ABSTRACT NUMERATION SYSTEMS ON BOUNDED LANGUAGES AND MULTIPLICATION BY A CONSTANT

Emilie Charlier
Institute of Mathematics, University of Liège, Grande Traverse 12 (B 37), B–4000 Liège, Belgium
echarlier@ulg.ac.be

Michel Rigo[1]
Institute of Mathematics, University of Liège, Grande Traverse 12 (B 37), B–4000 Liège, Belgium
M.Rigo@ulg.ac.be

Wolfgang Steiner[2]
LIAFA, CNRS, Université Paris Diderot – Paris 7, case 7014, 75205 Paris Cedex 13, France
steiner@liafa.jussieu.fr

Received: , Accepted: , Published:

Abstract

A set of integers is S-recognizable in an abstract numeration system S if the language made up of the representations of its elements is accepted by a finite automaton. For abstract numeration systems built over bounded languages with at least three letters, we show that multiplication by an integer $\lambda \geq 2$ does not preserve S-recognizability, meaning that there always exists a S-recognizable set X such that λX is not S-recognizable. The main tool is a bijection between the representation of an integer over a bounded language and its decomposition as a sum of binomial coefficients with certain properties, the so-called combinatorial numeration system.

1. Introduction

An *alphabet* is a finite set whose elements are called *letters*. For a given alphabet Σ, a *word* of length $n \geq 0$ over Σ is a map $w : \{1, \ldots, n\} \to \Sigma$. The length of a word w is denoted by $|w|$. The only word of length 0 is the *empty word* denoted by ε. The set of all words over Σ is Σ^*. The *concatenation* of the words u and v respectively of length m and n is the word $w = uv$ of length $m + n$ where $w(i) = u(i)$ for $1 \leq i \leq m$ and $w(i) = v(i - m)$ for $m + 1 \leq i \leq m + n$. Endowed with the concatenation product, Σ^* is a monoid with ε

[1]The first two authors were supported by an AutoMathA short visit grant (European Science Foundation).
[2]This author was supported by the French Agence Nationale de la Recherche, grant ANR–06–JCJC–0073.

as identity element. For a word u and $j \in \mathbb{N}$, u^j is the concatenation of j copies of u. In particular, we set $u^0 = \varepsilon$. We write $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$. A *language* over Σ is a subset of Σ^*. Since we use $|\cdot|$ to denote the length of a word, we have chosen to denote the cardinality of the set A by $\#A$ to avoid any misunderstanding.

Denote the *bounded language* over the alphabet $\Sigma_\ell = \{a_1, a_2, \ldots, a_\ell\}$ of size $\ell \geq 1$ by

$$\mathcal{B}_\ell = a_1^* a_2^* \cdots a_\ell^* := \{a_1^{j_1} a_2^{j_2} \cdots a_\ell^{j_\ell} \mid j_1, j_2, \ldots, j_\ell \geq 0\}.$$

We always assume that $(\Sigma_\ell, <)$ is totally ordered by $a_1 < a_2 < \cdots < a_\ell$. Let $x, y \in \Sigma_\ell^*$ be two words. Recall that x is *genealogically less than* y either if $|x| < |y|$ or if they have the same length and x is lexicographically smaller than y, i.e., there exist $p, x', y' \in \Sigma_\ell^*$ such that $x = pa_i x'$, $y = pa_j y'$ and $i < j$. We can enumerate the words of $\mathcal{B}_\ell$ using the increasing genealogical ordering (also called radix order or shortlex order) induced by the ordering $<$ of Σ_ℓ. For an integer $n \geq 0$, the $(n+1)$-st word of $\mathcal{B}_\ell$ is said to be the *$\mathcal{B}_\ell$-representation* of n and is denoted by $\mathrm{rep}_\ell(n)$. The reciprocal map $\mathrm{rep}_\ell^{-1} =: \mathrm{val}_\ell$ maps the n-th word of $\mathcal{B}_\ell$ onto its *numerical value* $n-1$. Notice that this map val_ℓ is a special case of a diagonal function as considered for instance in [9]. A set $X \subseteq \mathbb{N}$ is said to be *$\mathcal{B}_\ell$-recognizable* if $\mathrm{rep}_\ell(X)$ is a regular language over the alphabet Σ_ℓ, i.e., accepted by a finite automaton. This one-to-one correspondence between the words of $\mathcal{B}_\ell$ and the integers can be extended to any infinite regular language L over a totally ordered alphabet $(\Sigma, <)$. This leads to the general notion of abstract numeration system.

Definition 1. An *abstract numeration system* is a triple $S = (L, \Sigma, <)$ where L is an infinite regular language over the totally ordered alphabet $(\Sigma, <)$. We denote by $\mathrm{rep}_S(n)$ the $(n+1)$-st word in the genealogically ordered language L. A set X of integers is *S-recognizable* if $\mathrm{rep}_S(X)$ is a regular language.

For an abstract numeration system $S = (L, \Sigma, <)$ where $L = \mathcal{B}_\ell$ and $\Sigma = \Sigma_\ell$, the map rep_S is exactly rep_ℓ. Thus $\mathcal{B}_\ell$-recognizability is a special case of S-recognizability.

Note that the language $\mathcal{B}_\ell$ is recognized by the following automaton: the set of states is $\{q_1, \ldots, q_\ell\}$, each state is final, q_1 is initial, and for $1 \leq i \leq j \leq n$ we have a transition $q_i \xrightarrow{a_j} q_j$. The case $\ell = 4$ is depicted in Figure 1.

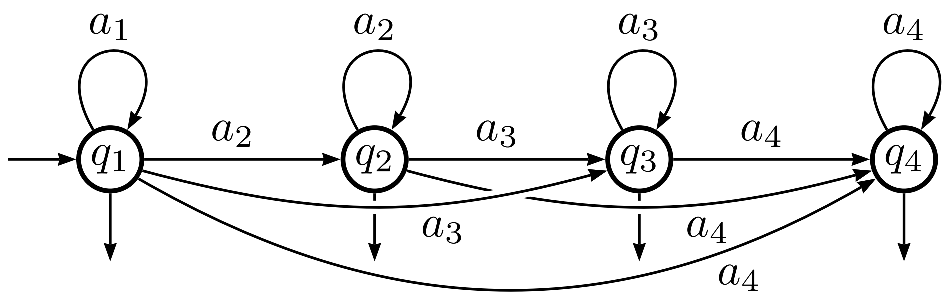

Figure 1: Automaton recognizing $\mathcal{B}_4$.

Example 1. Let $\Sigma_2 = \{a, b\}$ with $a < b$. The first words of $\mathcal{B}_2 = a^*b^*$ enumerated by genealogical order are

$$\varepsilon, a, b, aa, ab, bb, aaa, aab, abb, bbb, aaaa, \ldots$$

For instance, $\mathrm{rep}_2(5) = bb$ and $\mathrm{val}_2(a^*) = \{0, 1, 3, 6, 10, \ldots\}$ is a $\mathcal{B}_2$-recognizable subset of $\mathbb{N}$ (formed of all triangular numbers).

For details on bounded languages, see for instance [5] and for a reference on automata and formal languages theory, see [3].

In the framework of positional numeration systems, recognizable sets of integers have been extensively studied since the seminal work of A. Cobham in the late sixties (see for instance [3, Chap. V]). Since then, the notion of recognizability has been studied from various points of view (logical characterization, automatic sequences, ...). In particular, recognizability for generalized number systems like the Fibonacci system has been considered [2, 12]. Here we shall consider recognizable sets of integers in the general setting of *abstract numeration systems*. It is well-known that the class of regular languages L splits into two parts with respect to the behavior of the function $n \mapsto \#(L \cap \Sigma^n)$ [13]. This latter function is either bounded from above by n^k for some k or, infinitely often bounded from below by θ^n for some $\theta > 1$. In these cases, we speak respectively of *polynomial* and *exponential languages*.

Notice that usual positional numeration systems like integer base systems or the Fibonacci system are special cases of abstract numeration systems built on an exponential language. On the other hand, bounded languages are polynomial and this leads to new phenomena.

