Theoretical Computer Science 407 (2008) 117–133

Contents lists available at ScienceDirect

Theoretical Computer Science

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

On complexity functions of infinite words associated with generalized Dyck languages

Marion Le Gonidec *

Institut de Mathématiques de l'Université de Liège, Grande traverse 12, Bat B37, B-4000 Liège, Belgique

ARTICLE INFO

Article history:
Received 23 January 2008
Received in revised form 21 April 2008
Accepted 11 May 2008
Communicated by Z. Esik

Keywords:
Infinite words
Countable automata
Dyck language
Complexity

ABSTRACT

In this article, we construct a family of infinite words, generated by countable automata and also generated by substitutions over infinite alphabets, closely related to parenthesis languages and we study their complexity functions. We obtain a family of binary infinite words $m^{(b)}$, indexed on the number $b \geq 1$ of parenthesis types, such that the growth order of the complexity function of $m^{(b)}$ is $n(\log n)^2$ if $b = 1$ and $n^{1+\log_{2b} b}$ if $b \geq 2$.

1. Introduction

Infinite words over finite alphabets appear in many areas of mathematics and theoretical computer sciences. For example, an infinite word $m = m_0 m_1 m_2 \ldots$ over a finite alphabet can be associated with a set or a partition of integers, a real number, a language, an automaton or a dynamical system. The combinatorial and statistical properties of infinite words often reveal interesting properties of the associated objects (see for example [20,4,21] for overviews).

The structure of an infinite word, and especially the diversity of the factors occurring in an infinite word $m = m_0 m_1 m_2 \ldots$ over a finite alphabet can be well described by a sequence, called the *complexity function*, counting the subwords of m. This sequence, usually treated as a function and denoted by p_m, maps an integer n to the number $p_m(n)$ of different factors of length n occurring in m. As further motivations for studying this function, the sequence $\left(\frac{\log p_m(n)}{n}\right)_{n \geq 1}$ tends to the entropy of the dynamical system associated with the word (see for example [14]) and it is also linked with the Kolmogorov complexity of the word (see [22]). Studying complexity functions can also be useful in other mathematical areas. For example, if an infinite word m represents the expansion in some integer base of an irrational algebraic number, its complexity must satisfy $\liminf_{n \to +\infty} \frac{p_m(n)}{n} = +\infty$ (see [1]).

Even if complexity functions have many simple properties, their computation often remain awkward. Furthermore, various behaviour are possible, from constant functions to exponential growth functions and from simple to irregular increasing functions. Computing complexity functions has been the aim of many works (see [3] or [12] for surveys). Indeed, for a given word, there is no universal algorithm to compute its complexity exactly. The existing methods which are often based on the study of special factors of the infinite word and synchronization principles. We refer to [5,6] for formulas linking special factors and complexity. Special factors are really useful to compute complexity, especially when studying infinite words generated by simple algorithms. Methods involving special factors allow us to compute complexity functions exactly in many particular cases and also to obtain possible *growth orders* of the complexity functions for some large classes

* Tel.: +32 4 366 92 66.
E-mail address: M.LeGonidec@ulg.ac.be.

0304-3975/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2008.05.015

of infinite words. By possible growth orders, we mean functions f satisfying $f(n) = \mathcal{O}(p_m(n))$ and $p_m(n) = \mathcal{O}(f(n))$ for at least one word m of the class. For example, complexity functions of automatic words have growth orders equal to 1 or n [10] and fixed points of substitutions over finite alphabets have growth orders equal to 1, n, $n \log\log n$, $n \log n$ or n^2 [19].

Following these results of A. Cobham and J.-J. Pansiot, we might look for similar results for infinite words generated by some more complex algorithms such as pushdown automata or countable automata. This question can also be expressed using substitutions, for infinite words built by projection letter-to-letter morphism (called projections) of fixed points of substitutions of constant length over countable alphabets. An interesting sub-class of such words is made up of characteristic words of context-free languages of integer expansions in integer bases generated by deterministic and real time pushdown automata.

This topic is also close to the open problem of finding possible growth orders of complexity functions of substitutive words, that is, images by morphisms of fixed points of substitutions over finite alphabets. Indeed, the theorem of Pansiot [19] does not hold for substitutive words. A significant partial result, proved by Cassaigne and Nicolas [9] ensures that, for all integers $k \geq 1$, there exists a substitutive word with complexity function growth order $n^{1+\frac{1}{k}}$.

We have shown in [15] that the complexity functions of a large class of words over finite alphabets, generated by deterministic and countable automata of uniformly bounded degree, are at most polynomial, that is, $p_m(n) \leq Cn^{\alpha}$ for a constant α. Moreover, the constant α only depends on the number of transition labels of the automaton and on the uniform bound of its in-degree.

The first examples, made by using a simple countable automaton, are presented in [16]. These are examples of infinite words associated with context-free languages of integer binary expansions and have complexity functions equivalent to $n(\log_2 n)^2$. In an effort to find the possible growth orders of complexity functions for words associated with context-free languages generated by deterministic and real time pushdown automata and to answer the question of the reachability of the upper bound announced in [15], we have naturally considered words associated with parenthesis languages, as these languages are classical examples of context-free languages. This article presents this family whose set of growth order of complexity functions is $\{n^{1+\log_{2b} b},\ b \geq 2\} \cup \{n(\log n)^2\}$.

2. Well-parenthesized integers and infinite words associated with Dyck's languages

In this section, we build the family $\mathcal{F} = \left\{m^{(b)},\ b \in \mathbb{N}, b \geq 1\right\}$. It is parameterized by a positive integer b, of infinite words $m^{(b)}$ with alphabet $\{0, 1\}$. The word $m^{(b)}$ is the characteristic word of the subset of $\mathbb{N}$ made up of integers whose expansion in base $2b$ corresponds to a well-parenthesized expression over b sorts of parenthesis, using a letter-to-letter morphism.

In what follows, we will use usual notations about infinite words (See for example [20]). We denote, for any integer q greater than or equal to 2, by the word $\rho_q(n) = n_l \ldots n_1 n_0$ the q-ary expansion of the integer $n = \sum_{i=0}^{l} n_i q^i$ where $n_i \in \{0, \ldots, q-1\}$ and $n_l \neq 0$.

2.1. Definitions

Let us fix an integer $b > 0$ and let $P_b = \{p_0, \ldots, p_{b-1}\}$ be the set of b symbols corresponding to b different opening parenthesis. The related closing parenthesis set is noted $\overline{P}_b = \{\overline{p}_0, \ldots, \overline{p}_{b-1}\}$.

Let $\mathfrak{D}_b$ be the *restricted Dyck language over* $P_b \cup \overline{P}_b$, that is the language of well parenthesized expressions over $P_b \cup \overline{P}_b$:

$$\mathfrak{D}_b = \left\{w \in (P_b \cup \overline{P}_b)^*,\ \forall k \leq |w|, \forall i \in [\![0, b-1]\!], |\mathrm{Pref}_k(w)|_{\overline{p}_i} \leq |\mathrm{Pref}_k(w)|_{p_i} \text{ and } |w|_{\overline{p}_i} = |w|_{p_i}\right\}.$$

Remark 2.1. In some papers, restricted Dyck languages are called, for short, Dyck languages (see for example [25]) but the usual Dyck language over $P_b \cup \overline{P}_b$ is made of words of $\mathfrak{D}_b$ and mirror-images of words of $\mathfrak{D}_b$. This classical language appears for example in the Chomsky–Schützenberger theorem [11] which state that every context-free language is the image of the intersection of a Dyck language $\mathfrak{D}'_b$ and a rational language by a homomorphism.

Definition 2.2. For all integers $b \geq 1$, let $\mathcal{D}_b : [\![0, 2b-1]\!]^* \to (P_b \cup \overline{P}_b)^*$ be the morphism of monoïds defined over $[\![0, 2b-1]\!]$ by:

$$\mathcal{D}_b(i) = \begin{cases} \overline{p}_i & \text{if } i \in [\![0, b-1]\!], \\ p_{2b-1-i} & \text{if } i \in [\![b, 2b-1]\!] \end{cases}$$

and extended by concatenation to $[\![0, 2b-1]\!]^*$.

We say that an integer $n \in \mathbb{N}$ is a *well-parenthesized integer in base* $2b$ if $\mathcal{D}_b(\rho_{2b}(n))$ belongs to $\mathfrak{D}_b$, that is, if its image by $\mathcal{D}_b$ is a well parenthesized expression. We denote by I_b the set of well parenthesized integers in base $2b$, that is:

$$I_b = \{n \in \mathbb{N} \mid \mathcal{D}_b\left(\rho_{2b}(n)\right) \in \mathfrak{D}_b\},$$

and by L_b its associated language of $2b$-ary expansions:

$$L_b = \{\rho_{2b}(n) \mid n \in I_b\}.$$

For example, the integer $n = 3696$ is a well-parenthesized integer in base 4 as $\rho_4(3696) = 321\,300$ and the word $\mathcal{D}_2(321\,300) = p_0p_1\overline{p}_1p_0\overline{p}_0\overline{p}_0$ is a well-parenthesized expression. In the same way, the integer $n = 26\,244$ is not a well-parenthesized integer in base 6 as $\rho_6(26\,244) = 321\,300$ and $\mathcal{D}_3(321\,300) = p_2\overline{p}_2\overline{p}_1p_2\overline{p}_0\overline{p}_0$ is not a well-parenthesized expression.

Remarks 2.3. When $b = 1$, the correspondence between digits and parenthesis by morphism $\mathcal{D}_1$ is quite natural, but its generalization to $b \geq 2$ is less straightforward as it is not intrinsic. We chose to generalize the above statement as it well extends the balance property of expansions of well-parenthesized integers in base 2. Indeed, expansions of well-parenthesized integers in base 2 are balanced words, that is words containing as much occurrences of 1 as occurrences of 0. This property can also be expressed the following way: for every well-parenthesized integer n in base 2, if we set $\rho_{2b}(n) = n_{2l-1}\dots n_1n_0$, then we have $\sum_{i=0}^{2l-1} n_i = l$.

Using $\mathcal{D}_b$ as a generalization of $\mathcal{D}_1$, this property is well extended as, for all integers $b \geq 1$ and for every well-parenthesized integer n in base $2b$, if we set $\rho_{2b}(n) = n_{2l-1}\dots n_1n_0$, then we have

$$\sum_{i=0}^{2l-1} n_i = (2b-1)l.$$

This way, there is no favoured type of parenthesis. For any given length, all expansions of well-parenthesized integers have the same sum of digits. Of course, there are many other ways to generalize morphism $\mathcal{D}_1$ and we can ask whether results presented below are preserved or not if we change this extension.

Definition 2.4. For every integer $b \geq 1$, we define the infinite word $m^{(b)} = m_0^{(b)}m_1^{(b)}m_2^{(b)}\dots m_n^{(b)}\dots$ over $\{0, 1\}$ as the characteristic word of the set I_b, that is,

$$\forall n \geq 0, \quad m_n^{(b)} = 1 \iff n \in I_b \iff \mathcal{D}_b(\rho_{2b}(n)) \in \mathfrak{D}_b.$$

For example, we have:

$$m^{(1)} = 1010^7 1010^{29} 10^5 1010^3 10^{113} 1010^5 1010^3 10^{17} 1010^5 1010^3 10\dots$$
$$m^{(2)} = 10^8 10^2 10^{140} 10^2 10^8 10^{11} 10^{11} 10^{11} 10^2 10^{23} 10^{11} 10\dots.$$

In order to understand how the well-parenthesized integers in base $2b$ are distributed in $\mathbb{N}$, we study in this article the complexity function of their characteristic word $m^{(b)}$.

2.2. Main theorem and sketch of proof

We now state the main theorem.

Theorem 2.5. *The growth order of the complexity function of $m^{(b)}$, denoted by $f_b(n)$, is given by:*

$$f_b(n) = \begin{cases} n\,(\log_2 n)^2 & \text{if } b = 1, \\ n^{1+\log_{2b} b} & \text{if } b \geq 2, \end{cases}$$

where the notation $\log_{2b} n$ stands for $\frac{\ln n}{\ln 2b}$, for all $b \geq 1$.

To prove this theorem, we will construct, for each value of b, two constants $c(b)$ and $C(b)$ such that, for sufficiently large integers n,

$$c(b)f_b(n) \leq p_{m^{(b)}}(n) \leq C(b)f_b(n).$$

The first remark is that we have to isolate case $b = 1$, as the combinatorial and statistical properties of $m^{(1)}$ are quite different from the properties of words $m^{(b)}$ when $b \geq 2$ (see Paragraph 4.1). Thus, the growth order of $p_{m^{(1)}}$ will be studied in a separate part (Section 4). However, we use same sketch of proof in the section regarding cases $b \geq 2$ (Section 3) and in the section regarding case $b = 1$.

