Automata and Numeration Systems*

Véronique Bruyère
Université de Mons-Hainaut,
15 Avenue Maistriau,
B-7000 Mons, Belgium.
vero@sun1.umh.ac.be

1 Introduction

This article is a short survey on the following problem: given a set $X \subseteq \mathbb{N}$, find a "simple algorithm" accepting X and rejecting $\mathbb{N} \setminus X$. By simple algorithm, we mean a finite automaton, a substitution, a logical formula, ...

We will see that these algorithms strongly depend on the way one represents the integers. However, once the base of representation is fixed, these models are all equivalent.

This article is divided into two parts. Standard bases like base 10 are first considered. The properties are then generalized to nonstandard bases like the Fibonacci one. No proofs are given, but several examples try to explain the results.

The state of the art on this subject was presented at the "Séminaire Lotharingien de Combinatoire", Hesselberg, Germany, 4–6 October 1995.

2 Standard Numeration Systems

2.1 Dependence on the Base

We begin with an example: the set X of powers of two

$$X = \{2^n \mid n \in \mathbb{N}\} \ .$$

*This work was partially supported by ESPRIT-BRA Working Group 6317 *ASMICS*

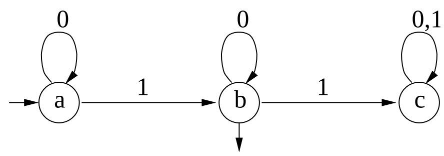

Figure 1: Automaton for the powers of two in base 2.

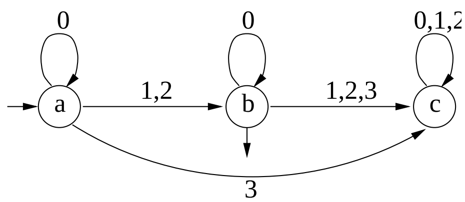

Figure 2: Automaton for the powers of two in base 4.

In base 2, X is represented as the set of binary words equal to $X_2 = \{1, 10, 100, 1000, \ldots\}$. The *finite automaton* of Figure 1 accepts X because the words of X_2 (with any number of leading zeroes) are exactly the labels of the paths going from the *initial state* a to the *final state* b. We say that X is 2-*recognizable*.

In base 4, there is also a finite automaton accepting X (see Figure 2). The set X is represented in base 4 as $X_4 = \{1, 2, 10, 20, 100, 200, \ldots\}$ and the words of X_4 over the alphabet $\{0, 1, 2, 3\}$ are the labels of the paths from state a to b in Figure 2 (leading zeroes are allowed). The set X is then 4-recognizable.

The automaton in base 4 is easily constructed from the automaton in base 2. The three states are identical, the arrows of the second automaton are exactly the paths with length 2 of the first one. Indeed letters 0, 1, 2, 3 in base 4 correspond to words 00, 01, 10, 11 in base 2.

The set X written in base 3 is $X_3 = \{1, 2, 11, 22, 121, 1012, \ldots\}$. No regularity appears inside the first words of X_3. Does this mean that X is *not* 3-recognizable?

To answer this question, we are going to state some results, in particular the famous Cobham's theorem.

Given a base $p \geq 2$, a set $X \subseteq \mathbb{N}$ is called *p-recognizable* if X written in base p is accepted by a finite automaton. Two bases $p, q \geq 2$ are said *multiplicatively dependent* if $p = r^k$, $q = r^l$ for some $r, k, l \in \mathbb{N}$. Otherwise, they are *multiplicatively independent*. The relation "to be multiplicatively dependent" divides $\mathbb{N} \setminus \{0,1\}$ into equivalence classes $\{\{2,4,8,\ldots\}$, $\{3,9,27,\ldots\}$, $\{5,25,125,\ldots\}$, $\{6,36,216\}$, $\ldots\}$.

It is rather easy to prove that *inside a class*, the existence of an automaton accepting $X \subseteq \mathbb{N}$ is independent of the base (remember the example above with bases 2 and 4).

Proposition 1 *Let $p, q \geq 2$ be two multiplicatively dependent bases. Then a set $X \subseteq \mathbb{N}$ is p-recognizable if and only if it is q-recognizable.*

Some sets X admit a finite automaton in *any* base. These are the *ultimately periodic* sets, equal to finite unions of arithmetic progressions, like the set $X = \{3n \mid n \in \mathbb{N}\} \cup \{3n+1 \mid n \in \mathbb{N}\}$.

Proposition 2 *Let $X \subseteq \mathbb{N}$ be an ultimately periodic set, then X is p-recognizable for any $p \geq 2$.*

Cobham's theorem [14] states that ultimately periodic sets are the only ones to be recognizable in every base. It states more: X is ultimately periodic as soon as it is recognizable in two bases chosen in two distinct equivalence classes. This result shows that the set X of powers of two cannot be 3-recognizable, as this set is not ultimately periodic.

