AUTOUR DES SYSTÈMES DE NUMÉRATION ABSTRAITS

Michel Rigo

http://www.discmath.ulg.ac.be/
http://orbi.ulg.ac.be/handle/2268/124195

In the Chomsky’s hierarchy, the simplest models of computation are finite automata accepting regular languages.

100100, 1000, 1000100, 0000001, ...
With this model in mind, \textit{what is a “simple” set of integers} ?

\textbf{Definition}

A set \( X \subset \mathbb{N} \) is \textit{k-recognizable}, if the set of base \( k \) expansions of the elements in \( X \) is accepted by some finite automaton, \textit{i.e.}, \( \text{rep}_k(X) \) is a regular language.

Much “simpler” than a \textit{recursive set} of integers for which there is an algorithm that decides whether or not a given number belongs to the set.
Some Examples

A 2-recognizable set

\[ X = \{ n \in \mathbb{N} \mid \exists i, j \geq 0 : n = 2^i + 2^j \} \cup \{1\} \]

\[ X = \{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 16, 17, 18, 20, 24, \ldots\} \]

\[ \text{rep}_2(X) = \{1, 10, 11, 100, 101, 110, 1000, 1001, 1010, 1100, \ldots\} \]
Some examples

- The set of even integers is 2-recognizable.
- The Prouhet–Thue–Morse set is 2-recognizable,

\[ X = \{ n \in \mathbb{N} \mid s_2(n) \equiv 0 \mod 2 \} \]
\[ X = \{0, 3, 5, 6, 9, 10, 12, 15, 17, 18, \ldots \} \]
\[ \text{rep}_2(X) = \{\varepsilon, 11, 101, 110, 1001, 1010, 1100, 1111, 10001, 10010, \ldots \} \]
- The set of powers of 2 is 2-recognizable.
Let $X = \{x_0 < x_1 < x_2 < \cdots \} \subseteq \mathbb{N}$. Define

\[
R_X := \limsup_{i \to \infty} \frac{x_{i+1}}{x_i} \quad \text{and} \quad D_X := \limsup_{i \to \infty} (x_{i+1} - x_i).
\]

**Gap Theorem (Cobham’72)**

Let $k \geq 2$. If $X \subseteq \mathbb{N}$ is a $k$-recognizable infinite subset of $\mathbb{N}$, then either $R_X > 1$ or $D_X < +\infty$.


**Corollary**

Let $k, t \geq 2$ be integers.

The set $\{n^t \mid n \geq 0\}$ is NOT $k$-recognizable.

S. Eilenberg, Automata, Languages, and Machines, 1974.
More Examples

Minsky–Papert 1966

The set $\mathcal{P}$ of prime numbers is not $k$-recognizable.

A proof using the gap theorem:
Since $n! + 2, \ldots, n! + n$ are composite numbers, $D_\mathcal{P} = +\infty$
Since $p_n \in (n \ln n, n \ln n + n \ln \ln n)$, $R_\mathcal{P} = 1$

E. Bach, J. Shallit, Algorithmic number theory, MIT Press

M.-P. Schützenberger (1968)

No infinite subset of $\mathcal{P}$ can be recognized by a finite automaton.
**Base Sensitivity**

*Is this notion of recognizability base dependent?*

- Is the set of even integers 3-recognizable? *(exercise)*
- Is the set of powers of 2 also 3-recognizable?

\[
2, 11, 22, 121, 1012, 2101, 11202, 100111, 200222, 1101221, 2210212, 12121201, 102020102, 211110211, 1122221122, 10022220021, 20122210112, 111022121001, 222122012002, 1222021101011, \ldots
\]
Two integers \( k, \ell \geq 2 \) are **multiplicatively independent** if \( k^m = \ell^n \Rightarrow m = n = 0 \), i.e., if \( \log k / \log \ell \) is irrational.

**Cobham’s theorem** (1969)

Let \( k, \ell \geq 2 \) be two multiplicatively independent integers. A set \( X \subseteq \mathbb{N} \) is \( k \)-rec. AND \( \ell \)-rec. IFF \( X \) is **ultimately periodic**, i.e., \( X \) is a finite union of arithmetic progressions.

V. Bruyère, G. Hansel, C. Michaux, R. Villemaire, Logic and \( p \)-recognizable sets of integers, *BBMS*’94.


**Tool** (**Kronecker’s theorem**)  

Let \( \theta \) be an irrational number. The sequence \( (\{n\theta\})_{n \geq 0} \) is dense in \([0, 1)\).
S. Eilenberg (p. 104): “The proof is correct, long and hard. It is a challenge to find a more reasonable proof of this fine theorem”

The easy part, e.g., conversion between base 2 and base 4,

\[
\begin{array}{l|l}
00 & 0 \\
01 & 1 \\
10 & 2 \\
11 & 3 \\
\end{array}
\]

- such a transformation preserves regularity
- \( L \) is regular IFF \( 0^* L \) is regular
Some consequences of Cobham’s theorem from 1969:

- $k$-recognizable sets are easy to describe but **non-trivial,**
- motivates **characterizations** of $k$-recognizability,
- motivates the study of “exotic” numeration systems,
- **generalizations** of Cobham’s result to various contexts: multidimensional setting, logical framework, extension to Pisot systems, substitutive systems, fractals and tilings, simpler proofs, . . .

There are three kinds of sets:

- Ultimately periodic sets are recognizable in all bases,
- Sets that are $k$-recognizable for some $k$, and only $k^m$-recognizable, $m \geq 1$,
- Sets that are not $k$-recognizable.

Multiplicative dependence is trivially an equivalence relation.
Logical characterization

**Büchi–Bruyère Theorem**

A set $X \subseteq \mathbb{N}^d$ is $k$-recognizable IFF it is definable by a first order formula in the extended Presburger arithmetic $\langle \mathbb{N}, +, V_k \rangle$.

$V_k(n)$ is the largest power of $k$ dividing $n \geq 1$, $V_k(0) = 1$.

$$\varphi_1(x) \equiv V_2(x) = x$$

$$\varphi_2(x) \equiv (\exists y)(V_2(y) = y) \land (\exists z)(V_2(z) = z) \land x = y + z$$

$$\varphi_3(x) \equiv (\exists y)(x = y + y + y + y + 3)$$

from formula to automata \hspace{1cm} from automata to formula...

**Restatement of Cobham’s thm.**

Let $k, \ell \geq 2$ be two multiplicatively independent integers. A set $X \subseteq \mathbb{N}$ is $k$-rec. AND $\ell$-rec. IFF $X$ is definable in $\langle \mathbb{N}, + \rangle$. 
Applications to decision problems and, in computer science, to model-checking and formal verification.

**Theorem (Bruyère 1985)**

The theory $⟨\mathbb{N}, +, V_k⟩$ is decidable.

**Example**

Let $X$ be a $k$-recognizable set of integers. Decide whether or not $X$ is ultimately periodic?

Let $\varphi(x)$ be a formula such that $a \in X$ IFF $\varphi(a)$ holds true. Consider the sentence

$$(\exists p)(\exists i)(\forall a \geq i)(\varphi(a) \iff \varphi(a + p)).$$
**MORPHIC CHARACTERIZATION**

**Theorem (Cobham 1972)**

An infinite word $x$ is morphic and generated by a $k$-uniform morphism $+ coding$ IFF $x$ is $k$-automatic, i.e., $\forall n \geq 0$, $x_n$ is generated by an automaton reading $\text{rep}_k(n)$.

$$f : A \mapsto AB, \quad B \mapsto BC, \quad C \mapsto CD, \quad D \mapsto DD$$

$$f^\omega(A) = ABBBCBCDBCCDCDDDBBCDCCDCCDDDCCDDDDDDDDDDDDDDD \cdots$$
A set $X \subseteq \mathbb{N}$ is $k$-recognizable IFF its characteristic sequence is $k$-automatic.

Link with combinatorics on words

$$f(0) = 01, \quad f(1) = 10$$

$$f^\omega(0) = 011010011001011001101001 \cdots$$

A. Thue (1912)

The Thue–Morse word is overlap free.
The \textit{k-kernel} of $x = (x_n)_{n \geq 0}$ is defined by

$$N_k(x) = \{(x_{ek^n+d})_{n \geq 0} \mid e \geq 0, \ 0 \leq d < k^e\}$$

\textbf{S. Eilenberg (1974)}

A sequence $x = (x_n)_{n \geq 0}$ is \textit{k-automatic} IFF $N_k(x)$ is finite.

\textbf{Definition (J.-P. Allouche, J. Shallit 1992)}

Let $R$ be a (possibly infinite) commutative ring. Let $x = (x_n)_{n \geq 0} \in R^\mathbb{N}$. If the $R$-module generated by all sequences in $N_k(x)$ is finitely generated then $x$ is said to be \textit{(R, k)-regular}.

Morphic characterization

A sequence of C. Mallows

There is a unique monotone sequence \((a(n))_{n \geq 0}\) of non-negative integers such that \(a(a(n)) = 2n\) for all \(n \neq 1\),

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
</tbody>
</table>

This sequence \((a(n))_{n \geq 0}\) is \((\mathbb{Z}, 2)\)-regular.

J.-P. Allouche, J. Shallit, The ring of \(k\)-regular sequences II.

J. Bell (2005)

Let \(R\) be a commutative ring. Let \(k, \ell\) be two multiplicatively independent integers. If a sequence \(x \in R^\mathbb{N}\) is both \((R, k)\)-regular and \((R, \ell)\)-regular, then it satisfies a linear recurrence over \(R\).
**Definition**

Consider an increasing sequence \((U_n)_{n \geq 0}\) of integers such that

- \(U_0 = 1\)
- \(\sup U_{n+1}/U_n\) is bounded

Any integer \(n\) can be written as

\[
    n = \sum_{i=0}^{\ell} c_i U_i, \quad c_i > 0.
\]

We choose the **greedy representation**: \(\text{rep}_U(n) = c_{\ell} \cdots c_0\).

