European Journal of Combinatorics 52 (2016) 59–73

Contents lists available at ScienceDirect

European Journal of Combinatorics

journal homepage: www.elsevier.com/locate/ejc

On prefixal factorizations of words

Aldo de Luca [a], Luca Q. Zamboni [c,b]

[a] Dipartimento di Matematica e Applicazioni, Università di Napoli Federico II, Italy

[b] FUNDIM, University of Turku, Finland

[c] Université de Lyon, Université Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 boulevard du 11 novembre 1918, F69622 Villeurbanne Cedex, France

ARTICLE INFO

Article history:
Received 9 May 2015
Accepted 30 August 2015
Available online 26 September 2015

ABSTRACT

We consider the class $\mathscr{P}_1$ of all infinite words $x \in \mathbb{A}^\omega$ over a finite alphabet $\mathbb{A}$ admitting a prefixal factorization, i.e., a factorization $x = U_0U_1U_2 \cdots$ where each U_i is a non-empty prefix of x. With each $x \in \mathscr{P}_1$ one naturally associates a "derived" infinite word $\delta(x)$ which may or may not admit a prefixal factorization. We are interested in the class $\mathscr{P}_\infty$ of all words x of $\mathscr{P}_1$ such that $\delta^n(x) \in \mathscr{P}_1$ for all $n \geq 1$. Our primary motivation for studying the class $\mathscr{P}_\infty$ stems from its connection to a coloring problem on infinite words independently posed by T. Brown and by the second author. More precisely, let $\mathbf{P}$ be the class of all words $x \in \mathbb{A}^\omega$ such that for every finite coloring $\varphi : \mathbb{A}^+ \to C$ there exist $c \in C$ and a factorization $x = V_0V_1V_2 \cdots$ with $\varphi(V_i) = c$ for each $i \geq 0$. In a recent paper (de Luca et al., 2014), we conjectured that a word $x \in \mathbf{P}$ if and only if x is purely periodic. In this paper we prove that $\mathbf{P} \subsetneq \mathscr{P}_\infty$, so in other words, potential candidates to a counter-example to our conjecture are amongst the non-periodic elements of $\mathscr{P}_\infty$. We establish several results on the class $\mathscr{P}_\infty$. In particular, we prove that a Sturmian word x belongs to $\mathscr{P}_\infty$ if and only if x is nonsingular, i.e., no proper suffix of x is a standard Sturmian word.

E-mail addresses: aldo.deluca@unina.it (A. de Luca), lupastis@gmail.com (L.Q. Zamboni).

http://dx.doi.org/10.1016/j.ejc.2015.08.007
0195-6698/

1. Introduction

Let **P** denote the class of all infinite words $x \in \mathbb{A}^\omega$ over a finite alphabet $\mathbb{A}$ such that for every finite coloring $\varphi : \mathbb{A}^+ \to C$ there exist $c \in C$ and a factorization $x = V_0V_1V_2\ldots$ with $\varphi(V_i) = c$ for all $i \geq 0$. Such a factorization is called φ-monochromatic. In [5] we conjectured:

Conjecture 1. *Let x be an infinite word. Then $x \in \mathbf{P}$ if and only if x is (purely) periodic.*

Various partial results in support of Conjecture 1 were obtained in [5,6,14]. Given $x \in \mathbb{A}^\omega$, it is natural to consider the binary coloring $\varphi : \mathbb{A}^+ \to \{0, 1\}$ defined by $\varphi(u) = 0$ if u is a prefix of x and $\varphi(u) = 1$ otherwise. Then any φ-monochromatic factorization is nothing more than a prefixal factorization of x, i.e., a factorization of the form $x = U_0U_1U_2\ldots$ where each U_i is a non-empty prefix of x. Thus a first necessary condition for a word x to belong to **P** is that x admits a prefixal factorization. Not all infinite words admit such a factorization, in fact as is shown in [5], square-free words and Lyndon words do not admit a prefixal factorization.