Theorem 1 *Let $p, q \geq 2$ be two multiplicatively independent bases. If a set $X \subseteq \mathbb{N}$ is p-recognizable and q-recognizable, then X is ultimately periodic.*

The reader is referred to Chapter 5 of Eilenberg's book [17] for properties of p-recognizable sets. Cobham's proof [14] is only 4 pages long, but Eilenberg says that "it is a challenge to find a more reasonable proof". The first comprehensible proof is Hansel's one [26, 38]. Other authors have recently found simple logical proofs [34]. An extension of Cobham's theorem to subsets X of N^m with $m \geq 2$ is due to Semenov [41]. The proof of Semenov is difficult; simpler proofs are available in [36, 10], [35] and [3]. Fagnot [20] has recently extended Cobham's theorem where the set X is replaced by two sets X, Y with the same "set of factors".

2.2 Equivalence of the Models

We again consider the example X of the powers of two and the (fixed) base 2. We already know that X is 2-recognizable.

It is also 2-*substitutive* in the following way. There exists a 2-*substitution* $f : A = \{a, b, c\} \to A^2$ and a projection $g : A \to \{0, 1\}$

$$f \;:\; \begin{array}{lcl} a & \to & ab \\ b & \to & bc \\ c & \to & cc \end{array} \qquad g \;:\; \begin{array}{lcl} a & \to & 0 \\ b & \to & 1 \\ c & \to & 0 \;. \end{array}$$

In this situation, the algorithm accepting X runs as follows: infinitely iterate f on letter a, apply g on the generated infinite word. Notice that this infinite word is a fixpoint of f because $f(a)$ begins with a.

$$\begin{array}{ccccccccccc} a & \to & ab & \to & abbc & \to & abbcbccc & \to & \dots & & \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & & & \\ 0 & \to & 01 & \to & 0110 & \to & 01101000 & \to & \dots & \to & \underline{X} \;. \end{array}$$

The generated binary infinite word is precisely the *characteristic word* $\underline{X}$ of X

$$n \in X \Leftrightarrow \underline{X}_n = 1 \;.$$

Another model to accept X uses formal series. With X we associate the *formal series*

$$X(y) = \sum_{n=0}^{\infty} y^{2^n} = y + y^2 + y^4 + \dots \;.$$

The set X is called 2-*algebraic* because the series $X(y)$ is algebraic over $\mathbb{F}_2[y]$, i.e., it is root of the polynomial $p(t) = t^2 + t + y$ with coefficients in $\mathbb{F}_2[y]$. Indeed, remember that in the field $\mathbb{F}_2$ with characteristic 2, one has $1+1 = 0$.

The last model uses logical formulas. Let $\langle \mathbb{N}, +, V_2 \rangle$ be the logical *structure* where $+$ denotes the usual addition in $\mathbb{N}$ and $V_2(x) = y$ means that y is the greatest power of two dividing x. The computation of $y = V_2(x)$ is easy if x is written in base 2: if the representation of x is $u10^n$, then the representation of y is 10^n. With this structure, we can write *first-order formulas* with variables describing $\mathbb{N}$, the equality $=$, the two functions $+$ and V_2, the connectives $\vee, \wedge, \neg, \to, \leftrightarrow$ and the quantifiers $\exists, \forall$ (on the variables).

We say that the set X of powers of two is 2-*definable* because X is defined by the formula $\varphi(x)$ equal to $V_2(x) = x$. Another example of 2-definable set is $\{3n \mid n \in \mathbb{N}\}$ using the formula $(\exists y)(x = y + y + y)$.

Let us now give the general definitions for a fixed base $p \geq 2$. The first one is the notion of p-recognizable set introduced in Subsection 2.1. A set $X \subseteq \mathbb{N}$ is called *p-substitutive* if there exist a *p-substitution* $f : A \to A^p$ (with $f(a)$ beginning with a for some $a \in A$) and a projection $g : A \to \{0, 1\}$ such that the characteristic infinite word $\underline{X}$ of X is equal to $g(f^\omega(a))$. If p is a prime number, X is said to be *p-algebraic* if the formal series $X(y) = \sum_{n \in X} y^n$ is algebraic over $K[y]$ with K a field with characteristic p. Finally, we say that X is *p-definable* if X equals $\{x \in \mathbb{N} \mid \varphi(x) \text{ is true }\}$ where $\varphi(x)$ is a first-order formula of the structure $\langle \mathbb{N}, +, V_p \rangle$. Recall that $V_p(x) = y$ means that y is the greatest power of p dividing x.

As shown on the example above, the four models are equivalent.

Theorem 2 *Let $p \geq 2$ and $X \subseteq N$. Are equivalent*
1. X is p-recognizable,
2. X is p-substitutive,
3. X is p-algebraic (if p is prime),
4. X is p-definable.

