Journal of Combinatorial Theory, Series A 128 (2014) 1–40

Contents lists available at ScienceDirect

Journal of Combinatorial Theory, Series A

www.elsevier.com/locate/jcta

Infinite self-shuffling words

Émilie Charlier[a], Teturo Kamae[b], Svetlana Puzynina[c,d,1], Luca Q. Zamboni[c,e,2]

[a] *Département de Mathématique, Université de Liège, Belgium*
[b] *Advanced Mathematical Institute, Osaka City University, Japan*
[c] *Department of Mathematics and Statistics, University of Turku, Finland*
[d] *Sobolev Institute of Mathematics, Novosibirsk, Russia*
[e] *Université de Lyon, Université Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France*

ARTICLE INFO

Article history:
Received 25 November 2013
Available online 9 August 2014

Keywords:
Word shuffling
Sturmian words
Lyndon words
Morphic words
Thue–Morse word

ABSTRACT

In this paper we introduce and study a new property of infinite words: An infinite word $x \in \mathbb{A}^{\mathbb{N}}$, with values in a finite set $\mathbb{A}$, is said to be *k-self-shuffling* ($k \geq 2$) if x admits factorizations: $x = \prod_{i=0}^{\infty} U_i^{(1)} \cdots U_i^{(k)} = \prod_{i=0}^{\infty} U_i^{(1)} = \cdots = \prod_{i=0}^{\infty} U_i^{(k)}$. In other words, there exists a shuffle of k-copies of x which produces x. We are particularly interested in the case $k = 2$, in which case we say x is self-shuffling. This property of infinite words is shown to be independent of the complexity of the word as measured by the number of distinct factors of each length. Examples exist from bounded to full complexity. It is also an intrinsic property of the word and not of its language (set of factors). For instance, every aperiodic word contains a non-self-shuffling word in its shift orbit closure. While the property of being self-shuffling is a relatively strong condition, many important words arising in the area of symbolic dynamics are verified to be self-shuffling. They include for instance the Thue–Morse word fixed by the morphism $0 \mapsto 01$, $1 \mapsto 10$. As another example we show that all Sturmian words of slope $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and intercept

E-mail addresses: echarlier@ulg.ac.be (É. Charlier), kamae@apost.plala.or.jp (T. Kamae), svepuz@utu.fi (S. Puzynina), lupastis@gmail.com (L.Q. Zamboni).

[1] Partially supported by the Academy of Finland under grant 251371, by Russian Foundation of Basic Research (grants 12-01-00448 and 12-01-00089).
[2] Partially supported by a FiDiPro grant (137991) from the Academy of Finland and by ANR grant *SUBTILE*.

http://dx.doi.org/10.1016/j.jcta.2014.07.008

$0 < \rho < 1$ are self-shuffling (while those of intercept $\rho = 0$ are not). Our characterization of self-shuffling Sturmian words can be interpreted arithmetically in terms of a dynamical embedding and defines an arithmetic process we call the *stepping stone model*. One important feature of self-shuffling words stems from their morphic invariance: The morphic image of a self-shuffling word is self-shuffling. This provides a useful tool for showing that one word is not the morphic image of another. In addition to its morphic invariance, this new notion has other unexpected applications particularly in the area of substitutive dynamical systems. For example, as a consequence of our characterization of self-shuffling Sturmian words, we recover a number theoretic result, originally due to Yasutomi, on a classification of pure morphic Sturmian words in the orbit of the characteristic.

1. Introduction

Let $\mathbb{A}$ be a finite non-empty set. We denote by $\mathbb{A}^*$ (resp. $\mathbb{A}^\mathbb{N}$) the set of all finite (resp. infinite) words $u = x_0x_1x_2\cdots$ with $x_i \in \mathbb{A}$.

Given k finite words $x^{(1)}, x^{(2)}, \ldots, x^{(k)} \in \mathbb{A}^*$ we let $\mathscr{S}(x^{(1)}, x^{(2)}, \ldots, x^{(k)}) \subseteq \mathbb{A}^*$ denote the collection of all words z for which there exists a factorization

$$z = \prod_{i=0}^{n} U_i^{(1)} U_i^{(2)} \cdots U_i^{(k)}$$

with each $U_i^{(j)} \in \mathbb{A}^*$ and with $x^{(j)} = \prod_{i=0}^{n} U_i^{(j)}$ for each $1 \le j \le k$. Intuitively, z may be obtained as a *shuffle* of the words $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$. For instance, it is readily checked that $011100110 \in \mathscr{S}(0010, 101, 11)$. Analogously, given k infinite words $x^{(1)}, x^{(2)}, \ldots, x^{(k)} \in \mathbb{A}^\mathbb{N}$ we define $\mathscr{S}(x^{(1)}, x^{(2)}, \ldots, x^{(k)}) \subseteq \mathbb{A}^\mathbb{N}$ to be the collection of all infinite words z for which there exists a factorization

$$z = \prod_{i=0}^{\infty} U_i^{(1)} U_i^{(2)} \cdots U_i^{(k)}$$

with each $U_i^{(j)} \in \mathbb{A}^*$ and with $x^{(j)} = \prod_{i=0}^{\infty} U_i^{(j)}$ for each $1 \le j \le k$.

Finite word shuffles were extensively studied in [12]. Given $x \in \mathbb{A}^*$, it is generally a difficult problem to determine whether there exists $y \in \mathbb{A}^*$ such that $x \in \mathscr{S}(y, y)$ (see Open Problem 4 in [12]). The problem has recently been shown to be NP-complete for sufficiently large alphabets [5,16]. However, in the context of infinite words, this question is essentially trivial: In fact, it is readily verified that if $x \in \mathbb{A}^\mathbb{N}$ and each symbol $a \in \mathbb{A}$ occurring in x occurs an infinite number of times in x, then there exists at least one (and typically infinitely many) $y \in \mathbb{A}^\mathbb{N}$ with $x \in \mathscr{S}(y, y)$. Instead, in the framework of infinite words, a far more delicate question is the following: