On complexity functions of infinite words associated to generalized Dyck languages

Marion Le Gonidec
Institut de Mathématiques de l'université de Liège
Grande traverse 12, Bat B37
B-4000 Liège
Belgique
M.LeGonidec@ulg.ac.be

Résumé

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 $b \geq 2$.

I Introduction

Infinite words over finite alphabets appear many aeras of mathematics and theoretical informatics. For example, an infinite word $m = m_0m_1m_2\ldots$ over a finite alphabet can be associated to 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 reveals interesting properties of the associated objects (see for example [PF02], [AS03] and [PP04] for overviews).

The structure of an infinite word, and especially the diversity of the factors occuring in an infinite word $m = m_0m_1m_2\ldots$ over a finite alphabet can be well described by a sequence measuring the subword complexity called *complexity function*. This sequence, usually treated as a function and noted p_m, maps an integer n to the number $p_m(n)$ of different factors of length n occuring 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 to the word (see for example [Kůr03]) and it is also linked with the Kolmogorov complexity of the word (see [Sta06]). From the number theory point of view, studying complexity functions can also be useful as, for example, if an infinite word m represents the expansion in some integer base of an algebraic number, its complexity must satisfies $\liminf_{n\to+\infty} \frac{p_m(n)}{n} = +\infty$ (see [AB07]).

Even if complexity functions must fit many simple properties, their computations often remain awkward and their possible behaviours are various, 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 [All94] or [Fer99] for surveys). Indeed, for a given word, there is no universal algorithm to compute its complexity exactly, there only exists methods, often based on the study of special factors of the infinite word and synchronization principles. We refer to [Cas96] and

[Cas97] for formulas holding between special factors and complexity. Special factors are really usefull to compute complexity, moreover when we consider infinite words generated by simple algorithms. Methods involving special factors allow to compute exactly complexity functions in many particular cases but also allow to obtain possible *growth orders* of the complexity functions for some large classes of infinite words, that is functions f such that, for at least one word m of the class, we have $f(n) = \mathcal{O}(p_m(n))$ and $p_m(n) = \mathcal{O}(f(n))$. For example, complexity functions of automatic words have growth orders equal to 1 or n [Cob72] 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 [Pan85].

One can aks, from these results of A. Cobham and J.-J. Pansiot, if we can obtain similar results for infinite words generated by some more complex algorithms as pushdown automata or countable automata. This question can also be expressed from the substitutive point of view, 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 contains characteristic words of context-free languages of integers representations in integer bases generated by deterministic and real time pushdown automata.

This topic is also close from 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 J.-J. Pansiot [Pan85] does not hold for substitutive words. A significant partial result, proved by J. Cassaigne and F. Nicolas [CN08] ensure that, for all integer $k \geq 1$, there exists a substitutive words with complexity function growth order $n^{1+\frac{1}{k}}$.

We have shown in [LG06] 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 some 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.

First examples, made by using a simple countable automaton are presented in [LG07]. Those are examples of infinite words associated to context-free languages of integer binary representations, have complexity functions equivalent to $n(\log_2 n)^2$. Trying to find the possible growth orders of complexity functions for words associated to context-free languages generated by deterministic and real time pushdown automata but also trying to answer the question of the reachability of the upper bound announced in [LG06], we have naturally considered words associated to 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 $\{n^{1+\log_{2b} b},\ b \geq 2\} \cup \{n(\log n)^2\}$.

II Well parenthesed integers and infinite words associated to Dyck's languages

In this section, we build the family $\mathcal{F} = \left\{m^{(b)},\ b \in \mathbb{N}, b \geq 1\right\}$, parametrized by a posivite integer b, of infinite words $m^{(b)}$ over $\{0,1\}$ such that $m^{(b)}$ is the characteristic word of the subset of $\mathbb{N}$ made of integers whose representations in base $2b$ correspond, using a letter-to-letter morphism, to well parenthesed expressions over b sorts of parenthesis.

In the following, we will use usual notations about infinite words (See for example [PF02]) and we denote, for some integer q greater than 2, by the word $\rho_q(n) = n_l \dots n_1 n_0$ the q-ary representation of the integer $n = \sum_{i=0}^{l} n_i q^i$ where $n_i \in \{0, \dots, q-1\}$ and $n_l \neq 0$.

II.1 Definitions

Let us fix some 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 $\overline{P}_b = \{\overline{p}_0, \ldots, \overline{p}_{b-1}\}$.

We note $\mathfrak{D}_b$ the *Dyck language over* $P_b \cup \overline{P}_b$, that is the language of well parenthesed 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 II.1. In some papers, the expression *Dyck language over* $P_b \cup \overline{P}_b$ is used for the language $\mathfrak{D}'_b$ made of words of $\mathfrak{D}_b$ and mirror-images of words of $\mathfrak{D}_b$. This is the case, for example, in the Chomsky-Schützenberger theorem's [CS66] stating that every context-free language is the image by a homomorphism of the intersection of one of some Dyck's language $\mathfrak{D}'_b$ and a rational language.

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

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

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

We say that an integer $n \in \mathbb{N}$ is a *well parenthesed 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 parenthesed expression. We also denote by I_b the set of well parenthesed integers in base $2b$, that is :

$$I_b = \{n \in \mathbb{N} \mid \mathcal{D}_b\left(\rho_2(n)\right) \in \mathfrak{D}_b\},$$

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

$$L_b = \{\rho_2(n) \mid n \in I_b\}.$$

For example, the integer $n = 3696$ is a well parenthesed integer in base 4 as $\rho_4(3696) = 321300$ and the word $\mathcal{D}_2(321300) = p_0p_1\overline{p}_1p_0\overline{p}_0\overline{p}_0$ is a well parenthesed expression. In the same way, the integer $n = 26244$ is not a well parenthesed integer in base 6 as $\rho_6(26244) = 321300$ and $\mathcal{D}_3(321300) = p_2\overline{p}_2\overline{p}_1p_2\overline{p}_0\overline{p}_0$ is not a well parenthesed expression.