Theorem 2 collects the works of several authors. Equivalence (1) $\Leftrightarrow$ (2) is proved in [15] where Cobham calls "tag systems" the p-substitutions.

Let us show on the example of the powers of two, how substitutions simulate automata. Remember that this set is 2-substitutive, thanks to f defined by $f(a) = ab$, $f(b) = bc$, $f(c) = cc$, and g defined by $g(a) = 0$, $g(b) = 1$, $g(c) = 0$. The 2-substitution f simulates the transitions of the automaton of Figure 1. Indeed the alphabet A used by f is the set of states $\{a, b, c\}$ of the automaton. The image of a by f is ab because there is an arrow from state a to state a labelled by 0 and an arrow from state a to state b labelled by 1. The iteration of f on the initial state a indicates the list of states reached by binary words of length 1, 2, 3, ... The projection g indicates, by a 1, which states are final.

		0	1		00	01	10	11		000	001	010	011	100	101	110	111
a	$\to$	a	b	$\to$	a	b	b	c	$\to$	a	b	b	c	b	c	c	c
$\downarrow$																	
0	$\to$	0	1	$\to$	0	1	1	0	$\to$	0	1	1	0	1	0	0	0

Equivalence (1) $\Leftrightarrow$ (3) is proved in [13].

Equivalence (1) $\Leftrightarrow$ (4) is due to Büchi [11], with a proof only for base 2 and structure $\langle \mathbb{N}, +, P_2 \rangle$ instead of $\langle \mathbb{N}, +, V_2 \rangle$ (predicate $P_2(x)$ means that

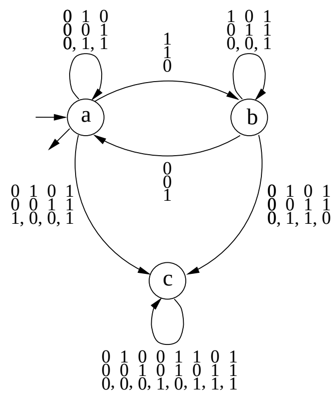

Figure 3: Automaton for the addition in base 2.

x is a power of p). In a review of Büchi's paper, MacNaughton [32] says that predicate P_2 is not correct and suggests predicate $e_2(x,y)$ meaning that y is a power of 2 which appears with digit 1 in the representation of x in base 2. In [8], Büchi's paper is corrected and generalized to any base p with structure $\langle \mathbb{N}, +, V_p \rangle$ instead of $\langle \mathbb{N}, +, P_p \rangle$. Moreover the proof of (4) $\Rightarrow$ (1) uses techniques developed by Hodgson in [28]. The other implication (1) $\Rightarrow$ (4) is first simplified in [33] and then in [46].

The characterization of automata by formulas is plenty of information for the following reasons. Addition can be checked by a finite automaton. The case of base 2 is given on Figure 3: the two states a, b indicate correct computation without and with carry respectively, the state c indicates uncorrect computation. Notice that words must be read from right to left.

Multiplication cannot be checked by a finite automaton (because it cannot be defined by a formula of $\langle \mathbb{N}, +, V_p \rangle$). Only multiplication by a constant is allowed since $c * y$, with c an integer constant, is defined in $\langle \mathbb{N}, +, V_p \rangle$ by $y + \cdots + y$ (c times). The set X whose "elements are numbers $27 * y$ with $y = z + 1$ and z divisible by 4" is p-recognizable for any $p \geq 2$. Indeed, X is definable by the following formula $\varphi(x)$ of $\langle \mathbb{N}, + \rangle$ which is a substructure of $\langle \mathbb{N}, +, V_p \rangle$

$$(\exists y)(\exists z)(\exists t)(x = 27 * y) \wedge (y = z + 1) \wedge (z = 4 * t)$$

(constants are definable in $\langle \mathbb{N}, + \rangle$). A proof based on automata would be rather tedious.

Hodgson's paper [28] shows that $\mathbb{N}$, $=$, $+$ and V_p are the "core" of the automata. It describes a generic method to translate a formula of $\langle \mathbb{N}, +, V_p \rangle$ into a finite automaton, as soon as the "basic" automata for $\mathbb{N}$, $\{(x,y) \in \mathbb{N}^2 \mid x = y\}$, $\{(x,y,z) \in \mathbb{N}^3 \mid x + y = z\}$ and $\{(x,y) \mid V_p(x) = y\}$ are known.

Finally, notice [33] that the structure $\langle \mathbb{N}, +, V_p \rangle$ has the same expressive power than the structure $\langle \mathbb{N}, +, e_p(x,y) \rangle$ proposed by MacNaughton. Notice also that Büchi has really made a mistake by using $\langle \mathbb{N}, +, P_p \rangle$ instead of $\langle \mathbb{N}, +, V_p \rangle$ which is more powerful than $\langle \mathbb{N}, +, P_p \rangle$ [16, 42, 12, 8].