In Paragraph 3.1, we will find upper bounds for $p_{m^{(b)}}(n)$ when $b \geq 2$, using methods developed in [15] to show that $p_{m^{(b)}}(n) = \mathcal{O}(f_b(n))$. In Paragraph 3.2, we will find lower bounds of $p_{m^{(b)}}(n)$ for all integers $b \geq 2$, showing a family of special factors of length n to prove that $f_b(n) = \mathcal{O}\left(p_{m^{(b)}}(n)\right)$ and some results presented in this paragraph will also be useful for case $b = 1$.

2.3. Alternative constructions of words $\{m^{(b)}, b \geq 1\}$

Before starting the proof of Theorem 2.5, we present two constructions of words $m^{(b)}$ which are more convenient than the construction presented in 2.1. These following two constructions underline different properties of $m^{(b)}$ and provide tools to study their complexity functions so these both points of view will be used in the rest of the article. These are based on the fact that $m^{(b)}$ is the indicative word of the context-free language L_b. Indeed, for any integer $b \geq 1$, the language L_b is recognized by the pushdown automaton $\mathcal{A}_b = (S, [\![0, 2b-1]\!], P_b, \phi)$ using empty stack acceptance mode, which means that,

$$w \in L_b \iff \phi_b(s_0\varepsilon, w) \in S\varepsilon,$$

where

- $S = \{s_0, s_1\}$ is the space of states,
- $[\![0, 2b-1]\!]$ is the input alphabet,
- P_b is the stack alphabet,
- $\phi_b : C \subset SP_b^* \times [\![0, 2b-1]\!] \to SP_b^*$ is the transition function, given by, for all W in P_b^+:

$$\begin{aligned}&\forall i \in [\![b, 2b-1]\!], \quad \phi_b(s_0\varepsilon, i) = s_1 p_{2b-1-i},\\ &\forall i \in [\![b, 2b-1]\!], \quad \phi_b(s_1 W, i) = s_1 p_{2b-1-i} W,\\ &\forall i \in [\![0, b-1]\!], \quad \phi_b(s_1 p_n W, i) = s_1 W,\end{aligned}$$

  where this function is naturally extended to a subset of $SP_b^* \times [\![0, 2b-1]\!]^*$ by

$$\phi_b(sW, w) = \phi_b\left(\dots\phi_b(\phi_b(sW, w_0), w_1)\dots), w_{|w|-1}\right),$$

- and $s_0\varepsilon$ is the initial configuration.

As the automaton $\mathcal{A}_b$ recognizes the language L_b, we have, for any integer n, $m_n^{(b)} = 1$ if and only if $\rho_{2b}(n)$ is accepted by $\mathcal{A}_b$.

Remarks 2.6. In this case, the space of states S of $\mathcal{A}_b$ can be reduced to a singleton (state s_1) but splitting it in two states s_0 and s_1 does not change the accepted language and allows us to feed the automaton with non proper expansions without changing exit states. Moreover, this will be useful to underline the equivalence between the both constructions of words $m^{(b)}$ we are going to use in the proof of Theorem 2.5. Indeed, to compute the complexity function of the word $m^{(b)}$, the pushdown automaton does not provide efficient tools so we do not develop it further. For further information about links between context-free languages and pushdown automata, we refer to [2]. Nevertheless, we can obtain the two ways of generating the words $m^{(b)}$ we will need in this article from the pushdown automaton $\mathcal{A}_b$. The first way is to construct $m^{(b)}$ by concatenation of the exit states of a countable automaton and the second is to construct $m^{(b)}$ as a projection (letter-to-letter morphism) of the infinite fixed point of a substitution of constant length over a countable alphabet.

2.3.1. *Construction by countable automaton*

The construction of words $m^{(b)}$ by pushdown automata $\mathcal{A}_b$ is not the best point of view to solve the problems of accessibility and co-accessibility between configurations that we will face in this article. Indeed, it does not allow to see easily the action of the transition function over configurations. A better point of view is to consider the transition graph $\mathcal{T}_b$ of the pushdown automaton $\mathcal{A}_b$. The transition graph $\mathcal{T}_b = (s_1 P_b^* \cup \{s_0\varepsilon\}, \varphi_b, s_0\varepsilon, S\varepsilon)$ acts here as an $2b$-automaton (see [17] or [15]) and by construction, recognizes the same language as $\mathcal{A}_b$. The $2b$-automata $\mathcal{T}_b$ is defined by the following elements:

- $P_b^* \cup \{s_0\}$ is the space of states of $\mathcal{T}_b$,
- $[\![0, 2b-1]\!]$ is the input alphabet of $\mathcal{T}_b$,
- φ_b is the transition function of $\mathcal{T}_b$, defined on a subset of $\left(s_1 P_b^* \cup \{s_0\varepsilon\}\right) \times [\![0, 2b-1]\!]$ to $s_1 P_b^* \cup \{s_0\varepsilon\}$, defined by:

$$\begin{aligned}&\forall i \in [\![b, 2b-1]\!], \quad \varphi_b(s_0\varepsilon, i) = s_1 p_{2b-1-i},\\ &\forall W \in P_b^*,\ \forall i \in [\![b, 2b-1]\!], \quad \varphi_b(s_1 W, i) = s_1 p_{2b-1-i} W,\\ &\forall W \in P_b^*,\ \forall i \in [\![0, b-1]\!], \quad \varphi_b(s_1 p_i W, i) = s_1 W.\end{aligned}$$

  This function is also naturally extended to a subset of $\left(s_1 P_b^* \cup \{s_0\varepsilon\}\right) \times [\![0, 2b-1]\!]^*$ by

$$\varphi_b(W, w) = \varphi_b\left(\dots\varphi_b(\varphi_b(W, w_0), w_1)\dots), w_{|w|-1}\right).$$

- $s_0\varepsilon$ is the initial state of $\mathcal{T}_b$,
- $S\varepsilon$ is the final states set of $\mathcal{T}_b$.

We refer to Muller and Schupp [18] and the works of Caucal [7,8] for literature on transition graphs; see also [24] or [23] for results on infinite graphs.

These deterministic countable automata $\mathcal{T}_b$ can be completed by adding a "trash state" x so that, for every state s of $s_1 P_b^* \cup \{s_0\varepsilon, x\}$, and for every integer i of $[\![0, 2b-1]\!]$, there is one and only one edge of $\mathcal{T}_b$ starting from s and labeled by i.

The automaton obtained this way is a countable deterministic $2b$-automaton and the infinite word $m^{(b)}$ over $\{0, 1\}$ is defined by:

$$\forall n \geq 0, \quad m_n^{(b)} = 1 \iff \varphi_b(\varepsilon, \rho_{2b}(n)) \in S\varepsilon.$$

The transition graphs $\mathcal{T}_1$ and $\mathcal{T}_2$ are represented in Figs. 1 and 2. For more visibility, the trash state has been removed from $\mathcal{T}_2$ and we have made the following simplifications of notation: $s_1 W \mapsto W$ and $s_0\varepsilon \mapsto s_0$.

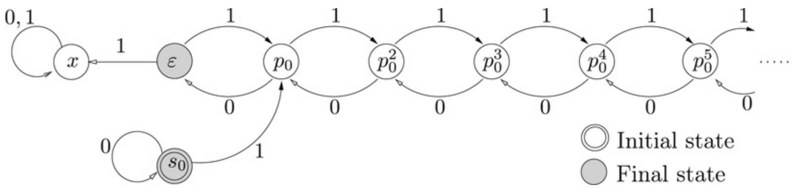

Fig. 1. Countable automaton generating the infinite word $m^{(1)}$.

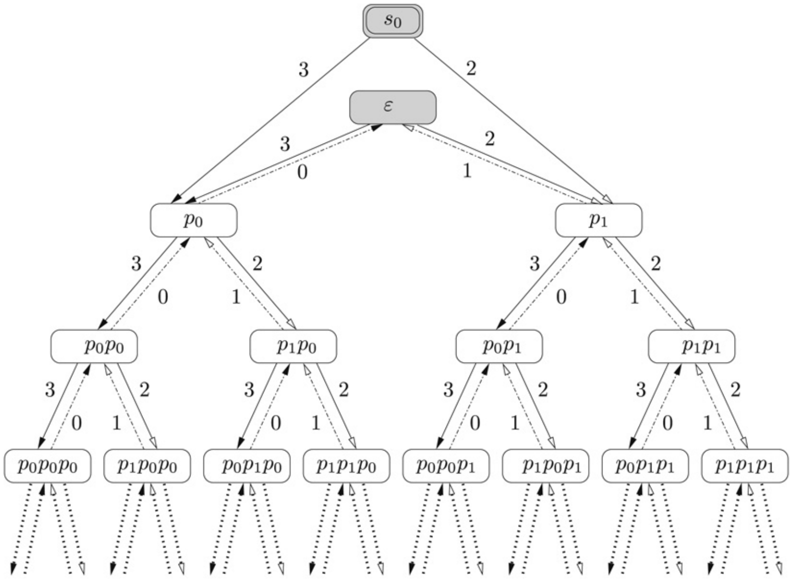

Fig. 2. Countable automaton generating the infinite word $m^{(2)}$.

2.3.2. Construction by substitution and projection

The other way to construct the word $m^{(b)}$ is as a projection (letter-to-letter morphism) of the fixed point of a substitution over a countable alphabet, due to the connection between automata and substitutions (see [17] for example).

Let us introduce a family of substitutions $\{{}^b\overline{\delta}, b \geq 1\}$ over countable alphabets:

- For $b = 1$, let $A_1 = \{p_0\}^* \cup \{s_0, x\}$, where s_0 and x are two symbols out of $\{p_0\}^*$. Let ${}^1\overline{\delta}$ be the following substitution over A_1:

$$
{}^1\overline{\delta}: \quad \begin{array}{rl}
A_1 & \to A_1^2 \\
x & \mapsto (x)^2 \\
s_0 & \mapsto (s_0)(p_0) \\
\varepsilon & \mapsto (x)(p_0) \\
p^n,\ n > 0 & \mapsto (p_0^{n-1})(p_0^{n+1}).
\end{array}
$$

- For any $b \geq 2$, let $A_b = P_b^* \cup \{s_0, x\}$, where s_0 and x are two symbols out of P_b^*. Let ${}^b\overline{\delta}$ be the following substitution over A_b:

$$
{}^b\overline{\delta}: \quad \begin{array}{rl}
A_b & \to A_b^{2b} \\
x & \mapsto (x)^{2b} \\
s_0 & \mapsto (s_0)(x)^{b-1}(p_{b-1})\dots(p_1)(p_0) \\
\varepsilon & \mapsto (x)^b(p_{b-1})\dots(p_1)(p_0) \\
s = p_i w & \mapsto (x)^i(w)(x)^{b-1-i}(p_{b-1}s)\dots(p_1 s)(p_0 s)
\end{array}
$$

For all integers $b \geq 1$, the substitution ${}^b\overline{\delta}$ is extended to A_b^* and $A_b^{\mathbb{N}}$ by concatenation. The substitution ${}^b\overline{\delta}$ has a unique infinite fixed point $\overline{m}^{(b)}$ in $A_b^{\mathbb{N}}$, starting by the letter s_0, that is, a unique infinite word such that

$$
{}^b\overline{\delta}\left(\overline{m}^{(b)}\right) = \overline{m}^{(b)}.
$$

Remark 2.7. The connection between the two constructions relies on the correspondence between the applications ${}^b\overline{\delta}_i : s \mapsto {}^b\overline{\delta}(s)_i$ and the applications $\varphi_b(\cdot, i)$. When the automaton $\mathcal{T}_b$ is fed with $\rho_{2b}(n)$, the outing state is exactly $\overline{m}_n^{(b)}$. See [17] or [15] for more details.

