1. Some definitions

The aim of this chapter is to introduce some concepts and to fix the notations we will use throughout this book. We will first introduce some terminology in combinatorics on words. These notions have their counterpart in terms of symbolic dynamics. We shall illustrate these definitions through the example of a particular sequence, the Morse sequence, generated by an algorithmic process we shall study in details in this book, namely a substitution. After recalling some basic notions on substitutions, we shall focus on the concept of automatic sequences. We conclude this chapter with Sturmian sequences.

Let us start by giving a short and incomplete list of basic references on the concepts developed here. Let us first mention the indispensable [285] on substitution dynamical systems. For detailed introductions to the symbolic dynamics, see [49, 69, 206, 223]. More generally, there are a number of excellent books on dynamical systems and ergodic theory [67, 103, 123, 164, 196, 271, 280, 367]. For references on word combinatorics, automata and formal languages, see [135, 178, 228, 229, 306] and the references therein.

We shall use the following usual notation: $\mathbb{R}$ denotes the set of *real numbers*; $\mathbb{Q}$ denotes the set of *rational numbers*; $\mathbb{Z}$ denotes the set of *relative integers*; $\mathbb{C}$ denotes the set of *complex numbers*; $\mathbb{N}$ denotes the set of *nonnegative integers* ($0 \in \mathbb{N}$); $\mathbb{N}^+$ denotes the set of positive integers ($0 \notin \mathbb{N}^+$). By x positive we mean $x > 0$ and by x nonnegative, we mean $x \geq 0$.

The cardinality of a set S shall be denoted by Card S.

1.1 Sequences, words, languages

Let us first introduce some terminology in word combinatorics.

Let $\mathcal{A}$ be a finite set that shall be called *alphabet*. Throughout this book the letters (also called *symbols*) of the alphabet $\mathcal{A}$ shall either be denoted as digits ($\mathcal{A} = \{0, 1, \ldots, d-1\}$) or as letters ($\mathcal{A} = \{a_1, \ldots, a_{d-1}\}$).

A (one-sided) *sequence* of elements of $\mathcal{A}$ (also called *(right) infinite word* on $\mathcal{A}$) shall be here an element $u = (u_n)_{n\in\mathbb{N}}$ in $\mathcal{A}^{\mathbb{N}}$.

Definition 1.1.1. *A* word *or* block *is a finite string of elements in $\mathcal{A}$. The set of all finite words over $\mathcal{A}$ is denoted by $\mathcal{A}^\star$. We denote by ε the empty*

word.
The concatenation *of two words* $V = v_1...v_r$ *and* $W = w_1...w_s$ *is the word* $VW = v_1...v_r w_1...w_s$*. This operation is associative and has a unit element, the empty word* ε*. The set* $\mathcal{A}^\star$ *is thus endowed with the structure of a monoid, and is called the* free monoid *or* free semi-group *generated by* $\mathcal{A}$*.*
If $W = w_1...w_s$*,* s *is called the* length *of* W *and denoted by* $|W|$*. We denote by* $|W|_a$ *the number of occurrences of a letter* $a \in \mathcal{A}$ *appearing in a word* $W \in \mathcal{A}^\star$*.*

Definition 1.1.2. *A word* $v_1...v_r$ *is said to* occur *at* place *(or* index*, or* rank*)* m *in a sequence* $(u_n)_{n\in\mathbb{N}}$ *or in a finite word* $u_0...u_s$*, if* $u_m = v_1, \ldots, u_{m+r-1} = v_r$*. We say also that the word* $v_1...v_r$ *is a* factor *of the sequence* u*. The* language *(respectively the language of length* n*) of the sequence* u*, denoted by* $\mathcal{L}_n(u)$ *(respectively* $\mathcal{L}(u)$*), is the set of all words of length* n *in* $\mathcal{A}^\star$ *which occur in* u*.*

Definition 1.1.3. *A sequence* u *is said to be* recurrent *if every factor occurs infinitely often.*

Definition 1.1.4. *A sequence* u *is* periodic *(respectively* ultimately periodic*) if there exists a positive integer* T *such that*

$$\forall n \in \mathbb{N},\ u_n = u_{n+T}, \quad (\textit{respectively } \exists n_0 \in \mathbb{N},\ \forall n \geq n_0,\ u_n = u_{n+T}).$$

1.2 Complexity function

There is a classical measure of disorder for sequences taking their values in a finite alphabet, the so-called complexity function.