3 Nonstandard Numeration Systems

3.1 Representations of Integers

The Fibonacci sequence can be used as a nonstandard base. For instance, the integer sixteen is represented as the following binary words

...	34	13	8	5	3	2	1
		1	0	0	1	0	0
		1	0	0	0	1	1
			1	1	1	0	0
			1	1	0	1	1

Among these four representations, one is more natural, the first one given by the Euclidean algorithm. We call it the *normalized representation* of sixteen in the Fibonacci base.

More generally [22], a *base* or *numeration system* is a striclty increasing sequence $U = (U_n)_{n \in \mathbb{N}}$ of integers such that

1. $U_0 = 1$ (in a way to represent all $n \in \mathbb{N}$),
2. $\sup \frac{U_{n+1}}{U_n} < \infty$ (in a way to have a finite alphabet of digits when using the Euclidean algorithm).

The *canonical alphabet* associated with U is therefore $A_U = \{0, 1, \ldots, c\}$ with c maximum such that $c < \sup \frac{U_{n+1}}{U_n}$. A *representation* of $n \in \mathbb{N}$ is any word $a_k \cdots a_0 \in A_U^*$ such that $n = \sum_{l=0}^{k} a_l U_l$. The particular representation given by the Euclidean algorithm is called *normalized representation*. These definitions naturally generalize the usual bases $p \geq 2$.

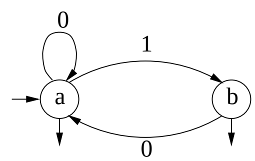

Figure 4: Automaton for $\mathbb{N}$ in Fibonacci base.

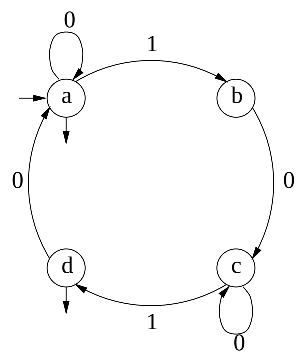

Figure 5: Automaton for Thue-Morse set in Fibonacci base.

3.2 U-Recognizable Sets

Let U be a numeration system. A set $X \subseteq \mathbb{N}$ is *U-recognizable* is the normalized representations of the elements of X are accepted by a finite automaton. Figures 4 and 5 show two examples for the Fibonacci base, respectively $X = \mathbb{N}$ and $X = \{n \in \mathbb{N} \mid$ the normalized representation of n has an even number of 1's $\}$.

To get a logical characterization of U-recognizable sets, as it was done for standard bases in Theorem 2, the first task is the construction of a finite automaton for the set $\mathbb{N}$, the equality $=$, the addition $+$ and the function V_U. This is done in the next sections.

3.3 U-Recognizability of $\mathbb{N}$

Contrarily to standard bases $p \geq 2$, for some numeration systems, a finite automaton for $\mathbb{N}$ does not exist.

Proposition 3 [44, 31] *Let U be a numeration system. If $\mathbb{N}$ is U-recognizable, then U satisfies a linear recurrence relation with integer coefficients.*

For instance, if U is the Fibonacci base, we know that $\mathbb{N}$ is U-recognizable (see Figure 4) and we have $U_n = U_{n-1} + U_{n-2}$, for all $n \geq 2$. The same holds for the usual base 10 for which $U_n = 10U_{n-1}$.

The converse of Proposition 3 is false, since $\mathbb{N}$ is not U-recognizable for the numeration system U defined by the recurrence $U_n = 2U_{n-2} + U_{n-3}$ and the initial conditions $U_0 = 1$, $U_1 = 2$ and $U_2 = 4$.

Hollander [29] has recently described which recurrence relations allow a finite automaton for $\mathbb{N}$, under the hypothesis that

$$\lim_{n\to\infty} \frac{U_{n+1}}{U_n} = \theta$$

for some real number $\theta > 1$. His description is strongly related to the θ-*expansion* $e_\theta(1)$ of 1 defined as follows [37]. Compute by the greedy algorithm

$$1 = \sum_{i=1}^{\infty} \frac{d_i}{\theta^i} \quad .$$

Then $e_\theta(1)$ is equal to $d_1 d_2 \cdots d_i \cdots$.

For instance, for the Fibonacci base, $\lim \frac{U_{n+1}}{U_n} = \varphi = \frac{1+\sqrt{5}}{2}$ and $1 = \frac{1}{\varphi} + \frac{1}{\varphi^2}$. For the decimal numeration system, $e_\theta(1) = 10$. For the numeration system $U_n = 3U_{n-1} - U_{n-2}$, with $U_0 = 1$, $U_1 = 3$, we have $\lim \frac{U_{n+1}}{U_n} = \varphi^2$ and $e_{\varphi^2}(1) = 21^\omega$. Observe that in these three examples, $\mathbb{N}$ is U-recognizable and the θ-expansion of 1 is either finite or ultimately periodic. This is always the case.