For example, we have:

$$\begin{aligned}
\overline{m}^{(1)} &= s_0(p_0)(\varepsilon)(p_0^2)(x)(p_0)(p_0)(p_0^3)(x)^2(\varepsilon)(p_0^2)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(x)^5(p_0)(p_0)(p_0^3)(x)(p_0)(p_0)(p_0^3)\\
&(p_0)(p_0^3)(p_0^3)(p_0^5)(x)^{10}(\varepsilon)(p_0^2)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(x)^2(\varepsilon)(p_0^2)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(p_0^2)(p_0^4)\\
&(p_0^4)(p_0^6)(x)^{21}(p_0)(p_0)(p_0^3)(x)(p_0)(p_0)(p_0^3)(p_0)(p_0^3)(p_0^3)(p_0^5)(x)^5(p_0)(p_0)(p_0^3)(x)(p_0)(p_0)(p_0^3)(p_0)\\
&(p_0^3)(p_0^3)(p_0^5)(x)(p_0)(p_0)(p_0^3)(p_0)(p_0^3)(p_0^3)(p_0^5)(p_0)(p_0^3)(p_0^3)(p_0^5)(p_0^3)(p_0^5)(p_0^5)(p_0^7)(x)^{42}(\varepsilon)(p_0^2)(\varepsilon)(p_0^2)\\
&(p_0^2)(p_0^4)(x)^2(\varepsilon)(p_0^2)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(p_0^2)(p_0^4)(p_0^4)(p_0^6)(x)^{10}(\varepsilon)\dots\\
\overline{m}^{(2)} &= s_0(x)(p_1)(p_0)(x)^5(\varepsilon)(p_1^2)(p_0p_1)(\varepsilon)(x)(p_1p_0)(p_0^2)(x)^{22}(p_1)(p_0)(x)(p_1)(p_1^3)(p_0p_1^2)(p_1)(x)\\
&(p_1p_0p_1)(p_0^2p_1)(x)^2(p_1)(p_0)(x)^5(p_0)(p_1^2p_0)(p_0p_1p_0)(p_0)(x)(p_1p_0^2)(p_0^3)(x)^{89}(\varepsilon)(p_1^2)(p_0p_1)(\varepsilon)(x)\\
&(p_1p_0)(p_0^2)(x)^5(\varepsilon)(p_1^2)(p_0p_1)(x)(p_1^2)(p_1^4)(p_0p_1^3)(p_1^2)(x)(p_1p_0p_1^2)(p_0^2p_1^2)(x)(\varepsilon)(p_1^2)(p_0p_1)(x)^5(p_0p_1)\\
&(p_1^2p_0p_1)(p_0p_1p_0p_1)(p_0p_1)(\varepsilon)(p_1p_0^2p_1)(p_0^3p_1)(x)^9(\varepsilon)(p_1^2)(p_0p_1)(\varepsilon)(x)(p_1p_0)(p_0^2)(x)^{20}(\varepsilon)(x)(p_1p_0)\\
&(p_0^2)(x)(p_1p_0)(p_1^3p_0)(p_0p_1^2p_0)(p_1p_0)(x)(p_1p_0p_1p_0)(p_0^2p_1p_0)(\varepsilon)(x)(p_1p_0)(p_0^2)\dots.
\end{aligned}$$

We define Π_b as the projection from A_b to $\{0, 1\}$ defined by:

$$\Pi_b(s) = \begin{cases} 1 & \text{if } s \in \{\varepsilon, s_0\},\\ 0 & \text{else.} \end{cases}$$

The projection Π_b is extended to $A_b^{\mathbb{N}}$ by concatenation and, for all integers $b \geq 1$, the infinite word $m^{(b)}$ satisfies $m^{(b)} = \Pi(\overline{m}^{(b)})$.

Remark 2.8. Construction by substitution and projections provides the main part of tools used to study complexity functions of words $m^{(b)}$ so we will use this construction for most of the article. However, the construction by automata remains helpful, often providing an easier understanding and allowing graphic interpretations.

3. Complexity function growth orders for $p_{m^{(b)}}(n)$ when $b \geq 2$

Notations 3.1. To prove these results, we will use the construction of the words $m^{(b)}$ by substitutions and projection. So, to make the proofs easier to read, we will denote the function $\Pi_b \circ {}^b\overline{\delta}^k$ by δ^k when no confusion is possible, for all integers $b \geq 1$ and positive integers k.

3.1. Upper bounds of $p_{m^{(b)}}(n)$ when $b \geq 2$

For every $b \geq 2$, finding an upper bound of the complexity function of $m^{(b)}$ relies on the study of factors of length two occurring in the fixed point $\overline{m}^{(b)}$ of ${}^b\overline{\delta}$. This study is the aim of the two following lemmas.

Lemma 3.2. *Let $b \geq 2$.*

Let k be a positive integer and s be an element of A_b.

The word $\delta^k(s)$ contains some occurrences of 1 if and only if $s = s_0$ or $s \in \cup_{i=0}^{k} P_b^i$ and $|s|$ and k have same parity.

So, for all integers $k = 2q + r$, with $r \in \{0, 1\}$ and $q \in \mathbb{N}^*$, we have $\frac{b^{k+2}-b^r}{b^2-1} + 1$ letters of A_b such that $\delta^k(s)$ is different of $0^{(2b)^k}$.

This lemma is straightforward using the following property of fixed points of substitutions of constant length (Proposition 3.1 of [15]): for all letters s occurring in $\overline{m}^{(b)}$ and for all non negative integers k, if we denote the word ${}^b\overline{\delta}^k(s) = u_0u_1\dots u_{(2b)^k-1}$, then

$$\forall n \in [\![0, (2b)^k - 1]\!], \quad u_n = {}^b\overline{\delta}_{n_0} \circ {}^b\overline{\delta}_{n_1} \circ \dots \circ {}^b\overline{\delta}_{n_l} \circ {}^b\overline{\delta}_0^{2b-(l+1)}(s),$$

where $\rho_{2b}(n) = n_l \dots n_1 n_0$ is the proper expansion of n in base $2b$.

We can also see this property on the graph of the automaton of $\mathcal{T}_b$. Indeed, the word $\delta^k(s)$ contain some occurrences of 1 if and only if ${}^b\overline{\delta}^k(s)$ contains some occurrences of ε or s_0, that is, it contains some occurrences of the letter 1 if and only if there exists a path of length k from s to ε or s_0 in the directed graph of $\mathcal{T}_b$.

Lemma 3.3. *Let $b \geq 2$. A word x_1x_2 of A_b^2 is a factor of $\overline{m}^{(b)}$ if and only if:*

1. $\exists s \in A_b$, $x_1x_2 = (s)(x)$,
2. $\exists s \in A_b \setminus \{s_0\}$, $x_1x_2 = (x)(s)$,
3. $x_1x_2 = (p_0p_1)(\varepsilon)$,
4. $\exists s \in P_b^*$, $\exists i \in [\![0, b-2]\!]$, $x_1x_2 = (p_{i+1}s)(p_is)$,
5. $\exists s \in P_b^*$, $x_1x_2 = (s)(p_{b-1}^2 s)$,
6. $\exists s \in P_b^*$, $\exists k \geq 1$, $x_1x_2 = (p_0^k p_1 p_0^{k-1} s)(s)$.

Proof of Lemma 3.3. This lemma also relies on a property of fixed points of substitutions of constant length (Proposition 3.3 of [15]). Indeed, as $\overline{m}^{(b)}$ is the fixed point of ${}^b\overline{\delta}$ starting with s_0, if we set ${}^b\overline{\delta}(s_0) = s_0s_1\ldots s_{2b-1}$, we have the following equality:

$$\overline{m}^{(b)} = s_0s_1\ldots s_{2b-1}{}^b\overline{\delta}(s_1){}^b\overline{\delta}(s_2)\ldots{}^b\overline{\delta}(s_{2b-1}){}^b\overline{\delta}^2(s_1){}^b\overline{\delta}^2(s_2)\ldots{}^b\overline{\delta}^2(s_{2b-1}){}^b\overline{\delta}^3(s_1)\ldots{}^b\overline{\delta}^3(s_{2b-1})\ldots.$$

From this statement, the word x_1x_2 is a factor of $\overline{m}^{(b)}$ if and only if one of the following cases happens:

(i) $\exists j \in [\![0, 2b-2]\!],\ x_1x_2 = s_js_{j+1}$,
(ii) $\exists k > 0, \exists j \in [\![1, 2b-2]\!], x_1 = {}^b\overline{\delta}^k_{2b-1}(s_j)$ and $x_2 = {}^b\overline{\delta}^k_0(s_{j+1})$,
(iii) $\exists k \geq 0, \exists j \in [\![1, 2b-2]\!],\ x_1 \in E \setminus \{s_0\},\ x_2 \in {}^b\overline{\delta}^k_0 \circ {}^b\overline{\delta}_{j+1} \circ {}^b\overline{\delta}^{-1}_j \circ {}^b\overline{\delta}^{-k}_{2b-1}(\{x_1\})$,
(iv) $\exists k > 0, x_1 = {}^b\overline{\delta}^k_{2b-1}(s_{2b-1})$ and $x_2 = {}^b\overline{\delta}^{k+1}_0(s_1)$.

Case (i) corresponds to words x_1x_2 extracted from the word $s_0s_1\ldots s_{2b-1}$, case (ii) corresponds to words x_1x_2 extracted at the junction of words ${}^b\overline{\delta}^k(s_j)$ and ${}^b\overline{\delta}^k(s_{j+1})$, case (iii) corresponds to words x_1x_2 extracted from words ${}^b\overline{\delta}^k(s_j)$ and case (iv) corresponds to words x_1x_2 extracted at the junction of words ${}^b\overline{\delta}^k(s_{2b-1})$ and ${}^b\overline{\delta}^{k+1}(s_1)$.

The different factors of $\overline{m}^{(b)}$ announced in Lemma 3.3 will progressively appear, analysing these different cases.

The factors of $\overline{m}^{(b)}$ obtained in case (ii) are

- $x_1x_2 = (x)(x)$, obtained when $k \geq 0$ and $j < b-1$ or $k > 0$ and $j = b-1$,
- $x_1x_2 = (x)(p_{b-1})$ obtained when $k = 0$ and $j = b-1$,
- $x_1x_2 = (p_{i+1})(p_i)$ for $i \in [\![0, b-2]\!]$, obtained when $k = 0$ and $j \geq b$,
- $x_1x_2 = (p_0^kp_i)(x)$ for $k \geq 1$ and $i \in [\![2, b-1]\!]$, obtained when $k \geq 1$ and $j \in [\![b, 2b-2]\!]$,
- $x_1x_2 = (p_0p_1)(\varepsilon)$ and $x_1x_2 = (p_0^kp_1)(x)$ for $k \geq 2$, obtained when $k \geq 1$ and $j = 2b-2$.

The factors of $\overline{m}^{(b)}$ obtained in case (iv) are the words $x_1x_2 = (p_0^k)x$ for any $k \geq 1$.

To describe the factors of $\overline{m}^{(b)}$ which can be reached in case (iii), it is useful to notice that all the letters of A_b occur in $\overline{m}^{(b)}$, and to detail the actions of the applications ${}^b\overline{\delta}^k_0 \circ {}^b\overline{\delta}_{j+1} \circ {}^b\overline{\delta}^{-1}_j \circ {}^b\overline{\delta}^{-k}_{2b-1}$ on singletons $\{s\}$ for $s \in A_b$ and for all $k \geq 1$. For all i in $[\![0, b-2]\!]$:

$${}^b\overline{\delta}^k_0 \circ {}^b\overline{\delta}_{i+1} \circ {}^b\overline{\delta}^{-1}_i \circ {}^b\overline{\delta}^{-k}_{2b-1}(\{s\}) = \begin{cases} A_b \setminus \{s_0\} & \text{if } s = x,\\ \{x\} & \text{if } \mathrm{Pref}_k(s) = p_0^kp_{2b-1-i},\\ \emptyset & \text{else,}\end{cases}$$

for $i = b-1$:

$${}^b\overline{\delta}^k_0 \circ {}^b\overline{\delta}_b \circ {}^b\overline{\delta}^{-1}_{b-1} \circ {}^b\overline{\delta}^{-k}_{2b-1}(\{s\}) = \emptyset,$$

for all i in $[\![b, 2b-2]\!]$:

$${}^b\overline{\delta}^k_0 \circ {}^b\overline{\delta}_{i+1} \circ {}^b\overline{\delta}^{-1}_i \circ {}^b\overline{\delta}^{-k}_{2b-1}(\{s\}) = \begin{cases} \{x\} & \text{if } \mathrm{Pref}_{k+1}(s) = p_0^kp_1,\\ \emptyset & \text{else,}\end{cases}$$

for $i = 2b-2$:

$${}^b\overline{\delta}^k_0 \circ {}^b\overline{\delta}_{i+1} \circ {}^b\overline{\delta}^{-1}_i \circ {}^b\overline{\delta}^{-k}_{2b-1}(\{s\}) = \begin{cases} \{v\} & \text{if } s = p_0^kp_1p_0^{k-1}v,\\ \{x\} & \text{if } \mathrm{Pref}_{k+1}(s) = p_0^kp_{2b-1-i} \text{ and } \mathrm{Pref}_{2k}(s) \neq p_0^kp_1p_0^{k-1},\\ \emptyset & \text{else.}\end{cases}$$

Using these equalities, we obtain:

- the words $(x)(x)$, $(x)(s)$ and $(s)(x)$ for all s of A_b are factors of $\overline{m}^{(b)}$ except for $(x)(s_0)$,
- the words $(s)(p_{b-1}^2s)$ for all s of A_b are factors of $\overline{m}^{(b)}$,
- the words $(p_{i+1}s)(p_is)$ for all s of P_b^* and all $i \in [\![0, b-2]\!]$ are factors of $\overline{m}^{(b)}$,
- the words $(p_0^kp_1p_0^{k-1}s)(s)$ for all $s \in P_b^*$ are factors of $\overline{m}^{(b)}$.