Definition 1.2.1. *Let* u *be a sequence; we call* complexity function *of* u*, and denote by* $p_u(n)$*, the function which with each positive integer* n *associates Card* $\mathcal{L}_n(u)$*, that is the number of different words of length* n *occurring in* u*.*

The complexity function is obviously non-decreasing. For any positive integer n, one has furthermore: $1 \leq p(n) \leq d^n$, where d denotes the cardinality of the alphabet $\mathcal{A}$. See [13] and [142] for many results linked to the notion of complexity. Note that the complexity function can be considered as a good measure of the disorder of a sequence as it is smallest for periodic sequences. Namely, a famous result of [255] is the following.

Proposition 1.2.1. *If* u *is a periodic or ultimately periodic sequence,* $p_u(n)$ *is a bounded function. If there exists an integer* n *such that* $p_u(n) \leq n$*,* u *is an ultimately periodic sequence.*

Proof. The first part is trivial. In the other direction, we have $p_u(1) \geq 2$ otherwise u would be constant, so $p_u(n) \leq n$ implies that $p_u(k+1) = p_u(k)$ for some k. We now write a fundamental equality:

$$p_u(k+1) - p_u(k) = \begin{cases} \sum_{W \in \mathcal{L}_k(u)} (\mathrm{Card}\{a \in A; Wa \in \mathcal{L}_{k+1}(u)\} - 1) \\ \sum_{W \in \mathcal{L}_k(u)} (\mathrm{Card}\{a \in A; aW \in \mathcal{L}_{k+1}(u)\} - 1). \end{cases}$$

And for every word W of length k occurring in u, there exists at least one word of the form Wa occurring in u, for some letter $a \in \mathcal{A}$; as $p_u(k+1) = p_u(k)$, there can be only one such word. Hence, if $u_i \ldots u_{i+k-1} = u_j \ldots u_{j+k-1}$, then $u_{i+k} = u_{j+k}$. As the set $\mathcal{L}_k(u)$ is finite, there exist $j > i$ such that $u_i \ldots u_{i+k-1} = u_j \ldots u_{j+k-1}$, and hence $u_{i+p} = u_{j+p}$ for every $p \geq 0$. ■

Let us introduce the notion of special word:

Definition 1.2.2. *Let W be a factor of the sequence $u \in \mathcal{A}^{\mathbb{N}}$. An* extension *of the factor W is a letter $x \in \mathcal{A}$ such that Wx is also a factor of the sequence u. A factor is said to be a* left special factor *(respectively* right special factor*) if it has more than one left (respectively right) extension.*

Let W^+ (respectively W^-) denote the number of right (respectively left extensions) of W. We have

$$p(n+1) - p(n) = \sum_{|W|=n} (W^+ - 1) = \sum_{|W|=n} (W^- - 1).$$

This equality which was fundamental in the previous proof, is a very useful tool for the computation of the complexity. See for instance Exercise 3.3.1 in Chap. 3, and also Chap. 6.

Remark. One can also introduce the notion of bispecial factors (these are the factors which are simultaneously right and left special factors). For a more detailed exposition, see [91].

1.3 Symbolic dynamical systems

Let us introduce some basic notions in symbolic dynamics. For expository books on the subject, see [49, 69, 285, 206, 223]. A more detailed exposition of this subject shall be made in Chap. 3.

Let us equip the set $\mathcal{A}^{\mathbb{N}}$ with the product topology of the discrete topology on each copy of $\mathcal{A}$. Thus, this set is a compact space. The topology defined on $\mathcal{A}^{\mathbb{N}}$ is equivalent to the topology defined by the following distance:

$$\text{for } u \neq v \in \mathcal{A}^{\mathbb{N}},\ d(u,v) = 2^{-\min\{n \in \mathbb{N}; u_n \neq v_n\}}.$$

Thus, two sequences are close to each other if their first terms coincide. Note that the space $\mathcal{A}^{\mathbb{N}}$ is complete as a metric compact space.

Let T denote the following map defined on $\mathcal{A}^{\mathbb{N}}$, called the *one-sided shift*:

$$T((u_n)_{n\in\mathbb{N}}) = (u_{n+1})_{n\in\mathbb{N}}.$$

The map T is uniformly continuous, onto but not necessarily one-to-one on $\mathcal{A}^{\mathbb{N}}$. (We shall denote the shift in this book either by T or by S according to the chapters.)

