

An Introduction to Abstract Numeration Systems

Émilie Charlier

Université libre de Bruxelles

Séminaire de vérification
Bruxelles, 16 mai 2012

Positional numeration systems

A **positional numeration system (PNS)** is given by a sequence of integers $U = (U_i)_{i \geq 0}$ such that

- ▶ $U_0 = 1$
- ▶ $\forall i \ U_i < U_{i+1}$
- ▶ $(U_{i+1}/U_i)_{i \geq 0}$ is bounded $\rightarrow C_U = \sup_{i \geq 0} [U_{i+1}/U_i]$

The **greedy U -representation** of a positive integer n is the unique word $\text{rep}_U(n) = c_{\ell-1} \cdots c_0$ over $\Sigma_U = \{0, \dots, C_U - 1\}$ satisfying

$$n = \sum_{i=0}^{\ell-1} c_i U_i, \quad c_{\ell-1} \neq 0 \quad \text{and} \quad \forall t \quad \sum_{i=0}^t c_i U_i < U_{t+1}.$$

Recognizable sets of integers

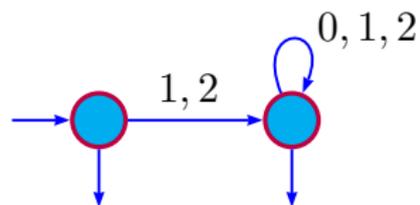
A set $X \subseteq \mathbb{N}$ is **U -recognizable** or **U -automatic** if the subset $\text{rep}_U(X) = \{\text{rep}_U(x) : x \in X\}$ of Σ_U^* is regular.

Integer base $b \geq 2$

$$U = (b^i)_{i \geq 0}$$

$$\Sigma_b = \{0, \dots, b-1\}$$

$$\mathcal{L}_b = \text{rep}_b(\mathbb{N}) = \Sigma_b^* \setminus 0 \Sigma_b^*$$



$\text{rep}_3(\mathbb{N})$

\mathbb{N} is 3-recognizable

$$\text{rep}_U, \Sigma_U \rightarrow \text{rep}_b, \Sigma_b$$

27	9	3	1	
			ε	0
			1	1
			2	2
		1	0	3
		1	1	4
		1	2	5
		2	0	6
		2	1	7
		2	2	8
	1	0	0	9

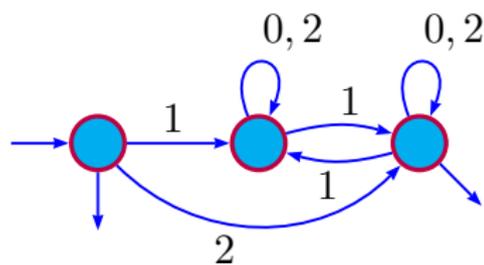
Integer base $b \geq 2$

$$U = (b^i)_{i \geq 0}$$

$$\Sigma_b = \{0, \dots, b-1\}$$

$$\mathcal{L}_b = \text{rep}_b(\mathbb{N}) = \Sigma_b^* \setminus 0 \Sigma_b^*$$

$$\text{rep}_U, \Sigma_U \rightarrow \text{rep}_b, \Sigma_b$$



$$\text{rep}_{U_3}(2\mathbb{N})$$

$2\mathbb{N}$ is 3-recognizable

27	9	3	1	
			ε	0
			1	1
			2	2
		1	0	3
		1	1	4
		1	2	5
		2	0	6
		2	1	7
		2	2	8
1	0	0		9

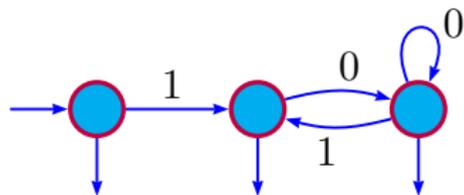
Fibonacci (or Zeckendorf) numeration system

Let $F = (F_i)_{i \geq 0} = (1, 2, 3, 5, 8, 13, 21, \dots)$ be defined by

$$F_0 = 1, F_1 = 2 \text{ and } \forall i \in \mathbb{N}, F_{i+2} = F_{i+1} + F_i.$$

$$\Sigma_F = \{0, 1\}$$

The factor **11** is forbidden :



$$\text{rep}_F(\mathbb{N}) = 1\{0, 01\}^* \cup \{\varepsilon\}$$

\mathbb{N} is F -recognizable

13	8	5	3	2	1		
					ε		0
					1		1
				1	0		2
			1	0	0		3
			1	0	1		4
	1	0	0	0	0		5
	1	0	0	1	0		6
	1	0	1	0	0		7
1	0	0	0	0	0		8

U -recognizability of arithmetic progressions

Proposition

Let $U = (U_i)_{i \geq 0}$ be a linear numeration system and let $p, q \in \mathbb{N}$.

