JACQUES JUSTIN

On a paper by Castelli, Mignosi, Restivo

Informatique théorique et applications, tome 34, nº 5 (2000), p. 373-377

<http://www.numdam.org/item?id=ITA_2000__34_5_373_0>

Theoretical Informatics and Applications
Theoret. Informatics Appl. **34** (2000) 373–377

ON A PAPER BY CASTELLI, MIGNOSI, RESTIVO

JACQUES JUSTIN[1, 2]

Abstract. Fine and Wilf's theorem has recently been extended to words having three periods. Following the method of the authors we extend it to an arbitrary number of periods and deduce from that a characterization of generalized Arnoux–Rauzy sequences or episturmian infinite words.

Résumé. Le théorème de Fine et Wilf a été récemment étendu aux mots ayant trois périodes. Nous l'étendons, en suivant la méthode des auteurs, à un nombre quelconque de périodes et en déduisons une caractérisation des suites d'Arnoux–Rauzy généralisées ou mots infinis épisturmiens.

AMS Subject Classification. 68R15.

INTRODUCTION

The Fine and Wilf theorem [5] is a well-known remarkable result in Combinatorics on words [7]. It states that if a word w has periods p and q and has length $|w| \geq p + q - \gcd(p,q)$ then it also has period $\gcd(p,q)$. Extensions in several directions has been made. In particular Castelli *et al.* in [2] generalize this theorem to words having three periods and deduce an interesting characterization of Arnoux–Rauzy sequences [1].

Here we extend Fine and Wilf's theorem to words having $t \geq 2$ periods and obtain a characterization of generalized Arnoux–Rauzy sequences (or episturmian words in our terminology of [4,6]).

Keywords and phrases: Fine and Wilf theorem, Arnoux–Rauzy sequence, episturmian word.

[1] LIAFA, Université Paris VII, Case 7014, 2 place Jussieu, 75251 Paris Cedex 05, France; e-mail: `justin@liafa.jussieu.fr`

[2] *Mailing address*: 19 rue de Bagneux, 92330 Sceaux, France.

Apart from the induction in our Theorem 1.2 and a part of Section 2, we follow [2] to which we refer.

1. Fine and Wilf's theorem for t periods

We use customary terminology for words [7]. If $w = a_1 \cdots a_n$, $a_i \in A$ is a word over alphabet A with length $|w| = n$ then w has period p, p a non-negative integer, if $a_i = a_{p+i}$ for $i \in [1, n-p]$. With this definition, 0 and all $p \geq n$ are periods of w. Let $\mathrm{pref}_k(w)$ denote the prefix of w of length k. The following lemma is taken from [2].

Lemma 1.1. *Let $p < q$ be positive integers. If the word w has periods p and q then $\mathrm{pref}_{n-p}(w)$ has period $q-p$ (and obviously p).*

Now consider a t-uple of non-negative integers, not all 0, $\underline{p} = (p_1, p_2, \ldots, p_t)$. If $p_1 \leq p_2 \leq \cdots \leq p_t$ we say that $\underline{p}$ is an ordered t-uple. The operator R on ordered t-uples is defined by

$$R(\underline{p}) = (0, \cdots, 0, p_r, p_{r+1} - p_r, \cdots, p_t - p_r)$$

where $r = \min\{i \geq 1 \mid p_i \neq 0\}$.

The operator O acting on a t-uple gives the corresponding ordered t-uple. Given an ordered t-uple $\underline{p}$ with $p_t > 0$, we define $\underline{p}^{(0)} = \underline{p}$ and, for $0 < i < t$, $\underline{p}^{(i)} = OR(\underline{p}^{(i-1)})$. We also define $|\underline{p}| = p_1 + p_2 + \cdots + p_t$ and

$$m(i, \underline{p}) = \min \left\{ k \in [0, t-1] \mid p_1^{(k)} = \cdots = p_i^{(k)} = 0 \right\}$$

and $h(i, \underline{p}) = |\underline{p}^{(m(i,\underline{p}))}|$. Thus $h(t-1, \underline{p}) = \gcd(p_1, \ldots, p_t) = \gcd(\underline{p})$. Then:

Theorem 1.2. *Let a word w have periods $p_1 \leq p_2 \leq \cdots \leq p_t$, $p_i \geq 0$ for $0 < i < t$, $p_t > 0$. If $|w| \geq f(\underline{p})$ where $\underline{p} = (p_1, \ldots, p_t)$ and*

$$f(\underline{p}) = \frac{|\underline{p}|}{t-1} + \sum_{i=1}^{t-2} \frac{h(i, \underline{p})}{(t-i)(t-1-i)} - \gcd(\underline{p}) \tag{1}$$

then w has period $\gcd(\underline{p})$.