The question addressed in the present paper deals with the preservation of the recognizability with respect to the operation of multiplication by a constant. *Let $S = (L, \Sigma, <)$ be an abstract numeration system, X be a S-recognizable set of integers and λ be a positive integer. What can be said about the S-recognizability of λX ?* This question is a first step before handling more complex operations such as addition of two arbitrary recognizable sets.

This question is rather difficult. For exponential languages, partial answers are known (see for instance [2]). The case of polynomial languages has not been considered yet (except for a^*b^* in [7]). Bounded languages are good candidates to start with. Indeed, an arbitrary polynomial language is a finite union of languages of the form $u_1v_1^*u_2v_2^* \cdots v_k^*u_{k+1}$ where the u_i's and v_i's are words [13], and the automata accepting these languages share the same properties as those accepting bounded languages. Therefore we hope that our results give the flavor of what could be expected for any polynomial languages.

Since rep_ℓ is a one-to-one correspondence between $\mathbb{N}$ and $\mathcal{B}_\ell$, the multiplication by a constant $\lambda \in \mathbb{N}$ can be viewed as a transformation $f_\lambda : \mathcal{B}_\ell \to \mathcal{B}_\ell$ acting on the language $\mathcal{B}_\ell$, the question being then to *study the preservation of the regularity of the subsets of $\mathcal{B}_\ell$ under this transformation.*

Example 2. Let $\ell = 2$, $\Sigma_2 = \{a, b\}$ and $\lambda = 25$. We have the following diagram.

$$\begin{array}{ccc} 8 & \xrightarrow{\times 25} & 200 \\ \mathrm{rep}_2 \downarrow & & \downarrow \mathrm{rep}_2 \\ a\,b^2 & \xrightarrow{\times 25} & a^9\,b^{10} \end{array} \qquad \begin{array}{ccc} \mathbb{N} & \xrightarrow{\times \lambda} & \mathbb{N} \\ \mathrm{rep}_2 \downarrow & & \downarrow \mathrm{rep}_2 \\ \mathcal{B}_\ell & \xrightarrow{f_\lambda} & \mathcal{B}_\ell \end{array}$$

Thus the multiplication by $\lambda = 25$ induces a mapping f_λ onto $\mathcal{B}_2$ such that for $w, w' \in \mathcal{B}_2$, $f_\lambda(w) = w'$ if and only if $\mathrm{val}_2(w') = 25\,\mathrm{val}_2(w)$.

This paper is organized as follows. In Section 2, we recall a few results related to our main question. In particular, we characterize the recognizable sets of integers for abstract numeration systems whose language is slender, i.e., has at most d words of each length for some constant d. We easily get that in this situation, the multiplication by a constant always preserves recognizability.

In Section 3, we compute $\mathrm{val}_\ell(a_1^{n_1} \cdots a_\ell^{n_\ell})$ and derive an easy bijective proof of the fact that any nonnegative integer can be written in a unique way as

$$n = \binom{z_\ell}{\ell} + \binom{z_{\ell-1}}{\ell-1} + \cdots + \binom{z_1}{1}$$

with $z_\ell > z_{\ell-1} > \cdots > z_1 \geq 0$. Fraenkel [4] called this system *combinatorial numeration system* and referred to Lehmer [8]. Even if this seems to be a folklore result, the only proof that we were able to trace out goes back to Katona [6] who developed different arguments to obtain the same decomposition.

In Section 4, we make explicit the regular subsets of $\mathcal{B}_\ell$ in terms of semi-linear sets of $\mathbb{N}^\ell$ and give an application to the $\mathcal{B}_\ell$-recognizability of arithmetic progressions.

In Section 5, we answer our main question about bounded languages and recognizability after multiplication by a constant. We get a formula which can be used to obtain estimates on the $\mathcal{B}_\ell$-representation of λn from the one of n. Therefore, thanks to a counting argument and to the results from Section 4, we show that for any constant λ, there exists a $\mathcal{B}_\ell$-recognizable set X such that λX is no more $\mathcal{B}_\ell$-recognizable, with $\ell \geq 3$. Consequently, our main result can be summarized as follows. Let ℓ, λ be positive integers. For the abstract numeration system $S = (a_1^* \cdots a_\ell^*, \{a_1 < \cdots < a_\ell\})$, multiplication by $\lambda \geq 2$ preserves S-recognizability if and only if either $\ell = 1$ or $\ell = 2$ and λ is an odd square.

We put in the last section some structural results concerning the effect of multiplication by a constant in the abstract numeration system built on $\mathcal{B}_\ell$.

2. First results about S-recognizability

In this section we collect a few results directly connected with our problem.

Theorem 1. *[7] Let $S = (L, \Sigma, <)$ be an abstract numeration system. Any arithmetic progression is S-recognizable.*

Let us denote by $\mathbf{u}_L(n)$ (resp. $\mathbf{v}_L(n)$) the number of words of length n (resp. at most n) belonging to L. The following result states that only some constants λ are good candidates for multiplication within $\mathcal{B}_\ell$.

Theorem 2. *[11] Let $L \subseteq \Sigma^*$ be a regular language such that $\mathbf{u}_L(n) = \Theta(n^k)$ for some $k \in \mathbb{N}$ and $S = (L, \Sigma, <)$. Preservation of S-recognizability after multiplication by λ holds only if $\lambda = \beta^{k+1}$ for some $\beta \in \mathbb{N}$.*

We write $f = \Theta(g)$ if there exist N and $C > 0$ such that for all $n \geq N$, $f(n) \leq C\, g(n)$ (i.e., $f = \mathcal{O}(g)$) and also if there exist $D > 0$ and an infinite sequence $(n_i)_{i\in\mathbb{N}}$ such that $f(n_i) \geq D\, g(n_i)$ for all $i \geq 0$.

As we shall see in the next section that $\mathbf{u}_{\mathcal{B}_\ell}(n) = \Theta(n^{\ell-1})$, we have to focus only on multipliers of the form β^ℓ. The particular case of $\mathbf{u}_L(n) = \mathcal{O}(1)$ (i.e., L is *slender*) is interesting in itself and is settled as follows. Let us first recall the definition from [1] and the characterization from [10, 12] of such languages.

Definition 2. The language L is said to be *d-slender* if $\mathbf{u}_L(n) \leq d$ for all $n \geq 0$. The language L is said to be *slender* if it is d-slender for some $d > 0$.

A regular language L is slender if and only if it is a *union of single loops*, i.e., if for some $k \geq 1$ and words x_i, y_i, z_i, $1 \leq i \leq k$,

$$L = \bigcup_{i=1}^{k} x_i\, y_i^* z_i.$$

Moreover, we can assume that the sets $x_i\, y_i^* z_i$ are pairwise disjoint. Notice that the regular expression $x_i\, y_i^* z_i$ is a shorthand to denote the language $\{x_i y_i^n z_i \mid n \geq 0\}$, again $x_i y_i^n z_i$ has to be understood as the concatenation of x_i, n copies of y_i and then followed by z_i.

Theorem 3. *Let $L \subseteq \Sigma^*$ be a slender regular language and $S = (L, \Sigma, <)$. A set $X \subseteq \mathbb{N}$ is S-recognizable if and only if X is a finite union of arithmetic progressions.*

Proof. By the characterization of slender languages, we have

$$L = \bigcup_{i=1}^{k} x_i\, y_i^* z_i \cup F,\ \ x_i, z_i \in \Sigma^*, y_i \in \Sigma^+$$

where the sets $x_i\, y_i^* z_i$ are pairwise disjoint and F is a finite set. The sequence $(\mathbf{u}_L(n))_{n\in\mathbb{N}}$ is ultimately periodic of period $C = \mathrm{lcm}_i |y_i|$. Moreover, for n large enough, if $x_i\, y_i^n\, z_i$ is the m-th word of length $|x_i\, z_i| + n\, |y_i|$ then $x_i\, y_i^{n+C/|y_i|}\, z_i$ is the m-th word of length $|x_i\, z_i| + n\, |y_i| + C$.