---


M. Lothaire, Algebraic Combinatorics on Words, Cambridge Univ. Press 2002, Chap. by Ch. Frougny

Combinatorics, Automata and Number Theory, V. Berthé, M. Rigo (Eds.), Cambridge Univ. Press 2010, Chap. 2 & 3
canonical alphabet $A_U = \{0, \ldots, \lceil \max U_{n+1}/U_n \rceil - 1 \}$

$\text{rep}_U : \mathbb{N} \rightarrow A_U^*$

for any alphabet $B \subset \mathbb{Z}$, $\text{val}_U : B^* \rightarrow \mathbb{Z}$

$$\text{val}_U(d_\ell \cdots d_0) = \sum_{i=0}^{\ell} d_i U_i.$$ 

**Remark**

We have *positional* numeration systems.
FIBONACCI (ZECKENDORF 1972)

\[ \text{rep}_F(11) = 10100 \text{ but } \text{val}_F(10100) = \text{val}_F(10011) = \text{val}_F(1111) \]
\[ U_0 = 1, \ U_1 = 2 \text{ and } U_{n+2} = U_{n+1} + U_n. \]


..., 610, 377, 233, 144, 89, 55, 34, 21, 13, 8, 5, 3, 2, 1

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>8</td>
<td>10000</td>
<td>15</td>
<td>100010</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>9</td>
<td>10001</td>
<td>16</td>
<td>100100</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>10</td>
<td>10010</td>
<td>17</td>
<td>100101</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>101</td>
<td>11</td>
<td>10100</td>
<td>18</td>
<td>101000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1000</td>
<td>12</td>
<td>10101</td>
<td>19</td>
<td>101001</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1001</td>
<td>13</td>
<td>100000</td>
<td>20</td>
<td>101010</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1010</td>
<td>14</td>
<td>100001</td>
<td>21</td>
<td>1000000</td>
<td></td>
</tr>
</tbody>
</table>
Can we extend Cobham’s theorem on recognizability into two integer base systems to non-standard numeration systems?

**Definition**

A set \( X \subset \mathbb{N} \) is *\( U \)-recognizable*, if the set of greedy expansions of the elements of \( X \) is accepted by some finite automaton, i.e., \( \text{rep}_U(X) \) is a regular language.

If \( X \subset \mathbb{N} \) is \( U \)-rec. and \( V \)-rec., \( U \) and \( V \) being “sufficiently independent”, does it imply that \( X \) is ultimately periodic?

We can also study the \( U \)-recognizable sets of integers for themselves!
It is natural to ask whether or not $\text{rep}_U(\mathbb{N})$ is regular... check with a DFA if a word over $A_U$ is a valid representation?

**Observation (G. Hansel, J. Shallit, ...)**

If $\mathbb{N}$ is $U$-recognizable, then $(U_n)_{n \geq 0}$ satisfies a linear recurrence relation with (constant) integer coefficients.

$\text{rep}_U(U_\ell) = 10^\ell$ for all $\ell \geq 0$. Amongst the words of length $\ell + 1$ in $\text{rep}_U(\mathbb{N})$, the smallest one for the genealogical ordering is $10^\ell$.

Consequently, $U_{\ell+1} - U_\ell = \#(\text{rep}_U(\mathbb{N}) \cap A^{\ell+1})$.

Since $\text{rep}_U(\mathbb{N})$ is regular, it is accepted by a DFA and the number of words of length $n$ in $\text{rep}_U(\mathbb{N})$ is equal to the number of paths of length $n$ from the initial state to the final ones (then use Cayley-Hamilton theorem).
\( \mathbb{N} \) being \( U \)-recognizable is somehow a minimal requirement,

**Proposition**

Let \( p, r \geq 0 \). If \((U_n)_{n \geq 0}\) is a numeration system satisfying a linear recurrence relation with integer coefficients, then

\[
\text{val}^{-1}_{A_U, U}(p \, \mathbb{N} + r) = \left\{ c_\ell \cdots c_0 \in A_U^* \mid \sum_{k=0}^{\ell} c_k \, U_k \in p \, \mathbb{N} + r \right\}
\]

is accepted by a DFA that can be effectively constructed.

**Corollary**

If \( \mathbb{N} \) is \( U \)-recognizable, then any ultimately periodic set is \( U \)-recognizable.
Satisfying a linear recurrence is not enough...

**Counter-example (Shallit 1994)**

Take \((U_n)_{n \geq 0}\) defined by \(U_n = (n + 1)^2\).
We have \(U_0 = 1\), \(U_1 = 4\), \(U_2 = 9\) and
\(U_{n+3} = 3U_{n+2} - 3U_{n+1} + U_n\). In that case,

\[
\text{rep}_U(\mathbb{N}) \cap 10^*10^* = \{10^a10^b \mid b^2 < 2a + 4\}
\]

showing with the pumping lemma that \(\mathbb{N}\) is not \(U\)-recognizable.

N. Loraud, \(\beta\)-shift, systèmes de numération et automates, JTNB 7 (1995), 473—498.

Consider a linear numeration system such that the characteristic polynomial of \((U_n)_{n \geq 0}\) is the minimal polynomial of a Pisot number (i.e., an algebraic integer \(\alpha > 1\) whose Galois conjugates have modulus less than 1).


\[
\lim_{n \to \infty} \frac{U_n}{c\alpha^n} = 1.
\]

For these systems, all the “nice” properties hold true:

- \(\text{rep}_U(\mathbb{N})\) is regular (for any reasonable initial conditions),
- for a precise choice of initial conditions, we have a Bertrand system (i.e., \(\nu \in \text{rep}_U(\mathbb{N}) \iff \nu 0 \in \text{rep}_U(\mathbb{N})\)),
- normalization is computable by some finite automaton,
- the logical characterization can be extended,
- the morphic characterization too.
Pisot numeration systems

A link with the expansions of real numbers $L(\beta)$ is the set of factors in some sequences $d_\beta(x)$, $x \in [0, 1]$

$$d_\beta(x) = x_1 x_2 \cdots, \quad x = \sum_{i=1}^{+\infty} x_i \beta^{-i}$$

A. Bertrand (1989)

Let $U$ be a numeration system. It is a Bertrand system if and only if there exists a real number $\beta > 1$ such that

$$\text{rep}_U(\mathbb{N}) = L(\beta).$$

In this case, if $U$ is linear, then $\beta$ is a root of the characteristic polynomial of $U$. 
$D_\beta$ is the set of greedy $\beta$-expansions of numbers of $[0, 1)$.

