Construction of regular languages and recognizability of polynomials

Michel Rigo

Institut de Mathématiques, Université de Liège,
Grande Traverse 12 (B 37), B-4000 Liège, Belgium.
M.Rigo@ulg.ac.be

Keywords: Numeration system; Recognizability; Regular language; Complexity function.

Abstract

A generalization of numeration systems in which $\mathbb{N}$ is recognizable by finite automata can be obtained by describing a lexicographically ordered infinite regular language. We show that if $P \in \mathbb{Q}[x]$ is a polynomial such that $P(\mathbb{N}) \subset \mathbb{N}$ then there exists a numeration system in which the set of representations of $P(\mathbb{N})$ is regular. The main issue is to construct a regular language with a complexity function equals to $P(n+1) - P(n)$ for n large enough.

---

1 Introduction

Recently, P. Lecomte and I introduced the concept of numeration system on a regular language [6]. A *numeration system* is a triple $S = (L, \Sigma, <)$ where L is an infinite regular language over a totally ordered finite alphabet $(\Sigma, <)$. The lexicographic ordering of L gives a one-to-one correspondence r_S between the set $\mathbb{N}$ of the natural numbers and the language L.

For each $n \in \mathbb{N}$, $\mathrm{r}_S(n)$ denotes the $(n+1)^{th}$ word of L with respect to the lexicographic ordering and is called the *S-representation* of n. For $w \in L$, we set $\mathrm{val}_S(w) = \mathrm{r}_S^{-1}(w)$ and we call it the *numerical value* of w.

One of the main issue about numeration systems is the study of recognizability. By recognizability, one means the following. Let S be a numeration system. A subset $X \subset \mathbb{N}$ is said to be *S-recognizable* if $\mathrm{r}_S(X)$ is recognizable by a finite automaton. Therefore we can consider two kinds of questions.

◇ For a given numeration system S, what are the S-recognizable subsets of $\mathbb{N}$?

◇ For a given subset X of $\mathbb{N}$, is it S-recognizable for some numeration system S ?

A partial but useful answer to the first question is that arithmetic progressions are always recognizable in any numeration system. Moreover if $X \subset \mathbb{N}$ is recognizable for some system S then for each $t \in \mathbb{N}$, $X+t$ is also S-recognizable [6]. (These two results will be used in this paper.)

On the other hand, there is no numeration system S for which the set of primes is S-recognizable [7]. In this paper, we will be mainly concerned with the second question when X is a polynomial image of $\mathbb{N}$.

For classical numeration systems with an integer base $k \geq 2$, it is well known that the set of the perfect squares is not k-recognizable (see [2] for a survey about classical numeration systems). However, the numeration system

$$S = (a^*b^* \cup a^*c^*, \{a, b, c\}, a < b < c)$$

is such that the set $\{n^2 : n \in \mathbb{N}\}$ is S-recognizable [6]. The choice of the language $a^*b^* \cup a^*c^*$ was given by some complexity considerations: this language has exactly $2n+1$ words of length n. (The complexity function of a language $L \subset \Sigma^*$ maps $n \in \mathbb{N}$ onto $\#(L \cap \Sigma^n)$.) In view of this result, J.-P. Allouche asked the following question. Is it possible to generalize the result on the S-recognizability of the perfect squares to the set $\{n^k : n \in \mathbb{N}\}$, $k > 2$? Moreover, if P is a polynomial belonging to $\mathbb{N}[x]$ (resp. $\mathbb{Z}[x]$ or $\mathbb{Q}[x]$) such that $P(\mathbb{N}) \subset \mathbb{N}$ then can one find a numeration system S such that $P(\mathbb{N})$ is S-recognizable ?

In all these cases, we answer affirmatively. For a given polynomial P, we give an explicit method to construct a numeration system S such that $\mathrm{r}_S(P(\mathbb{N}))$ is regular. For this purpose, we show how to obtain a regular language which contains exactly $P(n+1) - P(n)$ words of length n for n large enough. The construction of regular languages with a specified complexity function is a problem beyond the concern of numeration systems.

We are lucky enough to get more. Using the same technique, we show that $f(n) = \sum_i P_i(n)\, \alpha_i^n$ where $P_i \in \mathbb{Q}[x]$ and $\alpha_i \in \mathbb{N}$ is such that $f(\mathbb{N})$ is S-recognizable for some numeration system S.