Roughly speaking, for sufficiently large n, the structures of the ordered sets of words of length n and $n+C$ are the same.
The regular subsets of L are of the form

$$\bigcup_{j\in J} x_{i_j}\,(y_{i_j}^{\alpha_j})^* z_{i_j} \cup F' \tag{1}$$

where J is a finite set, $i_j \in \{1,\ldots,k\}$, $\alpha_j \in \mathbb{N}$ and F' is a finite subset of L.
We can now conclude. If X is S-recognizable, then $\mathrm{rep}_S(X)$ is a regular subset of L of the form (1). In view of the first part of the proof, it is clear that X is ultimately periodic with period length $\mathrm{lcm}(C, \mathrm{lcm}_j|y_{i_j}^{\alpha_j}|)$. The converse is immediate by Theorem 1. □

Example 3. Consider the language $L = ab^*c \cup b(aa)^*c$. It contains exactly two words of each positive even length: $ab^{2i}c < ba^{2i}c$ and one word for each odd length larger than 2: $ab^{2i+1}c$. The sequence $\mathbf{u}_L(n)$ is ultimately periodic of period two: $0,0,2,1,2,1,\ldots$.

Corollary 1. *Let S be a numeration system built on a slender language. If $X \subseteq \mathbb{N}$ is S-recognizable, then λX is S-recognizable for all $\lambda \in \mathbb{N}$.*

Finally, for a bounded language over a binary alphabet, the case is completely settled too, the aim of this paper being primarily to extend the following result.

Theorem 4. *[7] Let β be a positive integer. For the abstract numeration system $S = (a^*b^*, \{a < b\})$, multiplication by β^2 preserves S-recognizability if and only if β is odd.*

3. $\mathcal{B}_\ell$-representation of integers : combinatorial expansion

In this section we determine the number of words of a given length in $\mathcal{B}_\ell$ and we obtain an algorithm for computing $\mathrm{rep}_\ell(n)$. Interestingly, this algorithm is related to the decomposition of n as a sum of binomial coefficients of a specified form. Since we shall be mainly interested by the language $\mathcal{B}_\ell$, we use the following notation.

Definition 3. We set

$$\mathbf{u}_\ell(n) := \mathbf{u}_{\mathcal{B}_\ell}(n) = \#(\mathcal{B}_\ell \cap \Sigma_\ell^n) \quad \text{and} \quad \mathbf{v}_\ell(n) := \#(\mathcal{B}_\ell \cap \Sigma_\ell^{\leq n}) = \sum_{i=0}^{n} \mathbf{u}_\ell(i).$$

Let us also recall that the binomial coefficient $\binom{i}{j}$ vanishes for integers $i < j$.

Lemma 1. *For all $\ell \geq 1$ and $n \geq 0$, we have*

$$\mathbf{u}_{\ell+1}(n) = \mathbf{v}_\ell(n) \tag{2}$$

and

$$\mathbf{u}_\ell(n) = \binom{n+\ell-1}{\ell-1}. \tag{3}$$

Proof. Relation (2) follows from the fact that the set of words of length n belonging to $\mathcal{B}_{\ell+1}$ is partitioned according to

$$\bigcup_{i=0}^{n} \left(a_1^* \cdots a_\ell^* \cap \Sigma_\ell^i\right) a_{\ell+1}^{n-i}.$$

To obtain (3), we proceed by induction on $\ell \geq 1$. Indeed, for $\ell = 1$, it is clear that $\mathbf{u}_1(n) = 1$ for all $n \geq 0$. Assume that (3) holds for ℓ and let us verify it still holds for $\ell + 1$. Thanks to (2), we have

$$\mathbf{u}_{\ell+1}(n) = \sum_{i=0}^{n} \mathbf{u}_\ell(i) = \sum_{i=0}^{n} \binom{i+\ell-1}{\ell-1} = \sum_{i=0}^{n} \binom{i+\ell-1}{i} = \binom{n+\ell}{\ell}. \qquad \square$$

Lemma 2. *Let* $S = (a_1^* \cdots a_\ell^*, \{a_1 < \cdots < a_\ell\})$. *We have*

$$\mathrm{val}_\ell(a_1^{n_1} \cdots a_\ell^{n_\ell}) = \sum_{i=1}^{\ell} \binom{n_i + \cdots + n_\ell + \ell - i}{\ell - i + 1}. \tag{4}$$

Consequently, for any $n \in \mathbb{N}$,

$$|\mathrm{rep}_\ell(n)| = k \Leftrightarrow \underbrace{\binom{k+\ell-1}{\ell}}_{\mathrm{val}_\ell(a_1^k)} \leq n \leq \underbrace{\sum_{i=1}^{\ell} \binom{k+i-1}{i}}_{\mathrm{val}_\ell(a_\ell^k)}.$$

Proof. From the structure of the ordered language $\mathcal{B}_\ell$, one can show that

$$\mathrm{val}_\ell(a_1^{n_1} \cdots a_\ell^{n_\ell}) = \mathrm{val}_\ell(a_1^{n_1+\cdots+n_\ell}) + \mathrm{val}_{\{a_2,\ldots,a_\ell\}}(a_2^{n_2} \cdots a_\ell^{n_\ell}) \tag{5}$$

where notation like $\mathrm{val}_{\{a_2,\ldots,a_\ell\}}(w)$ specifies not only the size but the alphabet of the bounded language on which the numeration system is built. To understand this formula, an example is given below in the case $\ell = 3$. Notice that $\mathrm{val}_{\{a_2,\ldots,a_\ell\}}(a_2^{n_2} \cdots a_\ell^{n_\ell}) = \mathrm{val}_{\ell-1}(a_1^{n_2} \cdots a_{\ell-1}^{n_\ell})$. Using this latter observation and iterating the decomposition (5), we obtain

$$\mathrm{val}_\ell(a_1^{n_1} \cdots a_\ell^{n_\ell}) = \sum_{i=1}^{\ell} \mathrm{val}_{\ell-i+1}(a_1^{n_i+\cdots+n_\ell}).$$

Moreover, it is well known that $\mathrm{val}_\ell(a_1^n) = \mathbf{v}_\ell(n-1)$. Hence the conclusion follows using relations (2) and (3). $\square$

Example 4. Consider the words of length 3 in the language $a^*b^*c^*$,

$$aaa < aab < aac < abb < abc < acc < bbb < bbc < bcc < ccc.$$

We have $\mathrm{val}_3(aaa) = \binom{5}{3} = 10$ and $\mathrm{val}_3(acc) = 15$. If we apply the erasing morphism $\varphi : \{a,b,c\} \to \{a,b,c\}^*$ defined by $\varphi(a) = \varepsilon$, $\varphi(b) = b$ and $\varphi(c) = c$ on the words of length 3, we get

$$\varepsilon < b < c < bb < bc < cc < bbb < bbc < bcc < ccc.$$

So the ordered list of words of length 3 in $a^*b^*c^*$ contains an ordered copy of the words of length at most 2 in the language b^*c^* and to obtain $\mathrm{val}_3(acc)$, we just add to $\mathrm{val}_3(aaa)$ the position of the word cc in the ordered language b^*c^*. In other words, $\mathrm{val}_3(acc) = \mathrm{val}_3(aaa) + \mathrm{val}_2(cc)$ where val_2 is considered as a map defined on the language b^*c^*.

The following result is given in [6]. Here we obtain a bijective proof relying only on the use of abstract numeration systems on a bounded language.

Corollary 2 (Combinatorial numeration system). *Let ℓ be a positive integer. Any integer $n \geq 0$ can be uniquely written as*

$$n = \binom{z_\ell}{\ell} + \binom{z_{\ell-1}}{\ell-1} + \cdots + \binom{z_1}{1} \tag{6}$$

with $z_\ell > z_{\ell-1} > \cdots > z_1 \geq 0$.