**W. Parry (1960)**

Let $\beta > 1$ and let $s$ be an infinite sequence of non-negative integers. The sequence $s$ belongs to $D_\beta$ IFF

$$\forall k \geq 0, \quad \sigma^k(s) <_{\text{lex}} d^{*}_\beta(1)$$

and $s$ belongs to $S_\beta$, i.e., closure of $D_\beta$, IFF

$$\forall k \geq 0, \quad \sigma^k(s) \leq_{\text{lex}} d^{*}_\beta(1).$$

Let $\beta > 1$ be a real number. The language $L(\beta)$ is regular if and only if $\beta$ is a Parry number.

**Corollary**

The DFA accepting $\text{rep}_U(\mathbb{N})$ has a very special form.

The $\beta$-shift $S_\beta$ is a dynamical system which is

- sofic IFF $d_\beta(1)$ is ultimately periodic,
- of finite type IFF $d_\beta(1)$ is finite.

Pisot numeration systems

Integer base systems are special case of Pisot systems.

**Fibonacci**

\[ U_{n+2} = U_{n+1} + U_n \quad \text{with} \quad U_0 = 1 \quad \text{and} \quad U_1 = 2 \]

\[ P(X) = X^2 - X - 1 \quad \text{has roots} \quad \frac{1 + \sqrt{5}}{2}, \quad \frac{1 - \sqrt{5}}{2} \]

- \( d_\beta(1) = 11 \), \( \text{rep}_U(\mathbb{N}) \) is regular (no block 11)

- we have a Bertrand system
  (i.e., \( \nu \in \text{rep}_U(\mathbb{N}) \iff \nu0 \in \text{rep}_U(\mathbb{N}) \)),

---

0

\[ \quad 1 \]

\[ \quad 0 \]

\[ \quad 0 \]

\[ \quad 1 \]
(modified) Fibonacci

\[ U_{n+2} = U_{n+1} + U_n \]

with the initial conditions \( U_0 = 1, U_1 = 3 \)

\[ (U_n)_{n \geq 0} = 1, 3, 4, 7, 11, 18, 29, 47, \ldots \]
Normalization $\nu_U : B^* \to A_U^*$ seems to be an essential tool, $B \subset \mathbb{Z}$, if $\text{val}_U(w) \geq 0$, then $\nu_U(w) = \text{rep}_U(\text{val}_U(w))$.

Example for Fibonacci

$\nu_F : 11011 \mapsto 100100, \quad 11100 \mapsto 100100, \ldots, \quad 22 \mapsto 1001$

**Theorem (Ch. Frougny 1992)**

For any given alphabet $B$, for a Pisot system $U$, $\nu_U$ is realisable by a finite letter-to-letter transducer

**Corollary**

Addition is a $U$-recognizable ternary relation.


Ch. Frougny, J. Sakarovitch, Number representation and finite automata, CANT Ch. 2, Cambridge Univ. Press (2010).
Logical characterization

**Büchi–Bruyère–Hansel Theorem**

A set $X \subseteq \mathbb{N}$ is $U$-recognizable IFF it is definable by a first order formula in the extended Presburger arithmetic $\langle \mathbb{N}, +, V_U \rangle$.

$V_U(n)$ is the smallest $U_i$ occurring in $\text{rep}_U(n)$ with a non-zero coefficient.
What about a multidimensional context?

Everything works fine!

▷ automata reading \( n \)-tuples (with leading zeroes),
▷ morphisms with images being \( n \)-cubes of size \( k \),
▷ logical characterization in \( \langle \mathbb{N}, +, V_k \rangle \),
▷ extension to Cobham–Semenov’ theorem

\[ \text{COBHAM–SEMENOV’ THEOREM} \]

Let \( k, \ell \geq 2 \) be two multiplicatively independent integers. A set \( X^n \subseteq \mathbb{N} \) is \( k \)-rec. AND \( \ell \)-rec. IFF \( X \) is definable in \( \langle \mathbb{N}, + \rangle \).
WHAT ABOUT A MULTIDIMENSIONAL CONTEXT?
Looking at Cobham–Semenov’ theorem, the *right* extension of ultimate periodicity over \( \mathbb{N} \) could be definability in \( \langle \mathbb{N}, + \rangle \).

*cf.* local periodicity and Muchnik criterion

So far, we have seen

- integer base systems and recognizable sets
- Pisot numeration systems and recognizable sets

P. Lecomte (1997): “everyone is taking an increasing sequence of integers then look at the regularity of \( \text{rep}_U(\mathbb{N}) \).
We could proceed the other way round and start directly by taking a regular language!”

**Remark**

For positional numeration systems, \( \text{rep}_U \) is an increasing map:

\[
x < y \iff \text{rep}_U(x) < \text{rep}_U(y).
\]
**Definition ANS (P. Lecomte, M.R. 2001)**

An abstract numeration system $S = (L, A, <)$ is a regular language $L$ over a totally ordered finite alphabet $(A, <)$.