The fact that the set of primes is never recognizable and that the polynomial images of $\mathbb{N}$ are recognizable give another interpretation of a well-known result (see [5, Theorem 21]): *no non-constant polynomial $f(n)$ with integral coefficients can be prime for all n, or for all sufficiently large n.*

2 Recognizability of polynomials

The present section is organized as follows. First, we give an explicit iterative method to obtain regular languages L_k such that the number of words of length n is exactly n^k (in [9], it is said that such languages can be easily obtained but we need our method for later purposes). Next, we gradually increase the difficulty. We begin with the case $P \in \mathbb{N}[x]$ which is quite simple since we only deal with the operation of addition. Next we consider $P \in \mathbb{Z}[x]$ and the problem of subtraction must be resolved. Our proof in the case of negative coefficients rests on our construction of the languages L_k. Finally, we consider the most general case, $P \in \mathbb{Q}[x]$ and the problem of division. In each of these last three steps, we give an instructive short example of the construction.

i) Languages with complexity n^k

Let us recall some basic definitions.

Definition 1 The *complexity function* or *counting function* of a language $L \subseteq \Sigma^*$ is

$$\rho_L : \mathbb{N} \to \mathbb{N} : n \mapsto \#(\Sigma^n \cap L)$$

where $\#A$ denotes the cardinality of the set A.

Definition 2 If x and y are two words in Σ^* then the *shuffle* of x and y is the language $x \sqcup\!\sqcup\, y$ defined by

$$\{x_1 y_1 \dots x_n y_n : x = x_1 \cdots x_n, y = y_1 \cdots y_n, x_i, y_i \in \Sigma^*, 1 \le i \le n, n \ge 1\}.$$

If $L_1, L_2 \subseteq \Sigma^*$ then the *shuffle* of the two languages is the language

$$L_1 \sqcup\!\sqcup\, L_2 = \{w \in \Sigma^* : w \in x \sqcup\!\sqcup\, y, \text{for some } x \in L_1, y \in L_2\}.$$

Recall that if L_1, L_2 are regular then $L_1 \sqcup\!\sqcup\, L_2$ is also regular (see for instance [4, Proposition 3.5]).

Definition 3 Let $L \subseteq \Sigma^*$. Then Σ is the *minimal alphabet of* L if $\forall \sigma \in \Sigma$, $\exists w \in L : w = u\sigma v,\ u, v \in \Sigma^*$.

We want to construct regular languages L_k such that $\rho_{L_k}(n) = n^k$, $k \in \mathbb{N}$. To that end, we define regular languages M_k such that $\rho_{M_k}(n) = (n+1)^{k-1}$, $k \ge 2$. The first two languages L_0 and L_1 are, for example, $L_0 = a^*$ and $L_1 = a^+ b^*$.

Let $k \geq 2$. Assume that we have $L_0, \ldots, L_{k-1}$. One has

$$\rho_{M_k}(n) = \sum_{j=0}^{k-1} \binom{k-1}{j} n^j.$$

Therefore, M_k can be obtained as a finite union of regular languages L_j's over distinct alphabets, $j < k$. That is

$$M_k = \bigcup_{j=0}^{k-1} \bigcup_{i=1}^{\binom{k-1}{j}} L_{j,i} \qquad (1)$$

where $\rho_{L_{j,i}}(n) = n^j$. If σ_k does not belong to the minimal alphabet of M_k, then we can define L_k as

$$L_k = M_k \sqcup\!\sqcup \{\sigma_k\}. \qquad (2)$$

Indeed, for each of the $(n+1)^{k-1}$ words w of length n in M_k, $w \sqcup\!\sqcup \sigma_k$ contains $n+1$ words of length $n+1$. So there are exactly $(n+1)^k$ words of length $n+1$ in L_k.

As an example, we give the nine words of length 3 in L_2. First, we have $M_2 = a^* \cup b^+ c^*$ and Table 1 shows the situation.

$M_2 \cap \{a,b,c\}^2$	$L_2 = M_2 \sqcup\!\sqcup \{d\}$
aa	aad, ada, daa
bb	bbd, bdb, dbb
bc	bcd, bdc, dbc

Table 1
The nine words of length three in L_2.