For any integer $b \geq 1$, the language L_b is recognized by the pushdown automaton $\mathcal{A}_b = (S, [\![0, 2b-1]\!], P_b, \phi, s_0\varepsilon, E'\varepsilon)$ where

- $S = \{s_0, s_1\}$ is the space of states,
- $[\![0, 2b-1]\!]\}$ is the input alphabet,
- P_b is the stack alphabet,
- $\phi_b : C \subset SP_b^* \times [\![0, 3]\!] \to SP_b^*$ is the transition function, given by, for all W in P_b^* :

$$\begin{aligned} &\forall i \in [\![b, 2b-1]\!],\ \phi_b(s_0\varepsilon, i) = s_1p_{2b-1-i}, \\ &\forall i \in [\![b, 2b-1]\!],\ \phi_b(s_1W, i) = s_1p_{2b-1-i}W, \\ &\forall i \in [\![0, b-1]\!],\ \phi_b(s_1p_nW, i) = s_1W. \end{aligned}$$

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

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

– the mode of acceptance is by empty stack so the accepted configurations are $S\varepsilon$.
For more informations about links between context-free languages and pushdown automata, we refer to [ABB97].

Remark II.3. In this case, the space of states S can be reduced to a singleton (state s_1) but splitting the in two states does not change the accepted language and allows to feed the automata with non proper representations without changing outing states and will be useful further.

In order to understand how the well parenthesed integers of I_b are distributed in $\mathbb{N}$, we study the complexity function of their characterisitc word $m^{(b)}$:

Definition II.4. For some 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,

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

In particular, $m^{(b)}$ is the indicative word of the context-free language L_b.

For example, we have :

$$m^{(1)} = 10 10^7 10 10^{29} 10^5 10 10^3 10^{113} 10 10^5 10 10^3 10^{17} 10 10^5 10 10^3 10 \dots$$

$$m^{(2)} = 10^8 10^2 10^{140} 10^2 10^8 10^{11} 10^{11} 10^{11} 10^2 10^{23} 10^{11} 10 \dots$$

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

II.2 Main theorem and plan of proof

We now state the main theorem.

Theorem II.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 $\log_{2b} n$ represents $\frac{\ln n}{\ln 2b}$, for all $b \geq 1$.

To prove this theorem, we will construct, for each value of b, two functions of n surrounding $p_{m^{(b)}}(n)$ which are equivalent to f_b, up to multiplicative constants, for sufficiently large 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 IV.1). By the way, the growth order of $p_{m^{(1)}}$ will be studied in a separate part (Section IV). However, we use same scheme of proof in the section regarding cases $b \geq 2$ (Section III) and in the section regarding case $b = 1$.

In Paragraph III.1, we will find upper bounds of $p_{m^{(b)}}(n)$ when $b \geq 2$, using methods developped in [LG06] to show that $p_{m^{(b)}}(n) = \mathcal{O}(f_b(n))$. In Paragraph III.2, we will find lower bounds of $p_{m^{(b)}}(n)$ for all $b \geq 2$, showing up a family of special factors of length n to show that $f_b(n) = \mathcal{O}(p_{m^{(b)}}(n))$ and some results presented this paragraph will be also useful for case $b = 1$.

But first, we present two constructions of words $m^{(b)}$ wich are more convenient to prove Theorem II.5 than the construction by pushdown automata. These following constructions are equivalent but underline different properties of $m^{(b)}$ so the two points of view will be used in the rest of the article.

II.3 Alternative constructions of words $\{m^{(b)}, b \geq 1\}$

From the pushdown automaton $\mathcal{A}_b$ recognizing the language L_b, we can find two other ways of generating the words $m^{(b)}$, using countable objects. 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. these constructions are actually equivalent.

II.3.1 Construction by countable automaton

Construction of words $m^{(b)}$ by pushdown automata is not the better point of view to solve problems of accessibility and co-accessibility between configurations we will face in this article. Indeed, it does not allow to easily see the action of the transition functions 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 $\mathcal{T}_b = (P_b^* \cup \{s_0\}, \varphi_b, s_0, \{s_0, \varepsilon\})$ acts here as an $2b$-automaton (see [Mau06] or [LG06]) and recognize, by construction, the same langage as $\mathcal{A}_b$:

- $P_b^* \cup \{s_0\}$ 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 $(P_b^* \cup \{s_0\}) \times [\![0, 2b-1]\!]$ to $P_b^* \cup \{s_0\}$, defined by :

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

  This function is also naturally extended to a subset of $(P_b^* \cup \{s_0\}) \times [\![0, 2b-1]\!]^*$ 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 is the initial state of $\mathcal{T}_b$,
- $\{s_0, \varepsilon\}$ form the final states set of $\mathcal{T}_b$.

We refer to D. Muller and P. Schupp [MS85] and works of D. Caucal [Cau92] and [Cau03] for litterature about transition graphs ; see also [Tho02] or [Sti05] for results about infinite graphs.

These deterministic countable automata $\mathcal{T}_b$ can be completed by adding a “trash state” x so that every state s of $A_b = P_b^* \cup \{s_0, x\}$, and 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 :

$$\forall n \geq 0,\ m_n^{(b)} = 1 \Longleftrightarrow \varphi_b(\varepsilon, \rho_{2b}(n)) \in \{\varepsilon, s_0\}.$$

The transition graphs $\mathcal{T}_1$ and $\mathcal{T}_2$ are represented in Figure 1 and Figure 2 (the trash state has been removed from $\mathcal{T}_2$ for more visibility).

II.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 for example [Mau06]).

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

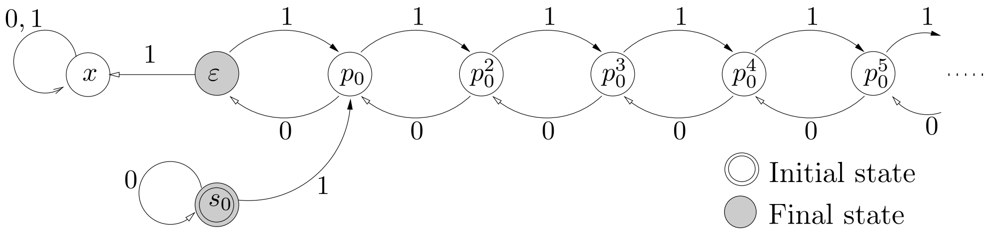

FIG. 1 – Countable automaton generating the infinite word $m^{(1)}$.