All different factors of $\overline{m}^{(b)}$ announced in Lemma 3.3 have been displayed. ■

Proof (**Upper bounds of** $p_{m^{(b)}}(n)$ **when** $b \geq 2$). Let us fix an integer $k = 2q + r$ for $r \in \{0, 1\}$ and $q \in \mathbb{N}$ and an integer n satisfying $(2b)^k \leq n < (2b)^{k+1}$.

First, we find an upper bound of $p_{m^{(b)}}((2b)^k)$ for all integers $k \geq 0$ and then we will obtain an upper bound of $p_{m^{(b)}}(n)$ for all integers $n \geq 1$.

As $\overline{m}^{(b)}$ is the fixed point of the substitution ${}^b\overline{\delta}$, we have $\overline{m}^{(b)} = {}^b\overline{\delta}^k\left(\overline{m}^{(b)}\right)$. According to this, every factor of length $(2b)^k$ of $m^{(b)}$ occurs as a subfactor of some $\delta^k(x_1)\delta^k(x_2)$ where x_1x_2 is a factor of length two of $\overline{m}^{(b)}$. In order to find a thin upper bound of the number of factors of length $(2b)^k$, we need to know how many factors x_1x_2 of $\overline{m}^{(b)}$ have the property that the projection Π_b does not maps both ${}^b\overline{\delta}^k(x_1)$ and ${}^b\overline{\delta}^k(x_2)$ on the word $0^{(2b)^k}$.

Let us divide this set of factors of $\overline{m}^{(b)}$ of length two in two parts:

(A) factors x_1x_2 for which only one among the words $\delta^k(x_1)$ and $\delta^k(x_2)$ is different from the word $0^{(2b)^k}$. Typically, these are words $(s)(x)$ or $(x)(s)$ for the letters $s \in P_b^*$ such that ε or s_0 occurs in ${}^b\overline{\delta}^k(s)$,

(B) factors x_1x_2 for which both $\delta^k(x_1)$ and $\delta^k(x_2)$ are different from the word $0^{(2b)^k}$.

From Lemma 3.2, we know that $\frac{b^{k+2}-b^r}{b^2-1} + 1$ letters s of A_b provide words ${}^b\delta^k(s)$ which are different from $0^{(2b)^k}$ thus we can construct at most $2\left(\frac{b^{k+2}-1}{b^2-1} + 1\right) - 1$ different factors x_1x_2 of type (A) as $(x)(s_0)$ never occurs in $m^{(b)}$.

Let us now count the factors x_1x_2 of type (B). From Lemma 3.3, those factors can be of the following types:

(i) $(p_0p_1)(\varepsilon)$,
(ii) $(p_{i+1}s)(p_is)$ for $s \in P_b^*$ and $i \in [\![0, b-1]\!]$,
(iii) $(s)(p_{b-1}^2 s)$ for $s \in P_b^*$,
(iv) $(p_0^\alpha p_1 p_0^{\alpha-1}s)(s)$ for $s \in P_b^*$ and an integer $\alpha \geq 1$.

(i) When $x_1x_2 = (p_0p_1)(\varepsilon)$, the words $\delta^k(x_1)$ and $\delta^k(x_2)$ are different from $0^{(2b)^k}$ only when k is even.

(ii) Factors $x_1x_2 = (p_{i+1}s)(p_is)$, for $s \in P_b^*$ are of type (B) only when the length $|s|$ of the word s is smaller than k and $|s|$ and k have different parity (see Lemma 3.2). So, in this way, we obtain $\frac{b^{k+2}-b^{1-r}}{b^2-1}$ factors of length two such that the projection Π_b maps neither ${}^b\overline{\delta}^k(x_1)$ nor ${}^b\overline{\delta}^k(x_2)$ to the word $0^{(2b)^k}$.

(iii) Factors $x_1x_2 = (s)(p_1^2 s)$, for $s \in P_b^*$ are of type (B) only when the length of s is strictly smaller than k and with the same parity as k (see Lemma 3.2). Thus we obtain $\frac{b^k-b^r}{b^2-1}$ factors of length two such that the projection Π_b maps neither ${}^b\overline{\delta}^k(x_1)$ nor ${}^b\overline{\delta}^k(x_2)$ to $0^{(2b)^k}$.

(iv) Factors of type $x_1x_2 = (p_0^\alpha p_1 p_0^{\alpha-1}s)(s)$ for $s \in P_b^*$ are of type (B) only when $|s| = k - 2\beta$, for a fixed integer $\beta \in [\![1, q]\!]$ and $\alpha \leq \beta$ (where $k = 2q + r$ with $r \in \{0, 1\}$). We obtain by this construction $N = \sum_{\beta=1}^{q} \beta b^{k-2\beta}$ additional factors of type (B). From the expression of N, we get:

$$N = \begin{cases} \frac{1}{(b^2-1)^2}\left(b^{k+2} - \frac{b^2-1}{2}k - 1\right) & \text{if } k \text{ is even,} \\ \frac{b}{(b^2-1)^2}\left(b^{k+1} - \frac{b^2-1}{2}k + b + \frac{b^2-1}{2}\right) & \text{if } k \text{ is odd.} \end{cases}$$

From this equality, we only need that $N \leq \frac{b}{(b^2-1)^2}\left(b^{k+1} - \frac{b^2-1}{2}k + b + \frac{b^2-1}{2}\right)$ for all integers k.

It appears from the enumeration of factors of type (A) and (B) that the number of different words $\delta^k(x_1)\delta^k(x_2)$ different from $0^{2(2b)^k}$ occurring in $m^{(b)}$ admits the following upper bound $U(k)$:

$$U(k) = \frac{1}{2(b^2-1)}\left(2(4b^2+1)b^k - b(b^2-1)k + b(b^2+6b-1) - 12\right).$$

It follows that $p_{m^{(b)}}\left((2b)^k\right) \leq (2b)^k U(k)$ and $p_{m^{(b)}}\left((2b)^{k+1}\right) \leq (2b)^{k+1}U(k+1)$. Using the fact that complexity function is an increasing function and that $\log_{2b}(n)$ is in $[\![k, k+1[\![$ for all integers n of $[\![(2b)^k, (2b)^{k+1}[\![$, we get:

$$\forall n \geq 1, \quad p_{m^{(b)}}(n) \leq \frac{2bn}{2(b^2-1)}\left(2b(4b^2+1)n^{\log_{2b} b} - b(b^2-1)\log_{2b} n + 6b^2 - 12\right).$$

and so $p_{m^{(b)}}(n) = \mathcal{O}\left(n^{1+\log_{2b} b}\right)$. ∎

3.2. Lower bounds of $p_{m^{(b)}}(n)$ when $b \geq 2$

To obtain lower bounds of complexity functions of words $m^{(b)}$, we need to know more about the structure of words $\delta^k(w)$ which are different from $0^{(2b)^k}$, for a fixed integer k. We first study the number of occurrences of the letter 1 and their repartition in such words. It will allow us to match special factors of fixed length and in this way, to find the lower bounds of complexity. Indeed, for a fixed binary word m, the difference $p_m(n+1) - p_m(n)$ corresponds exactly to the number of left special factors (and also to the number of right special factors).

The main difficulty to overcome in the proof arises from the construction of the words $m^{(b)}$ by substitution and projection. This problem also appears for substitutive words (see [9]). Indeed, the projection erases a lot of information and the following two phenomena may appear:

- two or infinitely many different special factors of $\overline{m}^{(b)}$ can project on the same special factor of $m^{(b)}$,
- two or infinitely many different non-special factors of $\overline{m}^{(b)}$ can project on a single special factor of $m^{(b)}$.

As a consequence, the difficulty is not really to find special factors but to ensure the special factors we have found are different. To do this, we will extract special factors from factors of type $\delta^k((x)(x)(s))$ of $m^{(b)}$ (see Lemmas 3.14 and 3.15). To ensure that words $\delta^k(s)$ are sufficiently different to produce different special factors, we first determine a set which indexes the occurrences of the letter 1 in words $\delta^k(s)$. This set will also give information about where these occurrences take place. This is the aim of Lemma 3.6 and Proposition 3.9. Then, Lemma 3.11 will count these occurrences and as it is not sufficient to differenciate them, we study further the repartition of letters 1 in words $\delta^k(s)$ in Lemma 3.13.

Notations 3.4. For a fixed integer $b \geq 1$, we introduce notations about integers and their $2b$-ary expansions, which will be useful to make the proofs easier to read.

For a given word w of $[\![0, 2b-1]\!]^*$, we denote by $[w]_{2b}$ the integer for which w is an expansion in base $2b$ (w can be a non proper expansion), that is $[w]_{2b} = \sum_{i=0}^{|w|-1} w_i (2b)^{|w|-1-i}$.

We also denote by ${}^b\overline{\delta}_w$ the function defined by:

$$ {}^b\overline{\delta}_w = {}^b\overline{\delta}_{w_{|w|-1}} \circ \cdots \circ {}^b\overline{\delta}_{w_1} \circ {}^b\overline{\delta}_{w_0}. $$

For a given word s of P_b^*,

- the image of s by the application $p_i \mapsto i$ is denoted $\tilde{s}$, so that $\tilde{s}_i = 2b - 1 - \mathcal{D}_b^{-1}(s_i)$,
- the image of s by the application $p_i \mapsto 2b-1-i$ and mirror image is denoted $\hat{s}$, so that $\hat{s}_i = \mathcal{D}_b^{-1}(s_{|s|-1-i})$.

Remarks 3.5. From these notations, we get:

- for some $s \in P_b^*$, the word $\hat{s}$ is the shortest word such that $\overline{\delta}_{\hat{s}}(\varepsilon) = s$.
- the word $\mathcal{D}_b(\hat{s})$ is the mirror image of s.
- the word $\mathcal{D}_b(\hat{s}\tilde{s})$ is the shortest well parenthesized expression with prefix $\mathcal{D}_b(\hat{s})$ as, for all integers $n \in [\![0, |s|-1]\!]$, $\mathcal{D}_b(\tilde{s})_n = \overline{p}_i$ if and only if $\mathcal{D}_b(\hat{s})_{|s|-1-n} = p_i$ (the word $\tilde{s}$ labels the shortest path from s to ε in the graph of $\mathcal{T}_b$).

For example, if we set $b = 2$ and $s = p_0p_0p_1p_0p_1$, then $\tilde{s} = 00\,101$ and $\hat{s} = 23\,233$ and the word $\mathcal{D}_2(\hat{s})\mathcal{D}_2(\tilde{s}) = p_1p_0p_1p_0p_0\overline{p}_0\overline{p}_0\overline{p}_1\overline{p}_0\overline{p}_1$ is the shortest well parenthesized expression over two parenthesis types with prefix $\mathcal{D}_2(\hat{s})$. On the other hand, if we set $b = 3$ and $s = p_0p_0p_1p_0p_1$, then $\tilde{s} = 00\,101$ and $\hat{s} = 45\,455$ and the word $\mathcal{D}_3(\hat{s})\mathcal{D}_3(\tilde{s}) = p_1p_0p_1p_0p_0\overline{p}_0\overline{p}_0\overline{p}_1\overline{p}_0\overline{p}_1$ is also the shortest well parenthesized expression over three parenthesis types with prefix $\mathcal{D}_3(\hat{s})$.

If we set $b = 1$, all these notations are really heavy but valid as, for any element $s = p_0^n$ of $\{p_0\}^*$, we have $\tilde{s} = 0^n$ and $\hat{s} = 1^n$.

Lemma 3.6. *Let $b \geq 1$. Let s be a word of P_b^* and w a word of $[\![0, 2b-1]\!]^*$.*

The word $\mathcal{D}_b(\hat{s}w)$ belongs to $\mathfrak{D}_b$ if and only if there exists $|s|+1$ words, denoted by $w^{(0)}, w^{(1)}, \ldots, w^{(|s|)}$ of $[\![0, 2b-1]\!]^*$ such that, for all i of $[\![0, |s|]\!]$, $\mathcal{D}_b(w^{(i)}) \in \mathfrak{D}_b$ and

$$ w = w^{(0)}\tilde{s}_0 w^{(1)}\tilde{s}_1 \ldots w^{(|s|-1)}\tilde{s}_{|s|-1} w^{(|s|)}. $$

Proof of Lemma 3.6. Let s be a word of P_b^* and w a word of $[\![0, 2b-1]\!]^*$.