Proof. The theorem is true for $t = 2$ (this is Fine and Wilf's theorem). By induction assume it is true for $2, 3, \ldots, t' = t-1$. If it is false for t then let n be the minimal integer such that for some word w and some ordered t-uple as in the theorem we have $n = |w| \geq f(\underline{p})$ and w has not period $\gcd(\underline{p})$. Using the fact that $h(i, \underline{p}) \geq \gcd(\underline{p})$ for $1 \leq i \leq t-2$ we see that $f(\underline{p}) \geq p_1$. If $p_1 > 0$ consider $\underline{q} = \underline{p}^{(1)} = O(p_1, p_2 - p_1, \ldots, p_t - p_1)$. Clearly $\underline{q}^{(1)} = \underline{p}^{(2)}$ whence easily $h(i, \underline{q}) = h(i, \underline{p})$ for $1 \leq i \leq t-1$, whence $f(\underline{q}) = f(\underline{p}) - p_1$. By Lemma 1.1, $v = \mathrm{pref}_{n-p_1}(w)$ has periods $p_1, p_2 - p_1, \ldots, p_t - p_1$. As $|v| \geq f(\underline{q})$ the minimality of n gives that v has period $\gcd(\underline{q}) = \gcd(\underline{p})$, hence w also has this period, a contradiction. Therefore $p_1 = 0$. Let now $\underline{q} = (p_2, p_3, \dots, p_t)$. We easily find $m(i, \underline{q}) = m(i+1, \underline{p})$ and $h(i, \underline{q}) = h(i+1, \underline{p})$ for $1 \le i \le t-2$. Also, as $m(1, \underline{p}) = 0$ we have $h(1, \underline{p}) = f(\underline{q})$. Reporting in (1) we find $f(\underline{p}) = f(\underline{q})$. Then as $|w| \ge f(\underline{q})$ and $\underline{q}$ is a $(t-1)$-uple, w has period $\gcd(\underline{q}) = \gcd(\underline{p})$, a contradiction. □

Remark. The bound $f(\underline{p})$ is tight as the following example shows. Consider the infinite "ω-Rauzy" or "ω-bonacci" word on $\mathbb{N}$

$$0102010301020104010201030102010 5 \cdots$$

Its prefix of length $2^t - 1$ has periods $p_i = 2^t - 2^{t-i}$, $1 \le i \le t$. For this t-uple $\underline{p}$ we easily find $f(\underline{p}) = 2^t$, therefore $|w| = f(\underline{p}) - 1$ and w has not period 1.

2. A characterization of episturmian words

Let us say that an ordered t-uple $\underline{p}$ is *good* if $m(1, \underline{p}) = m(2, \underline{p}) = \dots = m(t-1, \underline{p})$ and $\gcd(\underline{p}) = 1$. When $\underline{p}$ is good, easily $f(\underline{p}) = (p_1 + \dots + p_t - 1)/(t-1)$ and this value is an integer. Let A be a t-letter alphabet. We denote by t-PER the set of the $w \in A^*$ having periods $p_1, \dots p_t$ such that $\underline{p} = (p_1, \dots, p_t)$ is good and $|w| = f(\underline{p}) - 1$.

Now let us shortly present infinite episturmian words [4, 6]. An infinite word $\mathbf{t} \in A^\omega$ is *episturmian* if the set of its factors is closed under reversal and for any $\ell \in \mathbb{N}$ there exists at most one factor of length ℓ which is right special in $\mathbf{t}$ (a factor u is *right special* if ux, uy are factors for at least two different letters x, y).

As episturmian words are uniformly recurrent and as we are interested here only in factors, we limit ourselves to the consideration of *standard* episturmian words. Let $\mathbf{s}$ be a standard episturmian word and let $u_1 = \varepsilon$ (the empty word), $u_2, u_3, \dots$ be the sequence of its palindromic prefixes. Then there exists an infinite word $\Delta(\mathbf{s}) = x_1 x_2 \cdots, x_i \in A$ called its *directive word* such that for all $n \in \mathbb{N}_+$ (the set of positive integers),

$$u_{n+1} = (u_n x_n)^{(+)}$$

where the *right palindromic* closure $(+)$ is defined by: $w^{(+)}$ is the shortest palindrome having w as a prefix [3]. When every letter of A occurs infinitely many times in $\Delta(\mathbf{s})$, $\mathbf{s}$ is *A-strict* or is a characteristic Arnoux–Rauzy sequence over A. For $a \in A$ let ψ_a be the morphism given by

$$\psi_a(a) = a,\ \psi_a(x) = ax \quad \text{for} \quad x \in A, x \neq a.$$

Let

$$\mu_n = \psi_{x_1} \psi_{x_2} \cdots \psi_{x_n},\ \mu_0 = Id$$

and

$$h_n = \mu_n(x_{n+1}).$$

Then we have the useful formula $u_{n+1} = h_{n-1}u_n$ and more generally $(u_n x)^{(+)} = \mu_{n-1}(x)u_n$ for $x \in A$. We also denote by μ_v the morphism $\psi_{a_1}\psi_{a_2}\cdots\psi_{a_n}$ for any $v = a_1 \cdots a_n$, $a_i \in A$. These notations will be kept hereafter. We have:

