa$
$u_{n+1} yxu_n yx$	open	$aabaabaaab$
...	...	...
$u_{n+1} yxu_{n+1}$	open	$aabaabaaabaa$
$u_{n+1} yxu_{n+1} y$	closed	$aabaabaaabaab$

Table : The structure of the prefixes of a standard Sturmian word $w = aabaabaaabaabaa \cdots$ with respect to the u_n prefixes. Here $d_0 = d_1 = 2$ and $d_2 = 1$.

Standard Sturmian Words

prefix of w	open/closed	example
$u_n x y u_n$	open	*aaba*
$u_n x y u_n x$	closed	*aabaa*
$u_n x y u_n x y$	closed	*aabaab*
$\dots$	$\dots$	$\dots$
$u_n x y u_{n+1} = u_{n+1} y x u_n$	closed	*aabaabaa*
$u_{n+1} y x u_n y$	open	*aabaabaaa*
$u_{n+1} y x u_n y x$	open	*aabaabaaab*
$\dots$	$\dots$	$\dots$
$u_{n+1} y x u_{n+1}$	open	*aabaabaaabaa*
$u_{n+1} y x u_{n+1} y$	closed	*aabaabaaabaab*

Table : The structure of the prefixes of a standard Sturmian word $w = aabaabaaabaabaa \cdots$ with respect to the u_n prefixes. Here $d_0 = d_1 = 2$ and $d_2 = 1$.

Standard Sturmian Words

Theorem

Let $w = w_\alpha$ *be a standard Sturmian word, with* $0 < \alpha < 1/2$, *and let* $\alpha = [0; d_0 + 1, d_1, \ldots]$. *The word* $ba^{-1}w$, *obtained from* w *by swapping the first letter, can be written as an infinite product of squares of reversed standard words in the following way:*

$$ba^{-1}w = \prod_{n\geq 0}(u_n^{-1}u_{n+1})^2,$$

where $(u_n xy)_{n\geq -1}$ *is the standard sequence of* w.
In other words, one can write

$$w = a^{d_0}ba^{d_0-1}\prod_{n\geq 1}(u_n^{-1}u_{n+1})^2.$$

Example

Take the Fibonacci word F. Then,

$$u_1 = \varepsilon, \quad u_2 = a, \quad u_3 = aba, \quad u_4 = abaaba, \quad u_5 = abaababaaba, \ldots$$

So,

$$u_1^{-1}u_2 = a, \quad u_2^{-1}u_3 = ba, \quad u_3^{-1}u_4 = aba, \quad u_4^{-1}u_5 = baaba, \ldots$$

Example

Take the Fibonacci word F. Then,

$$u_1 = \varepsilon, \quad u_2 = a, \quad u_3 = aba, \quad u_4 = abaaba, \quad u_5 = abaababaaba, \dots$$

So,

$$u_1^{-1}u_2 = a, \quad u_2^{-1}u_3 = ba, \quad u_3^{-1}u_4 = aba, \quad u_4^{-1}u_5 = baaba, \dots$$

Indeed, $u_n^{-1}u_{n+1}$ is the reversal of the Fibonacci finite word f_{n-1}.

Example

Take the Fibonacci word F. Then,

$$u_1 = \varepsilon, \quad u_2 = a, \quad u_3 = aba, \quad u_4 = abaaba, \quad u_5 = abaababaaba, \ldots$$

So,

$$u_1^{-1}u_2 = a, \quad u_2^{-1}u_3 = ba, \quad u_3^{-1}u_4 = aba, \quad u_4^{-1}u_5 = baaba, \ldots$$

Indeed, $u_n^{-1}u_{n+1}$ is the reversal of the Fibonacci finite word f_{n-1}.
By the previous theorem, we have:

$$\begin{aligned} F &= ab\prod_{n\geq 1}(u_n^{-1}u_{n+1})^2 \\ &= ab\prod_{n\geq 0}(\widetilde{f_n})^2 \\ &= ab\cdot(a\cdot a)(ba\cdot ba)(aba\cdot aba)(baaba\cdot baaba)\cdots \end{aligned}$$

i.e., F can be obtained by concatenating ab and the squares of the reversals of the Fibonacci finite words f_n starting from $n = 0$.