In what follows, M_k and L_k will refer to the languages defined in (1) and (2) respectively.

Remark 4 Let u_k be the size of the minimal alphabet of L_k. The construction of L_k gives

$$\begin{cases} u_0 = 1, \;\; u_1 = 2, \\ u_k = 1 + \sum_{j=0}^{k-1} u_j \binom{k-1}{j}, \;\; \forall k \geq 2. \end{cases}$$

By direct inspection, one can check that $u_2 = 4$, $u_3 = 10$, $u_4 = 30$, $u_5 = 104 < 5!$ and for $n = 6, \ldots, 9$, $u_n < n!$. Let $k \geq 9$. One has, by induction on k, the following upper bound

$$u_k < \sum_{j=0}^{k-1} j! \binom{k-1}{j} = e\,\Gamma(k,1) < e\,(k-1)!$$

where $\Gamma(k,1)$ is the incomplete gamma function defined by

$$\Gamma(a,b) = \int_b^{+\infty} t^{a-1} e^{-t}\,dt.$$

There are certainly several ways to improve the size of the alphabet. For instance, $a^*b^* \sqcup\!\sqcup \{c\}$ has the same complexity function as $L_2 = (a^* \cup b^+ c^*) \sqcup\!\sqcup \{d\}$ but is over a smaller alphabet. This simple modification could change u_2 and thus u_n for all $n \geq 3$. Moreover, for all n, it is clear that $u_k^n \geq n^k$ (with an alphabet of size u_k, there are at the most u_k^n words of length n and we need at least n^k of them). Therefore we have a lower bound $2^{k/2}$ on the size of the alphabet of a language containing n^k words of length n. Also, there is a systematic construction to get a regular language K over an alphabet with 2^k letters such that $\rho_K(n) = n^k$. Consider the matrix

$$A_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

One has $(A_1^n)_{1,2} = n$. For $k \geq 2$, A_k is the direct product of the matrices A_1 and A_{k-1}, i.e. $A_k = A_1 \otimes A_{k-1}$. Then $(A_k^n)_{1,2^k} = n^k$. This matrix A_k can be viewed as the transition matrix of a deterministic finite automaton over an alphabet of 2^k letters. For instance,

$$A_2 = \begin{pmatrix} 1A_1 & 1A_1 \\ 0A_1 & 1A_1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

is the transition matrix of the automaton sketched in Figure 1. Consequently,

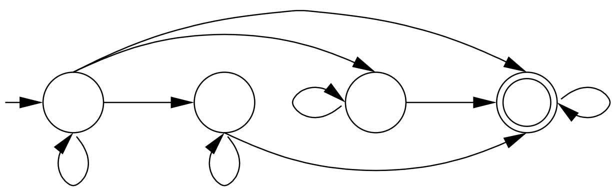

Fig. 1. An automaton with A_2 transition matrix.

there exists a regular language K over an alphabet of size 2^k such that $\rho_K(n) = n^k$.

Nevertheless, in what follows, the main thing is that L_k is constructed in (2) with one last operation of shuffle with a new letter σ_k.

Remark 5 After reading an earlier version of this paper, J. Shallit suggested another construction of a language K such that $\rho_K(n) = n^k$. It uses the following result (see [1, Section 6.5])

$$n^k = \sum_{t=0}^{k} t!\, S(k,t) \binom{n}{t}$$

where $S(k,t)$ are the Stirling numbers of the second kind. The language over $\{a,b\}$ with all strings of length n containing exactly t occurrences of the letter b is regular and has a complexity $\rho(n) = \binom{n}{t}$. Therefore a union of such languages over distinct alphabets gives the language K.

This construction is perhaps simpler than the construction of L_k but uses a larger alphabet. The size of the minimal alphabet is $\max_{t=0,\ldots,k} t!\, S(k,t)$ and a lower bound is given by $k!$. We will not use it in the following because the operation of shuffle given in (2) is needed in our proof of Lemma 10.

ii) Recognizability of polynomials belonging to $\mathbb{N}[x]$

The main idea is that we have to find a regular language L such that the positions of the first words of each length in the lexicographically ordered language L are the values taken by the polynomial. All the proofs of this paper rely on the following lemma.