Definition 1.3.1. *The* symbolic dynamical system *associated with a sequence u with values in $\mathcal{A}$ is the system $(\overline{\mathcal{O}(u)}, T)$, where $(\overline{\mathcal{O}(u)} \subset \mathcal{A}^{\mathbb{N}}$ is the closure of the* positive orbit *of the sequence u under the action of the shift T, i.e., $\{T^n u, n \in \mathbb{N}\}$.*

Remark.

More generally, a *topological dynamical system* (X, T) is usually defined as a compact set X together with a continuous map T acting on the set X.

Definition 1.3.2. *For a word $W = w_0...w_r$, the* cylinder set *$[W]$ is the set $\{x \in X_u; x_0 = w_0, ..., x_r = w_r\}$.*

Lemma 1.3.1. *The set $X_u := \overline{\mathcal{O}(u)}$ is a compact space and T is a continuous mapping from X_u to X_u. The cylinder sets are* clopen *(open and closed) sets and form a base of open sets for the topology of X_u.*

Proof. The set $\mathcal{A}^{\mathbb{N}}$ is compact, and X_u is a closed subset; $d(Tx, Ty) \leq 2d(x, y)$, hence the continuity. And this implies that if $x = \lim_{n\to+\infty} T^{k_n} u$, then $Tx = \lim_{n\to+\infty} T^{k_n+1} u$ and $Tx \in X_u$.
If the cylinder $[W]$ is nonempty and x is a point in it, $[W]$ is identified with both the open ball $\{y; d(x,y) < 2^{-n}\}$ and the closed ball $\{y; d(x,y) \geq 2^{-n-1}\}$, hence the last assertion. ■

Note that the topology extends in a natural way to $\mathcal{A}^{\mathbb{N}} \cup \mathcal{A}^{\star}$.

Lemma 1.3.2. *The following statements are equivalent:*

- *(1) $x \in X_u$.*
- *(2) There exists a sequence k_n such that $x_0...x_n = u_{k_n}...u_{k_n+n}$ for every $n \geq 0$.*
- *(3) $\mathcal{L}_n(x) \subset \mathcal{L}_n(u)$ for all n.*

Proof. Statement (1) is equivalent to $d(x, T^{k_n} u) < 2^{-n}$ for some sequence k_n, and that is exactly (2); (2) implies (3) as any word occurring in x must occur in some $x_0...x_n$, and (3) implies (2) because $x_0...x_n$ is in $\mathcal{L}_{n+1}(u)$. ■

Exercise 1.3.1. Prove that a sequence u is recurrent if and only if there exists a strictly increasing sequence $(n_k)_{k\in\mathbb{N}}$ such that $u = \lim_{k\to+\infty} T^{n_k} u$. Deduce that the sequence u is recurrent if and only if T is onto on $\overline{\mathcal{O}(u)}$.

Consider now the notion of *minimal* sequence (also called *uniform recurrence*):

Definition 1.3.3. *A sequence $u = (u_n)_{n\in\mathbb{N}}$ is minimal (or uniformly recurrent) if for every $i < j$, there exists s such that for every n, there exists $n < m < n + s$ such that $u_m = u_i$, ..., $u_{m-i+j} = u_j$, i.e., if every word occurring in u occurs in an infinite number of places with bounded gaps.*

Let us note that the term "minimality" comes from symbolic dynamics: a sequence u is minimal if and only if the dynamical system $(\overline{\mathcal{O}(u)}, T)$ is minimal. For more details, see Chap. 3 and for instance, [285].

1.4 Substitutions

Substitutions are maps defined over the set of words $\mathcal{A}^\star$, which generate in a natural way infinite sequences with low complexity function.

1.4.1 Definitions

Definition 1.4.1. *A* substitution *is an application from an alphabet $\mathcal{A}$ into the set $\mathcal{A}^\star$ of finite words on $\mathcal{A}$; it extends to a morphism of $\mathcal{A}^\star$ for the concatenation by $\sigma(WW') = \sigma(W)\sigma(W')$. It also extends in a natural way to a map defined over $\mathcal{A}^{\mathbb{N}}$.*
It is called of constant length k *if $\sigma(a)$ is of length k for any $a \in A$.*
A fixed point *of σ is an infinite sequence u with $\sigma(u) = u$.*
An n-word *for σ is any one of the words $\sigma^n(a)$ for $a \in \mathcal{A}$.*

Remark.