- Enumerating the words in $L$ using genealogical ordering provides a one-to-one correspondence between $\mathbb{N}$ and $L$:

  $\text{rep}_S : \mathbb{N} \rightarrow L, \quad \text{val}_S : L \rightarrow \mathbb{N}$.

- This generalizes any positional system $U$ for which $\mathbb{N}$ is $U$-recognizable.

**Definition**

A set $X \subseteq \mathbb{N}$ is $S$-recognizable, if $\text{rep}_S(X)$ is regular.
Example: consider a prefix-closed language $L = \{b, \varepsilon\} \{a, ab\}^*$.
A non-positional ANS \( L = a^*b^* \)
A non-positional ANS $L = a^*b^*$
Abstract numeration systems

A non-positional ANS $L = a^* b^*$
A non-positional ANS $L = a^* b^*$
A non-positional ANS $L = a^* b^*$
A non-positional ANS $L = a^* b^*$
A non-positional ANS $L = a^*b^*$
A non-positional ANS $L = a^* b^*$
A non-positional ANS $L = a^* b^*$

$$\text{val}_S(a^p b^q) = \frac{1}{2} (p + q)(p + q + 1) + q = \binom{p + q + 1}{2} + \binom{q}{1}$$

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>aa</th>
<th>ab</th>
<th>bb</th>
<th>aaa</th>
<th>⋮</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

$U_0 = 1$, $U_1 = 2$, $p(a) = 1$, $p(b) = 2$

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

$\forall n \in \mathbb{N}, \exists z_1, \ldots, z_\ell : n = \binom{z_\ell}{\ell} + \binom{z_{\ell-1}}{\ell - 1} + \cdots + \binom{z_1}{1}$

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

[Katona, Gel’fand, Lehmer, Fraenkel, Lew, Morales, ⋮]
\text{Abstract numeration systems}

\[
\text{val}(a^p b^q) \mod 8
\]
Theorem [P. Lecomte, M.R.]
Let $S$ be an ANS. Any ultimately periodic set of integers is $S$-recognizable.

Equivalent formulation [D. Krieger et al. TCS'09]
Let $L$ be a regular language. Any “periodic decimation” of $L$ is a regular language.

In general, this result does not hold for context-free languages.
Another example, an unambiguous positional system

\[ L = \{1, 2\}^* \]

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(\varepsilon)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>7</td>
<td>111</td>
</tr>
<tr>
<td>8</td>
<td>112</td>
</tr>
<tr>
<td>9</td>
<td>121</td>
</tr>
<tr>
<td>10</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ABSTRACT NUMERATION SYSTEMS

**Lemma**

\[
\text{val}_S(w) = \sum_{q \in Q} \sum_{i=1}^{|w|} b_{q,i}(w) u_q(|w| - i)
\]

with

\[
b_{q,i}(w) := \#\{a < w_i \mid q_0 \cdot w_1 \cdots w_{i-1} a = q\} + 1_{q_0,q}
\]

and

\[
u_q(n) = \#\{v \in A^n \mid q \cdot v \in F\}.
\]

\[
v_q(n) = \#\{v \in A^{\leq n} \mid q \cdot v \in F\} = \sum_{i=0}^n u_q(i).
\]
Let $S = (L, A, <)$ an ANS.

If $xy$ belongs to $L_q$, $y \neq \varepsilon$, then

$$\text{val}_q(xy) = \text{val}_{q,x}(y) + v_q(|xy| - 1) - v_{q,x}(|y| - 1) + \sum_{w < x \atop |w| = |x|} u_{q,w}(|y|).$$
If $xy$ belongs to $L_q$, $y \neq \varepsilon$, then

$$\text{val}_q(xy) = \text{val}_{q \cdot x}(y) + v_q(|xy| - 1) - v_{q \cdot x}(|y| - 1) + \sum_{|w| = |x|}^{w < x} u_{q \cdot w}(|y|).$$
Many natural questions. . .

- What about $S$-recognizable sets?
  - Are ultimately periodic sets $S$-recognizable for any $S$?
  - For a given $X \subseteq \mathbb{N}$, can we find $S$ s.t. $X$ is $S$-recognizable?
  - For a given $S$, what are the $S$-recognizable sets?

- Can we compute “easily” in these systems?
  - Addition, multiplication by a constant, . . .

- Are these systems equivalent to something else?

- Any hope for a Cobham’s theorem?

- Can we also represent real numbers?

- Number theoretic problems like additive functions?

- Dynamics, odometer, tilings, logic. . .
Recall that the set of squares is never recognizable in any integer base system.

**Example**

Let $L = a^* b^* \cup a^* c^*$, $a < b < c$.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\varepsilon$</td>
<td>$a$</td>
<td>$b$</td>
<td>$c$</td>
<td>$aa$</td>
<td>$ab$</td>
<td>$ac$</td>
<td>$bb$</td>
<td>$cc$</td>
<td>$aaa$</td>
<td>...</td>
</tr>
</tbody>
</table>

**Folklore**

If $L$ is a regular language, then the set $\min(L)$ of minimal words of each length is again regular.
**Theorem (M.R. 2002)**

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

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

There exists an ANS $S$ such that $f(\mathbb{N})$ is $S$-recognizable.

