# Self-generating sets, integers with missing blocks, and substitutions


J.-P. Allouche
CNRS, LRI
Bâtiment 490
F-91405 Orsay Cedex
France
allouche@lri.fr

J. Shallit
School of Computer Science
University of Waterloo
Waterloo, Ontario, N2L 3G1
Canada
shallit@graceland.uwaterloo.ca

G. Skordev
CEVIS, Universität Bremen
Universitätsallee 29
D-28359 Bremen
Germany
skordev@cevis.uni-bremen.de



**Abstract**

We give a new construction of the Kimberling sequence defined by:

(a) 1 belongs to $S$;

(b) if the positive integer $x$ belongs to $S$, then $2x$ and $4x-1$ belong to $S$; and

(c) nothing else belongs to $S$,

hence

$$S = 1\ 2\ 3\ 4\ 6\ 7\ 8\ 11\ 12\ 14\ 15\ 16\ \cdots$$

which is sequence `A052499` in the Sloane's On-line Encyclopedia of Integer Sequences, by proving that this sequence is equal to sequence 1+ `A003754`, the sequence of integers whose binary expansion does not contain the block of digits 00. We give a general framework for this sequence and similar sequences, in relation to automatic or morphic sequences and to nonstandard numeration systems such as the lazy Fibonacci expansion.



*Keywords*: Kimberling sequence, self-generating sets, morphic sequences, automatic sequences, Fibonacci sequence, Thue-Morse sequence, lazy Fibonacci expansion, integers with missing blocks.

*MSC*: 11B37, 11B85, 11A63, 68R15.

# 1 Introduction

In a recent paper, Kimberling [11] studied the set of positive integers $S$ defined as follows:

(a) 1 belongs to $S$;

(b) if the positive integer $x$ belongs to $S$, then $2x$ and $4x-1$ belong to $S$; and

(c) nothing else belongs to $S$.

Hence, writing the integers in $S$ in increasing order, we obtain

$$S = 1\ 2\ 3\ 4\ 6\ 7\ 8\ 11\ 12\ 14\ 15\ 16\ \cdots.$$

Kimberling found a close relationship between this set and the Fibonacci sequence

$$0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ \cdots,$$

defined as the fixed point of the morphism $0 \to 01$, $1 \to 0$.

Here we give a different proof of Kimberling's result, showing in particular that the Kimberling sequence (sequence `A052499` in [19]) is equal to 1+ `A003754`, where `A003754` is the sequence of integers whose binary expansion does not contain the block of digits 00. We show furthermore that Kimberling's result can be considered as a typical example of results for sequences defined by a "missing digits" or "missing blocks" property.

# 2 The sequence studied by Kimberling

## 2.1 Kimberling's result

The main statement in [11] can be summarized as follows.

**Theorem 1 (Kimberling)** *Let $S$ be the increasing sequence of integers whose corresponding set, also denoted $S$, is defined as follows:*

*(a) 1 belongs to $S$;*

*(b) if the positive integer $x$ belongs to $S$, then $2x$ and $4x-1$ belong to $S$; and*

*(c) nothing else belongs to $S$.*

*Hence*

$$S = (S(n))_{n\geq 0} = 1\ 2\ 3\ 4\ 6\ 7\ 8\ 11\ 12\ 14\ 15\ 16\ \cdots$$

*Then, omitting the first term, this sequence reduced modulo 2 is the Fibonacci sequence $0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ \cdots$, defined as the fixed point of the morphism $0 \to 01$, $1 \to 0$.*

*In particular the ranks of the even terms of this sequence, i.e., the terms of the sequence*

$$R = (R(n))_{n\geq 1} := 1\ 3\ 4\ 6\ 8\ 9\ 11\ \cdots,$$

*are given by $R(n) = \lfloor (\frac{1+\sqrt{5}}{2})n \rfloor$.*

**Remark 1** The occurrence of the golden ratio $\frac{1+\sqrt{5}}{2}$ is unexpected, since the condition defining $S$ is a condition on the binary expansion of the integers involved.

**Remark 2** For more about the sequence $R = (R(n))_{n\geq 0}$, see sequence `A000201` in [19].

## 2.2 Integers whose binary expansion has no 00

Let us change the set $S$ slightly by translating it. Define the set $T$ by $T := S - 1$. In other words,

$$T := (T(n))_{n\geq 0} = 0\ 1\ 2\ 3\ 5\ 6\ 7\ 10\ 11\ 13\ 14\ 15\ \cdots$$

The following lemma is straightforward.

**Lemma 1** *The set of integers in the increasing sequence $T = (T(n))_{n\geq 0}$, also denoted $T$, can be defined as follows:*

*(a) 0 belongs to $T$;*

*(b) if the non-negative integer $x$ belongs to $T$, then $2x+1$ and $4x+2$ belong to $T$; and*

*(c) nothing else belongs to $T$.*

The next lemma is an easy consequence of Lemma 1. We remark that we assume the binary expansion of a number does not start with any leading zeros.