- We shall restrict ourselves in this book to *non-erasing* substitutions, i.e., substitutions with values in $\mathcal{A}^\star - \{\varepsilon\}$.
- Substitutions are very efficient tools for producing sequences. Let σ be a substitution over the alphabet $\mathcal{A}$, and a be a letter such that $\sigma(a)$ begins by a and $|\sigma(a)| \geq 2$. Then there exists a unique fixed point u of σ beginning with a. This sequence is obtained as the limit in $\mathcal{A}^\star \cup \mathcal{A}^{\mathbb{N}}$ (when n tends toward infinity) of the sequence of n-words $\sigma^n(a)$, which is easily seen to converge.

Example. Let us consider a classical example of substitution of constant length, namely the Morse substitution.

Definition 1.4.2. *The* Morse sequence *u is defined as the fixed point beginning by a of the* Morse substitution *σ defined over the alphabet $\{a, b\}$ by $\sigma(a) = ab$ and $\sigma(b) = ba$.*

1.4.2 Abelianization

Note that a substitution σ (over $\mathcal{A}$) is determined by $\sigma(a)$, for $a \in \mathcal{A}$. There is a convenient way to represent a substitution under the form a matrix.

Let us define the incidence matrix of the substitution σ.

Definition 1.4.3. *Let σ be a substitution defined over the alphabet $\mathcal{A} = \{\mathbf{a}_1, \ldots, a_d\}$ of cardinality d. The $d \times d$ matrix $\mathbf{M}_\sigma$ the entry of indices (i, j) of which is $|\sigma(a_j)|_{a_i}$ is, by definition, the* incidence matrix *of the substitution σ.*

Let us note that for every $(i, j) \in \mathcal{A}^2$ and for every $n \in \mathbb{N}$, $|\sigma^n(j)|_i$ is equal to the coefficient of index (i, j) of the matrix $\mathbf{M}_\sigma^n$.

Example. The incidence matrix of the Morse substitution is the matrix:

$$\mathbf{M}_\sigma = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Definition 1.4.4. *Let σ be a substitution defined over the alphabet $\mathcal{A}$ of cardinality d. Let $f : \mathcal{A}^\star \to \mathbb{Z}^d$ denote the* canonical homomorphism *(also called* homomorphism of abelianization *) associated with σ. If $\{\boldsymbol{e}_i \mid 1 \le i \le d\}$ is the canonical basis of $\mathbb{Z}^d \subset \mathbb{R}^d$, the map f is defined as follows:*

$$\begin{array}{ll} f(i) & = \boldsymbol{e}_i, \quad i = 1, 2, 3, \\ f(\varepsilon) & = \mathbf{0}, \\ f(W) & = \sum_{i=1}^{k} f(w_i) \ \textit{for every}\ W = w_1 \cdots w_k \in \mathcal{A}^\star. \end{array}$$

We thus have: $\forall W \in \mathcal{A}^\star$, $f(W) = (|w|_i)_{1 \le i \le d} \in \mathbb{N}^d$.

The incidence matrix $\mathbf{M}_\sigma$ satisfies:

$$\mathbf{M}_\sigma = [f(\sigma(a_1)), \ldots, f(\sigma(a_d))].$$

Furthermore, we have the following commutative relation:

Lemma 1.4.1. *The commutative relation holds:*

$$\begin{array}{ccc} & \mathcal{A}^\star \xrightarrow{\sigma} \mathcal{A}^\star & \\ f & \downarrow \qquad \downarrow & f\,; \\ & \mathbb{Z}^3 \xrightarrow{\mathbf{M}_\sigma} \mathbb{Z}^3 & \end{array}$$

in other words,

$$\forall W \in \mathcal{A}^\star, \ f(\sigma(W)) = \mathbf{M}_\sigma f(W).$$

We shall see in Chap 7, Chap. 8 or Chap. 10 how convenient this notion can be.

1.4.3 Primitivity

Let us introduce the fundamental notion of primitivity. Primitive substitutions have very interesting properties.

Definition 1.4.5. *The substitution σ over the alphabet $\mathcal{A}$ is* primitive *if there exists a positive integer k such that, for every a and b in $\mathcal{A}$, the letter a occurs in $\sigma^k(b)$.*

Proposition 1.4.1. *If σ is primitive, any of its fixed points is a minimal sequence.*

Proof. We have $u = \sigma^k(u) = \sigma^k(u_0)\sigma^k(u_1)\ldots$; for any $b \in \mathcal{A}$, a occurs in $\sigma^k(b)$, hence a occurs in u an infinite number of times with bounded gaps; but then so does every $\sigma^n(a)$ in $u = \sigma^n(u)$, hence so does any word occurring in u. ∎

Remark. The converse of Proposition 1.4.1 is not true. Nevertheless, the following substitution is not primitive but its fixed point beginning by 0 is a minimal infinite sequence:

$$0 \mapsto 0010, \ 1 \mapsto 1.$$

This sequence is called the Chacon sequence; for more details, see Chap. 3.