Proof. The map $\mathrm{rep}_\ell : \mathbb{N} \to a_1^* \cdots a_\ell^*$ is a one-to-one correspondence. So any integer n has a unique representation of the form $a_1^{n_1} \cdots a_\ell^{n_\ell}$ and the conclusion follows from Lemma 2. □

The general method given in [7, Algorithm 1] has a special form in the case of the language $\mathcal{B}_\ell$. We derive an algorithm computing the decomposition (6) or equivalently the $\mathcal{B}_\ell$-representation of any integer.

Algorithm 1. Let `n` be an integer and `l` be a positive integer. The following algorithm produces integers `z(l)`,...,`z(1)` corresponding to the z_i's appearing in the decomposition (6) of `n` given in Corollary 2.

```
For i=l,l-1,...,1 do
   if n>0,
      find t such that (t choose i) ≤ n < (t+1 choose i)
      z(i)←t
      n←n-(t choose i)
   otherwise, z(i)←i-1
```

Consider now the triangular system having $n_1, \ldots, n_\ell$ as unknowns

$$n_i + \cdots + n_\ell = \mathtt{z}(\ell - i + 1) - \ell + i, \quad i = 1, \ldots, \ell.$$

One has $\mathrm{rep}_\ell(\mathtt{n}) = a_1^{n_1} \cdots a_\ell^{n_\ell}$.

Remark 1. To speed up the computation of `t` in the above algorithm, one can benefit from methods of numerical analysis. Indeed, for given `i` and `n`, $\binom{\mathtt{t}}{\mathtt{i}} - \mathtt{n}$ is a polynomial in `t` of degree `i` and we are looking for the largest root z of this polynomial. Therefore, $\mathtt{t} = \lfloor z \rfloor$.

Example 5. For $\ell = 3$, one gets for instance

$$12345678901234567890 = \binom{4199737}{3} + \binom{3803913}{2} + \binom{1580642}{1}$$

and solving the system

$$\left.\begin{array}{rcl} n_1 + n_2 + n_3 & = & 4199737 - 2 \\ n_2 + n_3 & = & 3803913 - 1 \\ n_3 & = & 1580642 \end{array}\right\} \Leftrightarrow (n_1, n_2, n_3) = (395823, 2223270, 1580642),$$

we have $\mathrm{rep}_3(12345678901234567890) = a^{395823} b^{2223270} c^{1580642}$.

4. Regular subsets of $\mathcal{B}_\ell$

To study preservation of recognizability after multiplication by a constant, one has to consider an arbitrary recognizable subset $X \subseteq \mathbb{N}$ and show that $\beta^\ell X$ is still recognizable.

Definition 4. If w is a word over Σ_ℓ, $|w|_{a_j}$ counts the number of letters a_j in w. The *Parikh mapping* Ψ maps a word $w \in \Sigma_\ell^*$ onto the vector $\Psi(w) := (|w|_{a_1}, \dots, |w|_{a_\ell})$.

Remark 2. In this setting of bounded languages, rep_ℓ and Ψ are both one-to-one correspondences. Therefore, in what follows we shall make no distinction between an integer n, its $\mathcal{B}_\ell$-representation $\mathrm{rep}_\ell(n) = a_1^{n_1} \cdots a_\ell^{n_\ell} \in \mathcal{B}_\ell$ and the corresponding Parikh vector $\Psi(\mathrm{rep}_\ell(n)) = (n_1, \dots, n_\ell) \in \mathbb{N}^\ell$. In examples, when considering cases $\ell = 2$ or 3, we shall use convenient alphabets like $\{a < b\}$ or $\{a < b < c\}$.

Definition 5. A set $Z \subseteq \mathbb{N}^\ell$ is *linear* if there exist $\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_k \in \mathbb{N}^\ell$ such that

$$Z = \mathbf{p}_0 + \mathbb{N}\,\mathbf{p}_1 + \cdots + \mathbb{N}\,\mathbf{p}_k = \{\mathbf{p}_0 + \lambda_1 \mathbf{p}_1 + \cdots + \lambda_k \mathbf{p}_k \mid \lambda_1, \dots, \lambda_k \in \mathbb{N}\}.$$

The vectors $\mathbf{p}_1, \dots, \mathbf{p}_k$ are said to be the *periods* of Z. The set Z is *k-dimensional* if it has exactly k linearly independent periods over $\mathbb{Q}$. A set is *semi-linear* if it is a finite union of linear sets. The set of periods of a semi-linear set is the union of the sets of periods of the corresponding linear sets. Let $\mathbf{e}_i \in \mathbb{N}^\ell$, $1 \le i \le \ell$, denote the vector having 1 in the i-th component and 0 in the other components.

Lemma 3. *A set $X \subseteq \mathbb{N}$ is $\mathcal{B}_\ell$-recognizable if and only if $\Psi(\mathrm{rep}_\ell(X))$ is a semi-linear set whose periods are integer multiples of canonical vectors $\mathbf{e}_i$.*

Proof. Observe that the regular subsets of $\mathcal{B}_\ell$ are exactly the finite unions of sets of the form $a_1^{s_1}(a_1^{t_1})^* \cdots a_\ell^{s_\ell}(a_\ell^{t_\ell})^*$ with $s_i, t_i \in \mathbb{N}$. □

With such a characterization, we obtain an alternative proof of Theorem 1.

Proposition 1. *Let $p, q \in \mathbb{N}$. The set $\Psi(\mathrm{rep}_\ell(q+\mathbb{N}\,p)) \subseteq \mathbb{N}^\ell$ is a finite union of linear sets of the form*

$$\mathbf{x}+\mathbb{N}\,P\,\mathbf{e}_1+\cdots+\mathbb{N}\,P\,\mathbf{e}_\ell \quad \textit{for some } P\in\mathbb{N}.$$

Proof. We use equation (4). For a given i, $1 \le i \le \ell$, the sequence $(\binom{n}{\ell-i+1} \bmod p)_{n\in\mathbb{N}}$ is periodic (see e.g. [14]). Denote the period lengths by π_i and set $P = \mathrm{lcm}_i\,\pi_i$. Then

$$\mathrm{val}_\ell(a_1^{x_1}\cdots a_i^{x_i}\cdots a_\ell^{x_\ell}) \equiv \mathrm{val}_\ell(a_1^{x_1}\cdots a_i^{x_i+P}\cdots a_\ell^{x_\ell}) \pmod p \quad \text{for all } i,\ 1\le i\le \ell.$$

We have just shown that $\mathbf{x}=(x_1,\ldots,x_\ell)\in\mathbb{N}^\ell$ belongs to $\Psi(\mathrm{rep}_\ell(q+\mathbb{N}\,p))$ if and only if $\mathbf{x}+n_1\,P\,\mathbf{e}_1+\cdots+n_\ell\,P\,\mathbf{e}_\ell$ belongs to the same set for all $n_1,\ldots,n_\ell\in\mathbb{N}$. Therefore

$$\Psi(\mathrm{rep}_\ell(q+\mathbb{N}\,p)) = \bigcup_{\substack{\mathrm{val}_\ell(a_1^{x_1}\cdots a_\ell^{x_\ell})\in q+\mathbb{N}\,p\\ 0\le\sup x_i<q+P}} (\mathbf{x}+\mathbb{N}\,P\,\mathbf{e}_1+\cdots+\mathbb{N}\,P\,\mathbf{e}_\ell). \qquad \square$$

Example 6. In Figure 2, the x-axis (resp. y-axis) counts the number of a_1's (resp. a_2's) in a word. The empty word corresponds to the lower-left corner. A point in $\mathbb{N}^2$ of coordinates (i, j) has its color determined by the value of $\mathrm{val}_2(a_1^i\,a_2^j)$ modulo p (with $p = 3, 5, 6$ and 8 respectively). There are therefore p possible colors. In this figure, we represent words $a_1^i\,a_2^j$ for $0 \le i, j \le 19$.