If \mathbb{N} is U -recognizable, then $p + \mathbb{N}q$ is U -recognizable and a DFA accepting $\text{rep}_U(p + \mathbb{N}q)$ can be obtained efficiently.

Consequently, any ultimately periodic set is U -recognizable.

U -recognizability of \mathbb{N}

Is the set \mathbb{N} U -recognizable? Otherwise stated, is the **numeration language** $\text{rep}_U(\mathbb{N})$ regular? Not necessarily:

Theorem (Shallit 1994)

*Let U be a PNS. If \mathbb{N} is U -recognizable, then U is **linear**, i.e., it satisfies a linear recurrence relation over \mathbb{Z} .*

Loraud (1995) and Hollander (1998) gave sufficient conditions for the numeration language to be regular : “The characteristic polynomial of the recurrence relation has a particular form”.

Abstract numeration systems

An **abstract numeration system** (ANS) is a triple $S = (L, \Sigma, <)$ where L is an infinite regular language over a totally ordered alphabet $(\Sigma, <)$.

By enumerating the words of L w.r.t. the radix order $<_{rad}$ induced by $<$, we define a bijection :

$$\text{rep}_S : \mathbb{N} \rightarrow L \quad \text{val}_S = \text{rep}_S^{-1} : L \rightarrow \mathbb{N}.$$

A set $X \subseteq \mathbb{N}$ is **S -recognizable** if $\text{rep}_S(X)$ is regular.

$$L = \{a, b\}^* \quad \Sigma = \{a, b\} \quad a < b$$

n	0	1	2	3	4	5	6	7	\dots
$\text{rep}(n)$	ε	a	b	aa	ab	ba	bb	aaa	\dots

$$L = a^*b^* \quad \Sigma = \{a, b\} \quad a < b$$

n	0	1	2	3	4	5	6	\dots
$\text{rep}(n)$	ε	a	b	aa	ab	bb	aaa	\dots

A generalization

ANS generalize PNS having a regular numeration language:

Let U be a PNS and let $x, y \in \mathbb{N}$. We have

$$x < y \Leftrightarrow \text{rep}_U(x) <_{\text{rad}} \text{rep}_U(y).$$

Example (Fibonacci)

$$\text{rep}_F(\mathbb{N}) = 1\{0, 01\}^* \cup \{\varepsilon\}$$

$6 < 7$ and $1001 <_{\text{rad}} 1010$
(same length)

$6 < 8$ and $1001 <_{\text{rad}} 10000$
(different lengths)

13	8	5	3	2	1	
					ε	0
					1	1
				1	0	2
			1	0	0	3
			1	0	1	4
		1	0	0	0	5
		1	0	0	1	6
		1	0	1	0	7
1	0	0	0	0	0	8

A generalization

ANS generalize PNS having a regular numeration language:

Let U be a PNS and let $x, y \in \mathbb{N}$. We have

$$x < y \Leftrightarrow \text{rep}_U(x) <_{\text{rad}} \text{rep}_U(y).$$

Example (Fibonacci)

$$\text{rep}_F(\mathbb{N}) = 1\{0, 01\}^* \cup \{\varepsilon\}$$

$6 < 7$ and $1001 <_{\text{rad}} 1010$
(same length)

$6 < 8$ and $1001 <_{\text{rad}} 10000$
(different lengths)

13	8	5	3	2	1	
					ε	0
					1	1
				1	0	2
			1	0	0	3
			1	0	1	4
		1	0	0	0	5
		1	0	0	1	6
		1	0	1	0	7
	1	0	0	0	0	8

A generalization

ANS generalize PNS having a regular numeration language:

Let U be a PNS and let $x, y \in \mathbb{N}$. We have

$$x < y \Leftrightarrow \text{rep}_U(x) <_{rad} \text{rep}_U(y).$$

Example (Fibonacci)

$$\text{rep}_F(\mathbb{N}) = 1\{0, 01\}^* \cup \{\varepsilon\}$$

$\text{rep}_F(n)$	n
ε	0
1	1
10	2
100	3
101	4
1000	5
1001	6
1010	7
10000	8

Decimation of languages

Let L be a language ordered w.r.t. the radix order.