Obviously, if $w = w^{(0)}\tilde{s}_0 w^{(1)}\tilde{s}_1 \ldots w^{(|s|-1)}\tilde{s}_{|s|-1} w^{(|s|)}$ where all the words $\mathcal{D}_b(w^{(i)})$ belong to $\mathfrak{D}_b$, then $\mathcal{D}_b(\hat{s}w)$ also belongs to $\mathfrak{D}_b$, as $\mathcal{D}(\hat{s}\tilde{s})$ is a well parenthesized expression too.

Let us now fix a word w such that $\mathcal{D}_b(\hat{s}w)$ belongs to $\mathfrak{D}_b$.

We proceed by induction on the length of the considered element s to prove that there exists a $(|s|+1)$-uple of words $w^{(i)}$ of $[\![0, 2b-1]\!]^*$ satisfying $\mathcal{D}_b(w^{(i)})$ belongs to $\mathfrak{D}_b$ for all i and such that $w = w^{(0)}\tilde{s}_0 w^{(1)}\tilde{s}_1 \ldots w^{(|s|-1)}\tilde{s}_{|s|-1} w^{(|s|)}$. For $s = \varepsilon$, this property is obvious.

Assuming this property is true for words of P_b^n, for a fixed $n \geq 0$, let us fix a word s in P_b^{n+1} and let w be a word of $[\![0, 2b-1]\!]^*$ such that $\mathcal{D}_b(\hat{s}w)$ belongs to $\mathfrak{D}_b$. Let us define the word u as the shortest prefix of w such that $\mathcal{D}_b(\hat{s}_n u)$ belongs to $\mathfrak{D}_b$. If $\mathcal{D}_b(\hat{s}_n) = p_i$ then the last letter of u must be $\overline{p}_i$, that is $\mathcal{D}_b(\tilde{s}_0)$ and if we note $w^{(0)} = \mathrm{Pref}_{|u|-1}(u)$ (this word can be the empty word ε), then the word $\mathcal{D}_b\left(w^{(0)}\right)$ also belongs to $\mathfrak{D}_b$.

Moreover, if we note $w' = \mathrm{Suff}_{|w|-|u|}(w)$ and $s' = \mathrm{Pref}_n(s)$, so that $\hat{s}w = \hat{s'}\hat{s}_n w^{(0)}\tilde{s}_0 w'$, then $\mathcal{D}_b(\hat{s'}w')$ belongs to $\mathfrak{D}_b$. Using the induction hypothesis on the element s', we obtain the following decomposition of w:

$$w = w^{(0)}\tilde{s}_0 w'^{(0)}\tilde{s'}_0 w'^{(1)}\tilde{s'}_1 \dots w'^{(n-1)}\tilde{s'}_{n-1} w'^{(n)},$$

where $w'^{(i)}$ belongs to $[\![0, 2b-1]\!]^*$ and $\mathcal{D}_b(w^{(i)})$ belongs to $\mathfrak{D}_b$ for every $i \in [\![0, n]\!]$.

As $\tilde{s'}_i = \tilde{s}_{i+1}$, there exists $n+2$ words $w^{(i)}$ in $[\![0, 2b-1]\!]^*$, such that $\mathcal{D}_b(w^{(i)})$ belongs to $\mathfrak{D}_b$ for every $i \in [\![0, n+1]\!]$ and satisfying:

$$w = w^{(0)}\tilde{s}_0 w^{(1)}\tilde{s}_1 \dots w^{(n)}\tilde{s}_n w^{(n+1)}.$$

This concludes the induction process and ends the proof of Lemma 3.6. ■

Notations 3.7. For a pair of positive integers (α, β), we call $\mathcal{U}_b(\alpha, \beta)$ the set of $(\beta+1)$-uples $(w^{(0)}, w^{(1)}, \dots, w^{(\beta)})$ of words over the alphabet $[\![0, 2b-1]\!]$, such that $|w^{(0)}w^{(1)}w^{(\beta)}| = 2\alpha$ and for all i in $[\![0, \beta]\!]$, $\mathcal{D}(w^{(i)})$ belongs to $\mathfrak{D}_b$, that is

$$\mathcal{U}_b(\alpha, \beta) = \left\{ \left(w^{(0)}, w^{(1)}, \dots, w^{(\beta)}\right),\ \mathcal{D}_b\left(w^{(i)}\right) \in \mathfrak{D}_b,\ \sum_{i=0}^{\beta} |w^{(i)}| = 2\alpha \right\}.$$

We also denote $N_b(k, l) = \mathrm{Card}\left(\mathcal{U}_b(\frac{k-l}{2}, l)\right)$, for all integers $k \geq 3$ and all integers $l \leq k$ with the same parity as k.

Definition 3.8. Let $w = w_0 w_1 \dots w_l$ be a finite word over an alphabet A and n be an integer of $[\![0, l]\!]$. We say that *a letter a of A occurs at rank n in the word w* if $w_n = a$.

Proposition 3.9. *Let $b \geq 1$. Let us fix an integer $k \geq 3$ and an element $s \in P_b^l$ where $l = k - 2a$ (with $a \leq \lfloor \frac{k}{2} \rfloor$).*

The letter 1 occurs in $\delta^k(s)$ at rank $n \in [\![0, (2b)^k - 1]\!]$ if and only if $a \geq 0$ and there exists $(l+1)$ words $w^{(0)}, w^{(1)}, \dots, w^{(l)}$ of $\mathcal{U}_b(a, l)$ such that $n = \left[w^{(0)}\tilde{s}_0 w^{(1)}\tilde{s}_1 \dots w^{(l-1)}\tilde{s}_{l-1} w^{(l)}\right]_{2b}$.

Proof of Proposition 3.9. Let us fix an integer $l = k - 2a$ with $a \leq \lfloor \frac{k}{2} \rfloor$ and an element s in P_b^l. Let n be an integer of $[\![0, (2b)^k - 1]\!]$ and the word $w \in [\![0, 2b-1]\!]^k$ defined by $w = 0^{k-|\rho_{2b}(n)|}\rho_{2b}(n)$ so that $n = [w]_{2b}$ and the letter of ${}^b\overline{\delta}^k(s)$ occurring at rang n is ${}^b\overline{\delta}_w(s)$.

We first notice that $\delta^k(s)_n = 1$ if and only if ${}^b\overline{\delta}_w(s) = \varepsilon$.

If we have $\delta^k(s)_n = 1$, then ${}^b\overline{\delta}_w(s) = \varepsilon$ because s_0 cannot be reached from any element s of P_b^*. As $s = {}^b\overline{\delta}_{\hat{s}}(\varepsilon)$, we also have ${}^b\overline{\delta}_w \circ {}^b\overline{\delta}_{\hat{s}}(\varepsilon) = \varepsilon$, that is ${}^b\overline{\delta}_{\hat{s}w}(\varepsilon) = \varepsilon$. As a consequence, the word $\mathcal{D}_b(\hat{s}w)$ is a well parenthesized expression and Lemma 3.6 gives the general form of w: there exists $(l+1)$ words $w^{(0)}, w^{(1)}, \dots, w^{(l)}$ of $\mathcal{U}_b(a, l)$ such that $w = w^{(0)}\tilde{s}_0 w^{(1)}\tilde{s}_1 \dots w^{(l-1)}\tilde{s}_{l-1} w^{(l)}$. In particular, when $|s| > k$, any word w can have this property as $|\tilde{s}| > k$ and $|w| = k$.

On the other hand, if $0 \leq a \leq \lfloor \frac{k}{2} \rfloor$ and there exists $(l+1)$ words $w^{(0)}, w^{(1)}, \dots, w^{(l)}$ of $\mathcal{U}_b(a, l)$ such that

$$w = w^{(0)}\tilde{s}_0 w^{(1)}\tilde{s}_1 \dots w^{(l-1)}\tilde{s}_{l-1} w^{(l)},$$

then it follows that ${}^b\overline{\delta}_w(s) = {}^b\overline{\delta}_{w^{(0)}\tilde{s}_0 w^{(1)}\tilde{s}_1 \dots w^{(l-1)}\tilde{s}_{l-1} w^{(l)}}(s) = {}^b\overline{\delta}_{w^{(l)}} \circ {}^b\overline{\delta}_{\tilde{s}_{l-1}} \circ {}^b\overline{\delta}_{w^{(l-1)}} \circ \dots \circ {}^b\overline{\delta}_{w^{(1)}} \circ {}^b\overline{\delta}_{\tilde{s}_0} \circ {}^b\overline{\delta}_{w^{(0)}}(s)$.

As words $\mathcal{D}_b\left(w^{(i)}\right)$ are well parenthesized expressions, the functions ${}^b\overline{\delta}_{w^{(i)}}$ are all equal to the identity function over P_b^*, so ${}^b\overline{\delta}_w(s) = {}^b\overline{\delta}_{\tilde{s}_{l-1}} \circ \dots \circ {}^b\overline{\delta}_{\tilde{s}_1} \circ {}^b\overline{\delta}_{\tilde{s}_0}(s) = {}^b\overline{\delta}_{\tilde{s}}(s)$. From Remarks 3.5, we get ${}^b\overline{\delta}_w(s) = \varepsilon$ and further $\delta^k(s)_n = 1$. ■

Corollary 3.10. *Let $b \geq 1$.*

Let us fix some integer $k \geq 3$ and an element s of P_b^*.

1. *If $|s| = k$ ($a = 0$), the word $\delta^k(s)$ contains a single occurrence of the letter 1 which occurs at rank $[\tilde{s}]_{2b}$ and if $|s| < k$ ($a > 0$), the word $\delta^k(s)$ holds at least two occurrences of the letter 1.*
2. *For all integers $k \geq 3$ and all integer $l \leq k$ with the same parity as k, the set $\mathcal{U}_b\left(\frac{k-l}{2}, l\right)$ provides an indexation of the ranks of occurrences of 1 in the words $\delta^k(s)$. This index only depends on k and l, so the number $|\delta^k(s)|_1$ of letters 1 in $\delta^k(s)$ is the same for all s in P^l.*

Lemma 3.11. *Let $b \geq 1$. For an integer $k \geq 3$ and an element $s \in P_b^l$ with $l \leq k$ of same parity as k, the number of occurrences of the letter 1 in $\delta^k(s)$ is given by $N_b(k, l)$ where*

$$N_b(k, l) = \mathit{Card}\left(\mathcal{U}_b\left(\frac{k-l}{2}, l\right)\right) = b^{\frac{k-l}{2}} \frac{l+1}{k+1} \binom{k+1}{\frac{k-l}{2}}.$$

Proof of Lemma 3.11. Let us begin the proof by recalling some combinatorial results. The number of well parenthesized expressions of length 2α, α in $\mathbb{N}$, over b sorts of parenthesis is $b^\alpha \mathcal{C}_\alpha$ where $\mathcal{C}_\alpha = \frac{1}{\alpha+1}\binom{2\alpha}{\alpha}$ is the α-th Catalan number. For all pairs of integers (α, β), we also get:

$$\operatorname{Card}\left(\mathcal{U}_b(\alpha,\beta)\right) = b^\alpha \sum_{n_0+\cdots+n_\beta=\alpha} \prod_{j=0}^{\beta} \mathcal{C}_{n_j}.$$

In order to obtain a simple expression for $N_b(\alpha, \beta)$, we use some methods from analytic combinatorics. For an introduction to these methods, one can consult the book by Flajolet and Sedgewick [13]. For a given pair of positive integers (α, β), we set:

$$C(\alpha,\beta) = \sum_{n_0+\cdots+n_\beta=\alpha} \prod_{j=0}^{\beta} \mathcal{C}_{n_j}.$$

We also consider the generating series of Catalan numbers, noted κ:

$$\kappa(t) = \sum_{n\geq 0} \mathcal{C}_n t^n.$$

This series satisfies the equation $\kappa(t) = 1+t\kappa(t)^2$ and $C(\alpha,\beta) = [t^\alpha]\,\kappa(t)^{\beta+1}$, where $[t^\alpha]\,\kappa(t)^{\beta+1}$ represents the coefficient of t^α in $\kappa(t)^{\beta+1}$, so computing the $C(\alpha,\beta)$ is equivalent to computing the coefficients of $\kappa(t)^{\beta+1}$.