Lemma 6 *[8] Let L be a regular language over a totally ordered alphabet. The set $\mathcal{I}(L)$ obtained by taking from all the words of L of the same length only the first one in the lexicographic order is regular.*

Proposition 7 *Let $P \in \mathbb{N}[x]$. If $P(\mathbb{N}) \subset \mathbb{N}$ then there exists a numeration system $S = (L, \Sigma, <)$ such that $P(\mathbb{N})$ is S-recognizable.*

PROOF. Since translation by a constant doesn't alter the recognizability of a set, as recalled in the introduction (see [6] for details), we can assume that $P(0) = 0$.

Since $P(n+1) - P(n)$ only contains powers of n with non-negative integral coefficients, the construction of a regular language $L \subset \Sigma^*$ such that $\rho_L(n) = P(n+1) - P(n)$ can be achieved by union of languages L_k over distinct alphabets Σ_k. We fix a total order $<$ on $\Sigma = \cup_k \Sigma_k$ and let $S = (L, \Sigma, <)$.

We can assume that $\varepsilon \in L$ and that the first word w of length 2 in the lexicographically ordered language L, i.e. $\mathcal{I}(L \cap \Sigma^2) = \{w\}$, is such that $\mathrm{val}_S(w) = P(2)$. Indeed, finite modifications of a regular language do not alter its regularity. Notice that we have to consider words of length 2, instead of words of length 1, because $P(1)$ is not necessarily equal to one and therefore cannot possibly be represented by the first word of length 1.

Let $n \geq 2$. Since $\rho_L(n) = P(n+1) - P(n)$, it is clear that if the numerical value of the first word of length n is $P(n)$ then the numerical value of the first word of length $n+1$ is $P(n+1)$. Consequently,

$$\mathrm{r}_S(P(\mathbb{N}) \setminus \{P(1)\}) = \mathcal{I}(L \setminus \Sigma).$$

By Lemma 6, $P(\mathbb{N})$ is S-recognizable (a single word should perhaps be added to a regular language for the S-representation of $P(1)$).

Example 8 Let $P(x) = 2\,x^2 + 3\,x$. Then

$$P(x+1) - P(x) = 4\,x + 5.$$

We consider the language $L \subset \Sigma^*$ which is formed by four copies of L_1 and five copies of L_0. Observe that with five copies of L_0, we obtain five words of any positive length but the only one empty word ε. To ensure that $\mathrm{r}_S(P(2)) = \mathrm{r}_S(14)$ is the first word of length 2 in L, we add to our language four new words of length 1 (we possibly have to add four letters to Σ). This remark applies for all the following constructions: if one uses n copies of L_0 then add $n-1$ words of length 1 and treat the case $n=1$ separately. So, here one can take

$$L = \bigcup_{i=1}^{4} a_i^+ b_i^* \cup \bigcup_{i=1}^{5} c_i^* \cup \{b_1, \dots, b_4\}$$

and $\mathrm{r}_S(\mathbb{N}) = \{b_1\} \cup a_1^* \setminus \{a_1\}$ if S is the numeration system on L induced by the ordering $a_1 < \cdots < a_5 < b_1 < \cdots < b_5 < c_1 < \cdots < c_4$.

Corollary 9 *Let $k \in \mathbb{N} \setminus \{0, 1\}$. There exists a numeration system S such that the set $\{x^k : x \in \mathbb{N}\}$ is S-recognizable.* □

iii) Recognizability of polynomials belonging to $\mathbb{Z}[x]$

The next lemma gets rid of the problem of the coefficients belonging to $\mathbb{Z}$ instead of $\mathbb{N}$.

Lemma 10 *Let k and α be two positive integers. There exist a regular language $\mathcal{L}$ such that $\rho_{\mathcal{L}}(n) = n^k - \alpha\, n^{k-1}$ for all $n \geq \alpha$.*

PROOF. Assume that $k \geq 2$. Let Σ be the minimal alphabet of M_k. From the construction given in (2), one has $L_k = M_k \sqcup\!\sqcup \{\sigma_k\}$ where $\sigma_k \notin \Sigma$. For $i = 1, \ldots, n$, L_k has exactly n^{k-1} words of length n with σ_k in position i. From this observation, the language

$$\mathcal{L} = L_k \setminus \bigcup_{i=0}^{\alpha-1} \Sigma^* \, \sigma_k \, \Sigma^i$$

has exactly $n^k - \alpha \, n^{k-1}$ words of length n for $n \geq \alpha$. Notice that $\rho_{\mathcal{L}}(n) = 0$ if $n < \alpha$.