Standard Sturmian Words

Example (continued)

Note that F is also obtained by concatenating the reversals of the Fibonacci finite words f_n starting from $n = 0$:

$$\begin{aligned} F &= \prod_{n \geq 0} \widetilde{f_n} \\ &= a \cdot ba \cdot aba \cdot baaba \cdot ababaaba \cdots \end{aligned}$$

Example (continued)

Note that F is also obtained by concatenating the reversals of the Fibonacci finite words f_n starting from $n = 0$:

$$\begin{aligned} F &= \prod_{n \geq 0} \widetilde{f_n} \\ &= a \cdot ba \cdot aba \cdot baaba \cdot ababaaba \cdots \end{aligned}$$

and also by concatenating ab and the words f_n starting from $n = 0$:

$$\begin{aligned} F &= ab \prod_{n \geq 0} f_n \\ &= ab \cdot a \cdot ab \cdot aba \cdot abaab \cdot abaababa \cdots \end{aligned}$$

Example (continued)

Note that F is also obtained by concatenating the reversals of the Fibonacci finite words f_n starting from $n = 0$:

$$\begin{aligned} F &= \prod_{n\geq 0} \widetilde{f_n} \\ &= a \cdot ba \cdot aba \cdot baaba \cdot ababaaba \cdots \end{aligned}$$

and also by concatenating ab and the words f_n starting from $n = 0$:

$$\begin{aligned} F &= ab \prod_{n\geq 0} f_n \\ &= ab \cdot a \cdot ab \cdot aba \cdot abaab \cdot abaababa \cdots \end{aligned}$$

So we have:

$$F = ab \prod_{n\geq 0} (\widetilde{f_n})^2 = \prod_{n\geq 0} \widetilde{f_n} = ab \prod_{n\geq 0} f_n.$$

Standard Sturmian Words

Question: Is it possible to characterize the sequence of open and closed prefixes of a standard Sturmian word w_α in terms of the continued fraction expansion of α?

Standard Sturmian Words

Question: Is it possible to characterize the sequence of open and closed prefixes of a standard Sturmian word w_α in terms of the continued fraction expansion of α?

Definition

The *continuants* of an integer sequence $(a_n)_{n\geq 0}$ are defined as $K[\ ] = 1$, $K[a_0] = a_0$, and, for every $n \geq 1$,

$$K[a_0, \ldots, a_n] = a_n K[a_0, \ldots, a_{n-1}] + K[a_0, \ldots, a_{n-2}].$$

Standard Sturmian Words

Question: Is it possible to characterize the sequence of open and closed prefixes of a standard Sturmian word w_α in terms of the continued fraction expansion of α?

Definition

The *continuants* of an integer sequence $(a_n)_{n\geq 0}$ are defined as $K[\ ] = 1$, $K[a_0] = a_0$, and, for every $n \geq 1$,

$$K[a_0, \ldots, a_n] = a_n K[a_0, \ldots, a_{n-1}] + K[a_0, \ldots, a_{n-2}].$$

Let w be a standard Sturmian word and $(s_n)_{n\geq -1}$ its standard sequence. Then:

$$|s_n| = K[1, d_0, \ldots, d_{n-1}].$$

From the previous theorem, we have:

Corollary

Let $w = w_\alpha$ be a standard Sturmian word, with $0 < \alpha < 1/2$, and $\alpha = [0; d_0 + 1, d_1, \ldots]$.

Let, for every $n \geq 0$, $k_n = K\,[1, d_0, \ldots, d_{n-1}, d_n - 1]$. Then

$$oc(w) = \prod_{n \geq 0} 1^{k_n} 0^{k_n}.$$

Thank You