**Lemma 2** *The set $T$ is the set of integers that do not contain the block 00 in their binary expansion.*

As an immediate corollary we have

**Corollary 1** *Sequences* `A052499` *and* `A003754` *in [19] are related by the formula* `A052499` $=$ $1+$ `A003754`.

## 2.3 The characteristic function of the set $T$ is 2-automatic

In this section we prove that the characteristic function of the sequence $T$ is 2-automatic, and also that the sequence $T$ reduced modulo any integer $\geq 2$ is morphic. For definitions and properties of automatic and morphic sequences we refer to [5]. It is known that the characteristic function of a set defined by conditions similar to those in Lemma 1 is 2-automatic. However, we give a quick proof for the characteristic function of $T$, since we will need an explicit uniform morphism.

**Proposition 1** *Define the morphism $\sigma$ over the alphabet $\{a, b, c, d\}$ as follows: $\sigma(a) = ab$, $\sigma(b) = cb$, $\sigma(c) = db$, and $\sigma(d) = dd$. Define the map $f : \{a, b, c, d\} \to \{0, 1\}$ by $f(a) = f(b) = f(c) = 1$, $f(d) = 0$. Then the image under $f$ of the fixed point of the morphism $\sigma$ beginning with $a$ is the characteristic function of the set $T$.*

*Proof.* Using Lemma 1 it is clear that the characteristic function of $T$, say $(u_n)_{n\geq 0}$, satisfies the relations

- $u_0 = 1$,
- $u_{2n+1} = u_n$, for all $n \geq 0$,
- $u_{4n} = \begin{cases} 0, & \text{if } n \geq 1; \\ 1, & \text{if } n = 0. \end{cases}$

Hence, defining $V(n) := \begin{pmatrix} u_n \\ u_{2n} \\ u_{4n} \end{pmatrix}$, $A_0 := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, and $A_1 := \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, we have

$$V(2n) = A_0 V(n), \quad \text{and} \quad V(2n+1) = A_1 V(n)$$

for all $n \geq 0$. Defining $a := V(0) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $b := \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $c := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, and $d := \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, and the morphism $\sigma$ by $\sigma(a) = ab$, $\sigma(b) = cb$, $\sigma(c) = db$, $\sigma(d) = dd$, we thus have

$$\sigma(V(n)) = (V(2n), V(2n+1)),$$

hence the result. □

## 2.4 Generating the sequence ($T$ modulo 2) by substitution

If we consider the 2-dimensional sequence $\left(\begin{pmatrix} u_n \\ n \bmod 2 \end{pmatrix}\right)_{n\geq 0}$, i.e., the sequence

$$\begin{pmatrix} u_0 \\ 0 \end{pmatrix} \begin{pmatrix} u_1 \\ 1 \end{pmatrix} \begin{pmatrix} u_2 \\ 0 \end{pmatrix} \begin{pmatrix} u_3 \\ 1 \end{pmatrix} \cdots$$

it is clear that it can be obtained by combining the two morphisms $\sigma$ and $\lambda$, where $\lambda$ is defined on $\{0,1\}$ by $\lambda(0) = 01$, $\lambda(1) = 01$. In other words we have just proved the following proposition.

**Proposition 2** *The sequence* $\left(\begin{pmatrix} u_n \\ n \bmod 2 \end{pmatrix}\right)_{n\geq 0}$ *can be obtained by taking the image under the map* $g$ *of the fixed point of the morphim* $\Lambda$ *defined on the alphabet* $A := \begin{pmatrix} a \\ 0 \end{pmatrix}$, $B := \begin{pmatrix} b \\ 1 \end{pmatrix}$, $C := \begin{pmatrix} c \\ 0 \end{pmatrix}$, $D := \begin{pmatrix} d \\ 0 \end{pmatrix}$, $E := \begin{pmatrix} d \\ 1 \end{pmatrix}$, *by:*

$$\Lambda(A) = AB, \ \Lambda(B) = CB, \ \Lambda(C) = DB, \ \Lambda(D) = DE, \ \Lambda(E) = DE,$$

*where*

$$g(A) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ g(B) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ g(C) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ g(D) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ g(E) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

An immediate corollary gives a way of generating the sequence $T \bmod 2$; it suffices, in the construction of Proposition 2, to erase the letters whose first coordinate is $d$, i.e., the letters $D$ and $E$. But this can also be done by restricting the morphism to the alphabet $\{A, B, C\}$ and erasing the letters $D$ and $E$ from the images of $A$, $B$ and $C$ by the original morphism.

**Proposition 3** *The sequence $(T \bmod 2)$ can be obtained as the image by the map $h$ of the fixed point of the morphism $\psi$ on the alphabet $\{A, B, C\}$, where $\psi$ and $h$ are defined by*

$$\psi(A) = AB$$
$$\psi(B) = CB$$
$$\psi(C) = B$$

*and $h(A) = 0$, $h(B) = 1$, $h(C) = 0$.*

Actually the same method gives a more general result.