The trick is to compute coefficients of the series $(\kappa(t)-1)^{\beta+1}$. Indeed, the series $K(t) = \sum_{n\geq 1} \mathcal{C}_n t^n$ satisfies the equation $K(t) = t(K(t)+1)^2$, which is a Lagrangian equation: it can be written under the form $K(t) = tf(K(t))$ for $f(x) = (x+1)^2$. In this context, we can use the inversion theorem, so we obtain for all pairs (α, n) of positive integers:

$$[t^\alpha]\,K(t)^n = \frac{n}{\alpha}\left[t^{\alpha-n}\right] f(t)^\alpha = \frac{n}{\alpha}\binom{2\alpha}{\alpha-n}.$$

From the functional equations satisfied by κ and K, we also get that $K(t) = t\kappa(t)^2$ and it allows us to obtain:

$$\kappa(t)^{2n} = \frac{K(t)^n}{t^n} \quad \text{and} \quad \kappa(t)^{2n+1} = \frac{K(t)^{n+1}}{t^n} + \frac{K(t)^n}{t^n}.$$

thus, we get

$$[t^\alpha]\,\kappa(t)^{2n} = \left[t^{\alpha+n}\right] K(t)^n = \frac{n}{\alpha+n}\binom{2(\alpha+n)}{m}$$

and

$$[t^\alpha]\,\kappa(t)^{2n+1} = \left[t^{\alpha+n}\right] K(t)^{n+1} + \left[t^{\alpha+n}\right] K(t)^n = \frac{n+1}{\alpha+n}\binom{2(\alpha+n)}{\alpha-1} + \frac{n}{\alpha+n}\binom{2(\alpha+n)}{\alpha}.$$

Simplifying, we obtain, for every integer n,

$$[t^\alpha]\,\kappa(t)^{2n+1} = \frac{2n+1}{2\alpha+2n+1}\binom{2(\alpha+n)+1}{\alpha},$$

Furthermore, for all pairs (α, β) of positive integers, we get:

$$C(\alpha,\beta) = [t^\alpha]\,\kappa(t)^{\beta+1} = \frac{\beta+1}{2\alpha+\beta+1}\binom{2\alpha+\beta+1}{\alpha},$$

and the expression of the $N_b(k,l)$ follows: $N_b(k,l) = b^{\frac{k-l}{2}} \frac{l+1}{k+1}\binom{k+1}{\frac{k-l}{2}}$. ■

Lemma 3.12. *for all $b \geq 2$ and $k \geq 4$, the application $\phi_k : a \mapsto N_b(k-1, k-2a-1)$ is strictly increasing on $[\![1, \lfloor \frac{k}{2} \rfloor - 1]\!]$.*

This technical lemma will provide that special factors we are going to display are all different from each other.

Proof of Lemma 3.12. For $k = 4$ and $k = 5$, ϕ_k is strictly increasing as $[\![1, \lfloor \frac{k}{2} \rfloor - 1]\!]$ is then reduced to the singleton $\{1\}$.

For $k \geq 6$, we form the quotient $\frac{\phi_k(a+1)}{\phi_k(a)}$ for a in $[\![1, \lfloor \frac{k}{2} \rfloor - 2]\!]$:

$$\frac{\phi_k(a+1)}{\phi_k(a)} = \frac{N_b(a+1, k-1-2(a+1))}{N_b(\beta, k-1-2a)} = b\frac{(k-2\beta-2)(k-\beta)}{(k-2\beta)(\beta+1)},$$

and we consider the following function:

$$\varphi : \begin{array}{ccc} \left[1, \lfloor \frac{k}{2} \rfloor - 2\right] & \to & \mathbb{R} \\ x & \mapsto & \frac{(k-2x-2)(k-x)}{(k-2x)(x+1)} \end{array}$$

we have:

$$\varphi'(x) = \frac{4k\left(-4x^2 + 4(k-1)x - k^2 + k - 2\right)}{(k-2x)^2(x+1)^2}.$$

The function φ' is negative on $[\![1, \lfloor \frac{k}{2} \rfloor - 2]\!]$, so φ is a decreasing function and for all integers β in $[\![1, \lfloor \frac{k}{2} \rfloor - 2]\!]$, we obtain:

$$\frac{N_b(\beta+1, k-1-2(\beta+1))}{N_b(\beta, k-1-2\beta)} \geq bf\left(\lfloor \frac{k}{2} \rfloor - 2\right) > \frac{b}{2},$$

whatever the parity of k and it implies that φ_k is strictly increasing on $[\![1, \lfloor \frac{k}{2} \rfloor - 1]\!]$ when $b \geq 2$. ■

In Proposition 3.9 and Lemma 3.11, we have established how many occurrences of the letter 1 the words $\delta^k(s)$ hold. Now, we have to know more about where these occurrences take place. This is the aim of the following lemma.

Lemma 3.13. *Let $b \geq 1$.*
Let us fix an integer $k \geq 2$, an element $s \in P_b^l$ for $l = k - 2a$ with a in $[\![1, \lfloor \frac{k}{2} \rfloor]\!]$ and an integer i in $[\![1, l-1]\!]$.

1. *For every element $(w^{(i+1)}, w^{(i+2)}, \ldots, w^{(l)})$ of $\mathcal{U}_b(a, l-i-1)$,*
$$\left[\tilde{s}_0 \ldots \tilde{s}_i w^{(i+1)} \tilde{s}_{i+1} w^{(i+2)} \ldots w^{(l-1)} \tilde{s}_{l-1} w^{(l)}\right]_{2b} \leq \left[\tilde{s}_0 \ldots \tilde{s}_i (2b-1)^a 0^a \tilde{s}_{i+1} \ldots \tilde{s}_{l-1}\right]_{2b}.$$
2. *If $(w^{(0)}, w^{(1)}, \ldots, w^{(l)})$ is in $\mathcal{U}_b(a, l)$ with at least one non empty word $w^{(j)}$ among the i first ones, then*
$$\left[\tilde{s}_0 \ldots \tilde{s}_i (2b-1)^a 0^a \tilde{s}_{i+1} \ldots \tilde{s}_{l-1}\right]_{2b} \leq \left[w^{(0)} \tilde{s}_0 w^{(1)} \tilde{s}_1 \ldots w^{(l-1)} \tilde{s}_{l-1} w^{(l)}\right]_{2b}.$$
3. *When $a \geq 1$, there are no occurrences of the letter 1 in $\delta^k(s)$ between the letters 1 occurring at ranks $\left[\tilde{s}_0 \ldots \tilde{s}_i (2b-1)^a 0^a \tilde{s}_{i+1} \ldots \tilde{s}_l\right]_{2b}$ and $\left[\tilde{s}_0 \ldots \tilde{s}_{i-1} (b)(b-1) \tilde{s}_i \ldots \tilde{s}_l ((b)(b-1))^{a-1}\right]_{2b}$.*
4. *The rank where the $N_b(k-i-1, l-i-1)$-th the letter 1 occurs in $\delta^k(s)$ is $\left[\tilde{s}_0 \ldots \tilde{s}_i (2b-1)^a 0^a \tilde{s}_{i+1} \ldots \tilde{s}_l\right]_{2b}$.*

Proof of Lemma 3.13. The first three items of this lemma easily follow from the two basic facts detailed below.

The words $\tilde{s}$ belong to $[\![0, b-1]\!]^*$ and all the words which $\mathcal{D}_b$ maps on non empty well parenthesized expressions are words of $[\![0, 2b-1]\!]^*$ starting with a letter of $[\![b, 2b-1]\!]$.

The second fact used is that, among the words w over $[\![0, 2b-1]\!]^*$ of length less than $2a$ which $\mathcal{D}_b$ maps on well parenthesized expressions, the one giving the greatest integer $[w]_{2b}$ is $w = (2b-1)^a 0^a$.

The fourth item can be deduced from previous items and the definition of the integers $N_b(\cdot, \cdot)$. ■

We are now ready to display the right special factors of $m^{(b)}$. By counting these special factors, we will find a lower bound of $p_{m^{(b)}}(n+1) - p_{m^{(b)}}(n)$ for sufficiently large positive integers n and, by adding these inequalities, we will obtain a lower bound for the complexity function $p_{m^{(b)}}(n)$.

Lemma 3.14. *Let $b \geq 2$.*
For every element s in P_b^*, the word $(x)(x)(s)$ occurs in $\overline{m}^{(b)}$.

Proof of Lemma 3.14. We first notice that every element s of P_b^* occurs in $\overline{m}^{(b)}$, for every $b \geq 2$.

If $b = 2$, the letter (p_1p_0s) occurs in $\overline{m}^{(2)}$ for every element s of P_2^*. As $\overline{m}^{(2)}$ is the fixed point of ${}^2\overline{\delta}$, the words ${}^2\overline{\delta}(p_1p_0s)$ and ${}^2\overline{\delta}^2(p_1p_0s)$ also occur in $\overline{m}^{(2)}$. As $\mathrm{Pref}_5\left({}^2\overline{\delta}^2(p_1p_0s)\right) = (x)^4(s)$, the word $(x)(x)(s)$ occurs in $\overline{m}^{(2)}$.

If $b \geq 3$, the letter $(p_{b-1}s)$ occurs in $\overline{m}^{(b)}$ for every element s of P_b^*. As $\overline{m}^{(b)}$ is the fixed point of ${}^b\overline{\delta}$, words ${}^b\overline{\delta}(p_1p_0s)$ also occur in $\overline{m}^{(b)}$. As $\mathrm{Pref}_b\left({}^b\overline{\delta}(p_{b-1}s)\right) = (x)^{b-1}(s)$, the word $(x)(x)(s)$ occurs in $\overline{m}^{(b)}$. ■

Lemma 3.15. *Let $b \geq 2$.*
Let us fix an integer $k \geq 3$, and an integer n satisfying $(2b)^k \leq n < (2b)^{k+1}$.
For each element $s \in P_b^{l-1}$ with $l = k - 2a$, where a belongs to $[\![1, \lfloor \frac{k}{2} \rfloor - 1]\!]$, and each integer j of $[\![1, b-1]\!]$, the word $\delta^k((x)(x)(p_js))$ contains a special factors of length n denoted $u^{(n,s,j)}$.

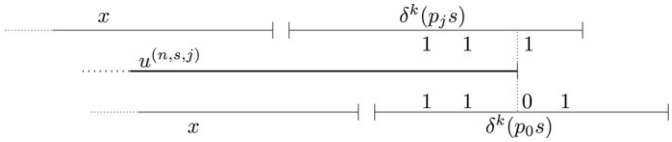

Fig. 3. Construction of the right special factor $u^{(s,n,j)}$.

Proof of Lemma 3.15. Let s be an element in P_b^{l-1} for $l = k - 2a$ with a in $[\![0, \lfloor \frac{k}{2} \rfloor - 1]\!]$ and j be an integer in $[\![1, b-1]\!]$.

From to Proposition 3.9, the following set of integers is the set of ranks where the letter 1 occurs in $\delta^k(p_0 s)$:

$$\mathcal{R}_k(p_0 s) = \left\{ \left[w^{(0)} 0 w^{(1)} s_0 w^{(2)} \dots w^{(l-1)} \tilde{s}_{l-1} w^{(l)} \right]_{2b},\ (w^{(0)}, w^{(1)}, \dots, w^{(l)}) \in \mathcal{U}_b\,(a, l) \right\},$$

and the following set of integers is the set of ranks where the letter 1 occurs in $\delta^k(p_j s)$:

$$\mathcal{R}_k(p_j s) = \left\{ \left[w^{(0)} j w^{(1)} s_0 w^{(2)} \dots w^{(l-1)} \tilde{s}_{l-1} w^{(l)} \right]_{2b},\ (w^{(0)}, w^{(1)}, \dots, w^{(l)}) \in \mathcal{U}_b\,(a, l) \right\}.$$

Moreover, of all l-uple $(w^{(1)}, \dots, w^{(l)})$ of $\mathcal{U}_b(a, l-1)$, we have the following equality:

$$\left[j w^{(1)} s_0 w^{(2)} \tilde{s}_1 \dots w^{(l-1)} \tilde{s}_{l-1} w^{(l)} \right]_{2b} - \left[0 w^{(1)} s_0 w^{(2)} \tilde{s}_1 \dots w^{(l-1)} \tilde{s}_{l-1} w^{(l)} \right]_{2b} = j(2b)^{k-1}.$$

Thus, using Lemma 3.13, the ranks where the $N_b(k-1, l-1)$ first occurrences of the letter 1 in $\delta^k(p_j s)$ can be deduced from the ranks where the $N_b(k-1, l-1)$ first occurrences of the letter 1 in $\delta^k(p_0 s)$ by adding $j(2b)^{k-1}$.

On the other hand, as we have

$$\left[b(b-1) j \tilde{s} (b(b-1))^{a-1} \right]_{2b} - \left[b(b-1) 0 \tilde{s} (b(b-1))^{a-1} \right]_{2b} = j(2b)^{k-3},$$

so the rank of the $(N_b(k-1, l-1)+1)$-th occurrence of the letter 1 in $\delta^k(p_j s)$ (according to Lemma 3.13, rank $\left[b(b-1) j \tilde{s} (b(b-1))^{a-1} \right]_{2b}$ according to Lemma 3.13) can be deduced from the rank of occurrence of the $(N_b(k-1, l-1)+1)$-th letter 1 in $\delta^k(p_0 s)$ (i. e. rank $\left[b(b-1) 0 \tilde{s} (b(b-1))^{a-1} \right]_{2b}$) by adding $j(2b)^{k-3}$.