**Definition of Growth Rate**

Let $A = (Q, q_0, F, \Sigma, \delta)$

$$u_{q_0}(n) = \#(L \cap \Sigma^n).$$

**Theorem (P. Lecomte, M.R. 2001)**

Let $S = (a^*b^*, a < b)$. Multiplication by $\lambda \in \mathbb{N}_{>0}$ preserves $S$-recognizability, i.e., for all $S$-recognizable set $X \subseteq \mathbb{N}$, $\lambda X$ is $S$-recognizable, IFF $\lambda$ is an odd square.

---

**Theorem (“Multiplication by a Constant”)**

<table>
<thead>
<tr>
<th>Language Type</th>
<th>Property $u_{q_0}(n)$</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>slender language</td>
<td>$u_{q_0}(n) \in O(1)$</td>
<td>OK</td>
</tr>
<tr>
<td>polynomial language</td>
<td>$u_{q_0}(n) \in O(n^k)$</td>
<td>NOT OK</td>
</tr>
<tr>
<td>exponential language with polynomial complement</td>
<td>$u_{q_0}(n) \in 2^{\Omega(n)}$</td>
<td>NOT OK</td>
</tr>
<tr>
<td>exponential language with exponential complement</td>
<td>$u_{q_0}(n) \in 2^{\Omega(n)}$</td>
<td>OK?</td>
</tr>
</tbody>
</table>

The successor function can be computed by means of finite automata: It is realized by a (left or right) letter-to-letter finite transducer.

**Theorem (P.-Y. Angrand, J. Sakarovitch 2010)**

The radix enumeration of a rational language is a finite union of co-sequential functions.

A cascade of (at most 2) sequential (right) transducers, that is, a first transducer reads the input and produces an output which is then taken as the input of second transducer which depends on the final state in the computation of the first one.

**Definition**

Let $f : \Sigma \to \Sigma^*$ and $g : \Sigma \to \Gamma^*$ be two morphisms such that $f(a) \in a\Sigma^+$. We define a *morphic word* (a.k.a. substitutive) over $\Gamma$,

$$w = g(\lim_{n \to \infty} f^n(a)) = g(f^\omega(a)).$$

We can assume $f$ non-erasing and $g$ is a coding.

**Example (Characteristic Sequence of Squares)**

$f : a \mapsto abcd, \ b \mapsto b, \ c \mapsto cdd, \ d \mapsto d, \ g : a, b \mapsto 1, \ c, d \mapsto 0$.

$$f^\omega(a) = abcdabcbdddbcdadbcdddbcddddbcdddbcdabcbdddbcdadbcdddbcddddbcdddbcdddbcdadbcdddbcddddbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcddddbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdddbcdadbcdddbcdd dbc

$$g(f^\omega(a)) = 1100100001000000100000000010 \cdots$$
What is the link between morphic words and ANS?

Recall this result (Cobham 1972)

An infinite word $x$ is morphic and generated by a $k$-uniform morphism + coding IFF $x$ is $k$-automatic, i.e., $\forall n \geq 0$, $x_n$ is generated by an automaton reading $\text{rep}_k(n)$.

We can introduce $S$-automatic sequences...
**Morphic words**

**Definition**
Let $S = (L, \Sigma, <)$ be an ANS and $\mathcal{M} = (Q, q_0, \Sigma, \delta, \Gamma, \tau)$ be a DFAO. Consider the $S$-automatic sequence

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

**Example**

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

![Diagram](attachment:image.png)

01023031200231010123023031203120231002303120123023031203120231002310123 ⋮
Extension of Cobham’s result

**Theorem (A. Maes, M.R. 2002)**

An infinite word $x$ is morphic IFF there exists some ANS $S$ such that $x$ is a $S$-automatic.

The set of $S$-automatic sequences (for all $S$) coincides with the set of morphic words.

**Remark**

A set $X \subseteq \mathbb{N}$ is $S$-recognizable IFF its characteristic sequence is $S$-automatic.
**Morphic Words**

$k$-automatic sequence \[\iff\] $k$-uniform morphism + coding  

[A. Cobham'72]

multidimensional setup  
\[x : \mathbb{N}^d \rightarrow A\]

$k$-automatic sequence \[\iff\] morphism \(g : A \rightarrow (A^q)^d\) + coding  

[O. Salon'87]

$S$-automatic sequence \[\iff\] non-erasing morphism + coding  

[A. Maes, M.R.'02]

$S$-automatic sequence \[\iff\] “shape-symmetric” morphism + coding  

[É. Charlier, T. Kärki, M.R.'09]
**MORPHIC WORDS**

\[ \varphi : a \mapsto \begin{array}{cc} a & b \\ c & d \end{array} \quad b \mapsto \begin{array}{c} i \\ e \end{array} \quad c \mapsto \begin{array}{cc} i & j \end{array} \quad d \mapsto \begin{array}{c} i \end{array} \quad e \mapsto \begin{array}{cc} f & b \end{array} \]

\[ f \mapsto \begin{array}{cc} g & b \\ h & d \end{array} \quad g \mapsto \begin{array}{cc} f & b \\ h & d \end{array} \quad h \mapsto \begin{array}{cc} i & m \end{array} \quad i \mapsto \begin{array}{cc} i & m \\ h & d \end{array} \]

\[ j \mapsto \begin{array}{c} k \\ c \end{array} \quad k \mapsto \begin{array}{cc} l & m \\ c & d \end{array} \quad l \mapsto \begin{array}{cc} k & m \\ c & d \end{array} \quad m \mapsto \begin{array}{c} i \\ h \end{array} \]