**Proposition 4** *Let $e$ be an integer $\geq 2$. Then the sequence $(T \bmod e)$ is morphic (i.e., it can be obtained as the image of the fixed point of a morphism on some finite alphabet).*

*Proof.* The proof is left to the reader. The main two ideas are to take a cartesian product of two morphisms where the second morphism is a 2-morphism that generates the periodic sequence $(0\ 1\ 2\ \cdots\ (e-2)\ (e-1)\ 0\ 1\ 2\ \cdots)$. One can take

if $e$ is even,

$$\begin{cases} 0 & \to & 01 \\ 1 & \to & 23 \\ 2 & \to & 45 \\ \vdots & & \\ \frac{e-2}{2} & \to & (e-2)(e-1) \\ \frac{e}{2} & \to & 01 \\ 1+\frac{e}{2} & \to & 23 \\ \vdots & & \\ (e-1) & \to & (e-2)(e-1) \end{cases}$$

if $e$ is odd,

$$\begin{cases} 0 & \to & 01 \\ 1 & \to & 23 \\ 2 & \to & 45 \\ \vdots & & \\ \frac{e-1}{2} & \to & (e-1)0 \\ \frac{e+1}{2} & \to & 12 \\ \frac{e+3}{2} & \to & 34 \\ \vdots & & \\ (e-1) & \to & (e-2)(e-1) \end{cases}$$

For example the patient reader can check that the sequence $(T \bmod 3)$ can be generated as the image under the map

$$p, s, t \to 0, \quad q, v \to 1, \quad r, u \to 2$$

of the fixed point of the morphism

$$p \to pq, \quad q \to rs, \quad s \to tq, \quad r \to u, \quad t \to q, \quad u \to vu, \quad v \to s.$$

$\square$

**Remark 3** The construction of a morphism obtained by "erasing" some letters in a given morphism was easy here, since the letters we erased had the property that their images under the original morphism were also erased. Nevertheless this restriction can be dropped (see Theorem 3 below).

## 2.5 Complements

Actually it happens that the sequence ($T$ mod 2) is a fixed point of a simpler morphism, namely the morphism defined on $\{0,1\}$ by $0 \to 01$, $1 \to 011$. Furthermore the shifted sequence obtained from ($T$ mod 2) by erasing the first 0 is the fixed point of the morphism $1 \to 10$, $0 \to 1$ (i.e., it is equal to the Fibonacci sequence beginning with 1).

**Proposition 5**

- *The sequence* $(T \bmod 2) = (T(n) \bmod 2)_{n \geq 0}$ *is the fixed point of the morphism* $\mu$ *defined on* $\{0,1\}$ *by* $\mu(0) = 01$, $\mu(1) = 011$.
- *The sequence* $(T(n+1) \bmod 2)_{n \geq 0}$ *is the fixed point of the morphism* $1 \to 10$, $0 \to 1$ *(i.e., it is equal to the Fibonacci sequence beginning with* 1*).*

*Proof.*

- We know from Proposition 3 that the sequence $(T(n) \bmod 2)_{n \geq 0}$ is equal to the sequence $\lim_{k \to \infty} h(\psi^k(A))$.

  We first prove by induction on $k \geq 0$ that $1\ h(\psi^{2k}(A)) = h(\psi^{2k}(BC))$. Namely the result clearly holds for $k = 0$. Then, if the result holds for $k$, we have:

$$\begin{aligned} 1\ h(\psi^{2k+2}(A)) &= 1\ h(\psi^{2k}(ABCB)) = 1\ h(\psi^{2k}(A))h(\psi^{2k}(BCB)) \\ &= h(\psi^{2k}(BC))h(\psi^{2k}(BCB)) = h(\psi^{2k}(BCBCB)) \\ &= h(\psi^{2k+2}(BC)). \end{aligned}$$

  Now we prove by induction on $k \geq 0$ that $\mu^{k+1}(0) = h(\psi^{2k}(A))\ 1$ and $1\ \mu^k(1) = h(\psi^{2k}(B))\ 1$. The result holds for $k = 0$ and $k = 1$. If it holds for $k \geq 1$, we have

$$\begin{aligned} 1\ \mu^{k+1}(1) &= 1\ \mu^k(011) = 1\ \mu^k(0)\mu^k(1)\mu^k(1) \\ &= 1\ h(\psi^{2k-2}(A))\ 1\ \mu^k(1)\mu^k(1) \\ &= 1\ h(\psi^{2k-2}(A))h(\psi^{2k}(B))\ 1\ \mu^k(1) \\ &= 1\ h(\psi^{2k-2}(A))h(\psi^{2k}(B))h(\psi^{2k}(B))\ 1 \\ &= h(\psi^{2k-2}(BC))h(\psi^{2k}(B))h(\psi^{2k}(B))\ 1 \\ &= h(\psi^{2k-2}(BC))h(\psi^{2k-2}(\psi^2(B)))h(\psi^{2k-2}(\psi^2(B)))\ 1 \\ &= h(\psi^{2k-2}(BCBCBBCB))\ 1 \\ &= h(\psi^{2k+2}(B))\ 1 \end{aligned}$$

  and

$$\begin{aligned} \mu^{k+2}(0) &= \mu^{k+1}(01) = h(\psi^{2k}(A))\ 1\ \mu^{k+1}(1) \\ &= h(\psi^{2k}(A))h(\psi^{2k+2}(B))\ 1 = h(\psi^{2k}(ABCB))\ 1 \\ &= h(\psi^{2k+2}(B))\ 1. \end{aligned}$$

This finally implies:

$$(T(n) \bmod 2)_{n\geq 0} = \lim_{k\to\infty} h(\psi^k(A)) = \lim_{k\to\infty} h(\psi^{2k}(A)) = \lim_{k\to\infty} \mu^{k+1}(0) = \lim_{k\to\infty} \mu^k(0).$$

- To prove the second claim, it clearly suffices to prove that we have $\lim_{k\to\infty} \psi^k(A) = A \lim_{k\to\infty} \tau^k(B)$, where $\tau$ is the morphism defined on the alphabet $\{B, C\}$ by $\tau(B) = BC$, $\tau(C) = B$.