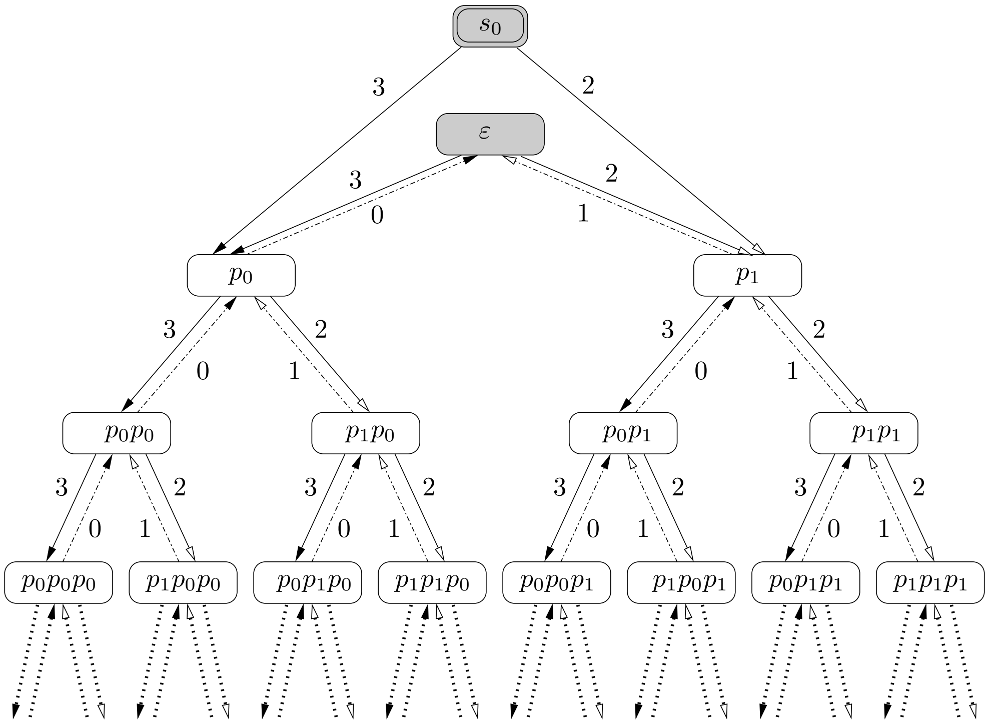

FIG. 2 – Countable automaton generating the infinite word $m^{(2)}$.

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

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

– 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 ${}^b\overline{\delta}$ the following substitution over A_b :

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

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

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

For example, we have :

$$\begin{aligned} \overline{m}^{(1)} = {} & s_0(p_0)(\varepsilon)(p_0^2)(x)(p_0)(p_0)(p_0^3)(x)^2(\varepsilon)(p_0^2)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(x)^5(p_0)(p_0)(p_0^3)(x)(p_0)(p_0)(p_0^3) \\ & (p_0)(p_0^3)(p_0^3)(p_0^5)(x)^{10}(\varepsilon)(p_0^2)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(x)^2(\varepsilon)(p_0^2)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(p_0^2)(p_0^4) \\ & (p_0^4)(p_0^6)(x)^{21}(p_0)(p_0)(p_0^3)(x)(p_0)(p_0)(p_0^3)(p_0)(p_0^3)(p_0^3)(p_0^5)(x)^5(p_0)(p_0)(p_0^3)(x)(p_0)(p_0)(p_0^3)(p_0) \\ & (p_0^3)(p_0^3)(p_0^5)(x)(p_0)(p_0)(p_0^3)(p_0)(p_0^3)(p_0^3)(p_0^5)(p_0)(p_0^3)(p_0^3)(p_0^5)(p_0^3)(p_0^5)(p_0^5)(p_0^7)(x)^{42}(\varepsilon)(p_0^2)(\varepsilon)(p_0^2) \\ & (p_0^2)(p_0^4)(x)^2(\varepsilon)(p_0^2)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(\varepsilon)(p_0^2)(p_0^2)(p_0^4)(p_0^2)(p_0^4)(p_0^4)(p_0^6)(x)^{10}(\varepsilon)\dots \end{aligned}$$

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

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 integer $b \geq 1$, the infinite word $m^{(b)}$ satisfies $m^{(b)} = \Pi(\overline{m}^{(b)})$.

Remark II.6. The connection between the two constructions relies on the correspondance 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 [Mau06] or [LG06] for the details.

The most convenient point of view to study the complexity functions of words $m^{(b)}$ is the substitutive one. For the most part of the article, we will used it for proofs but the automatic point of view is also interesting because it often provide an easier understanding, allowing graphic interpretations.

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

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

III.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 occuring in the fixed point $\overline{m}^{(b)}$ of ${}^b\overline{\delta}$. This study is the aim of the two following lemmas.

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

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

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

This lemma is obvious using the following property of fixed points of substitutions of constant length (Proposition 3.1 of [LG06]) : for all letter s occuring in $\overline{m}^{(b)}$ and for all integer k, if we denote the word ${}^b\overline{\delta}^k(s) = u_0 u_1 \dots u_{(2b)^k - 1}$, then

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

where $\rho_{2b}(n) = n_l \dots n_1 n_0$ is the proper representation 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 occurences of 1 if and only if ${}^b\overline{\delta}^k(s)$ contains some occurences of ε or s_0, that is, it contains some occurences 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 III.3. *Let $b \geq 2$. A word $x_1 x_2$ of A_b^2 is a factor of $\overline{m}^{(b)}$ if and only if :*

1. *$\exists s \in A_b$, $x_1 x_2 = (s)(x)$,*
2. *$\exists s \in A_b \setminus \{s_0\}$, $x_1 x_2 = (x)(s)$,*
3. *$x_1 x_2 = (p_0 p_1)(\varepsilon)$,*
4. *$\exists s \in P_b^*$, $\exists i \in [\![0, b-2]\!]$, $x_1 x_2 = (p_{i+1} s)(p_i s)$,*
5. *$\exists s \in P_b^*$, $x_1 x_2 = (s)(p_{b-1}^2 s)$,*
6. *$\exists s \in P_b^*$, $\exists k \geq 1$, $x_1 x_2 = (p_0^k p_1 p_0^{k-1} s)(s)$.*