**coding**

\[ \mu : e, g, j, l \mapsto 1, \quad a, b, c, d, f, h, i, k, m \mapsto 0 \]
### Morphic Words

The image illustrates the concept of morphic words using a process of transformation. Each step shows a matrix being transformed according to a specific rule. The rule given is:

\[ a \rightarrow \begin{array}{cc} a & b \\ c & d \end{array} \rightarrow \begin{array}{ccc} a & b & i \\ c & d & e \end{array} \rightarrow \begin{array}{cccc} a & b & i & i & m \\ c & d & e & h & d \\ i & j & i & f & b \\ i & m & k & i & m \\ h & d & c & h & d \end{array} \rightarrow \begin{array}{cccccc} a & b & i & i & m & i \\ c & d & e & h & d & h \\ i & j & i & f & b & i \\ i & m & k & i & m & g & b & i \\ h & d & c & h & d & e & e \end{array} \rightarrow \begin{array}{cccccc} a & b & i & i & m & i \\ c & d & e & h & d & h \\ i & j & i & f & b & i \\ i & m & i & l & m & i & m & i \\ h & d & h & c & d & h & d & h \\ i & m & i & i & j & i & m & i \end{array} \rightarrow \cdots \]
MORPHIC WORDS

From E. Duchêne, A. S. Fraenkel, R. Nowakowski, M.R., Extensions and restrictions of wythoff’s game preserving wythoff’s sequence as set of $\mathcal{P}$-positions, JCTA (2010).
From A. Maes Ph.D. thesis, *Prédicats morphiques et applications à la décidabilité de théories arithmétiques*
**Theorem (F. Durand 1998)**

Let \((f, g, a)\) (resp. \((f', g', a')\)) be a primitive substitution with a dominating eigenvalue \(\alpha > 1\) (resp. \(\beta > 1\)). Let \(\alpha\) and \(\beta\) be multiplicatively independent. If \(x = g(f^\omega(a)) = g'(f'^\omega(a'))\), then \(x\) is ultimately periodic.

Morphic words

An “application”

**Example**

The Fibonacci word 0100101001 · · · generated by
\[ f : 0 \mapsto 01, \quad 1 \mapsto 0 \]
is not \( k \)-automatic.

Indeed, this (Sturmian) word is not ultimately periodic and for all integers \( k \), \( k \) and \((1 + \sqrt{5})/2\) are multiplicatively independent.

Of course, one can also use this result about frequency

**Proposition**

In any \( k \)-automatic sequence, if the frequency of a symbol exists, then it is rational.
An “application”

**Example**

If $X \subseteq \mathbb{N}$ is both $S$- and $T$-recognizable where $S$ (resp. $T$) is built over an exponential (resp. a polynomial) language then $X$ is ultimately periodic.
A FEW WORDS ON $\omega$-HD0L ULTIMATE PERIODICITY

Question: given $f, g$, two morphisms, decide whether or not $g(f^\omega(a))$ is ultimately periodic.

- Trivial for $k$-automatic sequences, thanks to first order logic.
A FEW WORDS ON $\omega$-HD0L ULTIMATE PERIODICITY


Equivalent question: *Let given a $S$-recognizable set of integers, decide whether or not it is ultimately periodic.*

- F. Durand, Decidability of the HD0L ultimate periodicity problem, arXiv:1111.3268v1
- I. Mitrofanov, A proof for the decidability of HD0L ultimate periodicity, arXiv:1110.4780
Some open problems

- Give a proof based on ANS for the $\omega$-HD0L ultimate periodicity (based on automata, we could have a better view/complexity).
- If $g(f^\omega(a))$ is infinite, one can always assume that $f$ is non-erasing and $g$ is a coding [Cobham’68, Allouche–Shallit’03, Honkala’09], again give a proof based only on automata.
- Given a ANS, decide whether or not this system is a positional numeration system.
How to represent real numbers

Example (base 10)

\[ \pi - 3 = 0.14159265358979323846264338328 \cdots \]
\[ \frac{1}{10}, \frac{14}{100}, \frac{141}{1000}, \cdots, \frac{val(w_n)}{10^n}, \cdots \]

\[ \text{val}(w) \]
\[ \# \{ \text{words of length } \leq |w| \} \]

This deserves notation

\[ v_{q_0}(n) = \#(L \cap \Sigma^{\leq n}) = \sum_{i=0}^{n} u_{q_0}(i). \]
**Example (Avoid aa on three letters)**

| $w$    | $\text{val}(w)$ | $\mathbf{v}_{q_0}(|w|)$ | $\frac{\text{val}(w)}{\mathbf{v}_{q_0}(|w|)}$ |
|--------|-----------------|--------------------------|----------------------------------|
| $bc$   | 8               | 12                       | 0.66666666666666667              |
| $bac$  | 19              | 34                       | 0.55882352941176                 |
| $babc$ | 52              | 94                       | 0.55319148936170                 |
| $babac$| 139             | 258                      | 0.53875968992248                 |
| $bababc$| 380           | 706                      | 0.53824362606232                |

$$\lim_{n \to \infty} \frac{\text{val}((ba)^n c)}{\mathbf{v}_{q_0}(2n + 1)} = \frac{1}{1 + \sqrt{3}} + \frac{3}{9 + 5\sqrt{3}} \simeq 0.535898.$$
**How to Represent Real Numbers**

\[
\text{val}_S(w) = \sum_{q \in \mathbb{Q}} \sum_{i=1}^{\lfloor w \rfloor} b_{q,i}(w) u_q(\lfloor w \rfloor - i)
\]

with

\[
b_{q,i}(w) := \# \{ a < w_i \mid q_0 \cdot w_1 \cdots w_{i-1}a = q \} + 1_{q_0,q}
\]

**Hypotheses:** For all state \( q \) of \( M_L \), either

(i) \( \exists N_q \in \mathbb{N} : \forall n > N_q, u_q(n) = 0 \), or

(ii) \( \exists \beta_q \geq 1, P_q(x) \in \mathbb{R}[x], b_q > 0 : \lim_{n \to \infty} \frac{u_q(n)}{P_q(n)\beta_q^n} = b_q \).