  Clearly $\tau^2(B) = \tau(B)B$, so by applying $\tau^k$ to both sides, we obtain the relation $\tau^{k+2}(B) = \tau^{k+1}(B)\tau^k(B)$. We will now prove by induction on $k \geq 0$ the relations:

$$B\ \psi(\tau^k(C)) = \tau^k(B)\ B \text{ and } B\ \psi(\tau^k(B)) = \tau^{k+1}(B)\ B.$$

  These relations are clearly true for $k = 0$. If they are true for $k$, then

$$\begin{aligned} B\ \psi(\tau^{k+1}(B)) &= B\ \psi(\tau^k(BC)) = B\ \psi(\tau^k(B))\ \psi(\tau^k(C)) \\ &= \tau^{k+1}(B)\ B\ \psi(\tau^k(C)) = \tau^{k+1}(B)\tau^k(B)\ B \\ &= \tau^{k+2}(B)\ B. \end{aligned}$$

  and

$$B\ \psi(\tau^{k+1}(C)) = B\ \psi(\tau^k(B)) = \tau^{k+1}(B)\ B.$$

Our last step is to prove that for all $k \geq 0$ we have

$$\psi^k(A) = A\ \tau^k(B)\ \tau^{k-1}(B)\ \cdots\ \tau(B)\ B.$$

Again this is true for $k = 0$, and if it is true for $k$, then

$$\begin{aligned} \psi^{k+1}(A) &= \psi(\psi^k(A)) = \psi(A)\ \psi(\tau^k(B))\ \psi(\tau^{k-1}(B))\ \ \cdots\ \psi(\tau(B))\ \psi(B) \\ &= A\ B\psi(\tau^k(B))\ \psi(\tau^{k-1}(B))\ \ \cdots\ \psi(\tau(B))\ \psi(B) \\ &= A\ \tau^{k+1}(B)\ B\ \psi(\tau^{k-1}(B))\ \ \cdots\ \psi(\tau(B))\ \psi(B) \\ &= A\ \tau^{k+1}(B)\ \tau^k(B)\ B\ \ \cdots\ \psi(\tau(B))\ \psi(B) \\ &= \cdots \\ &= A\ \tau^{k+1}(B)\ \tau^k(B)\ \cdots\ \tau(B)\ B. \end{aligned}$$

It is now clear that

$$\lim_{k\to\infty} \psi^k(A) = A\ \lim_{k\to\infty} \tau^k(B).$$

□

## 2.6 The sequence $T$ and the lazy Fibonacci expansion

We have seen that the integers in $T$ are exactly the integers whose binary expansion does not contain the block 00. Writing these expansions in increasing order:

$$0,\ 1,\ 10,\ 11,\ 101,\ 110,\ 111,\ 1010,\ \ldots$$

we obtain exactly the representations of the integers $0,\ 1,\ 2,\ \ldots$ in the "lazy Fibonacci expansion" (i.e., as expansions in the Fibonacci base $F_2 = 1$, $F_3 = 2$, ..., $F_{n+2} = F_{n+1} + F_n$ that do not contain two consecutive 0's; see for example [10]). In other words, we have the following proposition:

**Proposition 6** *The binary representation of $T(n)$ is given by the lazy Fibonacci expansion of $n$.*

**Remark 4** The above proposition, which can be read between the lines of [11], contains in particular the equality (with our notation)

$$T(F_{n+3} - 2) = 2^n - 1,$$

due to Bottomley (see sequence `A052499` in [19]). Namely,

$$F_{n+3} - 2 = F_2 + F_3 + \cdots + F_{n+1}.$$

Actually, introducing the concept of generalized automatic sequences in [17] and the concept of generalized regular sequences in [2], we can prove the following theorem.

**Theorem 2** *The sequence $(T(n))_{n\geq 0}$ is Fibonacci-regular. For each $e \geq 2$, the sequence $(T(n) \bmod e)_{n\geq 0}$ is Fibonacci-automatic.*

*Proof.* As a consequence of [2, Theorem 5.2], it suffices to prove the first assertion. We define two maps and two languages as follows:
- the maps $x \to [x]_2$ and $x \to [x]_F$ are defined from $\{0,1\}^*$ to $\mathbb{N}$ by

$$\begin{aligned} [w_0 w_1 \cdots w_k]_2 &= w_0 2^k + w_1 2^{k-1} + \cdots + w_k 2^0 \\ [w_0 w_1 \cdots w_k]_F &= w_0 F_{k+2} + w_1 F_{k+1} + \cdots + w_k F_2. \end{aligned}$$

- the set $\mathcal{F}$ (resp. $\mathcal{L}$) is the subset of $\{0,1\}^*$ consisting of all words beginning with 1 and having no subword equal to 11 (resp., beginning with 1 and having no subword equal to 00). In other words, $\mathcal{F}$ is the set of Fibonacci expansions of integers, and $\mathcal{L}$ is the set of lazy Fibonacci expansions of integers.