If $k = 1$, it suffices to consider the language $\mathcal{L} = a^\alpha a^+ b^*$.

Proposition 11 *Let $P \in \mathbb{Z}[x]$. If $P(\mathbb{N}) \subset \mathbb{N}$ then there exists a numeration system $S = (L, \Sigma, <)$ such that $P(\mathbb{N})$ is S-recognizable.*

PROOF. We proceed as in Proposition 7 and consider the polynomial $Q(n) = P(n+1) - P(n)$. Observe that since $P(\mathbb{N}) \subset \mathbb{N}$, the coefficient of the dominant power in P is positive and thus the same remark holds for Q. By adding extra terms of the form $x^j - x^j$, if $\deg(Q) = k$ then $Q(x)$ can be written as

$$x^{i_1+1} - a_{i_1} \, x^{i_1} + \cdots + x^{i_r+1} - a_{i_r} \, x^{i_r} + \sum_{l=0}^{k} b_l \, x^l$$

where $i_1, \ldots, i_r \in \{0, \ldots, k-1\}$, $a_{i_1}, \ldots, a_{i_r} \in \mathbb{N} \setminus \{0\}$ and $b_0, \ldots, b_k \in \mathbb{N}$. Let $\alpha = \sup_{j=1,\ldots,r} a_{i_j}$. Using Lemma 10, for $j = 1, \ldots, r$ we construct languages $\mathcal{L}_j$'s such that for all $n \geq \alpha$, $\rho_{\mathcal{L}_j}(n) = n^{i_j+1} - a_{i_j} \, n^{i_j}$. By union of languages $\mathcal{L}_j$'s and L_l's, we can construct a regular language L such $\forall n \geq \alpha$, $\rho_L(n) = Q(n)$.

We can assume that L contains exactly $P(\alpha)$ words of length at the most $\alpha - 1$. This can be achieved by adding or removing a finite number of words from the language L (this operation doesn't alter the regularity of L). Let S be a numeration system constructed on the ordered regular language L. The first word of length α has a numerical value equal to $P(\alpha)$ and $\forall n \geq \alpha$, $\rho_L(n) = P(n+1) - P(n)$. Then one has

$$\mathrm{r}_S(\{P(n) : n \geq \alpha\}) = \mathcal{I}(L) \cap \Sigma^{\geq \alpha}.$$

Eventually we have to add a finite number of words for the representation of $P(0), \ldots, P(\alpha - 1)$ and

$$\mathrm{r}_S(P(\mathbb{N})) = (\mathcal{I}(L) \cap \Sigma^{\geq \alpha}) \cup \{\mathrm{r}_S(P(0)), \ldots, \mathrm{r}_S(P(\alpha - 1))\}.$$

By Lemma 6, $\mathrm{r}_S(P(\mathbb{N}))$ is regular.

Example 12 Let $P(x) = x^4 - 3\,x^2 - 2\,x + 5$. Then

$$\begin{aligned} Q(n) = P(n+1) - P(n) &= 4\,x^3 + 6\,x^2 - 2\,x - 4 \\ &= 4\,x^3 + 5\,x^2 + x^2 - 3\,x + x - 4. \end{aligned}$$

With four copies of L_3, five copies of L_2 and using Lemma 10, one can construct a regular language L such that

$$\rho_L(n) = \begin{cases} 4\,n^3 + 6\,n^2 - 2\,n - 4, \text{ if } n \geq 4; \\ 4\,n^3 + 5\,n^2, \qquad\qquad \text{otherwise.} \end{cases}$$

We have $P(4) = 205$ and the number of words of length at the most 3 belonging to L is 214. Thus we remove 9 words of length at the most 3 in L. Therefore, the first word of length 4 in L is the representation of $P(4)$ and

$$\mathrm{r}_S(\{P(n) : n \geq 4\}) = \mathcal{I}(L) \cap \Sigma^{\geq 4} \tag{3}$$

is a regular subset of L. Since $\{P(0), \ldots, P(3)\}$ is equal to $\{1, 5, 53\}$, we add the second, the 6^{th} and the 54^{th} word of L to (3) and obtain $\mathrm{r}_S(P(\mathbb{N}))$.

iv) Recognizability of polynomials belonging to $\mathbb{Q}[x]$

Finally, we obtain the theorem of recognizability in the general case.