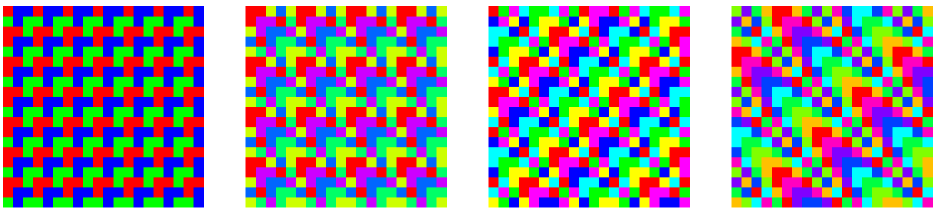

Figure 2: $\Psi(\mathrm{rep}_2(q+\mathbb{N}\,p))$ for $p = 3, 5, 6, 8$.

5. Multiplication by $\lambda = \beta^\ell$

In the case of a bounded language on ℓ letters, if multiplication by some constant preserves recognizability, then, by Theorem 2 and Lemma 1, this constant must be a ℓ-th power.

The next result gives a relationship between the length of the $\mathcal{B}_\ell$-representations of n and $\beta^\ell n$, roughly by a factor β.

Lemma 4. *For sufficiently large $n \in \mathbb{N}$, we have*

$$|\mathrm{rep}_\ell(\beta^\ell n)| = \beta\,|\mathrm{rep}_\ell(n)| + \left\lceil \frac{(\beta-1)(\ell+1)}{2} \right\rceil - i$$

for some $i \in \{0, 1, \ldots, \beta\}$.

Proof. Consider first $n = \mathrm{val}_\ell(a_\ell^q)$ for some sufficiently large $q \in \mathbb{N}$, and let

$$\beta^\ell \left(\binom{q+\ell-1}{\ell} + \binom{q+\ell-2}{\ell-1} + \cdots + \binom{q}{1} \right) = \binom{z_\ell+\ell-1}{\ell} + \binom{z_{\ell-1}+\ell-2}{\ell-1} + \cdots + \binom{z_1}{1}$$

for some integers $z_\ell \geq z_{\ell-1} \geq \cdots \geq z_1 \geq 0$ (depending on q). Then we have

$$\beta^\ell \left(\frac{q^\ell}{\ell!} + \frac{(\ell+1)\, q^{\ell-1}}{2\,(\ell-1)!} + \mathcal{O}(q^{\ell-2}) \right) = \frac{z_\ell^\ell}{\ell!} + \frac{(\ell-1)\, z_\ell^{\ell-1}}{2\,(\ell-1)!} + \frac{z_{\ell-1}^{\ell-1}}{(\ell-1)!} + \mathcal{O}(z_\ell^{\ell-2}),$$

thus $z_\ell = \beta q + \mathcal{O}(1)$. Since $z_\ell \geq z_{\ell-1}$, we have $z_{\ell-1} = d\beta q + o(q)$ with $0 \leq d \leq 1$ and we obtain

$$\frac{\beta^\ell(\ell+1)}{2\,(\ell-1)!} q^{\ell-1} = \frac{\beta^{\ell-1}}{(\ell-1)!} \left((z_\ell - \beta q) + \frac{\ell-1}{2} + d^{\ell-1} \right) q^{\ell-1} + \mathcal{O}(q^{\ell-2}),$$

$$z_\ell = \beta q + \frac{(\beta-1)(\ell+1)}{2} + 1 - d^{\ell-1}.$$

Set $c = (\beta-1)(\ell+1)/2$ and assume first $c \notin \mathbb{Z}$. Then we have $d^{\ell-1} = 1/2$, hence

$$|f_{\beta^\ell}(a_\ell^q))| = z_\ell = \beta q + \lceil c \rceil.$$

Since $\mathrm{val}_\ell(a_1^q) = \mathrm{val}_\ell(a_\ell^{q-1}) + 1$, we have

$$|\mathrm{rep}_\ell(\beta^\ell \mathrm{val}_\ell(a_1^q))| \geq \beta(q-1) + \lceil c \rceil = \beta q + \lceil c \rceil - \beta.$$

If $|\mathrm{rep}_\ell(n)| = q$, then $|\mathrm{rep}_\ell(\beta^\ell n)|$ is clearly between these two values.
Assume now $c \in \mathbb{Z}$. Then we have $d \in \{0, 1\}$. Similarly to the computation of $c_{\ell-2}$ achieved in Remark 3 below, we obtain that

$$\begin{aligned} &\binom{\beta q + c + \ell}{\ell} - \beta^\ell \binom{q+\ell}{\ell} \\ &\quad = \left(\frac{c^2}{2} + \frac{(\ell+1)c}{2} + \frac{(1-\beta^2)(3\ell+2)(\ell+1)}{24} \right) \frac{(\beta q)^{\ell-2}}{(\ell-2)!} + \mathcal{O}(q^{\ell-3}) \\ &\quad = \frac{c(\beta+1)}{12} \frac{(\beta q)^{\ell-2}}{(\ell-2)!} + \mathcal{O}(q^{\ell-3}). \end{aligned}$$

This means that the numerical value of the first word of length $\beta q + c + 1$ is larger than $\beta^\ell \mathrm{val}_\ell(a_1^{q+1})$ for large enough q. We infer that $d = 1$ since

$$z_\ell = |\mathrm{rep}_\ell(\beta^\ell \mathrm{val}_\ell(a_\ell^q))| \leq |\mathrm{rep}_\ell(\beta^\ell \mathrm{val}_\ell(a_1^{q+1}))| < \beta q + c + 1.$$

As above, we have $|\mathrm{rep}_\ell(\beta^\ell \mathrm{val}_\ell(a_1^q))| \geq \beta q + c - \beta$, and the lemma is proved. □

In certain cases, we can give a formula for the entire expansion of $\beta^\ell \mathrm{val}_\ell(a_\ell^q)$.

Lemma 5. *Define $c_{\ell-1}, c_{\ell-2}, \dots, c_0$ recursively by*

$$c_k = k!\,(\beta^{\ell-k}-1)\sum_{i=k}^{\ell}\frac{S_1(i,k)}{i!} - \sum_{i=k+2}^{\ell}\sum_{j=k+1}^{i}\frac{S_1(i,j)\,j!}{i!\,(j-k)!}c_{i-1}^{j-k}$$

where $S_1(i,j)$ are the unsigned Stirling numbers of the first kind. Then we have

$$\beta^\ell\left(\binom{q+\ell-1}{\ell}+\binom{q+\ell-2}{\ell-1}+\cdots+\binom{q}{1}\right) = \binom{\beta q+c_{\ell-1}+\ell-1}{\ell}+\binom{\beta q+c_{\ell-2}+\ell-2}{\ell-1}+\cdots+\binom{\beta q+c_0}{1}. \quad (7)$$

Moreover, if all c_k's, $0 \le k < \ell$, are integers and $c_{\ell-1} \ge c_{\ell-2} \ge \cdots \ge c_0$, then

$$\mathrm{rep}_\ell(\beta^\ell \mathrm{val}_\ell(a_\ell^q)) = a_1^{c_{\ell-1}-c_{\ell-2}} a_2^{c_{\ell-2}-c_{\ell-3}} \cdots a_{\ell-1}^{c_1-c_0} a_\ell^{\beta q+c_0}$$

for all $q \ge -c_0/\beta$, hence $\mathrm{rep}_\ell(\beta^\ell \mathrm{val}_\ell(a_\ell^*))$ is regular.

