Another characterization of Sturmian words
(one more)

Gwénaël Richomme
LaRIA, Université de Picardie Jules Verne
5 rue du Moulin Neuf, 80000 Amiens, France
e-mail : richomme@laria.u-picardie.fr

Abstract

Sturmian words (balanced, non ultimately periodic, infinite words) have been widely studied since the works of Morse and Hedlund [5]. Many equivalent definitions have been established. Here, we give another one related to the notion of n-chains of Morse and Hedlund.

We consider here the alphabet $\Sigma = \{a, b\}$. Given a word w, i.e. an element of the free monoid Σ^*, and given a letter x in Σ, we denote $|w|_x$ the number of occurrences of x in w.

For a subset X of Σ^*, we denote by $Card(X)$ the number of elements in X and by X^ω the set of (right) infinite words obtained by concatenation of words of X.

Given an infinite word w, we denote by $F(w)$ the set of (finite) factors of w, and by $F_n(w)$ the set of factors of w of length n. An infinite word is *ultimately periodic* if there exist two finite words u and v such that $w = uv^\omega$.

A word w over Σ is *balanced*, if for any factors u and v of w, $|u| = |v|$ implies $||u|_a - |v|_a| \leq 1$. Observe that if w is balanced, than there exists an integer $k \geq 0$ such that $F(w) \cap ab^*a$ is included in $\{ab^ka, ab^{k+1}a\}$. Moreover, if w is an infinite non ultimately periodic word, this inclusion is an equality.

Theorem 1 [2] *Let w be an infinite word. Are equivalent:*

(i) $\forall n \in \mathbb{N}$, $Card(F_n(w)) = n + 1$,

(ii) *w is balanced and non ultimately periodic.*

If an infinite word verify one assertion of the previous proposition, it is called *Sturmian*. There are many other characterizations of Sturmian words (see [3] for a recent survey).

A good tool to study Sturmian words are *Sturmian morphisms* namely those morphisms that preserve Sturmian words (the image of a Sturmian word by a Sturmian morphism is a Sturmian word). Given the morphisms E, φ, $\tilde{\varphi}$ defined by $E(a) = b$, $E(b) = a$, $\varphi(a) = ab$, $\varphi(b) = a$ and $\tilde{\varphi}(a) = ba$, $\tilde{\varphi}(b) = a$, we have:

Theorem 2 [1, 4] *The set of Sturmian morphisms is $\{E, \varphi, \tilde{\varphi}\}^*$.*

Let w be an infinite word over the alphabet Σ. We denote:

$$X_n(w) = \{x \in F(w) \cap a\Sigma^* \cap \Sigma^* a \mid |x|_a = n\}.$$

$$Y_n(w) = \{x \in F(w) \cap b\Sigma^* \cap \Sigma^* b \mid |x|_b = n\}.$$

When there is no confusion, we write X_n (resp. Y_n) instead of $X_n(w)$ (resp. $Y_n(w)$).

Here we prove:

Proposition 3 *An infinite word w is Sturmian iff $\forall n \geq 0$, $Card(X_n(w)) = Card(Y_n(w)) = n$.*

This result can be related with the works of Morse and Hedlund [5] who called n-chain any word beginning and ending with a, and containing exactly $n+1$ occurrences of a. For $n \geq 1$, the set X_n is the set of $(n-1)$-chains of w.

The following Lemma 4 is the "only if" part of Proposition 3. This can be deduced from results in [5], but in order to be self-contained, we give here a short proof of it.

Lemma 4 *If w is a Sturmian word than $\forall n \geq 0$, $Card(X_n) = Card(Y_n) = n$.*

Proof. Let w be a Sturmian word. We will prove that, for all $n \in \mathbb{N}$, $Card(X_n) = n$ (the proof for Y_n is similar exchanging a and b).

Since $X_0 = \emptyset$ and $X_1 = \{a\}$, the condition is verified for $n = 0$ and $n = 1$.

Let $n \geq 1$ be an integer such that $Card(X_n) = n$.

Since w is a non ultimately periodic infinite word, each word in X_n can be extended to the right in a word of X_{n+1}, and there exists at least one word in X_n that can be extended in at least two different words of X_{n+1}. Thus $Card(X_{n+1}) \geq n+1$.

Now assume that there exist two different words u and v in X_n that can be extended in at least two words in X_{n+1}. Since w is Sturmian, there exists an integer $k \geq 0$ such that $ab^*a \cap F(w) = \{ab^ka, ab^{k+1}a\}$. Then necessarily, the words $ub^ka, ub^{k+1}a, vb^ka, vb^{k+1}a$ are factors of w. Let z be the longest common suffix of u and v. The words azb^ka and bzb^{k+1} are factors of w: a contradiction with w balanced. This shows that $Card(X_{n+1}) = n+1$. $\square$

Remark that for an infinite word w to be Sturmian, the two conditions $Card(X_n) = n$ and $Card(Y_n) = n$, $n \in \mathbb{N}$, are needed. Indeed, consider the word $F = abaababa\ldots$ unique fix point of φ. This word is a Sturmian word (see [3] for instance) called the Fibonacci word. The word aaF is not Sturmian since it contains the two factors aaa and bab, but we have for all $n \geq 0$, $Card(Y_n(aaF)) = n$, because $Y_n(F) = Y_n(aaF)$.