These facts allow us to build a special factor parameterized by the element s of P_b^{l-1} and the integer $j \in [\![1, b-1]\!]$. For all $s \in P_b^{l-1}$, we define the word $u^{(n,s,j)}$ of length n extracted from $\delta^k(x)\delta^k(x)\delta^k(p_j s)$ so that the last letter of $u^{(n,s,j)}$ is the $\left[b(b-1) j \tilde{s} (b(b-1))^{a-1} \right]_{2b}$-th letter of $\delta^k(p_j s)$ (see Fig. 3), that is, if we set $\delta^k(x)\delta^k(x)\delta^k(p_j s) = v_0 v_1 \dots v_{3(2b)^k - 1}$ and $R_b(s, j) = \left[b(b-1) j \tilde{s} (b(b-1))^{a-1} \right]_{2b}$, then

$$u^{(n,s,j)} = v_{2(2b)^k + R_b(s,j) - 2 - n} v_{2(2b)^k + R_b(s,j) - 1 - n} \cdots v_{2(2b)^k + R_b(s,j) - 2}.$$

The words $u^{(n,s,j)}$ are right special factors of $m^{(b)}$. Indeed, as $(x)(x)(p_0 s)$ and $(x)(x)(p_j s)$ occur in $\overline{m}^{(b)}$, the words $\delta^k(x)\delta^k(x)\delta^k(p_0 s)$ and $\delta^k(x)\delta^k(x)\delta^k(p_j s)$ occur in $m^{(b)}$ and the word $u^{(n,s,j)}$ can be extended on the right by the letter 1 as a subword of $\delta^k(x)\delta^k(x)\delta^k(p_j s)$ and by the letter 0 as a subword of $\delta^k(x)\delta^k(x)\delta^k(p_0 s)$. ■

Lemma 3.16. *Let $b \geq 2$.*

Let $k \geq 4$ be an integer and n be an integer satisfying $(2b)^k \leq n < (2b)^{k+1}$. The function $\mathcal{S}_n$, defined below is injective.

$$\begin{array}{llll} \mathcal{S}_n : & \bigcup_{a \in [\![1, q-1]\!]} (P_b \setminus \{p_0\}) \times P_b^{k-2a-1} & \rightarrow & \{0, 1\}^n \\ & (n, s, j) & \mapsto & u^{(n,s,j)} \end{array}$$

Proof of Lemma 3.16 . Let $u^{(n,s,j)}$ and $u^{(n,s',j')}$ be two words of $\mathcal{S}_n \left(\bigcup_{a \in [\![1, q-1]\!]} (P_b \setminus \{p_0\}) \times P_b^{k-2a-1} \right)$.

Assume $u^{(n,s,j)} = u^{(n,s',j')}$. As the letter 1 occurs $N_b(k-1, |s|)$ times in $u^{(n,s,j)}$, then $N_b(k-1, |s|) = N_b(k-1, |s'|)$ and Lemma 3.12 ensures that $N_b(k-1, |s|)$ injectively determines the integer $|s|$ when k is fixed, so $|s| = |s'|$.

We now fix $|s| = |s'| = l = k - 2a - 1$. As the first occurrence of 1 in $u^{(n,s,j)}$ is also the first occurrence of 1 in $\delta^k(p_j s)$, the number of letters from this first occurrence of 1 to the end of $u^{(n,s,j)}$ is given by:

$$\left[b(b-1) j \tilde{s} (b(b-1))^{a-1} \right]_{2b} - \left[j \tilde{s} (b(b-1))^{a} \right]_{2b} = \left[b(b-1) j' \tilde{s'} (b(b-1))^{a-1} \right]_{2b} - \left[j' \tilde{s'} (b(b-1))^{a} \right]_{2b}.$$

From this equality, we get $[j\tilde{s}]_{2b} = [j'\tilde{s'}]_{2b}$ and furthermore $j = j'$ (as j and j' are different from 0) and $s = s'$.

Thus, the number of occurrences of the letter 1 and their repartition in $u^{(n,s,j)}$ uniquely characterize the elements s and j so $\mathcal{S}_n$ is injective. ■

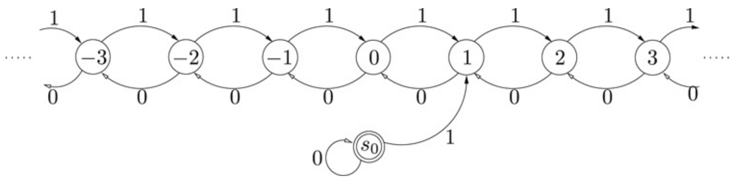

Fig. 4. Automaton generating the drunken man infinite words.

Proof (**Lower bounds of $p_{m^{(b)}}(n)$ when** $b \geq 2$). Let us fix an integer $k \geq 4$, $k = 2q + r$ with r in $\{0, 1\}$ and an integer n satisfying $(2b)^k \leq n < (2b)^{k+1}$.

In Lemma 3.15, we have identified a family of right special factors $u^{(n,s,j)}$ of length n. From Lemma 3.16, the displayed special factors are different when s varies in P_b^{l-1}, when $a = \frac{k-l}{2}$ varies in $[\![1, \lfloor\frac{k}{2}\rfloor - 1]\!]$ and when j varies in $[\![1, b-1]\!]$.

The family of special factors $\left\{u^{(n,s,j)}\right\}$ is in bijection with the set $\cup_{a\in[\![1,q-1]\!]} (P_b \setminus \{p_0\}) \times P_b^{k-2a-1}$, so, we have at least $\frac{b^{k-1}-b^{r+1}}{b+1}$ special factors of length n in $m^{(b)}$. Then, we obtain

$$\forall k \geq 4, \forall n \in [\![(2b)^k, (2b)^{k+1} - 1]\!], \quad p_{m^{(b)}}(n+1) - p_{m^{(b)}}(n) \geq \frac{b^{k-1} - b^2}{b+1}.$$

Adding these inequalities, we get for all integers $k \geq 5$ and for every integer $n \in [\![(2b)^k, (2b)^{k+1} - 1]\!]$,

$$p_{m^{(b)}}(n) - p_{m^{(b)}}\left((2b)^4\right) \geq \frac{1}{b+1}\sum_{i=4}^{k-1}(2b)^i(b^{i-1} - b^2) + \frac{1}{b+1}\left(n - (2b)^k\right)(b^{k-1} - b^2),$$

and it follows, for all $k \geq 5$ and all n in $[\![(2b)^k, (2b)^{k+1} - 1]\!]$:

$$p_{m^{(b)}}(n) - p_{m^{(b)}}\left((2b)^4\right) \geq \frac{1}{b+1}\left((b^{k-1} - b^2)\left(n - (2b)^k\right) + \frac{\left((2b^2)^k - (2b^2)^4\right)}{b(2b^2-1)} + b^2\frac{\left((2b)^k - (2b)^4\right)}{2b-1}\right).$$

With lower bounds $\frac{n^{\log_{2b} b}}{b}$ for b^k and $\frac{n}{2b}$ for $(2b)^k$, we obtain, for all $n \geq (2b)^4$:

$$p_{m^{(b)}}(n) \geq An^{1+\log_{2b} b} + Bn + C,$$

where the constants A, B and C only depend on b. This concludes the proof of Theorem 2.5 for $b \geq 2$ as we proved $n^{1+\log_{2b} b} = \mathcal{O}\left(p_{m^{(b)}}(n)\right)$. ■

4. Complexity function growth order of $p_{m^{(1)}}(n)$

4.1. About the difference of behaviour

As we will see, the combinatorial and statistical properties of the word $m^{(1)}$ are closer to the drunken man infinite words properties than to the words $m^{(b)}$ ones, when $b \geq 2$. Drunken man infinite words have been presented in [16]. Indeed, the automaton generating drunken man infinite words (see Fig. 4) is very similar to the automaton generating $m^{(1)}$, except for the trash state x, of course.

The difference between the behaviour of $m^{(1)}$ and $m^{(b)}$ for $b \geq 2$ comes from the differences between the transition graphs associated with pushdown automata generating them. Indeed, for $b \geq 2$, the number of states of the transition graph at fixed length from the final states grows exponentially, whereas it grows linearly for $b = 1$. Using substitutions, it can be translated as follows: the number of letters s in $\mathcal{A}_b$ such that the iterate ${}^b\overline{\delta}^k(s)$ contains some occurrences of the letter 1 grows linearly for $b = 1$ whereas it grows exponentially for $b \geq 2$ (as functions of k).

However, the proofs regarding case $b = 1$ use the same process as for case $b \geq 2$. The methods used to find upper bounds of complexity functions in cases $b = 1$ and $b \geq 2$ are the same but to obtain lower bounds in case $b = 1$, we need to exhibit a family of special factors which is really different from the displayed families when $b \geq 2$. The family of special factors of $m^{(1)}$ we are going to show is quite close to the family of special factors of drunken man infinite words, displayed in [16].

Moreover, despite using quite heavy notations when $b = 1$, some propositions and lemmas detailed for case $b \geq 2$ in Section 3 are also available for $b = 1$. To lighten the arguments and the proofs of this section, we will use simplifications of notations introduced in Notations 4.1.

Notations 4.1. We identify P_1^* with $\mathbb{N}$, A_1 with $\mathbb{N}\cup\{s_0,x\}$ so that:

$$\begin{array}{rrcl} {}^1\overline{\delta}: & A_1 & \to & A_1^2 \\ & x & \mapsto & (x)^2 \\ & s_0 & \mapsto & (s_0)(1) \\ & 0 & \mapsto & (x)(1) \\ & n,\ n>0 & \mapsto & (n-1)(n+1). \end{array}$$

The fixed point of this substitution is:

$$\overline{m}^{(1)} = s_0(1)(0)(2)(x)(1)(1)(3)(x)^2(0)(2)(0)(2)(2)(4)(x)^5(1)(1)(3)(x)(1)(1)(3)(1)(3)(3)(5)(x)^{10}\\ (0)(2)(0)(2)(2)(4)(x)^2(0)(2)(0)(2)(2)(4)(0)(2)(2)(4)(2)(4)(4)(6)(x)^{21}(1)(1)(3)(x)(1)(1)(3)(1)\\ (3)(3)(5)(x)^5(1)(1)(3)(x)(1)(1)(3)(1)(3)(3)(5)(x)(1)(1)(3)(1)(3)(3)(5)(1)(3)(3)(5)(3)(5)(5)(7)\\ (x)^{42}(0)(2)(0)(2)(2)(4)(x)^2(0)(2)(0)(2)(2)(4)(0)(2)(2)(4)(2)(4)(4)(6)(x)^{10}(0)(2)(0)(2)(2)(4)\ldots$$

and if we define $\Pi_1 : \mathbb{N}\cup\{s_0,x\} \to \{0,1\}$ by $\Pi_1^{-1}(\{1\}) = \{s_0, 0\}$, then $m^{(1)} = \Pi_1\left(\overline{m}^{(1)}\right)$.

4.2. An upper bound of $p_{m^{(1)}}(n)$

To find an upper bound of the complexity function, we use the following two lemmas which are the equivalent to Lemmas 3.2 and 3.3 for case $b = 1$.

Lemma 4.2. *For a given positive integer k, the word $\delta^k(s)$ contains occurrences of 1 if and only if $s = s_0$ or $s \in \{k-2i,\ i \in [\![0, \lfloor \frac{k}{2} \rfloor]\!]\}$, so we have $\lfloor \frac{k}{2}\rfloor + 1$ letters of A_1 such that $\delta^k(s)$ is different from 0^{2^k}.*

Lemma 4.3. *The word x_1x_2 is a factor of $\overline{m}^{(1)}$ if and only if:*

1. $x_1x_2 \in \{(s_0)(1), (x)(x), (x)(1), (x)(0), (0)(2), (1)(0)\}$,
2. $\exists n \in \mathbb{N}$, $x_1x_2 = (n)(x)$,
3. $\exists s \in \mathbb{N}$, $s \geq 1$, $\exists q \in [\![-1, \lfloor \frac{n}{2} \rfloor]\!]$, $x_1x_2 = (s)(s-2q)$,

These two propositions can be proved using the same arguments as for Lemmas 3.2 and 3.3. Using these two results in the same way as for $b \geq 2$, we can show that:

$$\forall n \geq 1, \quad p_{m^{(1)}}(n) \leq \frac{n}{4}\left((\log_2 n)^2 + 22\,(\log_2 n + 2)\right),$$

which implies that $p_{m^{(1)}}(n) = \mathcal{O}\left(n(\log_2 n)^2\right)$.