Let us add now the following extra assumptions on substitutions:

1. there exists a letter $a \in \mathcal{A}$ such that $\sigma(a)$ begins by a (this condition guarantees the existence of a fixed point);
2. $$\lim_{n\to\infty} |\sigma^n(b)| = \infty, \text{ for every letter } b \in \mathcal{A}.$$

Then under these assumptions, we have the following equivalence:

Proposition 1.4.2. *Let σ be a substitution satisfying the above conditions. The substitution σ is primitive if and only if the fixed point of σ beginning by a is minimal.*

The proof is left as an exercise (see also [285]).

1.4.4 The Perron–Frobenius theorem

The minimality of a substitution can easily be expressed in terms of the incidence matrix. Indeed the substitution σ is primitive if and only if the matrix $\mathbf{M}_\sigma$ is *positive*, i.e., there exists an integer k such that the k-th power of the matrix has (strictly) positive entries.

Positive matrices satisfy the Perron–Frobenius theorem:

Theorem 1.4.1 (Perron–Frobenius' theorem). *Let M be a primitive matrix with nonnegative entries. Then M admits a strictly positive eigenvalue α which dominates in modulus the other eigenvalues λ: $\alpha > |\lambda|$. The eigenvalue α is a simple eigenvalue and there exists an eigenvector with positive entries associated with α.*

This fundamental property of positive matrices implies the existence of frequencies for every factor of a fixed point of a primitive substitution.

Definition 1.4.6. *Let u be a sequence. The* frequency f_W *of a factor W of u is defined as the limit (when n tends towards infinity), if it exists, of the quotient of the number of occurrences of the factor W in $u_0 u_1 \ldots u_{n-1}$ divided by n.*

Theorem 1.4.2. *Let σ be a primitive substitution satisfying the conditions mentioned above in Sec. 1.4.3. Let u be a fixed point of σ. Then every factor has a frequency. Furthermore, all the frequencies are (strictly) positive.*

For a proof of this result, see Chap. 3 and [285].

The case where the dominant eigenvalue is a Pisot number is of particular interest.

Definition 1.4.7. *A substitution is* of Pisot type *if the eigenvalues of the incidence matrix satisfy the following: there exists a dominant eigenvalue α such that for every other eigenvalue λ, one gets:*

$$|\alpha| > 1 > |\lambda| > 0.$$

Thus, the eigenvalue α is a Pisot number.

Remark. Following the previous definition, 0 cannot be an eigenvalue of the incidence matrix of a Pisot substitution. Hence, the Morse substitution is not Pisot. This implies that the characteristic polynomial of the incidence matrix of a Pisot type substitution is irreducible over $\mathbb{Q}$; moreover, all the eigenvalues are simple (for a proof, see Proposition 7.2.1 below).

It is worth noticing the following result due to V. Canterini:

Theorem 1.4.3. *Any Pisot type substitution is primitive.*

We shall focus on primitive substitutions of Pisot type in Chap. 7, 8 and 12.

1.5 Automata

Among substitutions, substitutions of constant length play an important role. The aim of this section is to prove that they are connected in a fundamental way with automatic sequences.

1.5.1 Definitions

Let k be an integer greater than or equal to 2. Let us introduce the notion of k-automaton: such an automaton is a finite complete deterministic automaton (also called 2-*tape automaton or transducer*). For more details, the reader is referred to Chap. 2, 3, 10, see also [101, 135, 178] and the surveys [8, 19].

Definition 1.5.1. *A k-automaton is represented by a directed graph defined by:*

- *a finite set of vertices called* states $\mathcal{S} = \{i = a_1, a_2, \cdots, a_d\}$*, one of these states i is called the* initial state*;*
- *k oriented edges called* transition maps *from the set of states $\mathcal{S}$ into itself, labelled by integers between* 0 *and* $k-1$
- *a set Y and a map φ from $\mathcal{S}$ into a set Y, called* output function *or* exit map.