Proposition 4 [29] *Let U be a numeration system such that* $\lim \frac{U_{n+1}}{U_n} = \theta > 1$. *If $\mathbb{N}$ is U-recognizable, then $e_\theta(1)$ is finite or ultimately periodic (*in this case, θ is called a β-number [37]*).*

Now, assume that the θ-expansion of 1 is ultimately periodic

$$e_\theta(1) = d_1 \cdots d_l (d_{l+1} \cdots d_{l+p})^\omega \quad .$$

One verifies that $e_\theta(1)$ is root of the polynomial

$$P_{l,p}(x) = x^{l+p} - \sum_{i=1}^{l+p} d_i x^{l+p-i} - x^l + \sum_{i=1}^{l} d_i x^{l-i} \quad .$$

If l,p are minimal, then $P_{l,p}(x)$ is called *θ-polynomial*, otherwise *extended θ-polynomial*.

For example, for the previous system $U_n = 3U_{n-1} - U_{n-2}$, with 21^ω as θ-expansion of 1, we get $P_{1,1}(x) = x^2 - 3x + 1$ as θ-polynomial and $P_{2,2}(x) = x(x+1)(x^2-3x+1)$ as an extended θ-polynomial. One can notice that $P_{1,1}(x)$ is exactly the polynomial of the recurrence $U_n = 3U_{n-1} - U_{n-2}$ defining U.

We can now state the main theorem [29].

Theorem 3 *Let U be a numeration system such that $\lim \frac{U_{n+1}}{U_n} = \theta > 1$. Assume that $e_\theta(1)$ is ultimately periodic. Then $\mathbb{N}$ is U-recognizable if and only if U satisfies a recurrence relation whose polynomial is a (extended) θ-polynomial.*

A little more complex characterization holds, when $e_\theta(1)$ is finite, instead of being ultimately periodic.

3.4 Bertrand Numeration Systems

In [2], particular numeration systems are introduced, which can be seen as the most natural numeration systems in the following sense. A system U is called *Bertrand numeration system* if

$$\begin{array}{ll} & w \text{ is a normalized representation} \\ \Leftrightarrow & w0^n \text{ is a normalized representation.} \end{array}$$

For example, the Fibonacci base is a Bertrand numeration system. However, the system U with the same reccurence $U_n = U_{n-1} + U_{n-2}$, but with different initial conditions $U_0 = 1$ and $U_1 = 4$, is not a Bertrand numeration system.

The next property characterizes Bertrand numeration systems.

Theorem 4 *A system U is a Bertrand numeration system if and only if*

$$\begin{array}{l} U_0 = 1 \\ U_n = d_1 U_{n-1} + d_2 U_{n-2} + \ldots + d_n U_0 + 1 \end{array}$$

where $d_1 d_2 \cdots d_i \cdots$ is the θ-expansion $e_\theta(1)$ of 1, for some $\theta > 1$.

Moreover, with respect to Theorem 3, one can prove that for a Bertrand numeration system U, $\mathbb{N}$ is U-recognizable if and only if the polynomial of the recurrence defining U is *the θ-polynomial* of θ and the initial conditions are those described in the previous theorem.

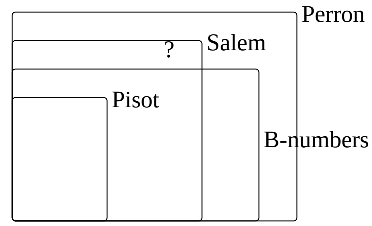

Figure 6: β-numbers inside the family of Perron numbers.

3.5 β-Numbers

Theorem 3 suggests the following question: "which numbers $\theta > 1$ are β-numbers?". As we will see, these numbers are related to *Pisot* (resp. *Salem*, *Perron*) numbers defined as algebraic integers $\theta > 1$ whose conjugates have modulus < 1 (resp. ≤ 1, $< \theta$). For instance, $\frac{1+\sqrt{5}}{2}$ is a Pisot number, $\frac{5+\sqrt{5}}{2}$ is a Perron number which is not a Salem number. The following strict inclusions hold

$$\mathbb{Z} \subset \text{ Pisot } \subset \text{ Salem } \subset \text{ Perron } .$$

The relation between β-numbers and these numbers is described in Figure 6 summarizing the next results.

Theorem 5 *Let θ be a β-number. Then the conjugates of θ have modulus respectively less than* 2 [37], $\frac{1+\sqrt{5}}{2}$ [45, 21], θ [30].

In [5], Blanchard notes that if θ is a Perron number with some real conjugate > 1, then θ is not a β-number.

Theorem 6 *Let θ be respectively a Pisot number* [1, 40]*, a Salem number with degree* 4 [6]*, then θ is a β-number.*

Boyd [7] has made several experimentations on Salem (not Pisot) numbers which suggest that, apart Salem numbers with degree 4, the only Salem numbers being β-numbers should have degree 6. In [45], Solomyak has given the exact description of a compact subset of the plane which is the closure of the set of all conjugates of all β-numbers.