■ **Proof of lemma III.3.** This lemma also relies on a property of fixed points of substitutions of constant length (Proposition 3.3 of [LG06]). Indeed, as $\overline{m}^{(b)}$ is the fixed point of ${}^b\overline{\delta}$ starting with s_0, if we set ${}^b\overline{\delta}(s_0) = s_0 s_1 \dots s_{2b-1}$, we have the following equality :

$$\overline{m}^{(b)} = s_0 s_1 \dots s_{2b-1} {}^b\overline{\delta}(s_1) {}^b\overline{\delta}(s_2) \dots {}^b\overline{\delta}(s_{2b-1}) {}^b\overline{\delta}^2(s_1) {}^b\overline{\delta}^2(s_2) \dots {}^b\overline{\delta}^2(s_{2b-1}) {}^b\overline{\delta}^3(s_1) \dots {}^b\overline{\delta}^3(s_{2b-1}) \dots$$

From this statement, the word $x_1 x_2$ is a factor of $\overline{m}^{(b)}$ if and only if one of the following cases happens :

(i) $\exists j \in [\![0, 2b-2]\!],\ x_1 x_2 = s_j s_{j+1}$,

(ii) $\exists k > 0, \exists j \in [\![1, 2b-2]\!]$, $x_1 = {}^b\overline{\delta}_{2b-1}^k(s_j)$ and $x_2 = {}^b\overline{\delta}_0^k(s_{j+1})$,

(iii) $\exists k \geq 0, \exists j \in [\![1, 2b-2]\!],\ x_1 \in E \setminus \{s_0\},\ x_2 \in {}^b\overline{\delta}_0^k \circ {}^b\overline{\delta}_{j+1} \circ {}^b\overline{\delta}_j^{-1} \circ {}^b\overline{\delta}_{2b-1}^{-k}(\{x_1\})$,

(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 $s_0s_1\dots s_{2b-1}$, 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 some word ${}^b\overline{\delta}^k(s_j)$ 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 $\overline{m}^{(b)}$ annonced in lemma III.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 $j \geq b$,
- $x_1x_2 = (p_0^k p_i)(x)$ 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 $x_1x_2 = (p_0^k p_1)(x)$ 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 some $k \geq 1$:
for all i in $[\![0, b-2]\!]$:

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

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 $[\![b, 2b-2]\!]$:

$$ {}^b\overline{\delta}^k_0 \circ {}^b\overline{\delta}_{i+1} \circ {}^b\overline{\delta}^{-1}_i \circ {}^b\overline{\delta}^{-k}_{2b-1}(\{s\}) = \begin{cases} \{x\} & \text{if } \mathrm{Pref}_{k+1}(s) = p_0^k p_1, \\ \emptyset & \text{else,} \end{cases} $$

for $i = 2b-2$:

$$ {}^b\overline{\delta}^k_0 \circ {}^b\overline{\delta}_{i+1} \circ {}^b\overline{\delta}^{-1}_i \circ {}^b\overline{\delta}^{-k}_{2b-1}(\{s\}) = \begin{cases} \{v\} & \text{if } s = p_0^k p_1 p_0^{k-1} v, \\ \{x\} & \text{if } \mathrm{Pref}_{k+1}(s) = p_0^k p_{2b-1-i} \text{ and } \mathrm{Pref}_{2k}(s) \neq p_0^k p_1 p_0^{k-1}, \\ \emptyset & \text{else.} \end{cases} $$

Using these equalities, we obtain :

- the words $(x)(x)$, $(x)(s)$ and $(s)(x)$ for all s of A_b are factors of $\overline{m}^{(b)}$ exepted $(x)(s_0)$,
- the words $(s)(p_{b-1}^2 s)$ for all s of A_b are factors of $\overline{m}^{(b)}$,
- the words $(p_{i+1}s)(p_i s)$ for all s of P_b^* and all $i \in [\![0, b-2]\!]$ are factors of $\overline{m}^{(b)}$,
- the words $(p_0^k p_1 p_0^{k-1} s)(s)$ for all $s \in P_b^*$ are factors of $\overline{m}^{(b)}$.

All different factors of $\overline{m}^{(b)}$ announced in Lemma III.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 some $r \in \{0, 1\}$ and $q \in \mathbb{N}$ and some integer n satisfying $(2b)^k \leq n < (2b)^{k+1}$.

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

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 it, 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 $\overline{m}^{(b)}$. We need, in order to find a thin upper bound of the number of factors of length $(2b)^k$, to know how many factors x_1x_2 of $\overline{m}^{(b)}$ are such 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 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 $\overline{{}^b\delta}^k(s)$,

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

As, by lemma III.2, we know that $\frac{b^{k+2}-b^r}{b^2-1}+1$ letters s of A_b provide words ${}^b\overline{\delta}^k(s)$ wich are different from $0^{(2b)^k}$, 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). By Lemma III.3, those factors can be of the following types :

(i) $(p_0p_1)(\varepsilon)$,

(ii) $(p_{i+1}s)(p_is)$ for some $s \in P_b^*$ and some $i \in [\![0, b-1]\!]$,

(iii) $(s)(p_{b-1}^2 s)$ for some $s \in P_b^*$,

(iv) $(p_0^\alpha p_1 p_0^{\alpha-1}s)(s)$ for some $s \in P_b^*$ and some 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 III.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 same parity as k (see lemma III.2). From it, 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 some fixed integer $\beta \in [\![1, q]\!]$ ($k = 2q+r$ with $r \in \{0, 1\}$) and $\alpha \leq \beta$. We obtain by this construction $N = \sum_{\beta=1}^q \beta b^{k-2\beta}$ additionnal 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 inegality, we only keep that, for all k, $N \leq \frac{b}{(b^2-1)^2}\left(b^{k+1} - \frac{b^2-1}{2}k + b + \frac{b^2-1}{2}\right)$.

It comes 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 $0^{2(2b)^k}$ occuring 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 using the fact that, for all integer n of $[\![(2b)^k, (2b)^{k+1}[\![$, $\log_{2b}(n)$ is in $[\![k, k+1[\![$, we get :

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

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

III.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 $\delta^k(w)$ wich are different from $0^{(2b)^k}$, for a fixed integer k. We first study the number of occurences of the letter 1 and their repartition in such words. It will allow us to match special factors of fixed length and, 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 principal difficulty to overcome in the proof arise from the construction of the words $m^{(b)}$ by substitution and projection. This problem also appears for substitutive words (see [CN08]). Indeed, the projection erases a lot of informations so two or infinitely many different special factors of $\overline{m}^{(b)}$ can project on the same special factor of $m^{(b)}$ or two non special factors of $m^{(b)}$ can project on a special factor of $m^{(b)}$.