Proof. The second part of the lemma is obvious. Thus we only have to show (7). Recall that the unsigned Stirling numbers of the first kind are defined by

$$i!\binom{x+i-1}{i} = x(x+1)\cdots(x+i-1) = \sum_{j=1}^{i} S_1(i,j)x^j$$

and satisfy the recursion

$$S_1(i+1,j) = S_1(i,j-1) + i\,S_1(i,j) \quad \text{for } 1 \le j \le i$$

with $S_1(i,j) = 0$ if $i < j$ or $j = 0$. Therefore we can write (7) as

$$\beta^\ell\left(\sum_{k=1}^{\ell}\frac{S_1(\ell,k)}{\ell!}q^k + \sum_{k=1}^{\ell-1}\frac{S_1(\ell-1,k)}{(\ell-1)!}q^k + \cdots + q\right) = \sum_{j=1}^{\ell}\frac{S_1(\ell,j)}{\ell!}(\beta q+c_{\ell-1})^j + \sum_{j=1}^{\ell-1}\frac{S_1(\ell-1,j)}{(\ell-1)!}(\beta q+c_{\ell-2})^j + \cdots + \beta q + c_0,$$

$$\beta^\ell\sum_{i=1}^{\ell}\sum_{k=1}^{i}\frac{S_1(i,k)}{i!}q^k = \sum_{i=1}^{\ell}\sum_{j=1}^{i}\frac{S_1(i,j)}{i!}\sum_{k=0}^{j}\binom{j}{k}c_{i-1}^{j-k}\beta^k q^k,$$

$$\beta^{\ell-k}\sum_{i=k}^{\ell}\frac{S_1(i,k)}{i!} = \sum_{i=k}^{\ell}\sum_{j=k}^{i}\frac{S_1(i,j)\,j!}{i!\,(j-k)!\,k!}c_{i-1}^{j-k} \quad \text{for } 0 \le k \le \ell.$$

Since the last equation holds for $k = \ell$ and

$$\beta^{\ell-k}\sum_{i=k}^{\ell}\frac{S_1(i,k)}{i!} = \sum_{i=k}^{\ell}\frac{S_1(i,k)}{i!} + \frac{c_k}{k!} + \sum_{i=k+2}^{\ell}\sum_{j=k+1}^{i}\frac{S_1(i,j)\,j!}{i!\,(j-k)!\,k!}c_{i-1}^{j-k}$$

for $0 \le k < \ell$ by the definition of c_k, the lemma is proved. ☐

Remark 3. The formula for c_k can be simplified using

$$\sum_{i=k}^{\ell} \frac{S_1(i,k)}{i!} = \begin{cases} S_1(\ell+1,k+1)/\ell! & \text{for } k \geq 1, \\ 0 & \text{for } k = 0. \end{cases}$$

Note that $c_{\ell-1}$ is the constant c in the proof of Lemma 4,

$$c_{\ell-1} = (\beta-1)\frac{S_1(\ell+1,\ell)}{\ell} = \frac{(\beta-1)(\ell+1)}{2} \quad \text{for } \ell \geq 2.$$

Since $S_1(\ell+1,\ell-1) = S_1(\ell,\ell-2) + \ell\frac{\ell(\ell-1)}{2} = \frac{(3\ell+2)(\ell+1)\ell(\ell-1)}{24}$, we have

$$\begin{aligned} c_{\ell-2} &= (\beta^2-1)\frac{(3\ell+2)(\ell+1)}{24} - \frac{\ell-1}{2}c_{\ell-1} - \frac{1}{2}c_{\ell-1}^2 \\ &= c_{\ell-1}\left(1 - \frac{\beta+1}{12}\right) = \frac{(\beta-1)(\ell+1)}{2} - \frac{(\beta^2-1)(\ell+1)}{24} \quad \text{for } \ell \geq 3. \end{aligned}$$

We now turn to our main counting argument that will be used to obtain that recognizability is not preserved through multiplication by a constant λ. Recall that $f_\lambda : \mathcal{B}_\ell \to \mathcal{B}_\ell$ is defined by $f_\lambda(w) = \mathrm{rep}_\ell(\lambda\, \mathrm{val}_\ell(w))$.

Lemma 6. *Let A be a k-dimensional linear subset of $\mathbb{N}^\ell$ for some integer $k < \ell$ and $B = \Psi^{-1}(A) \cap \mathcal{B}_\ell$ be the corresponding subset of $\mathcal{B}_\ell$. If $\Psi(f_{\beta^\ell}(B))$ contains a sequence $x^{(n)} = (x_1^{(n)}, \ldots, x_\ell^{(n)})$ such that $\min(x_{j_1}^{(n)}, x_{j_2}^{(n)}, \ldots, x_{j_{k+1}}^{(n)}) \to \infty$ as $n \to \infty$ for some $j_1 < j_2 < \cdots < j_{k+1}$, then $f_{\beta^\ell}(B)$ is not regular.*

Proof. Since A is a k-dimensional linear subset of $\mathbb{N}^\ell$, we clearly have

$$\#\{w \in B : |w| \leq n\} = \#\{x \in A : x_1 + \cdots + x_\ell \leq n\} = \Theta(n^k)$$

and, by Lemma 4, $\#\{w \in f_{\beta^\ell}(B) : |w| \leq n\} = \Theta(n^k)$. Thus $f_{\beta^\ell}(B)$ is regular if and only if $\Psi(f_{\beta^\ell}(B))$ is a finite union of at most k-dimensional sets as in Lemma 3. Since the sequence $x^{(n)}$ cannot occur in such a finite union, $f_{\beta^\ell}(B)$ is not regular. □

The coefficients $c_{\ell-1}$ and $c_{\ell-2}$ (explicitely given in Remark 3) are rational numbers. In the next two propositions, we discuss the fact that these coefficients could be integers and we rule out all the possible cases.

Proposition 2. *If $\frac{(\beta-1)(\ell+1)}{2} \notin \mathbb{Z}$ or $\frac{(\beta^2-1)(\ell+1)}{24} \notin \mathbb{Z}$ (and $\ell \geq 3$, $\beta \geq 2$), then $f_{\beta^\ell}(a_\ell^*)$ is not regular.*

Proof. We use notation of the proof of Lemma 4.

First case : $c_{\ell-1} = \frac{(\beta-1)(\ell+1)}{2} \notin \mathbb{Z}$
We have $z_\ell = \beta q + c_{\ell-1} + 1/2$, $z_{\ell-1} = 2^{-1/(\ell-1)}\beta q + o(q)$, hence

$$|f_{\beta^\ell}(a_\ell^q)|_{a_1} = (1 - 2^{-1/(\ell-1)})\beta q + o(q),$$
$$\sum_{j=2}^{\ell} |f_{\beta^\ell}(a_\ell^q)|_{a_j} = 2^{-1/(\ell-1)}\beta q + o(q),$$

and $f_{\beta^\ell}(a_\ell^*)$ is not regular by Lemma 6.

Second case : $c_{\ell-1} = \frac{(\beta-1)(\ell+1)}{2} \in \mathbb{Z}$
We have $z_\ell = \beta q + c_{\ell-1}$, $z_{\ell-1} = \beta q + \mathcal{O}(1)$ and $z_{\ell-2} = d\beta q + o(q)$ with $0 \le d \le 1$. By comparing the coefficients of $q^{\ell-2}$, we obtain

$$z_{\ell-1} = \beta q + c_{\ell-2} + 1 - d^{\ell-2}$$

Since in this case $c_{\ell-2} = \frac{(\beta-1)(\ell+1)}{2} - \frac{(\beta^2-1)(\ell+1)}{24} \notin \mathbb{Z}$, we have $0 < d < 1$, hence

$$|f_{\beta^\ell}(a_\ell^q)|_{a_2} = (1-d)\beta q + o(q), \quad \sum_{j=3}^{\ell} |f_{\beta^\ell}(a_\ell^q)|_{a_j} = d\beta q + o(q),$$

and $f_{\beta^\ell}(a_\ell^*)$ is not regular by Lemma 6. □

Proposition 3. *If $\frac{(\beta-1)(\ell+1)}{2} \in \mathbb{Z}$ and $\frac{(\beta^2-1)(\ell+1)}{24} \in \mathbb{Z}$ (and $\ell \ge 3$, $\beta \ge 2$), then $f_{\beta^\ell}(a_1^* a_\ell^*)$ is not regular.*