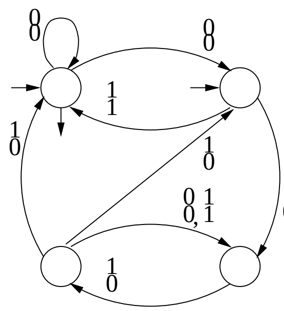

Figure 7: Normalizer for the Fibonacci base.

3.6 U-Recognizability of the Addition

Let us now turn to the U-recognizability of the addition. The addition in standard bases $p \geq 2$ is well-known. However the addition in the Fibonacci base, can be surprizing, since left carry *and* right carry happen!

…	13	8	5	3	2	1
	0	0	1	0	1	0
	0	1	0	0	1	0
	1	0	0	1	0	1

A way to perform the addition of two numbers is to add them digit by digit without carry, and then to normalize the result [23].

…	13	8	5	3	2	1	
	0	0	1	0	1	0	
	0	1	0	0	1	0	
	0	1	1	0	2	0	addition without carry
	1	0	0	1	0	1	normalization

Given a numeration system U, its canonical alphabet A_U and some alphabet $B \subset \mathbb{Z}$, the normalization $\eta_{U,B}$ is a map replacing any $w \in B^*$ by the corresponding normalized representation (if it exists). Hence, if $\eta_{U,B}$ is computable by a finite automaton, with $B = 2A_U$, then the addition is U-recognizable. Figure 7 shows a finite automaton for the normalization $\eta_{U,B}$ with U the Fibonacci base and $B = \{0, 1\}$.

About the recognizability by a finite automaton of the normalization, and therefore of the addition, the known results only concern β-numbers $\theta > 1$ whose minimal polynomial is the polynomial of the recurrence satisfied by the numeration system U. The case of non minimal polynomials is open.

Theorem 7 *[25] Let U be a numeration system given by a linear recurrence whose polynomial is the minimal polynomial of a β-number.*
1. If θ is a Pisot number, then $\eta_{U,B}$ is computable by a finite automaton, for any finite alphabet $B \subseteq \mathbb{Z}$,
2. if θ is a Perron number which is not Pisot, then $\eta_{U,B}$ is not computable by a finite automaton, for any finite alphabet $B \supseteq \{0, 1, \ldots, c+1\}$, if $A_U = \{0, 1, \ldots, c\}$.

Another proof of the first part of Theorem 7 can be found in [9].

3.7 U-Definable Sets and Cobham's Theorem

As for standard bases $p \geq 2$, we define the logical structure $\langle \mathbb{N}, +, V_U \rangle$ where $V_U(x) = y$ means that y is the smallest U_n appearing in the normalized representation of x with a non null digit. We say that $X \subseteq \mathbb{N}$ is U-definable if there exists a first-order formula of $\langle \mathbb{N}, +, V_U \rangle$ defining X. The following logical characterization uses the results of Subsection 3.3 and 3.6.

Theorem 8 [9] *Let U be a numeration system defined by a recurrence whose polynomial is the minimal polynomial of a Pisot number, then $X \subseteq \mathbb{N}$ is U-recognizable if and only if X is U-definable.*

Remember Cobham's theorem (Theorem 1) which states that the only sets recognizable in all standard bases $p \geq 2$ are the ultimately periodic ones. As ultimately periodic sets are all definable in $\langle \mathbb{N}, + \rangle$, they are also definable in $\langle \mathbb{N}, +, V_U \rangle$, showing that they are U-recognizable for any system U of Theorem 8. This result was first proved in [23] by other techniques.

The contrary also holds under some hypotheses. This gives a Cobham's theorem for non standard bases. In [39], if $X \subseteq \mathbb{N}$ is p-definable and U-definable with $p \geq 2$ and U a nonstandard numeration system as in Theorem 8, then it is proved that X is ultimately periodic. The result is still true in N^m. Another proof with θ a unitary Pisot number and in dimension 1 can already be found in [19]. Very recently, Cobham's theorem has been extended to two nonstandard numeration systems U and U' in any dimension

$m \geq 1$ [4, 27]. Fabre's proof [19] uses U-substitutions (see Subsection 3.8), the proof in [39] generalizes techniques developed in [36], while Hansel's and Bès' proofs [27, 4] rely on two deep logical theorems established in [34, 35]. It seems that this last approach is the most powerful one.