Lemma 2.1. *The ordered t-uple* $\underline{p}$ *different from* $(0,\dots,0,1)$ *is good if and only if, for some* $v \in A^*$, $\underline{p} = O(|\mu_v(x)| \mid x \in A)$.

Proof. Firstly let $\underline{p}$ be good. If $\underline{p}^{(1)} = (0,\dots,0,1)$ then $\underline{p} = O(|\mu_\varepsilon(x)| \mid x \in A))$. Otherwise by induction on $m(t-1,\underline{p})$ we have $\underline{p}^{(1)} = O(|\mu_w(x)| \mid x \in A)$ for some $w \in A^*$. Thus $p_1 = |\mu_w(y)| = |\mu_{wy}(y)|$ for some $y \in A$ and, for $i > 1$, $p_i - p_1 = |\mu_w(z)|$ for some letter $z \neq y$. Thus $p_i = |\mu_{wy}(z)|$ and $\underline{p}$ has the desired form. Conversely if $\underline{p} = O(|\mu_v(x)| \mid x \in A)$, let $v = wy$, $y \in A$, then $|\mu_v(z)| \geq |\mu_v(y)|$, $z \in A$ whence $p_1 = |\mu_v(y)| = |\mu_w(y)|$ and easily for $i > 1$, $p_i - p_1 = |\mu_v(z)|$ for some letter $z \neq y$. Thus $\underline{p}^{(1)}$ is good whence $\underline{p}$ is good. □

Indeed this is an application of the Rauzy rules, each $(\mu_v(x) \mid x \in A)$ is a Rauzy t-uple or standard t-uple.

Let t-ST be the set of all palindromic prefixes of standard episturmian words in A^ω. Then:

Theorem 2.2. *The sets t-PER and t-ST are equal.*

Proof. Firstly let $u_{n+1} \in t\text{-}ST$ be a palindromic prefix of $\mathbf{s}$ standard episturmian. For any $x \in A$ $(u_{n+1}x)^{(+)} = \mu_n(x)u_{n+1} = u_{n+1}\widetilde{\mu_n(x)}$. Thus u_n has periods the $|\mu_n(x)|$. By Lemma 2.1 the ordered t-uple $\underline{p}$ of the $|\mu_n(x)|$ is good. It remains to show that $|u_{n+1}| = f(\underline{p}) - 1$. This follows easily from Proposition 3.16 in [6] and more precisely from the formulas

$$|u_{n+1}| = |h_{n-1}| + |h_{n-2}| + \cdots + |h_0|$$

and for $x \in A$

$$|\mu_n(x)| = 1 + \sum_{i \in [0,n-1],\, x_{i+1} \neq x} |h_i|.$$

Thus $u_{n+1} \in t\text{-}PER$.

Conversely suppose $w \in t\text{-}PER$. Thus there is a good ordered t-uple $\underline{p}$ giving t periods for w. By remark 5.1 of [2] extended to t letters there exists, up to a renaming of the letters, exactly one word in t-PER corresponding to $\underline{p}$. By Lemma 2.1 $\underline{p} = O(|\mu_n(x)| \mid x \in A)$ for some standard episturmian word $\mathbf{s}$. Thus as proved just above u_{n+1} has periods the $|\mu_n(x)|$ and is in t-PER. Therefore u_{n+1} is w up to a renaming, thus $w \in t\text{-}ST$.

□

References

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

[2] M.G. Castelli, F. Mignosi and A. Restivo, Fine and Wilf's theorem for three periods and a generalization of Sturmian words. *Theoret. Comput. Sci.* **218** (1999) 83-94.

[3] A. de Luca, Sturmian words, structure, combinatorics and their arithmetics. *Theoret. Comput. Sci.* **183** (1997) 45-82.

[4] X. Droubay, J. Justin and G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.* (to appear).

[5] N.J. Fine and H.S. Wilf, Uniqueness Theorem for Periodic Functions. *Proc. Am. Math. Soc.* **16** (1965) 109-114.

[6] J. Justin and G. Pirillo, *Episturmian words and episturmian morphisms.* Preprint.

[7] M. Lothaire, *Combinatorics on Words.* Addison–Wesley, Reading, MA (1983).

Communicated by Ch. Choffrut.
Received June, 2000. Accepted November, 2000.