Proof. If we choose q large enough with respect to p, e.g. $q = p^3$, then we have

$$\beta^\ell \left(\binom{p+q+\ell-1}{\ell} + \binom{q+\ell-2}{\ell-1} + \binom{q+\ell-3}{\ell-2} + \cdots + \binom{q}{1} \right)$$
$$= \binom{\beta(p+q) + c_{\ell-1} + \ell - 1}{\ell} + \binom{\beta q - (\beta-1)\beta p + c_{\ell-2} + \ell - 2}{\ell-1}$$
$$+ \binom{\beta q - \frac{(\beta-1)\beta}{2}(\beta p)^2 + \mathcal{O}(p)}{\ell-2} + \mathcal{O}\big(q^{\ell-3}\big).$$

Indeed, this equation holds for $p = 0$ by Lemma 5. Therefore the coefficients of $q^\ell p^0$, $q^{\ell-1}p^0$ and $q^{\ell-2}p^0$ on the left-hand side are equal to those on the right-hand side. It is easy to see that the same holds for $q^{\ell-1}p^1$, $q^{\ell-2}p^2$ and $q^{\ell-3}p^3$. For $q^{\ell-2}p^1$ and $q^{\ell-3}p^2$, consider the following equations:

$$(\ell-2)!\,\beta^{1-\ell}\big[q^{\ell-2}p^1\big]: \quad \beta\frac{\ell-1}{2} = c_{\ell-1} + \frac{\ell-1}{2} - (\beta-1),$$
$$(\ell-3)!\,\beta^{1-\ell}\big[q^{\ell-3}p^2\big]: \quad \beta\frac{\ell-1}{4} = \frac{c_{\ell-1}}{2} + \frac{\ell-1}{4} + \frac{(\beta-1)^2}{2} - \frac{(\beta-1)\beta}{2}.$$

If the $\mathcal{O}(p)$ term is chosen properly, then the coefficient of $q^{\ell-3}p^1$ vanishes as well and $\mathcal{O}\big(q^{\ell-3}\big)$ remains. Since $c_\ell, c_{\ell-1} \in \mathbb{Z}$, we have thus

$$|f_{\beta^\ell}(a_1^p a_\ell^q)|_{a_1} = \beta^2 p + \mathcal{O}(1),$$
$$|f_{\beta^\ell}(a_1^p a_\ell^q)|_{a_2} = \frac{(\beta-1)\beta^3}{2}p^2 + \mathcal{O}(p),$$
$$\sum_{j=3}^{\ell} |f_{\beta^\ell}(a_1^p a_\ell^q)|_{a_j} = \beta q + \mathcal{O}(p^2),$$

and $f_{\beta^\ell}(a_1^* a_\ell^*))$ is not regular by Lemma 6. □

Example 7. We just illustrate some of the above computations. If $\ell = 3$, then we have $c_2 = 2(\beta - 1)$, $c_1 = 2(\beta - 1) - (\beta^2 - 1)/6$ and

$$c_0 = -\frac{c_1}{2} - \frac{c_1^2}{2} - \frac{c_2}{3} - \frac{c_2^2}{2} - \frac{c_2^3}{6} = -\frac{(\beta^2-1)^2}{72} - (\beta^3 - 1) - \frac{\beta^2-1}{4} + 2(\beta - 1).$$

If $\beta \equiv \pm 1 \pmod 6$, then this gives

$$f_{\beta^3}(a_3^q) = a_1^{\frac{\beta^2-1}{6}} a_2^{\frac{(\beta^2-1)^2}{72}+\beta^3-1+\frac{\beta^2-1}{12}} a_3^{\beta q - \frac{(\beta^2-1)^2}{72} - (\beta^3-1) - \frac{\beta^2-1}{4} + 2(\beta-1)}.$$

In particular, this latter formula shows that a_3^* cannot be used to prove that multiplication by β^3 does not preserve recognizability when $\beta \equiv \pm 1 \pmod 6$. Thanks to Proposition 2, $f_{\beta^3}(a_3^q)$ is regular if and only if $\beta \equiv \pm 1 \pmod 6$.

Otherwise, i.e., if $1 - \beta^2 \equiv j \pmod 6$ with $j \in \{1, 3, 4\}$, then $z_3 = \beta q + c_2$, $z_2 = \beta q + c_1 + 1 - j/6$ and

$$z_1 = \frac{j}{6}\beta q + c_0 - \frac{(1-j/6)^2}{2} - (1 - j/6)c_1 - \frac{1-j/6}{2}.$$

If we collect results from Theorems 2, 3, 4 and Propositions 2 and 3, we obtain the main result about multiplication by a constant.

Theorem 5. *Let ℓ, λ be positive integers. For the abstract numeration system*

$$S = (a_1^* \cdots a_\ell^*, \{a_1 < \cdots < a_\ell\}),$$

multiplication by $\lambda \geq 2$ preserves S-recognizability if and only if one of the following condition is satisfied :

- $\ell = 1$
- $\ell = 2$ *and λ is an odd square.*

Proof. The case $\ell = 1$ is ruled out by Theorem 3, the case $\ell = 2$ is given by Theorem 4. Consider $\ell \geq 3$. Thanks to Theorem 2, it suffices to consider λ of the β^ℓ and the conclusion follows from Propositions 2 and 3. □

6. Structural properties of $\mathcal{B}_\ell$ seen through f_{β_ℓ}

In this independent section, we inspect closely how a word is transformed when applying f_{β^ℓ}. To that end, $\mathcal{B}_\ell$ (or equivalently $\mathbb{N}$) is partitioned into regions where f_{β^ℓ} acts differently. Thanks to our discussion, we are able to detect some kind of pattern occurring periodically within these regions. To have a flavor of the computations involved in this section, the reader could first have a look at Example 8. According to Lemma 4, we define a partition of $\mathbb{N}$.

Definition 6. For all $i \in \{0, 1, \dots, \beta\}$ and $k \in \mathbb{N}$ large enough, we define

$$\mathcal{R}_{i,k} := \left\{ n \in \mathbb{N} : |\mathrm{rep}_\ell(n)| = k \text{ and } |\mathrm{rep}_\ell(\beta^\ell n)| = \beta\, k + \left\lceil \frac{(\beta-1)(\ell+1)}{2} \right\rceil - i \right\}.$$

Lemma 7. *If $\beta = \prod_{i=1}^k p_i^{\theta_i}$ where $p_1, \dots, p_k$ are prime numbers greater than ℓ and the θ_i's are positive integers, then for any $u \geq \ell$, we have*

$$\binom{u}{\ell} \equiv \binom{u+\beta^\ell}{\ell} \pmod{\beta^\ell}.$$

Proof. Let $u, v \geq \ell$. One has

$$\binom{v}{\ell} - \binom{u}{\ell} = \frac{v(v-1)\cdots(v-\ell+1) - u(u-1)\cdots(u-\ell+1)}{\ell!}.$$

The numerator on the r.h.s. is an integer divisible by $\ell!$. Moreover, this numerator is also clearly divisible by $v-u$ (indeed, it is of the form $P(v) - P(u)$ for some polynomial P). Notice that for $v = u + \beta^\ell$, the corresponding numerator is divisible by $\ell!$ and also by β^ℓ. But since any prime factor of β is larger than ℓ, $\ell!$ and β^ℓ are relatively prime. Consequently, the corresponding numerator is divisible by $\beta^\ell \ell!$. ☐

An inspection of multiplication by β^ℓ using the partition induced by Lemma 4 provides us with the following observation.