As a consequence, the difficulty is not really to find special factors but to ensure special factors we found are different. To do it, we will extract special factors from factors of type $\delta^k((x)(x)(s))$ of $m^{(b)}$ (see Lemma III.14 and III.15). To ensure words $\delta^k(s)$ are sufficiently different to produce different special factors, we first determine a set wich indexes the occurrences of the letter 1 in words $\delta^k(s)$. This set will aslo give informations about where these occurences takes place. This is the aim of Lemma III.6 and Proposition III.9. Then, Lemma III.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 III.13.

Notations III.4. For a fixed integer $b \geq 1$, we introduce notations about integers and $2b$-ary representations, which will be useful to ligthen the proofs :

For a given word w of $[\![0, 2b-1]\!]^*$, we denote by $[w]_{2b}$ the integer for which w is a representation in base $2b$ (w can be a non proper representation), 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 III.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 parenthesed expression with prefix $\mathcal{D}_b(\hat{s})$ as, for all integer $n \in [\![0, |s|-1]\!]$, $\mathcal{D}_b(\tilde{s})_n = \overline{p}_i$ 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 $\tilde{s} = 00101$ and $\hat{s} = 23233$ 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 parenthesed 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 $\tilde{s} = 00101$ and $\hat{s} = 45455$ 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 parenthesed 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 III.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 $w^{(0)}$, $w^{(1)}$, $\dots$, $w^{(|s|)}$ of $[\![0, 2b-1]\!]^*$ such that, for all i of $[\![0, |s|]\!]$, $\mathcal{D}_b(w^{(i)}) \in \mathfrak{D}_b$ and

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

■ **Proof of lemma III.6.** Let s be a word of P_b^* and w a word of $[\![0, 2b-1]\!]^*$.

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

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

We proceed by recurrence over the length of the considered element s to prove that there exists a $(|s|+1)$-uplet 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 $w = w^{(0)}\tilde{s}_0 w^{(1)}\tilde{s}_1 \dots w^{(|s|-1)}\tilde{s}_{|s|-1}w^{(|s|)}$. For $s = \varepsilon$, this property is obvious.

Assuming this property is true for words of P_b^n, for some $n \geq 0$, 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 recurrence 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 existe $n+2$ words $w^{(i)}$ in $[\![0, 2b-1]\!]^*$, such that for every $i \in [\![0, n+1]\!]$, $\mathcal{D}_b(w^{(i)})$ belongs to $\mathfrak{D}_b$ 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 recurrence process and ends the proof of Lemma III.6. ■

Notations III.7. For a couple of positive integers (α,β), we call $\mathcal{U}_b(\alpha,\beta)$ the set of $(\beta+1)$-uplets $(w^{(0)},w^{(1)},\dots,w^{(\beta)})$ 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 integer $k\geq 3$ and all integer $l\leq k$ with same parity as k.

Definition III.8. Let $w=w_0w_1\dots w_l$ 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 III.9. *Let $b\geq 1$. Let us fix some 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 $w^{(0)}$, $w^{(1)}$, …, $w^{(l)}$ of $\mathcal{U}_b(a,l)$ such that $n=\left[w^{(0)}\tilde{s}_0w^{(1)}\tilde{s}_1\dots w^{(l-1)}\tilde{s}_{l-1}w^{(l)}\right]_{2b}$.

■ **Proof of proposition III.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 also 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 ${}^b\overline{\delta}^k(s)$ occuring 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 parenthesed expression and Lemma III.6 gives the general form of w : there exists $(l+1)$ words $w^{(0)}$, $w^{(1)}$, …, $w^{(l)}$ of $\mathcal{U}_b(a,l)$ such that $w=w^{(0)}\tilde{s}_0w^{(1)}\tilde{s}_1\dots w^{(l-1)}\tilde{s}_{l-1}w^{(l)}$. In particular, when $|s|>k$, then 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 $w^{(0)}$, $w^{(1)}$, …, $w^{(l)}$ of $\mathcal{U}_b(a,l)$ such that

$$w=w^{(0)}\tilde{s}_0w^{(1)}\tilde{s}_1\dots w^{(l-1)}\tilde{s}_{l-1}w^{(l)},$$

then it follows that ${}^b\overline{\delta}_w(s)={}^b\overline{\delta}_{w^{(0)}\tilde{s}_0w^{(1)}\tilde{s}_1\dots w^{(l-1)}\tilde{s}_{l-1}w^{(l)}}(s)={}^b\overline{\delta}_{w^{(l)}}\circ{}^b\overline{\delta}_{\tilde{s}_{l-1}}\circ{}^b\overline{\delta}_{w^{(l-1)}}\circ\dots\circ{}^b\overline{\delta}_{w^{(1)}}\circ{}^b\overline{\delta}_{\tilde{s}_0}\circ{}^b\overline{\delta}_{w^{(0)}}(s)$.

As words $\mathcal{D}_b\left(w^{(i)}\right)$ are well parenthesed 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 III.5, we get ${}^b\overline{\delta}_w(s)=\varepsilon$ and further $\delta^k(s)_n=1$. ■

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

Lemma III.11. *Let $b \geq 1$. For an integer $k \geq 3$, and some element $s \in P_b^l$ for some integer $l \leq k$, with same parity as k, the number of occurences 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 III.11.**

Let us begin the proof by recalling some combinatorial results. The number of well parenthesed 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. We also get, for all couple of integers (α, β) :

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

In order to obtain a simple expression for the $N_b(\alpha, \beta)$, we use some methods from analytic combinatorics. For an introduction to those methods, one can consults the book of P. Flajolet and R. Sedgewick [FS07]. 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 serie of Catalan numbers, noted κ :

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

This serie 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 compute the coefficients of $\kappa(t)^{\beta+1}$.