If $w_0 < w_1 < \dots$ are the elements of L and $X \subseteq \mathbb{N}$, then

$$L[X] = \{w_n : n \in X\}.$$

If $S = (L, \Sigma, <)$, then $L[X] = \text{rep}_S(X)$.

If $L[X]$ is accepted by a finite automaton, what does it imply on X ? What conditions on X insures that $L[X]$ is regular?

Motivation for ANS

ANS are a generalization of all usual PNS like integer base numeration systems or linear numeration systems, and even rational numeration systems.

Thanks to this general point of view on numeration systems, we try to distinguish results that deeply depend on the algorithm used to represent the integers from results that only depend on the set of representations.

Due to the general setting of ANS, some new questions concerning languages arise naturally from this numeration point of view.

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

Some questions around ANS

- ▶ Rec. sets in a given ANS?
- ▶ Rec. sets in all ANS?
- ▶ Are there subsets of \mathbb{N} that are never recognizable?
- ▶ Given a subset of \mathbb{N} can we build an ANS for which it is rec.?
- ▶ How do rec. depend on the choice of the numeration?
- ▶ For which ANS do arithmetic operations preserve rec.?
- ▶ Operations preserving rec. in a given ANS?
- ▶ How to represent real numbers?
- ▶ Can we define automatic sequences in that context?
- ▶ Logical characterization of rec. sets?
- ▶ Extensions to the multidimensional setting?
- ▶ ...

S-automatic words

b -automatic words

An infinite word $x = (x_n)_{n \geq 0}$ is **b -automatic** if there exists a DFAO $\mathcal{A} = (Q, q_0, \Sigma_b, \delta, \Gamma, \tau)$ s.t. for all $n \geq 0$,

$$x_n = \tau(\delta(q_0, \text{rep}_b(n))).$$

Theorem (Cobham 1972)

*Let $b \geq 2$. An infinite word is **b -automatic** iff it is the image under a coding of an infinite fixed point of a **b -uniform** morphism.*

S-automatic words

Let $S = (L, \Sigma, <)$ be an ANS.

An infinite word $x = (x_n)_{n \geq 0}$ is **S-automatic** if there exists a DFAO $\mathcal{A} = (Q, q_0, \Sigma, \delta, \Gamma, \tau)$ s.t. for all $n \geq 0$,

$$x_n = \tau(\delta(q_0, \text{rep}_S(n))).$$

Theorem (Rigo-Maes 2002)

*An infinite word is **S-automatic** for some ANS S iff it is the image under a coding of an infinite fixed point of a morphism, i.e. a **morphic** word.*

Corollary

The set of primes is never S -recognizable.

Its characteristic sequence is not morphic (Maudivit 1988).

Corollary

The factor complexity of an S -automatic word is $O(n^2)$.

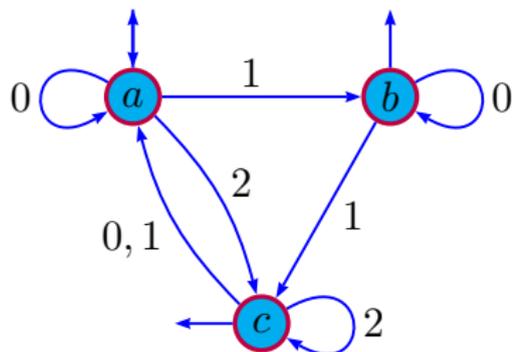
Idea of the proof

Example (Morphic \rightarrow S -Automatic)

Consider the morphism μ defined by $a \mapsto abc$; $b \mapsto bc$; $c \mapsto aac$.

We have $\mu^\omega(a) = abc**bc**caac**bc**caacabcabcaac**bc**caacabcabc \cdots$.