4.3. A lower bound of $p_{m^{(1)}}(n)$

To obtain a lower bound of the complexity function of $\overline{m}^{(1)}$, we will use some results presented in Section 3.2. Indeed, Proposition 3.9, Lemmas 3.11 and 3.13 are valid for $b = 1$. With new notations, Proposition 3.9 and Lemma 3.11 become the following two results.

Proposition 4.4. *Let us fix an integer $k \geq 3$ and an element $s = k - 2a$ of $\mathbb{N}$ with ($a \leq \lfloor \frac{k}{2} \rfloor$).*

The letter 1 occurs in $\delta^k(s)$ at rank $n \in [\![0, 2^k - 1]\!]$ if and only if $a \geq 0$ and there exists $(s+1)$ words $w^{(0)}, w^{(1)}, \ldots, w^{(s)}$ of $\mathcal{U}_1(a,s)$ such that $n = \left[w^{(0)}0w^{(1)}0\ldots w^{(s-1)}0w^{(s)}\right]_2$.

Lemma 4.5. *Let $b \geq 1$. For an integer $k \geq 3$ and an element $s = k-2a$ of $\mathbb{N}$ with $0 \leq a \leq \lfloor \frac{k}{2} \rfloor$, the number of occurrences of the letter 1 in $\delta^k(s)$ is given by $N_1(k,s)$ where*

$$N_1(k,s) = Card\left(\mathcal{U}_1(a,s)\right) = \frac{s+1}{k+1}\binom{k+1}{a}.$$

From Proposition 4.4 and Lemma 3.13, we also get the following proposition:

Proposition 4.6. *Let $k \geq 2$ be a positive integer and $s = k - 2a$ be an element of $\mathbb{N}$ for a in $[\![1, \lfloor \frac{k}{2} \rfloor]\!]$.*

The word $\delta^k(s)$ contains at least two occurrences of 1. Moreover, the rank of the first occurrence of the letter 1 in $\delta^k(s)$ is given by $[0^s(10)^a]_2$ and the rank of the last two occurrences of the letter 1 in $\delta^k(s)$ are given by $\left[1^{a-1}010^{k-a-1}\right]_2$ and $\left[1^a0^{k-a}\right]_2$ respectively.

Starting from these three results, we are going to find a family of special factors of $m^{(1)}$. A first remark is that we cannot copy what we have already done for $b \geq 2$. Indeed, special factors displayed in case $b \geq 2$ come from pairs of words of type $\delta^k((x)(x)(s))$ and $\delta^k\left((x)(x)(s')\right)$, where $(s,s') \in P_b$ are of same length, that is s and s' are at same distance from ε. It is not possible to use the same construction when $b = 1$ as there is only one letter s in $P_1 = \mathbb{N}$ at each distance from 0. Fortunately, we can find a family of special factors using the phenomenon presented in the following proposition.

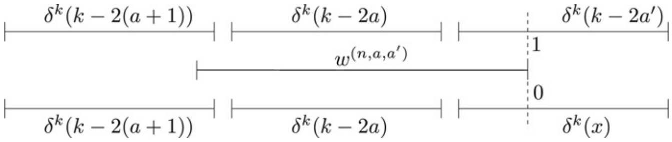

Fig. 5. Construction of the special factor $w^{(n,a,a')}$.

Proposition 4.7. *For all s in $\mathbb{N}$ with $s \geq 2$, for all $q \in [\![0, \lfloor \frac{s}{2} \rfloor]\!]$, the words $(s-2)(s)(s-2q)$ and $(s-2)(s)(x)$ occur in $\overline{m}^{(1)}$.*

This proposition can be shown easily by noticing that the words $(s-1)(s-q+1)$ and $(s-1)(x)$ occur in $\overline{m}^{(1)}$ (see Lemma 4.3) so their images by ${}^1\overline{\delta}$ also occur in $\overline{m}^{(1)}$ as $\overline{m}^{(1)}$ is the fixed point of ${}^1\overline{\delta}$. Their images ${}^1\overline{\delta}((s-1)(s-q+1))$ and ${}^1\overline{\delta}((s-1)(x))$ contain the word $(s-2)(s)(s-2q)$ and the word $(s-2)(s)(x)$ respectively.

From Proposition 4.7, we can display a family of special factors of $\overline{m}^{(1)}$ using pairs of words of type $\delta^k((s-2)(s)(s-2q))$ and $\delta^k((s-2)(s)(x))$. This is very similar to drunken man infinite words case (see [16]).

For all integers $k \geq 2$ and all integers n such that $2^k \leq n < 2^{k+1}$, we construct a special factor $w^{(n,a,a')}$ for each pair of integers (a, a') such that $1 \leq a \leq a' \leq \lfloor \frac{k}{2} \rfloor$.

The factor $w^{(n,a,a')}$ is extracted from $\delta^k\left((k-2(a+1))(k-2a)(k-2a')\right)$ such that the last letter of $w^{(n,a,a')}$ is the letter before the first occurrence of 1 in $\delta^k\left(k-2a'\right)$. The factor $w^{(n,a,a')}$ can also be extracted from $\delta^k((k-2(a+1))(k-2a)(x))$, followed by the letter 0 so the factor $w^{(n,a,a')}$ is a special factor (see Fig. 5).

If the word $w^{(n,a,a')}$ contains at least two occurrences of the letter 1, from Proposition 4.6, the length for the block of letters 0 between the last two occurrences of the letter 1 in $w^{(n,a,a')}$ which are the last two occurrences of the letter 1 in $\delta^k\left(p^{k-2a}\right)$, that is $\left[1^a 0^{k-a}\right]_2 - \left[1^{a-1}010^{k-a-1}\right]_2$ totally characterize a, and the length of the ending block of letters 0 of $w^{(n,a,a')}$, that is $2^k - \left[1^a 0^{k-a}\right]_2 + \left[0^{k-2a}(10)^{a'}\right]_2$ totally characterize a'. To use this argument, we only need to be sure that the last two occurrences of the letter 1 of $\delta^k\left(p^{k-2a}\right)$ actually occur in $w^{(n,a,a')}$, so we need the following technical condition:

$$2^k - \left[1^{a-1}010^{k-a-1}\right]_2 + \left[0^{k-2a'}(10)^{a'}\right]_2 \leq n,$$

that is,

$$2^{k-a} + 2^{k-a-1} + \frac{2^{2a'+3}-2}{3} \leq n.$$

This condition is actually realized if $1 \leq a \leq a' \leq \lfloor \frac{k}{2} \rfloor - 2$ when $k \geq 6$.

Thereby, we have found a family of special factors of length n, given by $\{w^{(n,a,a')},\ 1 \leq a \leq a' \leq \lfloor \frac{k}{2} \rfloor - 2\}$, which are different from each other when $k \geq 6$. As a consequence, $m^{(1)}$ has at least $\frac{\left(\lfloor \frac{k}{2} \rfloor - 2\right)\left(\lfloor \frac{k}{2} \rfloor - 1\right)}{2}$ different special factors of length $n \in [\![2^k, 2^{k+1}-1]\!]$ when $k \geq 6$.

It follows a lower bound of $p_{m^{(1)}}(n+1) - p_{m^{(1)}}(n)$, for all integers $k \geq 6$ and all integers n such that $2^k \leq n < 2^{k+1}$,

$$p_{m^{(1)}}(n+1) - p_{m^{(1)}}(n) \geq \frac{\left(\lfloor \frac{k}{2} \rfloor - 2\right)\left(\lfloor \frac{k}{2} \rfloor - 1\right)}{2} \geq \frac{(k-5)(k-3)}{8},$$

and, by adding these inequalities, we also obtain a lower bound of $p_{m^{(1)}}(n)$, for all integers $n \geq 2^6$, under the form $p_{m^{(1)}}(n) \geq n\left(A(\log_2 n)^2 + B \log_2 n + C\right) + D$, that is $n(\log_2 n)^2 = \mathcal{O}\left(p_{m^{(1)}}(n)\right)$.

5. Open questions

The family presented here was first studied to obtain examples of q^∞-automatic words with complexity more than $n^{1+\varepsilon}$ with $\varepsilon > 0$. In view of the results, the next step is to establish whether or not we can find examples of q^∞-automatic words, with complexity satisfying $p_m(n) \gg n^2$, and also to see if it is actually possible to reach the upper bound announced in [15].

Another motivation for this article is the nature of involved languages of integer expansions, which are context-free. Indeed, it would be interesting to know the possible growth orders of complexity functions of characteristic words of context-free languages in general. This question will probably need other tools, as many context-free languages cannot be constructed with the same kind of countable automata used in this article, for example because of the existence of ε-transitions in the associated pushdown automaton or for determinism issues.

The property of recognizability by countable automata is, in itself, an interesting topic. Indeed, the notion of recognizability by countable automata is different whether we choose to read integer expansions from the most to the least significant digit as we have done in this article, or from the least to the most significant digit (examples presented here are recognizable in both reading directions).

References

[1] B. Adamczewski, Y. Bugeaud, On the complexity of algebraic numbers I. expansions in integer bases, Annals of Math. 165 (2007) 547–565.
[2] J.-M. Autebert, J. Berstel, L. Boasson, Context-free languages and pushdown automata, in: Handbook of formal languages, vol. 1, Springer, 1997, pp. 111–174.
[3] J.-P. Allouche, Sur la complexité des suites infinies, Bull. Belg. Math. Soc. Simon Stevin 1 (2) (1994) 133–143.
[4] J.-P. Allouche, J. Shallit, Automatic Sequences. Theory, Applications, Generalizations, Cambridge University Press, 2003.
[5] J. Cassaigne, Special factors of sequences with linear subword complexity, in: Developments in language theory (Magdeburg, 1995), World Sci. Publishing, 1996, pp. 25–34.
[6] J. Cassaigne, Complexité et facteurs spéciaux, Bull. Belg. Math. Soc. Simon Stevin 4 (1) (1997) 67–88.
[7] D. Caucal, On the regular structure of prefix rewriting, Theoret. Comput. Sci. 106 (1992) 61–86.
[8] D. Caucal, On infinite transition graphs having a decidable monadic theory, Theoret. Comput. Sci. 290 (2003) 79–115.
[9] J. Cassaigne, F. Nicolas, Subword complexity, 2008, Preprint.
[10] A. Cobham, Uniform-tag sequences, Math. Syst. Theory 6 (1972) 164–192.
[11] N. Chomsky, M.-P. Schürtzenberger, The algebraic theory of context-free languages, in: P. Braffort, D. Hirshberg (Eds.), Computer Programming and Formal Systems, North-Holland, 1966, pp. 195–242.
[12] S. Ferenczi, Complexity of sequences and dynamical systems, Discrete Math. 206 (1-3) (1999) 145–154.
[13] P. Flajolet, R. Sedgewick, Analytic combinatorics, Preliminary version, Available on: http://algo.inria.fr/flajolet/Publications/books.html. 2007.
[14] P. Kůrka, Topological and Symbolic Dynamics, in: Cours spécialisés, vol. 11, SMF, 2003.
[15] M. Le Gonidec, Sur la complexité de mots infinis engendrés par des q-automates dénombrables, Ann. Inst. Fourier 56 (7) (2006) 2463–2491.
[16] M. Le Gonidec, Drunken man infinite words complexity, RAIRO/ITA, 2007 (in press).
[17] C. Mauduit, Propriétés arithmétiques des substitutions et automates infinis, Ann. Inst. Fourier 56 (7) (2006) 2525–2549.
[18] D. Muller, P. Schupp, The theory of ends, pushdown automata and second-order logic, Theoret. Comput. Sci. 37 (1985) 51–75.
[19] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, Lecture Notes in Comp. Sci. 172 (1985) 380–389, Automata, Languages and Programming (Antwerp, 1984).
[20] N. Pytheas Fogg, Substitutions in dynamics, arithmetics and combinatorics, in: V. Berthé, S. Ferenczi, C. Mauduit, A. Siegel (Eds.), in: Lect. Notes in Math., vol. 1794, Springer, 2002.
[21] D. Perrin, J.-E. Pin, Infinite Words. Automata, Semi-groups, Logic and Games, in: Pure and Applied Math., vol. 141, Elsevier, 2004.
[22] L. Staiger, Kolmogorov complexity of infinite words, CDMTCS Research Report Series 279 (2006).
[23] C. Stirling, Language theory and infinite graphs, Marktoberdorf summerschool, 2005, Notes for Logical Aspects of Secure Computer Systems.
[24] W. Thomas, An short introduction to infinite automata, in: Proceedings of the International Conference of Developments in Language Theory, in: Lecture Notes in Comput. Sci., vol. 2295, Springer, 2002, pp. 130–144.
[25] G. Viennot, Enumerative combinatorics and algebraic languages, in: Fundamentals of Computation Theory, in: Lecture Notes in Comput. Sci., vol. 199, Springer, 1985, pp. 450–464.