We then define a map $\lambda$ from $\mathcal{F}$ to $\mathcal{L}$ by noting that for all $w \in \mathcal{F}$, there exists a unique $\lambda(w) \in \mathcal{L}$ such that $[w]_F = [\lambda(w)]_F$ (by the uniqueness of lazy Fibonacci expansions of integers). For $w$ in $\mathcal{F}$, it is clear that $\lambda(w)$ can be computed from $w$ by repeatedly applying the rewriting rule $100 \to 011$ and erasing leading zeros if any; this process converges (look at the total number of 0's at each step), and the order in which the rewriting rules are applied does not matter, since the final result must be the unique lazy Fibonacci expansion of the integer with Fibonacci expansion equal to $w$.

Note that, applying the rewriting rules to $w \in \mathcal{F}$ without erasing the leading zeros if any appears during the process, leads to some word $\lambda'(w)$ with the same length as $w$, and which must be either $\lambda(w)$ or $0\lambda(w)$ (easy induction on the length of $w$). In particular if $w$ and $z$ are two words beginning with 1, we have $\lambda(wz) = \lambda(\lambda(w)\lambda'(z))$.

Notice that, with this notation, Proposition 6 can be written as follows: for all $w \in \mathcal{L}$ we have $T([w]_F) = [\lambda(w)]_2$.

To prove the claim in our Theorem 2 it suffices to prove the following relations that were discovered experimentally:

$$\begin{array}{lcl}
T([10000x]_F) & = & -2T([10x]_F) + T([100x]_F) + 2T([1000x]_F) \\
T([100010x]_F) & = & -2T([1000x]_F) + T([1010x]_F) + 2T([10010x]_F) \\
T([10010x]_F) & = & -2T([10x]_F) + 3T([1000x]_F) \\
T([101000x]_F) & = & -6T([10x]_F) + 7T([1000x]_F) \\
T([1010010x]_F) & = & -6T([1000x]_F) + 7T([10010x]_F) \\
T([10101x]_F) & = & -4T([1x]_F) + 5T([101x]_F)
\end{array}$$

We will prove only the first one; the remaining relations are proved analogously. We also will prove the first relation only in the case $x$ begins with 1 (if $x$ begins with $0\cdots 01$ the proof is the same, it is also similar if $x = 0\cdots 0$). We will use the fact that $x$ begins with 1, and the above reformulation of Proposition 6. Let us distinguish two cases.

- First case: $\lambda'(x) = \lambda(x)$.

  Then

$$\begin{array}{lclclcl}
T([10000x]_F) & = & T([\lambda(10000x)]_F) & = & T([1011\lambda(x)]_F) & = & [1011\lambda(x)]_2 \\
T([10x]_F) & = & T([\lambda(10x)]_F) & = & T([10\lambda(x)]_F) & = & [10\lambda(x)]_2 \\
T([100x]_F) & = & T([\lambda(100x)]_F) & = & T([11\lambda(x)]_F) & = & [11\lambda(x)]_2 \\
T([1000x]_F) & = & T([\lambda(1000x)]_F) & = & T([110\lambda(x)]_F) & = & [110\lambda(x)]_2
\end{array}$$

  Looking at the binary expansions, it is then apparent that

$$[1011\lambda(x)]_2 - [11\lambda(x)]_2 = 2[110\lambda(x)]_2 - 2[10\lambda(x)]_2,$$

  which is exactly the relation we want to prove.

- Second case: $\lambda'(x) = 0\lambda(x)$.

  Then

$$\begin{array}{lclclcl}
T([10000x]_F) & = & T([\lambda(10000x)]_F) & = & T([10110\lambda(x)]_F) & = & [10110\lambda(x)]_2 \\
T([10x]_F) & = & T([\lambda(10x)]_F) & = & T([11\lambda(x)]_F) & = & [11\lambda(x)]_2 \\
T([100x]_F) & = & T([\lambda(100x)]_F) & = & T([110\lambda(x)]_F) & = & [110\lambda(x)]_2 \\
T([1000x]_F) & = & T([\lambda(1000x)]_F) & = & T([1011\lambda(x)]_F) & = & [1011\lambda(x)]_2
\end{array}$$

  Looking at the binary expansions, it is then apparent that

$$[10110\lambda(x)]_2 - [110\lambda(x)]_2 = 2[1011\lambda(x)]_2 - 2[11\lambda(x)]_2,$$

  which is exactly the relation we want to prove.

□

# 3 A general framework

The sequence studied by Kimberling can be seen as a typical example of a family of similar sequences. Before giving the general result, we need some intermediate statements. The first one, due to Cobham [8] (see also [5, Theorem 7.7.4]), states that erasing given symbols in automatic (or, more generally, morphic) sequences essentially gives morphic sequences.

**Theorem 3 (Cobham)** *Let $U$ be a morphic sequence defined on the finite alphabet $\Sigma$. Let $\Gamma \subseteq \Sigma$. Let $V$ be the sequence obtained from $U$ by erasing all occurrences of the letters belonging to $\Gamma$. Then the sequence $V$ is either finite or morphic.*

Now we recall that two-dimensional sequences whose coordinates are $k$-automatic are also $k$-automatic, and that periodic sequences are $k$-automatic.