One canonically associates the DFA $\mathcal{A}_{\mu,a}$



$$L_{\mu,a} = \{\varepsilon, 1, 2, 10, 11, 20, 21, 22, 100, 101, 110, 111, 112, 200, \dots\}$$

If $S = (L_{\mu,a}, \{0, 1, 2\}, 0 < 1 < 2)$, then

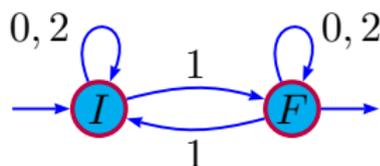
$$(\mu^\omega(a))_n = \delta_\mu(a, \text{rep}_S(n)) \text{ for all } n \geq 0.$$

Idea of the proof

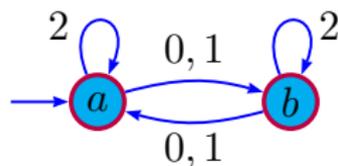
Example (S -Automatic \rightarrow Morphic)

$S = (L, \{0, 1, 2\}, 0 < 1 < 2)$ where $L = \{w \in \Sigma^* : |w|_1 \text{ is odd}\}$

minimal automaton of L

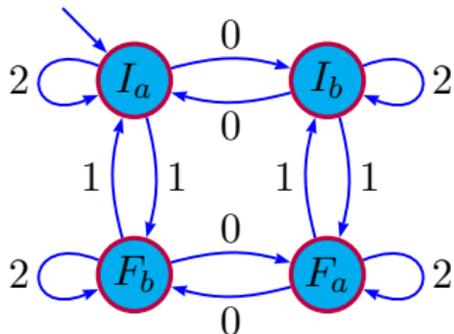


DFAO generating x



n	0	1	2	3	4	5	6	7	8	...
$\text{rep}_S(n)$	1	01	10	12	21	001	010	012	021	...
x	b	a	a	b	b	b	b	a	a	...

Example (Continued)



$$f: \alpha \mapsto \alpha I_b F_b I_a$$

$$I_a \mapsto I_b F_b I_a$$

$$I_b \mapsto I_a F_a I_b$$

$$F_a \mapsto F_b I_b F_a$$

$$F_b \mapsto F_a I_a F_b$$

$$g: \alpha, I_a, I_b \mapsto \varepsilon$$

$$F_a \mapsto a$$

$$F_b \mapsto b$$

$L \subseteq \Sigma^*$	ε	0	1	2	00	01	02	10	11	12	20
$f^\omega(\alpha)$	α	I_b	F_b	I_a	I_a	F_a	I_b	F_a	I_a	F_b	I_b
x			b			a		a		b	

$$g(f^\omega(\alpha)) = x$$

Multidimensional Case

A **d -dimensional infinite word** over an alphabet Σ is a map $x : \mathbb{N}^d \rightarrow \Sigma$. We use notation like x_{n_1, \dots, n_d} or $x(n_1, \dots, n_d)$ to denote the value of x at (n_1, \dots, n_d) .

If w_1, \dots, w_d are finite words over the alphabet Σ ,

$$(w_1, \dots, w_d)^\# := (\#^{m-|w_1|}w_1, \dots, \#^{m-|w_d|}w_d)$$

where $m = \max\{|w_1|, \dots, |w_d|\}$.

Example

$$(ab, bbaa)^\# = (\#\#ab, bbaa) = (\#, b)(\#, b)(a, a)(b, a)$$

A d -dimensional infinite word over an alphabet Γ is **b -automatic** if there exists a DFAO

$$\mathcal{A} = (Q, q_0, (\Sigma_b)^d, \delta, \Gamma, \tau)$$

s.t. for all $n_1, \dots, n_d \geq 0$,

$$\tau \left(\delta \left(q_0, (\text{rep}_b(n_1), \dots, \text{rep}_b(n_d))^0 \right) \right) = x_{n_1, \dots, n_d}.$$

Theorem (Salon 1987)

Let $b \geq 2$ and $d \geq 1$. A d -dimensional infinite word is **b -automatic** iff it is the image under a coding of a fixed point of a **b -uniform** d -dimensional morphism.