3.8 U-Substitutive Sets

Let us begin with an example. Figure 4 is the minimal automaton for $\mathbb{N}$ in the Fibonacci base. This automaton can be "simulated" by a substitution as follows. Let

$$\begin{array}{rrcl} f : & a & \to & ab \\ & b & \to & a \end{array}$$

This substitution describes the transition function on both states a and b. The iteration of f on the initial state a works as follows.

$$\begin{array}{ccccccccccccccc} & & 0 & 1 & & 00 & 01 & 10 & & 000 & 001 & 010 & 100 & 101 & \\ a & \to & a & b & \to & a & b & a & \to & a & b & a & a & b & \to \dots \end{array}$$

The constructed infinite word $abaababaabaababa \cdots$ is such that the state at position n equals a (resp. b) if and only if the normalized representation of n in Fibonacci base goes from the initial state to state a (resp. b).

More generally, we have the next property.

Proposition 5 *Let U be a numeration system such that $\mathbb{N}$ is U-recognizable. Then there exists a canonical U-substitution f_U which simulates the minimal automaton accepting $\mathbb{N}$.*

This notion of canonical substitution is introduced in [18] for Bertrand numeration systems. It has been generalized to any nonstandard base in [9]. For classical bases $p \geq 2$, the canonical substitution is $f_p : a \to a^p$.

We now proceed with another example in the Fibonacci base U. The automaton of Figure 5 is simulated by the *U-substitution*

$$\begin{array}{rrcl} f : & a & \to & ab \\ & b & \to & c \\ & c & \to & cd \\ & d & \to & a \end{array} \qquad \begin{array}{rrcl} g : & a & \to & 1 \\ & b & \to & 0 \\ & c & \to & 0 \\ & d & \to & 1 \end{array}$$

such that f describes the transitions and g the final states of the automaton.

Notice that the automaton of Figure 5 is a "splitting" of the automaton of Figure 4: merging states a, c, respectively b, d of the second automaton leads to the first automaton. In the same way, function f above is a "splitting" of function f_U defined at the beginning of this subsection.

Based on this example, given a numeration system U such that $\mathbb{N}$ is U-recognizable, we call *U-substitution* any map f which is a splitting of the canonical U-substitution f_U associated with U.

For standard bases $p \geq 2$, as the canonical substitution is the map $f_p : a \rightarrow a^p$, all the splittings of f_p are maps whose images are words with length p. These are exactly the p-substitutions introduced in Subsection 2.2.

The next theorem generalizes the characterization of p-recognizable sets by p-substitutive sets.

Theorem 9 *Let U be a numeration system such that $\mathbb{N}$ is U-recognizable. A set $X \subseteq \mathbb{N}$ is U-recognizable if and only if there exist a U-substitution $f : B \rightarrow B^*$, a letter $a \in B$ and a map $g : B \rightarrow \{0, 1\}$ such that*

$$\underline{X} = g(f^\omega(a)).$$

This equivalence is proved in [18] for Bertrand numeration systems. The general case is solved in [9]; the statement of Theorem 9 has been simplified to be more comprehensible. Another notion of substitution can be found in [43].

References

[1] A. Bertrand, Développements en base de Pisot et répartition modulo 1, *C. R. Acad. Sci. Paris* **285** (1977) 419–421.

[2] A. Bertrand-Mathis, Comment écrire les nombres entiers dans une base qui n'est pas entière, *Acta Math. Acad. Sci. Hungar.* **54** (1989) 237–241.

[3] A. Bès, Undecidable extensions of Büchi arithmetic and Cobham-Semenov theorem, *Technical report* **44**, LLAIC1, University of Auvergne, France (1996) 17 pages.

[4] A. Bès, An extension of the Cobham-Semenov theorem, *preprint* (1996) 10 pages.

[5] F. Blanchard, β-expansions and symbolic dynamics, *Theoret. Comput. Sci.* **65** (1989) 131–141.

[6] D. Boyd, Salem numbers of degree 4 have periodic expansions, *Théorie des Nombres – Number Theory*, J.M. de Koninck and C. Levesque, Eds., Walter de Gruyter & Co., Berlin and New York (1989) 57–64.

[7] D. Boyd, On the beta-expansion for Salem numbers of degree 6, to appear in *Mathematics of Computation* (1995) 16 pages.

[8] V. Bruyère, Entiers et automates finis, *Mémoire de fin d'études*, University of Mons, Belgium (1985).

[9] V. Bruyère, G. Hansel, Bertrand numeration systems and recognizability, submitted (1995) 22 pages.

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

[11] J.R. Büchi, Weak second-order arithmetic and finite automata, *Z. Math. Logik Grundlag. Math.* **6** (1960) 66–92.

[12] G. Cherlin, F. Point, On extensions of Presburger arithmetic, Proc. 4th Easter Model Theory conference, Gross Köris 1986 Seminarberichte 86, Humboldt Universität zu Berlin (1986) 17–34.

[13] G. Christol, T. Kamae, M. Mendès France, G. Rauzy, Suites algébriques, automates et substitutions, *Bull. Soc. Math. France* **108** (1980) 401–419.

[14] A. Cobham, On the base-dependence of sets of numbers recognizable by finite automata, *Math. System Theory* **3** (1969) 186–192.