**Lemma 3** *Let $k$ be an integer $\geq 2$.*

- *If the sequences $(u_n)_{n\geq 0}$ and $(v_n)_{n\geq 0}$ are $k$-automatic, then the sequence $((u_n, v_n))_{n\geq 0}$ is $k$-automatic.*
- *Any periodic sequence is $k$-automatic.*

*Proof.* The first assertion is clear: look at the $k$-kernel of the sequence $((u_n, v_n))_{n\geq 0}$ (for the definition of the $k$-kernel, see for example [5, p. 185 and p. 409]).

For the second assertion, let $e \geq 1$ be the period of the sequence and let $w$ be the prefix of this sequence of length $e$, say $w = a_0 a_1 \cdots a_{e-1}$. Let $\mathcal{A} := \{A_0, A_1, \ldots, A_{e-1}\}$ be an $e$-letter alphabet. Write the word $w^k$ as the concatenation of $e$ words of length $k$, say $w^k := z_0 z_1 \cdots z_{e-1}$. It is clear that our periodic sequence is the image under the map $A_i \to a_i$ of the fixed point beginning with $A_0$ of the morphism

$$A_0 \to z_0,\ A_1 \to z_1,\ \ldots,\ A_{e-1} \to z_{e-1}.$$

□

Now we can prove a general theorem similar to Propositions 3 and 4.

**Theorem 4** *Let $k$ be an integer $\geq 2$. Let $(u_n)_{n\geq 0}$ be a $k$-automatic sequence. Let $a$ be an element of the set of values of $(u_n)_{n\geq 0}$. Let $i_0 < i_1 < i_2 < \cdots < i_n < \cdots$ be the increasing sequence of the integers $j$ such that $u_j = a$. Let $e$ be an integer $\geq 2$. Then either the sequence $(i_n \bmod e)_{n\geq 0}$ is finite or it is the image of the fixed point of a (not necessarily uniform) morphism.*

*Proof.* The proof goes as above. First, the characteristic function of the set of the integers $j$ such that $u_j = a$ is $k$-automatic. Then, if the sequence $(v_n)_{n\geq 0}$ is the periodic sequence $0\ 1\ 2\ \cdots\ (e-1)\ 0\ 1\ 2 \cdots$, the sequence $((u_n, v_n))_{n\geq 0}$ is $k$-automatic (use Lemma 3).

Finally, if the sequence $(i_n \bmod e)_{n\geq 0}$ is not finite, then it is morphic from Theorem 3: it suffices to erase in the sequence $((u_n, v_n))_{n\geq 0}$ all terms for which $(u_n, v_n)$ is equal to $(b, \ell)$ for $b \neq a$ and $\ell \in [0, e-1]$.

Actually, even more is true. By using a result of Dekking [9], we can prove the following result.

**Theorem 5** *Let $(u_n)_{n\geq 0}$ be a morphic sequence. Let $a$ be an element of the set of values of $(u_n)_{n\geq 0}$. Let $i_0 < i_1 < i_2 < \cdots < i_n < \cdots$ be the increasing sequence of the integers $j$ such that $u_j = a$. Let $e$ be an integer $\geq 2$. Then either the sequence $(i_n \bmod e)_{n\geq 0}$ is finite or it is the image of the fixed point of a (not necessarily uniform) morphism.*

*Proof.* (sketch) Dekking proved [9] that any transduction of a morphic sequence is morphic. The sequence $(u_n)_{n\geq 0}$ is transduced by just looking at the value of each letter to see whether it is equal to $a$, while counting the rank of this letter modulo $e$. As output take the residue modulo $e$ in this count. □

# 4 Other examples in the literature

Actually, other examples of the situation described in the previous section can be found in the literature. We describe some of them below.

## 4.1 A result of Cateland

The following result is due to Cateland (see the chapter "Sur la représentation en base $(q, d)$" pages 90–108 in [7]). It shows how sequences of integers obtained by forbidding digits are $p$-regular (in the sense of [3]) while their characteristic functions are known to be $q$-automatic.

A particular case of the sequences studied by Cateland is given by Kimberling in [12]: take $q = 3$, $d = 0$, and look at the integers that contain only the digits $\{0, 1\}$ in their ternary expansion. Then the increasing sequence $S$ of these integers can be defined by: 0 belongs to $S$; if the non-negative integer $x$ belongs to $S$, then $3x$ and $3x + 1$ belong to $S$; nothing else belongs to $S$ (see below).

**Theorem 6 (Cateland)** *Let $q$ and $d$ be two integers such that $q \geq 3$, and $2 - q \leq d \leq 0$. Let $\Sigma_q^{(d)} := \{d, d+1, \cdots, d+q-1\}$. Every integer $n \geq 1$ can be uniquely written as*

$$n = \sum_{j=0}^{L} \varepsilon_j(n) q^j$$

*with $\varepsilon_j(n) \in \Sigma_q^{(d)}$, for all $j \in [0, L]$ and $\varepsilon_L(n) > 0$. Call this expansion the $(q, d)$-expansion of $n$. Fix $p$ digits in the set $\Sigma_q^{(d)}$ where at least one of these digits belongs to $\{1, 2, \cdots, d+q-1\}$. Then, the increasing sequence of the integers whose $(q, d)$-expansion contains only these $p$ digits is $p$-regular.*

An immediate corollary is:

**Corollary 2** *With the above notation, reducing modulo $e \geq 2$ the increasing sequence of the integers whose $(q, d)$-expansion contains only these $p$ digits gives a $p$-automatic sequence.*

## 4.2 The examples of Kimberling

In the paper [12] Kimberling studies sets of integers defined as follows. Let $F$ be a countable set of affine functions with integer coefficients. The set of integers $S_F$ is the smallest set of integers that contains 1 and is closed under any function in $F$. We now show how the examples of sets $S_F$ given by Kimberling enter our framework.