The trick is to compute coefficients of the serie $(\kappa(t)-1)^{\beta+1}$. Indeed, the serie $K(t) = \sum_{n\geq 1} \mathcal{C}_n t^n$ satisfy the equation $K(t) = t(K(t)+1)^2$, wich 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 have, for all pair (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 functionnal equations verified by κ and K, we also get that $K(t) = t\kappa(t)^2$ and it allows 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 all integer n,

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

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

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

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 III.12. *for all $b \geq 2$ and $k \geq 4$, the application $\phi_k : a \mapsto N_b(k-1, k-2a-1)$ is striclty 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 one from each other.

■ **Proof of lemma III.12.** For $k=4$ and $k=5$, ϕ_k is obviously 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 :

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

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 we obtain, for all β in $[\![1, \lfloor \frac{k}{2} \rfloor - 2]\!]$:

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

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

■

Now we have set, through Proposition III.9 and Lemma III.11, how many occurrences of ones the words $\delta^k(s)$ hold, we have to know more about where do these occurrences take place. This is the aim of the following lemma.

Lemma III.13. *Let $b \geq 1$.*

Let us fix some integer $k \geq 2$, some element $s \in P_b^l$ for some $l = k-2a$ with a in $[\![1, \lfloor \frac{k}{2} \rfloor]\!]$ an 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 is no occurrences of the letter* 1 *in* $\delta^k(s)$ *between the letters* 1 *occurring at ranks* $[\tilde{s}_0 \ldots \tilde{s}_i (2b-1)^a 0^a \tilde{s}_{i+1} \ldots \tilde{s}_l]_{2b}$ *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 occurrence of the letter* 1 *occurs in* $\delta^k(s)$ *is* $[\tilde{s}_0 \ldots \tilde{s}_i (2b-1)^a 0^a \tilde{s}_{i+1} \ldots \tilde{s}_l]_{2b}$*.*

■ **Proof of lemma III.13.** The three first 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 parenthesed 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 parenthesed expressions, the one giving the geatest 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 $m^{(b)}$, in order to find a lower bound of $p_{m^{(b)}}(n+1) - p_{m^{(b)}}(n)$ for positive integers n, and, by summing, to find a lower bound of its complexity function. From now, the presented results of the section are only valid for $b \geq 2$.

Lemma III.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 III.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 $(p_1 p_0 s)$ 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 ${}^2\overline{\delta}(p_1 p_0 s)$ and ${}^2\overline{\delta}^2(p_1 p_0 s)$ also occurs in $\overline{m}^{(2)}$. As $\mathrm{Pref}_5\left({}^2\overline{\delta}^2(p_1 p_0 s)\right) = (x)^4(s)$, 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 ${}^b\overline{\delta}(p_1 p_0 s)$ also occurs 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 III.15. *Let* $b \geq 2$*.*

Let us fix some integer $k \geq 3$*, and some integer* n *satisfying* $(2b)^k \leq n < (2b)^{k+1}$*.*

For each element $s \in P_b^{l-1}$ *with some* $l = k - 2a$*, where* a *belongs to* $[\![1, \lfloor \frac{k}{2} \rfloor - 1]\!]$*, and each integer* j *of* $[\![1, b-1]\!]$*, the word* $\delta^k((x)(x)(p_j s))$ *contains a special factors of length* n *denoted* $u^{(n,s,j)}$*.*

■ **Proof of lemma III.15.** Let s be some element in P_b^{l-1} for some $l = k - 2a$ with a in $[\![0, \lfloor \frac{k}{2} \rfloor - 1]\!]$ and j be some integer of$[\![1, b-1]\!]$.

Thanks to Proposition III.9, the following set of integers is the set of ranks where the occurrences of the letter 1 occur in $\delta^k(p_0 s)$:

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

and the following set of integers is the set of ranks where the occurrences of the letter 1 occur 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-uplet $(w^{(1)}, \dots, w^{(l)})$ of $\mathcal{U}_b(a, l-1)$, we have the following equality :

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

so, using lemma III.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 $\delta^k(p_j s)$, that is, according to Lemma III.13, rank $\left[b(b-1) j \tilde{s} (b(b-1))^{a-1} \right]_{2b}$ can be deduced from the rank of occurrence of the $(N_b(k-1, l-1)+1)$-th letter 1 in $\delta^k(p_0 s)$, that is rank $\left[b(b-1) 0 \tilde{s} (b(b-1))^{a-1} \right]_{2b}$, by adding $j(2b)^{k-3}$.

Those facts allow us to build a special factor parametrized 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 extract 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 figure 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

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

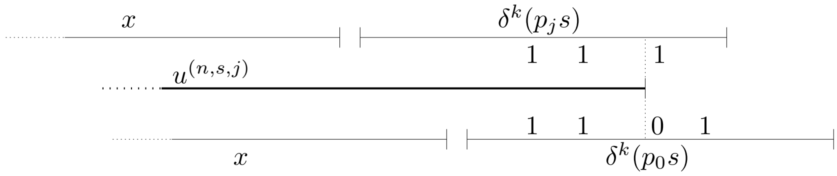

FIG. 3 – Construction of the right special factor $u^{(s,n,j)}$.

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 III.16. *Let $b \geq 2$.*

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

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

■ **Proof of lemma III.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 III.12 provide that $N_b(k-1,|s|)$ injectively determine 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. ■

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

Let us set some integer $k \geq 4$, $k = 2q + r$ with r in $\{0,1\}$ and some integer n satisfying $(2b)^k \leq n < (2b)^{k+1}$.

We have shown off in Lemma III.15 a family of right special factors $u^{(n,s,j)}$ of length n. Thanks to Lemma III.16, the displayed special factors are all different when s goes all over P_b^{l-1}, when $a = \frac{k-l}{2}$ goes all over $[\![1, \lfloor\frac{k}{2}\rfloor - 1]\!]$ and when j goes all over $[\![1, b-1]\!]$.

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

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

Summing this inequalities, we get, for all $k \geq 5$ and for all $n \in [\![(2b)^k, (2b)^{k+1}-1]\!]$,

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

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. It conclude the proof of Theorem II.5 for $b \geq 2$ as we proved $n^{1+\log_{2b} b} = \mathcal{O}\left(p_{m^{(b)}}(n)\right)$. ■

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

IV.1 About the difference of behaviour

As we will see, the infinite word $m^{(1)}$ is closer, from combinatorial and statistical points of view, from the drunken man infinite words than from the words $m^{(b)}$, when $b \geq 2$. Drunken man infinite words have been presented in [LG07]. Indeed, the automaton generating drunken man infinite words (see Figure 4) is very similar to the automaton generating $m^{(1)}$, exepted from the trash state x, of course.