From automata theory, we have

\[
\beta_{q_0} \geq \beta_q \text{ and } \beta_q = \beta_{q_0} \Rightarrow \text{deg}(P_q) \leq \text{deg}(P_{q_0})
\]
Let $\beta = \beta_{q_0}$ and for any state $q$, define

$$\lim_{n \to \infty} \frac{u_q(n)}{P_{q_0}(n)\beta^n} = a_q \in \mathbb{Q}(\beta), \quad a_{q_0} > 0 \text{ and } a_q \text{ could be zero.}$$

**If** $(w_n)_{n \in \mathbb{N}}$ **is converging to** $W = W_1W_2\cdots$ **then**

$$\lim_{n \to \infty} \frac{\text{val}(w_n)}{v_{q_0}(|w_n|)} = \frac{\beta - 1}{\beta^2} \sum_{j=0}^{\infty} \sum_{q \in Q} \frac{a_q}{a_{q_0}} b_{q,j+1}(W) \beta^{-j} = x.$$

We say that $W$ is a representation of $x$. 
HOW TO REPRESENT REAL NUMBERS

A real number can have

- a unique expansion
- finitely many expansions
- countably many expansions

$x \in I_w$, if there exists an infinite word having $w$ as prefix and representing $x$.

$W_\ell := \text{set of words of length } \ell \text{ that are prefixes of infinitely many words in } L$. Let $w \in W_\ell$,

$$I_w = \left[ \frac{1}{\beta} + \frac{\beta - 1}{\beta^{\ell+1}} \sum_{v < w, \ v \in W_\ell} \frac{a_{q_0.v}}{a_{q_0}}, \frac{1}{\beta} + \frac{\beta - 1}{\beta^{\ell+1}} \sum_{v \leq w, \ v \in W_\ell} \frac{a_{q_0.v}}{a_{q_0}} \right].$$
HOW TO REPRESENT REAL NUMBERS

**In base 10**

$L = \{\varepsilon\} \cup \{1, \ldots, 9\}\{0, \ldots, 9\}^*$, we represent $[1/10, 1]$.
**How to Represent Real Numbers**

### Notation (for all states q)

Ratio of words starting with $a$, $b$ or $c$...

\[
[0, 1] = [0, \lim \frac{u_{q,a}(n-1)}{u_q(n)}] \cup [\lim \frac{u_{q,a}(n-1)}{u_q(n)}, \lim \frac{u_{q,a}(n-1)+u_{q,b}(n-1)}{u_q(n)}] \cup [\lim \frac{u_{q,a}(n-1)+u_{q,b}(n-1)}{u_q(n)}, 1] \quad (A_{q,a}, A_{q,b}, A_{q,c})
\]

If $I = [a, b] \ni x$, then $f_I : [a, b] \to [0, 1] : x \mapsto (x - a)/(b - a)$
HOW TO REPRESENT REAL NUMBERS

Algorithm

Let $x \in [1/\beta, 1]$

Initialization

$q \leftarrow q_0$

$w \leftarrow \varepsilon$

$I \leftarrow [1/\beta, 1]$

$x \leftarrow f_I(x)$

repeat

Find the letter $\sigma \in \Sigma$ s.t. $x \in A_{q,\sigma}$.

$q \leftarrow q \cdot \sigma$

$w \leftarrow \text{concat}(w, \sigma)$

$I \leftarrow A_{q,\sigma}$

$x \leftarrow f_I(x)$

until some halt condition.
**NOTATION**

$h : Q \times [0, 1] \rightarrow Q \times [0, 1] : (q, x) \mapsto (q', x')$

There exists a unique letter $\sigma$ s.t. $x \in A_{q,\sigma}$ hence

\[
\begin{align*}
q' &= q \cdot \sigma \\
x' &= f_{A_{q,\sigma}}(x)
\end{align*}
\]

**GENERAL QUESTION**

are there $i < j$ such that $h^i(q_0, x) = h^j(q_0, x)$?
Fibonacci
FIBONACCI

\[\begin{array}{c}
0 & A & 1/\tau & B & 1 \\
A & B
\end{array}\]
FIBONACCI

\[ 0 \quad A \quad 1/\tau \quad B \quad 1 \]

A

B

A

B
Fibonacci

\[0 \quad A \quad 1/\tau \quad B \quad 1\]

A

B

A

B

A

B
FIBONACCI

\[
0 \quad A \quad 1/\tau \quad B \quad 1
\]

A

B

A

B
FIBONACCI
FIBONACCI
FIBONACCI
IN GENERAL, QUITE DIFFICULT


Fibred systems, see for instance, M. Madritsch