– Example 1 of [12] considers the sequence $S_F$ where $F := \{2t+1, 4t, 4t+1\}$ which is sequence `A003159` in [19]:

$$1\ 3\ 4\ 5\ 7\ 9\ 11\ 12\ 13\ 15\ 16\ 17\ 19\ 20\ 21\ 23\ 25\ \cdots$$

It is not difficult to see from Kimberling's definition that this sequence consists exactly of the integers whose binary expansion does not end with an odd number of 0's. The reader can check that the characteristic sequence $\chi$ of `A003159` (with a 0 inserted at the beginning) satisfies

$$\forall n \geq 0,\ \chi(2n+1) = 1,\ \chi(4n) = \chi(n),\ \chi(4n+2) = 0.$$

It is an easy exercise to prove that the sequence $(\chi(n))_{n\geq 0}$ is 2-automatic: it is the image by the map $a \to 1$, $b \to 1$, $c \to 0$, of the fixed point of the 2-morphism

$$a \to ab, \quad b \to cb, \quad c \to bb.$$

Note that it easily proved that, for $n \geq 1$, we have $\chi(n) = |M(n) - M(n-1)|$, where $(M(n))_{n\geq 0}$ is the celebrated Thue-Morse sequence

$$0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ \cdots$$

(the Thue-Morse sequence can be defined as the fixed point beginning with 0 of the 2-morphism $0 \to 01$, $1 \to 10$, see for example [4]). In other words, $\chi(n) = 1$ if and only if $M(n-1) \neq M(n)$. This implies that the sequence `A003159` is the summatory function of the sequence

$$1\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 1\ 2\ 1\ 1\ 2\ 2\ \cdots,$$

the fixed point of the morphism $1 \to 121$, $2 \to 12221$. (Note that the shifted sequence

$$2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 1\ 2\ 1\ 1\ 2\ 2\ \cdots,$$

is the fixed point of the morphism $2 \to 211$, $1 \to 2$; see [6]; for more on the links between the Thue-Morse sequence and the sequence `A003159` see also [1]).

– Example 2 of [12] considers the sequence $S_F$ where $F := \{3t, 3t+1\}$ which, up to inserting 0, is sequence `A005836` in [19]:

$$0\ 1\ 3\ 4\ 9\ 10\ 12\ 13\ 27\ 28\ 30\ 31\ 36\ 37\ 39\ 40\ 81\ \cdots$$

As noted by Kimberling, this sequence consists exactly of the integers whose ternary representation does not contain the digit 2.

Theorem 6 implies that this sequence $(U(n))_{n\geq 0}$ is 2-regular: it satisfies the relations $U(2n) = 3U(n)$, $U(2n+1) = 3U(n) + 1$. Note that the reduction modulo 2 of this sequence is the Thue-Morse sequence.

– Example 3 of [12] considers the sequence $S_F$ where $F := \{2t, 4t+1\}$ which, up to inserting a 0, is sequence `A003714` in [19]:

$$0\ 1\ 2\ 4\ 5\ 8\ 9\ 10\ 16\ 17\ 18\ 20\ 21\ 32\ 33\ 34\ 36\ \cdots$$

As noted by Kimberling, this sequence consists exactly of the integers whose binary expansion has no two adjacent 1's (sometimes called "Fibbinary numbers"; see [19]). The reader can easily check that the characteristic sequence $\chi$ of `A003714` is 2-automatic: it is the image by the map $a \to 1$, $b \to 1$, $c \to 0$, of the fixed point of the 2-morphism

$$a \to ab, \quad b \to ac, \quad c \to cc.$$

Reasoning analogous to the proof of Proposition 6 gives a way to compute this sequence by showing that, with the notation of Section 2, we have for all $w \in \mathcal{F}$ the relation $V([w]_F) = [w]_2$. In particular this sequence reduced modulo 2 is nothing but the Fibonacci sequence.

– Example 4 of [12] considers the sequence $S_F$ where $F := \{2t, 4t+3\} \cup \{2^{k+1}t + 2^k + 1 : \ k \geq 2\}$ which is sequence `A000069` in [19]:

$$1\ 2\ 4\ 7\ 8\ 11\ 13\ 14\ 16\ 19\ 21\ 22\ 25\ 26\ 28\ 31\ 32\ \cdots$$

As noted by Kimberling, this sequence consists exactly of the integers whose binary expansion has an odd number of 1's. It is not hard to deduce that the characteristic sequence $\chi$ of sequence `A000069` is 2-automatic: this is the fixed point of the 2-morphism

$$0 \to 01, \quad 1 \to 10$$

hence $(\chi(n))_{n\geq 0}$ is the Thue-Morse sequence beginning in 0. As noted in [19], sequence `A000069` is the sequence $2n + 1 - M(n)$, where $(M(n))_{n\geq 0}$ is the Thue-Morse sequence beginning with 0. Hence sequence `A000069` is 2-regular. We see that sequence (`A000069` modulo 2) is *also* the Thue-Morse sequence up to changing 0's into 1's and 1's into 0's.