Theorem 13 *Let $P \in \mathbb{Q}[x]$. If $P(\mathbb{N}) \subset \mathbb{N}$ then there exists a numeration system $S = (L, \Sigma, <)$ such that $P(\mathbb{N})$ is S-recognizable.*

PROOF. Let

$$P(x) = \frac{a_k}{b_k}\,x^k + \frac{a_{k-1}}{b_{k-1}}\,x^{k-1} + \cdots + \frac{a_0}{b_0}$$

with $b_0, \ldots, b_k, a_k \in \mathbb{N} \setminus \{0\}$ and $a_0, \ldots, a_{k-1} \in \mathbb{Z}$. Let s be the least common multiple of $b_0, \ldots, b_k$. One has

$$P = \frac{P'}{s}$$

with $P' \in \mathbb{Z}[x]$. By hypothesis $P(\mathbb{N}) \subset \mathbb{N}$; thus $P'(\mathbb{N}) \subset s\,\mathbb{N}$. As in Proposition 11, there exist a constant α and a regular language $L' \subset \Sigma^*$ such that $\forall n \geq \alpha$,

$$\rho_{L'}(n) = P'(n+1) - P'(n) = s[P(n+1) - P(n)].$$

We modify L' (by adding or removing a finite number of words) to have

$$\sum_{i=0}^{\alpha-1} \rho_{L'}(i) = s\,P(\alpha) = P'(\alpha).$$

In other words, if $\{w\} = \mathcal{I}(L' \cap \Sigma^\alpha)$ then $\mathrm{val}_{S'}(w) = P'(\alpha)$ for the numeration system $S' = (L', \Sigma, <)$ where $<$ is a total ordering of Σ. The arithmetic progression $s\,\mathbb{N}$ is S'-recognizable [6]. Consequently, $L = \mathrm{r}_{S'}(s\,\mathbb{N})$ is a regular language such that

$$\sum_{i=0}^{\alpha-1} \rho_L(i) = P(\alpha) \text{ and } \forall n \geq \alpha,\ \rho_L(n) = P(n+1) - P(n).$$

Indeed, to obtain L one takes in the lexicographically ordered language L' the words at position $is+1$, $i \in \mathbb{N}$. Since the first word of length α in L' is the first word of length α in L and its position in the lexicographically ordered language L is $P(\alpha)$, then we conclude as in Proposition 7, by using Lemma 6.

Example 14 Let

$$\begin{aligned} P(x) &= \frac{x^4}{3} - 2\,x^3 + \frac{37}{6}\,x^2 - \frac{17}{2}\,x + 4 \\ &= \frac{1}{3}\,(x-7)x^2(x+1) + \frac{17}{2}x(x-1) + 4. \end{aligned}$$

It is clear that $P(\mathbb{N}) \subset \mathbb{N}$ since one of the numbers x, $x-7$ or $x+1$ must be divisible by 3 and one of the numbers x or $x-1$ must be divisible by 2. We have $s = 6$ and

$$\begin{aligned} P'(n+1) - P'(n) &= 8\,n^3 - 24\,n^2 + 46\,n - 24 \\ &= 7\,n^3 + 45\,n + n^3 - 24\,n^2 + n - 24. \end{aligned}$$

Using seven copies of L_3, 45 copies of L_1 and applying Lemma 10 twice, we construct a language L' such that

$$\rho_{L'}(n) = \begin{cases} 6\,(P(n+1) - P(n)), & \text{if } n \geq 24; \\ 7\,n^3 + 45\,n, & \text{otherwise.} \end{cases}$$

The number of words of length at the most 23 in L' is 545652 and $6\,P(24) = 517776$. Thus we remove 27876 words from $L' \cap \Sigma^{\leq 23}$. In this new lexicographically ordered language, we only take the words at position $6i+1$, $i \in \mathbb{N}$, to obtain the regular language L. Thus the $[P(24)+1]^{th}$ word of L is the first word of length 24 belonging to L and

$$\rho_L(n) = P(n+1) - P(n) \text{ if } n \geq 24.$$

Hence,

$$\mathrm{r}_S(\{P(n) : n \geq 24\}) = \mathcal{I}(L) \cap \Sigma^{\geq 24}.$$

Eventually we have as usual to add a finite number of words for the representation of $P(0), \ldots, P(23)$.