Proposition 4. *Let $m_{i,k} = \min \mathcal{R}_{i,k}$ for $k \geq 0$ and $i \in \{0, \dots, \beta\}$. If β satisfies the condition of Lemma 7, then*

$$|\mathrm{rep}_\ell(\beta^\ell m_{i,k})|_{a_j} = |\mathrm{rep}_\ell(\beta^\ell m_{i,k+\beta^{\ell-1}})|_{a_j}$$

for all k large enough and $j \in \{2, \dots, \ell\}$. Furthermore,

$$|\mathrm{rep}_\ell(\beta^\ell m_{i,k+\beta^{\ell-1}})|_{a_1} = |\mathrm{rep}_\ell(\beta^\ell m_{i,k})|_{a_1} + \beta^\ell.$$

If $i < \beta$, then $m_{i,k} = \lceil C_i(k)/\beta^\ell \rceil$ with

$$C_i(k) = \mathrm{val}_\ell\left(a_1^{\beta\, k + \frac{(\beta-1)(\ell+1)}{2} - i}\right) = \binom{\beta\, k + \frac{(\beta-1)(\ell+1)}{2} - i + \ell - 1}{\ell}.$$

Proof. For $i = \beta$, we clearly have $m_{\beta,k} = \mathrm{val}_\ell(a_1^k)$ if $\mathcal{R}_{\beta,k}$ is non-empty, and it is easily verified that $\mathcal{R}_{\beta,k}$ is non-empty if k is large enough (and $\ell \geq 2$).
For $i < \beta$, note first that $(\beta - 1)(\ell + 1)$ is even since β satisfies the condition of Lemma 7. Thus we have

$$C_i(k) \leq \beta^\ell m_{i,k} < C_{i-1}(k)$$

Since $m_{i,k} - 1 \in \mathcal{R}_{i+1,k}$, we also obtain

$$C_{i+1}(k) + \beta^\ell \leq \beta^\ell m_{i,k} < C_i(k) + \beta^\ell.$$

Therefore $m_{i,k} = \lceil C_i(k)/\beta^\ell \rceil$ and there exists a unique integer $\mu_i(k)$ such that

$$\beta^\ell m_{i,k} = C_i(k) + \mu_i(k) \quad \text{and} \quad 0 \leq \mu_i(k) < \beta^\ell.$$

In particular, there exists also a unique integer $\mu_i(k + \beta^{\ell-1})$ such that

$$\beta^\ell m_{i,k+\beta^{\ell-1}} = C_i(k + \beta^{\ell-1}) + \mu_i(k + \beta^{\ell-1}) \quad \text{and} \quad 0 \leq \mu_i(k + \beta^{\ell-1}) < \beta^\ell.$$

From Lemma 7, we deduce that $C_i(k) \equiv C_i(k + \beta^{\ell-1}) \pmod{\beta^\ell}$ and consequently, $\mu_i(k) = \mu_i(k + \beta^{\ell-1})$. From Lemma 2, we deduce that

$$\mathrm{rep}_\ell(\beta^\ell m_{i,k}) = a_1^t \,\mathrm{rep}_{\{a_2,\ldots,a_\ell\}}(\mu_i(k)),$$

where t is such that $|\mathrm{rep}_\ell(\beta^\ell m_{i,k})| = \beta\, k + \frac{(\beta-1)(\ell+1)}{2} - i$, and

$$\mathrm{rep}_\ell(\beta^\ell m_{i,k+\beta^{\ell-1}}) = a_1^{t+\beta^\ell} \,\mathrm{rep}_{\{a_2,\ldots,a_\ell\}}(\mu_i(k)). \qquad \square$$

Remark 4. In the previous proposition, we were interested in the first word in $\mathcal{R}_{i,k}$ but we can even describe how multiplication by β^ℓ affects representations inside $\mathcal{R}_{i,k}$. With notation of the previous proof, for any $n \in \mathcal{R}_{i,k}$ (and k large enough), we have

$$\mathrm{rep}_\ell(\beta^\ell n) = a_1^t \,\mathrm{rep}_{\{a_2,\ldots,a_\ell\}}(\mu_i(k) + \beta^\ell(n - m_{i,k}))$$

with t such that $|\mathrm{rep}_\ell(\beta^\ell n)| = \beta\, k + \frac{(\beta-1)(\ell+1)}{2} - i$.

Example 8. Let $\ell = 3$ and $\beta = 5$. The number 171717 (resp. 172739) is the first element belonging to $\mathcal{R}_{4,100}$ (resp. $\mathcal{R}_{3,100}$). We have

$$\mathrm{rep}_3(171717) = a^{95}b^3c^2 \text{ and } \mathrm{rep}_3(5^3\, 171717) = a^{490}\underline{b^{14}c^0},$$

$$\mathrm{rep}_3(172739) = a^{55}b^{41}c^4 \text{ and } \mathrm{rep}_3(5^3\, 172739) = a^{493}\underline{b^0c^{12}}.$$

Therefore $\mu_4(100) = \mathrm{val}_{\{b,c\}}(b^{14}) = 105$ (resp. $\mu_3(100) = \mathrm{val}_{\{b,c\}}(c^{12}) = 90$). The number 333396 (resp. 334986) is the smallest element in $\mathcal{R}_{4,125}$ (resp. $\mathcal{R}_{3,125}$),

$$\mathrm{rep}_3(333396) = a^{119}b^6c^0 \text{ and } \mathrm{rep}_3(5^3\, 333396) = a^{615}\underline{b^{14}c^0},$$

$$\mathrm{rep}_3(334986) = a^{69}b^{41}c^{15} \text{ and } \mathrm{rep}_3(5^3\, 334986) = a^{618}\underline{b^0c^{12}}.$$

We have $\#\mathcal{R}_{4,100} = 1022$, $\#\mathcal{R}_{4,125} = 1590$ and get the following table.

j	$\Psi(\mathrm{rep}_3(5^3(m_{4,100}+j)))$	$\Psi(\mathrm{rep}_3(5^3(m_{4,125}+j)))$	$\Psi(\mathrm{rep}_{\{b,c\}}(\mu_4(100)+5^3j))$
0	$(490, 14, 0)$	$(615, 14, 0)$	$(14, 0)$
1	$(484, 0, 20)$	$(609, 0, 20)$	$(0, 20)$
2	$(478, 22, 4)$	$(603, 22, 4)$	$(22, 4)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1021	$(0, 34, 470)$	$(125, 34, 470)$	$(34, 470)$
1022	$\times$	$(124, 415, 90)$	$(415, 90)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1589	$\times$	$(0, 34, 595)$	$(34, 595)$

Acknowledgments

We thank P. Lecomte for fruitful discussions during the elaboration of this paper.

References

[1] M. Andraşiu, G. Păun, J. Dassow, A. Salomaa, Language-theoretic problems arising from Richelieu cryptosystems, *Theoret. Comput. Sci.* **116** (1993), 339–357.

[2] V. Bruyère, G. Hansel, Bertrand numeration systems and recognizability, *Theoret. Comput. Sci.* **181** (1997) 17–43.

[3] S. Eilenberg, *Automata, Languages, and Machines* vol. A, Academic Press, New York, (1974).

[4] A. Fraenkel, Systems of numeration, *Amer. Math. Monthly* **92** (1985), 105–114.

[5] S. Ginsburg, E. H. Spanier, Bounded ALGOL-like languages, *Trans. Am. Math. Soc.* **113** (1964), 333–368.

[6] G. Katona, A theorem on finite sets, Theory of Graphs, Proc. Colloquium, Tihany, Hungary (1966), 187–207.

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

[8] D. H. Lehmer, The machine tools of combinatorics, in *Applied Combinatorial Mathematics* (E. F. Beckenbach Ed.), Wiley, New York, (1964), 5–31.

[9] J. S. Lew, L. B. Morales, A. Sánchez-Flores, Diagonal polynomials for small dimensions, *Math. Systems Theory* **29** (1996), 305–310.

[10] G. Păun, A. Salomaa, Thin and slender languages, *Discrete Appl. Math.* **61** (1995), 257–270.

[11] M. Rigo, Numeration systems on a regular language: arithmetic operations, recognizability and formal power series, *Theoret. Comput. Sci.* **269** (2001), 469–498.

[12] J. Shallit, Numeration systems, linear recurrences, and regular sets, *Inform. and Comput.* **113** (1994), 331–347.

[13] A. Szilard, S. Yu, K. Zhang, J. Shallit, *Characterizing regular languages with polynomial densities*, Mathematical foundations of computer science 1992 (Prague, 1992), Springer, Berlin, 1992, pp. 494–503.

[14] Ś. Ząbek, *Sur la périodicité modulo m des suites de nombres* $\binom{n}{k}$, Ann. Univ. Mariae Curie-Skłodowska, Sect. A **10** (1958), 37–47.