The "if" part of Proposition 3, is given by the following:

Lemma 5 *Let w be an infinite word such that $\forall n \geq 0$, $Card(X_n(w)) = Card(Y_n(w)) = n$.*

There exist a Sturmian morphism f and a Sturmian word W such that $w = f(W)$. Consequently, w is Sturmian.

Before proving Lemma 5, we observe the following:

Remark 6 *If an infinite word w can be decomposed in $w = uv^\omega$ (i.e. w is ultimately periodic) with u and v two words, for all $l > |uv|_a$, $Card(X_l(w)) \leq |uv|_a < l$.*

Proof of Lemma 5. Let w be an infinite word and, for all $n \in \mathbb{N}$, $Card(X_n) = Card(Y_n) = n$.

We first prove that there exist integers $n_0 \geq 0$ and k, $0 \leq k \leq n_0 + 1$ such that

$$w \in b^k.\{ab^{n_0}, ab^{n_0+1}\}^\omega$$

Since $X_2 = 2$, there exist two different integers m and n such that $F(w) \cap ab^*a = \{ab^ma, ab^na\}$. Moreover since $Card(X_3) = 3$, exactly three of the four following words are factors of w: $x_1 = ab^nab^na$, $x_2 = ab^mab^ma$, $x_3 = ab^nab^ma$, $x_4 = ab^mab^na$.

Observe that x_3 and x_4 are necessarily in $F(w)$: for instance, if x_3 is not a factor of w, since ab^na is a factor of w, we have $w \in b^*a(b^ma)^*(b^na)^*((b^na)^\omega \cup b^\omega)$, and in particular w is ultimately periodic. From Remark 6, we have a contradiction.

Now assume that $n = m+2+p$ with $p \geq 0$. We have $x_3 = ab^pb^2b^mab^ma$, $x_4 = ab^mab^mb^2b^pa$. Moreover there exists an integer $j \geq 0$ such that x_3a^jb is a factor of w. Thus, if $m \neq 0$, the words $bab^ma^{1+j}b$, bb^mb, $bb^qab^{m-q}b$ for $0 \leq q \leq m$ are $m+3$ different words in Y_{m+2}: a contradiction with $Card(Y_{m+2}) = m+2$. If $m = 0$, since $Card(Y_2) = 2$ and since bb is a factor of w, there exists an integer $q > 1$ such that $Y_2 = \{bb, ba^qb\}$. Observe, moreover that a and b appear infinitely often in w (if w contains only i occurrences of a, $Card(X_{i+1}) = 0$). Thus, $w = u(b^na^q)^\omega$ i.e. is ultimately periodic: we have a contradiction from Remark 6.

In any case, $n = m+1$ or $m = n+1$, and thus integers n_0 and k exist such that $w \in b^k.\{ab^{n_0}, ab^{n_0+1}\}^\omega$.

As previously, we can see that $k \leq n_0 + 1$: otherwise, we can obtain contradictions with $Card(Y_{n_0+2}) = n_0 + 2$ when $n_0 \neq 0$ and with w non ultimately periodic when $n_0 = 0$.

Let us now end the proof of Lemma 5.

- If $k < n_0 + 1$, let f be the morphism defined by $f(a) = b^kab^{n_0-k}$ and $f(b) = b^kab^{n_0-k+1}$. From Theorem 2, the morphism f is Sturmian: $f = (Eo\tilde{\varphi})^{n_0-k}o(Eo\varphi)^k o\varphi oE$. Now, from what preceeds, $w \in \{b^kab^{n_0-k+1}, b^kab^{n_0-k}\}^\omega$. Thus, there exists an infinite word W such that $w = f(W)$.

  To a non empty factor U of W, we associate the word $u = b^{-k}f(U)b^ka$ (i.e. u is the word obtained from $f(U)b^ka$ by erasing the k first b of $f(U)$). By construction, $u \in X_{|U|+1}(w)$. Moreover, since f is injective, different words u, u' in $X_{n+1}(w)$ $(n \geq 0)$ are bijectively associated with different factors U, U' of W, and $|U| = |U'| = n$.

  Thus, for all $n \geq 0$, since $Card(X_{n+1}(w)) = n+1$, W has exactly $n+1$ factors of length n (i.e. $Card(F_n(W)) = n+1$). From Theorem 1, W is Sturmian and from the definition of Sturmian morphisms, w is Sturmian.

- If $k = n_0 + 1$, we obtain the result of Lemma 5 using the morphism f defined by $f(a) = b^{n_0}a$ and $f(b) = b^{n_0+1}a$ $(f = (Eo\varphi)^{n_0}o\tilde{\varphi}oE)$. □

Acknowledgement. The result presented here was conjectured during a discussion with J. Berstel. I also thank Patrice Séébold whose remarks greatly improved the paper.

References

[1] J. Berstel and P. Séébold, *A characterization of Sturmian morphisms*, in A. Borzyskowski, S. Sokolowski (Eds), MFCS'93, Lecture Notes in Computer Science 711, p281-290, 1993.

[2] E.M. Coven and G. Hedlund, *Sequences with minimal block growth*, Math. Systems theory 7, p376-382, 1974.

[3] Lothaire, *Algebraic combinatorics on words*, Cambridge University Press, to appear.

[4] F. Mignosi and P. Séébold, *Morphismes Sturmiens et Rgles de Rauzy*, J. de Thorie des Nombres de Bordeaux, vol 5, p221-233, 1993.

[5] M. Morse, G.A. Hedlund, *Symbolic Dynamics II: Sturmian trajectories*, Amer. J. Math. 62, p1-42, 1940.