Remark 15 In [6], we studied the problem of changing the ordering of the alphabet and we exhibited a subset X of $\mathbb{N}$ and two numeration systems S and S' which differ only by the ordering of the alphabet such that $\mathrm{r}_S(X)$ is regular and $\mathrm{r}_{S'}(X)$ not.

This kind of phenomenon doesn't appear here. For a given polynomial P, we have shown how to construct a particular numeration system $S = (L, \Sigma, <)$ such that $P(\mathbb{N})$ is S-recognizable. By construction, one can easily check that $P(\mathbb{N})$ is also T-recognizable for any system $T = (L, \Sigma, \prec)$ where $\prec$ is a reordering of Σ. Indeed, our construction relies only on the complexity function of L and Lemma 6 holds true for any ordering of the alphabet.

3 Exponential polynomial functions

Proceeding in the same way as in the previous section, we show that for any function of the form

$$f(n) = \sum_{i=1}^{k} P_i(n)\, \alpha_i^n,$$

where the P_i's are polynomials with rational coefficients such that $P_i(\mathbb{N})$ is included in $\mathbb{N}$ and the α_i's are non-negative integers, there exists a numeration system S, such that $f(\mathbb{N})$ is S-recognizable. It is interesting to note that each predicate $\{Q(n)\,\alpha^n \,|\, n \in \mathbb{N}\}$ for $\alpha \geq 0$ and Q a polynomial with non-negative integer values is morphic [3].

Proposition 16 *Let $\alpha \in \mathbb{N} \setminus \{0, 1\}$. There exists a numeration system S such that the set $\{\alpha^n : n \in \mathbb{N}\}$ is S-recognizable.*

PROOF. We have to construct a regular language L such that

$$\rho_L(n) = \alpha^{n+1} - \alpha^n = (\alpha - 1)\, \alpha^n.$$

This can be achieved by using $\alpha - 1$ distinct copies of Σ^*, where Σ is an alphabet of cardinality α.

Proposition 17 *Let $\alpha \in \mathbb{N}\setminus\{0,1\}$ and $P \in \mathbb{N}[x]$ such that $P(\mathbb{N}) \subset \mathbb{N}$. There exists a numeration system S such that the set*

$$\{P(n)\,\alpha^n : n \in \mathbb{N}\}$$

is S-recognizable.

PROOF. We have to construct a regular language L such that

$$\rho_L(n) = P(n+1)\,\alpha^{n+1} - P(n)\,\alpha^n = [\alpha\,P(n+1) - P(n)]\,\alpha^n.$$

It is obvious that $\alpha\,P(n+1) - P(n) \in \mathbb{N}[x]$. It is enough to show how to construct a regular language $L^{(k,\alpha)}$ containing exactly $n^k\,\alpha^n$ words of length n; $k \geq 1$. First, we construct $L^{(1,\alpha)}$. Let Σ be such that $\#\Sigma = \alpha$. With α distinct copies of Σ^*, we obtain a language $M_{1,1}$ such that $\rho_{M_{1,1}}(n-1) = \alpha^n$ (for each copy of Σ^*, one has $\rho_{\Sigma^*}(n-1) = \alpha^{n-1}$ and there are α copies). If a does not belong to the minimal alphabet of $M_{1,1}$ then $L^{(1,\alpha)}$ can be defined as

$$L^{(1,\alpha)} = M_{1,1} \sqcup\!\sqcup \{a\}.$$

Indeed, for each of the α^n words w of length $n-1$ in $M_{1,1}$, $w \sqcup\!\sqcup a$ contains n words of length n. Next, we construct $L^{(2,\alpha)}$. With α^2 distinct copies of Σ^*, we obtain a language $M_{2,1}$ such that $\rho_{M_{2,1}}(n-2) = \alpha^n$. If a_1 does not belong to the minimal alphabet of $M_{2,1}$ then $\rho_{M_{2,1} \sqcup\!\sqcup \{a_1\}}(n-1) = (n-1)\,\alpha^n$. Therefore,