Theorem (C-Kärki-Rigo 2010)

Let $d \geq 1$. The d -dimensional infinite word is *S-automatic* for some ANS $S = (L, \Sigma, <)$ where $\varepsilon \in L$ iff it is the image under a coding of a *shape-symmetric* infinite d -dimensional word.

Shape-symmetric

$$\mu(a) = \mu(f) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}; \mu(b) = \begin{bmatrix} e \\ c \end{bmatrix}; \mu(c) = \begin{bmatrix} e & b \end{bmatrix}; \mu(d) = \begin{bmatrix} f \end{bmatrix}$$

$$\mu(e) = \begin{bmatrix} e & b \\ g & d \end{bmatrix}; \mu(g) = \begin{bmatrix} h & b \end{bmatrix}; \mu(h) = \begin{bmatrix} h & b \\ c & d \end{bmatrix}.$$

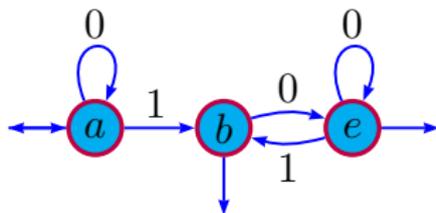
$$\mu^\omega(a) = \begin{array}{|c|c|c|c|c|c|c|} \hline a & b & e & e & b & e & b & e & \dots \\ \hline c & d & c & g & d & g & d & c & \\ \hline e & b & f & e & b & h & b & f & \\ \hline e & b & e & a & b & e & b & e & \\ \hline g & d & c & c & d & g & d & c & \\ \hline e & b & e & e & b & a & b & e & \\ \hline g & d & c & g & d & c & d & c & \\ \hline h & b & f & e & b & e & b & f & \\ \hline \vdots & & & & & & & \ddots & \\ \hline \end{array}$$

Consider the morphism μ_1 defined by

$$a \mapsto ab ; b \mapsto e ; e \mapsto eb.$$

We have $\mu_1^\omega(a) = abeebebebebebebebebebebeeb \dots$.

One canonically associates the DFA $\mathcal{A}_{\mu_1, a}$



$$L_{\mu_1, a} = \{\epsilon, 1, 10, 100, 101, 1000, 1001, 1010, 10000, \dots\}$$

Open question

- ▶ If S and T are two ANS, (S, T) -automatic words are bidimensional infinite words $(x_{m,n})_{m,n \geq 0}$ for which there exists a DFAO $\mathcal{A} = (Q, (\Sigma \cup \{\#\})^d, \delta, q_0, \Gamma, \tau)$ s.t. $\forall m, n \in \mathbb{N}$,

$$x_{m,n} = \tau(\delta(q_0, (\text{rep}_S(m), \text{rep}_T(n))^\#)).$$

Can these (S, T) -automatic words be characterized by iterating morphisms?

b -kernel

An infinite word $(x_n)_{n \geq 0}$ is b -automatic iff its b -kernel

$$\{(x_{b^e n + r})_{n \geq 0} : e, r \in \mathbb{N}, r < b^e\}$$

is finite. The b -kernel can be rewritten

$$\{(x_{b^{|w|}n + \text{val}_b(w)})_{n \geq 0} : w \in \Sigma_b^*\}.$$

8	4	2	1		8	4	2	1		8	4	2	1		
			ε	0		1	1	0	6		1	1	0	0	12
			1	1		1	1	1	7		1	1	0	1	13
		1	0	2		1	0	0	8		1	1	1	0	14
		1	1	3		1	0	0	9		1	1	1	1	15
	1	0	0	4		1	0	1	10		1	0	0	0	16
	1	0	1	5		1	0	1	11		1	0	0	0	17

NB: $b^{|w|}n + \text{val}_b(w)$ is the base- b value of the $(n + 1)$ -th word in \mathcal{L}_b having w as a suffix.

Open question

The S -kernel of $(x_n)_{n \geq 0}$ is

$$\{(x_{f_w(n)})_{n \geq 0} : w \in \Sigma^*\}$$

where $f_w(n)$ is the S -value of the $(n+1)$ -th word in L having w as a suffix.

Theorem (Rigo-Maes 2002)

An infinite word is S -automatic iff its S -kernel is finite.

- ▶ Does a similar characterization hold in the multidimensional setting?