Thus in the study of Conjecture 1, one can restrict to the class of words $\mathscr{P}_1$ admitting a prefixal factorization. But in fact more is true. We prove that if $x \in \mathscr{P}_1$, then x has only finitely many distinct unbordered prefixes and admits a unique factorization in terms of its unbordered prefixes. This allows us to associate with each $x \in \mathscr{P}_1$ a new infinite word $\delta(x)$ on an alphabet corresponding to the finite set of unbordered prefixes of x. In turn, the word $\delta(x)$ may or may not admit a prefixal factorization. In case $\delta(x) \notin \mathscr{P}_1$, then $\delta(x) \notin \mathbf{P}$ and from this one may deduce that x itself does not belong to **P**. This is for instance the case of the famous Thue–Morse infinite word $t = t_0t_1t_2\cdots \in \{0, 1\}^\omega$ where t_n is defined as the sum modulo 2 of the digits in the binary expansion of n,

$$t = 0110100110010110100 10\ldots.$$

The origins of t go back to the beginning of the last century with the works of A. Thue [15,16] in which he proves amongst other things that t is *overlap-free*, i.e., contains no word of the form uuu' where u' is a non-empty prefix of u. It is readily checked that t admits a prefixal factorization, in particular t may be factored uniquely as $t = V_0V_1V_2\ldots$ where each $V_i \in \{0, 01, 011\}$. On the other hand, as is established later (see Example 4), the *derived word* $\delta(t)$ is the square-free ternary Thue–Morse word fixed by the morphism $1 \mapsto 123$, $2 \mapsto 13$, $3 \mapsto 1$. Hence $\delta(t) \notin \mathscr{P}_1$. This in turn implies that $t \notin \mathbf{P}$. Concretely, consider the coloring $\varphi' : \{0, 1\}^+ \to \{0, 1, 2\}$ defined by $\varphi'(u) = 0$ if u is a prefix of t ending with 0, $\varphi'(u) = 1$ if u is a prefix of t ending with 1, and $\varphi'(u) = 2$ otherwise. We claim that t does not admit a φ'-monochromatic factorization. In fact, suppose to the contrary that $t = V_0V_1V_2\ldots$ is a φ'-monochromatic factorization. Since V_0 is a prefix of t, it follows that there exists $a \in \{0, 1\}$ such that each V_i is a prefix of t terminating with a. Pick $i \geq 1$ such that $|V_i| \leq |V_{i+1}|$. Then $aV_iV_i \in \mathrm{Fact}(t)$. Writing $V_i = ua$, (with u empty or in $\{0, 1\}^+$), we see $aV_iV_i = auaua$ is an overlap, contradicting that t is overlap-free.

In the study of Conjecture 1, one can further restrict to the subset $\mathscr{P}_2$ of $\mathscr{P}_1$ consisting of all $x \in \mathscr{P}_1$ for which $\delta(x) \in \mathscr{P}_1$. In this case, one can define a second derived word $\delta^2(x) = \delta(\delta(x))$ which again may or may not belong to $\mathscr{P}_1$. In case $\delta^2(x) \notin \mathscr{P}_1$, then not only is $\delta^2(x) \notin \mathbf{P}$, but as we shall see neither are $\delta(x)$ and x. Continuing in this way, we are led to consider the class $\mathscr{P}_\infty$ of all words x in $\mathscr{P}_1$ such that $\delta^n(x) \in \mathscr{P}_1$ for all $n \geq 1$. We prove that $\mathbf{P} \subseteq \mathscr{P}_\infty$, so in other words any potential counter-example to our conjecture is amongst the non-periodic words belonging to $\mathscr{P}_\infty$. However, $\mathbf{P} \neq \mathscr{P}_\infty$. In fact, we prove in Section 6 that a large class of Sturmian words belong to $\mathscr{P}_\infty$, while as proved in [5], no Sturmian word belongs to **P**.

The paper is organized as follows: In Section 2 we give a brief overview of some basic definitions and notions in combinatorics on words which are relevant to the subsequent material. In Section 3 we study the basic properties of words admitting a prefixal factorization and in particular prove each admits a unique factorization in terms of its finite set of unbordered prefixes. From this we define the derived word $\delta(x)$. We prove amongst other things that if x is a fixed point of a morphism, then the same is true of $\delta(x)$.

In Section 4 we recursively define a nested sequence $\cdots \subset \mathscr{P}_{n+1} \subset \mathscr{P}_n \subset \cdots \subset \mathscr{P}_1$ where $\mathscr{P}_{n+1} = \{x \in \mathscr{P}_n \mid \delta(x) \in \mathscr{P}_n\}$, and study some basic properties of the set $\mathscr{P}_\infty = \bigcap_{n\geq 1} \mathscr{P}_n$.