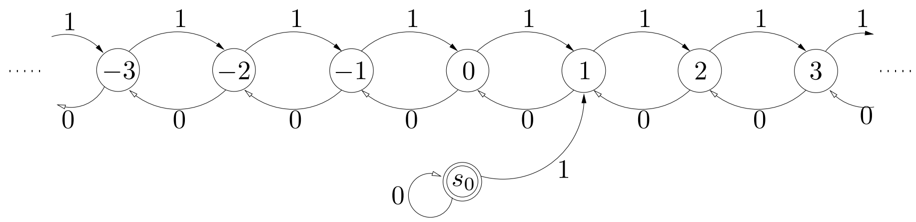

FIG. 4 – Automaton generating the drunken man infinite words

The difference between the behaviours of $m^{(1)}$ and $m^{(b)}$ for $b \geq 2$ comes from the difference of geometry between the transition graphs associated to pushdown automata generating them. Indeed, for $b \geq 2$, the quantity of states of the transition graph at fixed length from the final states grows exponentially, whereas for $b = 1$, it grows linearly. From the substitutive point of view, it can be translated as followed : the quantity of letters s in $\mathcal{A}_b$ such that the iterate ${}^b\overline{\delta}^k(s)$ contains some occurences of ones grows linearly for $b = 1$ and exponentially for $b \geq 2$ (as functions of k).

However, proofs regarding case $b = 1$ use same processings as for case $b \geq 2$. Methods used to find upper bounds of complexity functions in the cases $b = 1$ and $b \geq 2$ are the same but the way we obtain lower bound in case $b = 1$ need to exhibit a different family of special factors from the displayed families when $b \geq 2$. The family of special factors of $m^{(1)}$ we are going to show up is also really close from the family of special factors of drunken man infinite words, displayed in [LG07].

Moreover, despite using quite heavy notations when $b = 1$, some propositions and lemmas detailed for case $b \geq 2$ in section III are also available for $b = 1$.

Indeed, the states set $\{p_0\}^* \cup \{s_0, x\}$ of $\mathcal{T}_1$, could be identified to $\mathbb{N} \cup \{s_0, x\}$, simplifying the expression of the applications ${}^1\overline{\delta}_i$ and lightening the arguments. Simplifications are presented in the following notations.

Notation IV.1. To lighten the proofs of this section, 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 :

$$\begin{aligned}&\overline{m}^{(1)} = s_0(1)(0)(2)(x)(1)(1)(3)(x)^2(0)(2)(0)(2)(2)(4)(x)^5(1)(1)(3)(x)(1)(1)(3)(1)(3)(3)(5)(x)^{10}\\&(0)(2)(0)(2)(2)(4)(x)^2(0)(2)(0)(2)(2)(4)(0)(2)(2)(4)(2)(4)(4)(6)(x)^{21}(1)(1)(3)(x)(1)(1)(3)(1)\\&(3)(3)(5)(x)^5(1)(1)(3)(x)(1)(1)(3)(1)(3)(3)(5)(x)(1)(1)(3)(1)(3)(3)(5)(1)(3)(3)(5)(3)(5)(5)(7)\\&(x)^{42}(0)(2)(0)(2)(2)(4)(x)^2(0)(2)(0)(2)(2)(4)(0)(2)(2)(4)(2)(4)(4)(6)(x)^{10}(0)(2)(0)(2)(2)(4)\dots\end{aligned}$$

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

IV.2 An upper bound of $p_{m^{(1)}}(n)$

To find an upper bound of the complexity function, we use the two following lemmas wich are the equivalent to Lemmas III.2 and III.3 for case $b = 1$.

Lemma IV.2. *For a given positive integer k, the word $\delta^k(s)$ contains occurences 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 IV.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 same arguments as for Lemma III.2 and III.3. Using those two results the same way it has been done for $b \geq 2$, one can show that :

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

wich allows to say that $p_{m^{(1)}}(n) = \mathcal{O}\left(n(\log_2 n)^2\right)$.

IV.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 III.2. Indeed, Proposition III.9, Lemma III.11 and Lemma III.13 are valid for $b = 1$. With new notations, Proposition III.9 and Lemma III.11 become the two following results.

Proposition IV.4. *Let us fix some 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 $w^{(0)}$, $w^{(1)}$, ..., $w^{(s)}$ of $\mathcal{U}_1(a, s)$ such that $n = \left[w^{(0)}0w^{(1)}0\dots w^{(s-1)}0w^{(s)}\right]_2$.

Lemma IV.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 occurences of the letter 1 in $\delta^k(s)$ is given by $N_1(k, s)$ where*

$$N_1(k, s) = \mathit{Card}(\mathcal{U}_1(a, s)) = 2^a \frac{s + 1}{k + 1}\binom{k + 1}{a}.$$

From Proposition IV.4 and Lemma III.13, we also get the following Proposition.

Proposition IV.6. *Let $k \geq 2$ be a positive integer and $s = k - 2a$ be an element of $\mathbb{N}$ for some 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 two last occurrences of the letter 1 in $\delta^k(s)$ are respectively given by $\left[1^{a-1}010^{k-a-1}\right]_2$ and $\left[1^a 0^{k-a}\right]_2$.

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, the special factors displayed in case $b \geq 2$ come from couples of words of type $\delta^k((x)(x)(s))$ and $\delta^k((x)(x)(s'))$, 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 property 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.

Proposition IV.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 IV.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 ${}^1\overline{\delta}((s-1)(x))$ respectively contain the word $(s-2)(s)(s-2q)$ and the word $(s-2)(s)(x)$.

Thanks to Proposition IV.7, we can display a family of special factors of $\overline{m}^{(1)}$ using couples 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 [LG07]).

For all integer $k \geq 2$ and all integer 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 $\delta^k((k-2(a+1))(k-2a)(k-2a'))$ such that the last letter of $w^{(n,a,a')}$ is the letter before the first occurrence of 1 in $\delta^k(k-2a')$. 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 Figure 5).