– Example 5 of [12] considers the sequence $S_F$ where $F := \{2t+1, 4t\} \cup \{2^{k+1}t + 2^k - 2 : \ k \geq 2\}$ which is sequence `A059010` in [19] (note that the integer 16 is incorrectly missing in the sequence given in [12]):

$$1\ 3\ 4\ 7\ 9\ 10\ 12\ 15\ 16\ 19\ 21\ 22\ 25\ 26\ 28\ 31\ 33\ 34\ \cdots$$

As noted by Kimberling, this sequence consists exactly of the integers whose binary expansion has an even number of 0's. It is easy to show that the characteristic sequence of sequence `A059010` with a 0 inserted is 2-automatic: it is the image by the map $a \to 1$, $b \to 1$, $c \to 0$, of the fixed point of the 2-morphism

$$a \to ab, \quad b \to cb, \quad c \to bc.$$

Note that it can be proved that sequence `A059010` is the sequence $(2n - 1 - Q(n-1))_{n\geq 1}$, where $Q(n)$ is the parity of the number of 0's in the integer $n$, hence sequence `A059010` is 2-regular.

# 5 Questions

We list below some related questions.

- How far can the sequence proposed by Kimberling be generalized? One possible answer is given in the paper [12] with the sets $S_F$ quoted above. Under which conditions is the corresponding characteristic sequence automatic? More precisely if we have "mixed base" rules (for example, $S_F := \{2t+1, 3t\}$), then the corresponding characteristic sequence is probably not $k$-automatic for any $k$.

- Are the sequences of integers with missing blocks always particular cases of sequences defined à la Kimberling as above?

- Is it possible to study a more general situation where notions of "automaticity" or "regularity" in the sense of [17] or of [14, 16] and of [2] are considered? In particular the papers [18, 15, 13] might prove useful in this context.

**Acknowledgments**. Part of this work was done while JPA visited Cevis at Bremen University. JPA wants to thank all the colleagues of Cevis (with a special mention to H.-O. Peitgen) for their warm welcome.

**Note added October 28, 2004**. The equality `A052499` $=1+$ `A003754` has also been noted by C. Kimberling, see p. 144 of C. Kimberling, *Ordering words and sets of numbers: the Fibonacci case*, in Applications of Fibonacci numbers, Vol. 9, 2004, Kluwer Acad. Publ., Dordrecht, 137–144.

# References

[1] J.-P. Allouche, A. Arnold, J. Berstel, S. Brlek, W. Jockush, S. Plouffe, B. E. Sagan, *A relative of the Thue-Morse sequence*, Discrete Math. **139** (1995) 455–461.

[2] J.-P. Allouche, K. Scheicher, R. F. Tichy, *Regular maps in generalized number systems*, Math. Slovaca **50** (2000) 41–58.

[3] J.-P. Allouche, J. Shallit, *The ring of k-regular sequences*, Theoret. Comput. Sci. **98** (1992) 163–187.

[4] J.-P. Allouche, J. Shallit, *The ubiquitous Prouhet-Thue-Morse sequence*, in Sequences and their applications, Proceedings of SETA'98, C. Ding, T. Helleseth and H. Niederreiter (Eds.), 1999, Springer Verlag, 1–16.

[5] J.-P. Allouche, J. Shallit, Automatic Sequences. Theory, Applications, Generalizations, Cambridge University Press, 2003.

[6] S. Brlek, *Enumeration of factors in the Thue-Morse word*, Discrete Appl. Math. **24** (1989) 83–96.

[7] E. Cateland, Suites digitales et suites $k$-régulières, Thèse, Université Bordeaux I, 1991.

[8] A. Cobham, *On the Hartmanis-Stearns problem for a class of tag machines*, in IEEE Conference Record of 1968 Ninth Annual Symposium on Switching and Automata Theory, 1968, pp. 51-60.

[9] F. M. Dekking, *Iteration of maps by an automaton*, Discrete Math. **126** (1994) 81–86.

[10] P. Erdős, I. Joó, V. Komornik, *Characterization of the unique expansions* $1 = \sum_{i=1}^{\infty} q^{-n_i}$ *and related problems*, Bull. Soc. Math. France **118** (1990) 377–390.

[11] C. Kimberling, *A self-generating set and the golden mean*, J. Int. Seq. **3** (2000) Article 00.2.8.

http://www.math.uwaterloo.ca/JIS/VOL3/goldentext.html

[12] C. Kimberling, *Affinely recursive sets and orderings of languages*, Discrete Math. **274** (2004) 147–159.

[13] M. Hollander, *Greedy numeration systems and regularity*, Theory Comput. Syst. **31** (1998) 111–133.

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

[15] N. Loraud, *$\beta$-shift, systèmes de numération et automates*, J. Théor. Nombres Bordeaux **7** (1995) 473–498.

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

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

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

[19] N. J. A. Sloane, On-line Encyclopedia of Integer Sequences, published electronically at

http://www.research.att.com/∼njas/sequences/