[15] A. Cobham, Uniform tag systems, *Math. Systems Theory* **6** (1972) 164–192.

[16] F. Delon, Personal communication (1991).

[17] S. Eilenberg, *Automata, Languages and Machines*, Vol. A, Academic Press, New-York (1974).

[18] S. Fabre, Substitutions et indépendence des systèmes de numération, *Thèse*, University Aix-Marseille II (1992).

[19] S. Fabre, Une généralisation du théorème de Cobham, *Acta Arithmetica* **67** (1994) 197–208.

[20] I. Fagnot, Sur les facteurs des mots automatiques, to appear in *Theoret. Comput. Sci.* **172** (1995).

[21] L. Flatto, J.C. Lagarias, B. Poonen, The zeta function of the beta transformation, *Ergod. Theory Dynam. Sys.* **14** (1994) 237–266.

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

[23] C. Frougny, Systèmes de numération linéaires et automates finis, *Thèse d'Etat*, University Paris 7, France (1989).

[24] C. Frougny, Representations of numbers and finite automata, *Math. Systems Theory* **25** (1992) 37–60.

[25] C. Frougny, B. Solomyak, On representations of integers in linear numeration systems, to appear in *Ergod. Th. & Dynam. Sys.*, M. Pollicott and K. Schmidt, Eds., London Mathematical Society Lecture Note Series, Cambridge University Press, and *LITP report* **94/54**, University Paris 7 (1995) 18 pages.

[26] G. Hansel, A propos d'un théorème de Cobham, in: *Actes de la fête des mots*, D. Perrin, Ed., Greco de Programmation, CNRS, Rouen (1982).

[27] G. Hansel, personal communication (1996).

[28] B.R. Hodgson, Décidabilité par automate fini, *Ann. Sci. Math. Québec* **7** (1983) 39–57.

[29] M. Hollander, Greedy numeration systems and regularity, *preprint* (1995) 21 pages.

[30] D. Lind, The entropies of topological Markov shifts and a related class of algebraic integers *Ergod. Th. & Dynam. Sys.* **4** (1984) 283–300.

[31] N. Loraud, Beta-shift, systèmes de numération et automates, *preprint* (1995).

[32] R. MacNaughton, Review of [11], *J. Symbolic Logic* **28** (1963) 100–102.

[33] C. Michaux, F. Point, Les ensembles k-reconnaissables sont définissables dans $\langle \mathbb{N}, +, V_k \rangle$, *C. R. Acad. Sci. Paris* **303** (1986) 939–942.

[34] C. Michaux, R. Villemaire, Cobham's theorem seen through Büchi theorem, Proc. Icalp'93, *Lecture Notes inn Comput. Sci.* **700** (1993) 325–334.

[35] C. Michaux, R. Villemaire, Presburger arithmetic and recognizability of natural numbers by automata: new proofs of Cobham's and Semenov's theorems, *Annals of Pure and Applied Logic* **77** (1996) 251–277.

[36] A. Muchnik, Definable criterion for definability in Presburger Arithmetic and its application, *preprint in Russian*, Institute of New Technologies (1991).

[37] W. Parry, On the β-expansions of real numbers, *Acta Math. Acad. Sci. Hungar.* **11** (1960) 401–416.

[38] D. Perrin, Finite automata, in: *Handbook of Theoretical Computer Science*, vol. B, J. Van Leeuwen, Ed., Elsevier (1990) 2–57.

[39] F. Point, V. Bruyère, On the Cobham-Semenov theorem, to appear in *Math. System Theory* (1996) 25 pages.

[40] K. Schmidt, On periodic expansions of Pisot numbers and Salem numbers, *Bull. London. Math. Soc.* **12** (1980) 269–278.

[41] A. L. Semenov, Presburgerness of predicates regular in two number systems (in Russian), *Sibirsk. Mat. Zh.* **18** (1977) 403–418, English translation, *Siberian Math. J.* **18** (1977) 289-299.

[42] A. L. Semenov, On certain extensions of the arithmetic of addition of natural numbers (in Russian), *Izv. Akad. Nauk. SSSR ser. Mat.* **47** (1983) 623–658, English translation, *Math. USSR-Izv.* **22** (1984) 587–618.

[43] J. Shallit, A generalization of automatic sequences, *Theoret. Comput. Sci.* **61** (1988) 1–16.

[44] J. Shallit, Numeration systems, linear recurrences and regular sets, *Lect. Notes Comput. Sci.* **623** (1992) 89–100.

[45] B. Solomyak, Conjugates of beta-numbers and the zero-free for a class of analytic functions, *Proc. London Math. Soc.* **68** (1994) 477–498.

[46] R. Villemaire, The theory of $\langle \mathbb{N}, +, V_k, V_l \rangle$ is undecidable, *Theoret. Comput. Sci.* **106** (1992) 337–349.