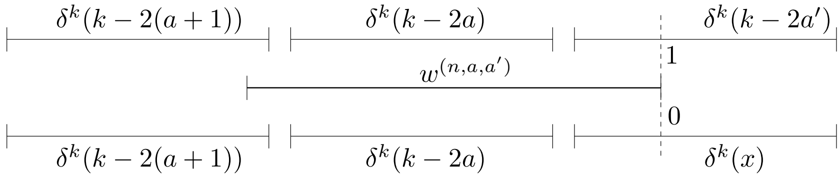

FIG. 5 – Construction of the special factor $w^{(n,a,a')}$.

If the word $w^{(n,a,a')}$ contains at least two occurences of the letter 1, thanks to Proposition IV.6, the length for the block of letters 0 between the two last occurrences of the letter 1 in $w^{(n,a,a')}$ which are the two last 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 two last 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 : $\{w^{(n,a,a')},\ 1 \leq a \leq a' \leq \lfloor \frac{k}{2} \rfloor - 2\}$, which are different one from each other when $k \geq 6$. As a consequence, $m^{(1)}$ has at least $\frac{(\lfloor \frac{k}{2} \rfloor - 2)(\lfloor \frac{k}{2} \rfloor - 1)}{2}$ 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 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 summing those inequalities, we also obtain a lower bound of $p_{m^{(1)}}(n)$, for all integer $n \geq 2^6$, under the form $p_{m^{(1)}}(n) \geq n\left(A(\log_2 n)^2 + B\log_2 n + C\right) + D$, that is $n(\log_2 n)^2 = \mathcal{O}\left(p_{m^{(1)}}(n)\right)$.

V Open questions

The family presented here was first studied to obtain examples of q^∞-automatic words with complexity more than some $n^{1+\varepsilon}$ with $\varepsilon > 0$. Seeing the results, the following step is to know if one can find (or not) examples of q^∞-automatic words, with complexity satisfying $p_m(n) \gg n^2$ and further to know if it is actually possible to reach the upper bound annonced in [LG06].

Another motivation for this article relies on the nature of involved languages of integer representations, which are context-free. Indeed, it would be interesting to known the possible growth orders of complexity functions of characteristic words of context-free languages in general. This question will probably need other tools, as some 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 other determinism questions).

The property of recognizability by countable automata in itself is also a interesting topic. Indeed, the notion of recognizability by countable automata is different whether we choose to read integers representations 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).

Références

[AB07] B. Adamczewski and Y. Bugeaud. On the complexity of algebraic numbers I. expansions in integer bases. *Annals of Math.*, 165 : 547–565, 2007.

[ABB97] J.-M. Autebert, J. Berstel, and L. Boasson. Context-free languages and pushdown automata. In *Handbook of formal languages, Vol. 1*, pages 111–174. Springer, 1997.

[All94] J.-P. Allouche. Sur la complexité des suites infinies. *Bull. Belg. Math. Soc. Simon Stevin*, 1(2) : 133–143, 1994.

[AS03] J.-P. Allouche and J. Shallit. *Automatic sequences. Theory, applications, generalizations.* Cambridge University Press, 2003.

[Cas96] J. Cassaigne. Special factors of sequences with linear subword complexity. In *Developments in language theory (Magdeburg, 1995)*, pages 25–34. World Sci. Publishing, 1996.

[Cas97] J. Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin*, 4(1) : 67–88, 1997.

[Cau92] D. Caucal. On the regular structure of prefix rewriting. *Theor. Comp. Sci.*, 106 : 61–86, 1992.

[Cau03] D. Caucal. On infinite transition graphs having a decidable monadic theory. *Theor. Comp. Sci.*, 290 : 79–115, 2003.

[CN08] J. Cassaigne and F. Nicolas. Subword complexity. *Preprint*, 2008.

[Cob72] A. Cobham. Uniform-tag sequences. *Math. Syst. Theory*, 6 : 164–192, 1972.

[CS66] N. Chomsky and M.-P. Schürtzenberger. The algebraic theory of context-free languages. In *Computer Programming and Formal Systems*, pages 195–242. North-Holland, 1966. P. Braffort and D. Hirshberg Eds.

[Fer99] S. Ferenczi. Complexity of sequences and dynamical systems. *Discrete Math.*, 206(1-3) : 145–154, 1999.

[FS07] P. Flajolet and R. Sedgewick. *Analytic combinatorics.* Preliminary version. available on `http ://algo.inria.fr/flajolet/Publications/books.html`, 2007.

[Kůr03] P. Kůrka. *Topological and symbolic dynamics*, volume 11 of *Cours spécialisés.* SMF, 2003.

[LG06] M. Le Gonidec. Sur la complexité de mots infinis engendrés par des q-automates dénombrables. *Ann. Inst. Fourier*, 56(7) :2463–2491, 2006.

[LG07] M. Le Gonidec. Drunken man infinite words complexity. *to appear in RAIRO/ITA*, 2007.

[Mau06] C. Mauduit. Propriétés arithmétiques des substitutions et automates infinis. *Ann. Inst. Fourier*, 56(7) : 2525–2549, 2006.

[MS85] D. Muller and P. Schupp. The theory onf ends, pushdown automata and second-order logic. *Theor. Comp. Sci.*, 37 : 51–75, 1985.

[Pan85] J.-J. Pansiot. Complexité des facteurs des mots infinis engendrés par morphismes itérés. *Lecture Notes in Comp. Sci.*, 172 : 380–389, 1985. Automata, languages and programming (Antwerp, 1984).

[PF02] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lect. Notes in Math.* Springer, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[PP04] D. Perrin and J.-E. Pin. *Infinite words. Automata, semi-groups, logic and games*, volume 141 of *Pure and Applied Math.* Elsevier, 2004.

[Sta06] L. Staiger. Kolmogorov complexity of infinite words. *CDMTCS Research Report Series*, 279, 2006.

[Sti05] C. Stirling. Language theory and infinite graphs. *Marktoberdorf summerschool*, 2005. Notes for *Logical Aspects of Secure Computer Systems.*

[Tho02] W. Thomas. An short introduction to infinite automata. In *Proceedings of the International Conference of Developments in Language Theory*, volume 2295 of *Lecture Notes in Comput. Sci.*, pages 130–144. Springer, 2002.