$$M_{2,2} = (M_{2,1} \sqcup\!\sqcup \{a_1\}) \cup M_{1,1} \text{ is such that } \rho_{M_{2,2}}(n-1) = n\,\alpha^n$$

where the union is made from languages over distinct alphabets. If a_2 is a new symbol,

$$L^{(2,\alpha)} = M_{2,2} \sqcup\!\sqcup \{a_2\}.$$

Continuing this way, we can construct $L^{(k,\alpha)}$ using the previously defined languages $M_{i,j}$'s and k operations of shuffle with new letters. For instance, one has for $L^{(3,\alpha)}$ the following table:

	description	*complexity*
$M_{3,1}$	α^3 copies of Σ^*	$\rho(n-3) = \alpha^n$
$M_{3,2}$	$M_{3,1} \sqcup\!\sqcup \{a_1\} \bigcup$ 2 copies of $M_{2,1}$	$\rho(n-2) = n\,\alpha^n$
$M_{3,3}$	$M_{3,2} \sqcup\!\sqcup \{a_2\} \bigcup$ 1 copy of $M_{2,2}$	$\rho(n-1) = n^2\,\alpha^n$
$L^{(3,\alpha)}$	$M_{3,3} \sqcup\!\sqcup \{a_3\}$	$\rho(n) = n^3\,\alpha^n$

Remark 18 Observe that the last step in the construction of $L^{(k,\alpha)}$ is the shuffle of $M_{k,k}$ and a new symbol that does not belong to the minimal alphabet of $M_{k,k}$. Moreover,

$$\rho_{M_{k,k}}(n-1) = n^{k-1}\,\alpha^n.$$

So, with the same construction as in Lemma 10 and Proposition 11, we can consider polynomials belonging to $\mathbb{Z}[x]$. Proceeding as in Theorem 13, we can assume that the polynomials belong to $\mathbb{Q}[x]$. Thus the following result is obvious.

Theorem 19 *Let P_i be polynomials belonging to $\mathbb{Q}[x]$ such that $P_i(\mathbb{N}) \subset \mathbb{N}$ and α_i be non-negative integers, $i = 1, \ldots, k$, $k \geq 1$. Set*

$$f(n) = \sum_{i=1}^{k} P_i(n)\,\alpha_i^n.$$

There exists a numeration system S such that $f(\mathbb{N})$ is S-recognizable.

4 Acknowledgments

The author would like to thank J.-P. Allouche and P. Lecomte for their support and fruitful conversations. We also thank J. Shallit for his valuable suggestions. We are grateful to the referees for advice in improving the general presentation of this paper, for the lower bound on the size u_k of the alphabet and for the construction of a language K over an alphabet of size 2^k.

References

[1] R. A. Brualdi, *Introductory Combinatorics*, North-Holland, New York-Oxford-Amsterdam, 1977.

[2] V. Bruyère, G. Hansel, C. Michaux, R. Villemaire, Logic and p-recognizable sets of integers, *Bull. Belg. Math. Soc.* **1** (1994) 191–238.

[3] O. Carton, W. Thomas, The monadic theory of morphic infinite words and generalizations, preprint (2000).

[4] S. Eilenberg, *Automata, Languages and Machines*, Vol. A, Academic Press, New York, 1974.

[5] C. H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, 4th ed. (1965).

[6] P. B. A. Lecomte, M. Rigo, Numeration systems on a regular language, *Theory Comput. Syst.* **34** (2001) 27–44.

[7] M. Rigo, Generalization of automatic sequences for numeration systems on a regular language, *Theoret. Comp. Sci.* **244** (2000) 271–281.

[8] J. Shallit, Numeration systems, linear recurrences, and regular sets, *Information and Computation*, **113** No 2 (1994) 331–347.

[9] A. Szilard, S. Yu, K. Zhang, J. Shallit, Characterizing regular languages with polynomial densities, *Proceedings of the 17th International Symposium on Mathematical Foundations of Computer Science*, *Lect. Notes in Comp. Sci.* **629** (1992) 494–503.