A sequence $(u_n)_{n\in\mathbb{N}}$ with values in Y is called k-*automatic* (it is also called a k-*uniform tag sequence*) if it is generated by a k-automaton as follows: let $\omega(n)$ be the base k expansion of the integer n; starting from the initial state one feeds the automaton with the sequence $\omega(n)$, the digits being read in increasing order of powers; after doing this the automaton is in the state $a(n)$; the automaton is said to be fed in *reverse reading*. Then put $u_n = \varphi(a(n))$. One can similarly give another definition of k-automaticity by reading the digits in the reverse order, i.e., by starting with the least significant digit, but these two notions are easily seen to be equivalent (see for instance Chap. 4). The automaton is said to be fed in *direct reading*. In the construction of the automaton in direct reading one must be careful about the problem of leading zeros, i.e., the edge 0 has to map the initial state onto itself. Except otherwise stated, we shall consider automata in reverse reading throughout this book.

Remark.

- An automaton is not considered here as a machine recognizing a language, as usual, but as a machine producing a sequence.
- The set of automatic sequences is countable.

1.5.2 Substitutions of constant length

There is a fundamental connection between automatic sequences and fixed points of substitutions of constant length. The following characterization of automatic sequences is due to Cobham (see [101]).

Definition 1.5.2. *A* letter-to-letter projection *is a map from a finite alphabet $\mathcal{A}$ to a finite alphabet, say $\mathcal{B}$, which extends in a natural way to a map from $\mathcal{A}^\star \cup \mathcal{A}^\mathbb{N}$ to $\mathcal{B}^\star \cup \mathcal{B}^\mathbb{N}$. A sequence, deduced from a fixed point of a substitution by a letter-to-letter projection, is called a* substitutive sequence.

Proposition 1.5.1. *A sequence u is k-automatic if and only if u is the image by a letter-to-letter projection of a fixed point of a substitution of constant length k.*

Proof.

Let us give a constructive proof of this equivalence by considering an automaton in direct reading: the image by a letter-to-letter projection p of the fixed point of a substitution of constant length is generated by an automaton with as set of states, the alphabet, with transition maps given by the substitution, and with as output function, the projection p.

- Let u be the image by a letter-to-letter projection p of the fixed point $v = (v(n))_{n\in\mathbb{N}}$ of a substitution σ of constant length k defined on the alphabet $\mathcal{S}$. Define $\sigma = \sigma_0 \cdots \sigma_{q-1}$, where σ_i is a mapping from $\mathcal{S}$ to $\mathcal{S}$.
  Let us construct a q-automaton A in direct reading which recognizes the sequence u. Consider as set of states the alphabet $\mathcal{S}$ in which v takes values. Put as initial state i the first letter of v. If $v(n) = v(m)$, then $v(qn + r) = \sigma_r(v(n)) = \sigma_r(v(m)) = v(qm + r)$. Hence we can define, for any r in $[0, q-1]$, the map $f_r \colon \mathcal{S} \to \mathcal{S}$, $s \mapsto v(qn+j)$, if there exists n such that $s = v(n)$, and else $s \mapsto i$. In other words there is an edge from a to b if b occurs in $\sigma(a)$, labelled by i if b is the $i+1$-th letter of $\sigma(a)$. Define as output function the projection p. It is easily seen that the q-automaton A generates in direct reading the sequence u. Indeed, if $n = \sum_{i=0}^{t} n_i k^i$: we start from v_0, then go to the $n_t + 1$-th letter of $\sigma(v_0)$, denoted by a_1, then go to the $n_{t-1}+1$-th letter of $\sigma(a_1)$, which is also the $kn_t+n_{t-1}+1$-th letter of $\sigma^2(v_0)$, and, after n steps, we arrive to the $n+1$-th letter of $\sigma^{t+1}(v_0)$, which is v_n.
- Conversely, let u be a sequence generated by a k-automaton A in direct reading. Let $\mathcal{S}$ be the set of states of the automaton A and let $f_0, \cdots, f_{k-1}$ be the transition maps. Define the substitution of constant length $\sigma = f_0 \cdots f_{k-1}$ over $\mathcal{S}$. Let v be the fixed point of σ beginning with the initial state i (note that the edge f_0 maps i onto itself by definition, hence the existence of a fixed point). It is easily checked that the sequence u is the image by the output function φ of the fixed point v. ■

Example. One deduces from the proposition above that the following 2-automaton with initial state a and exit map $Id_{\{a,b\}}$ generates the Morse sequence:

1.6 Sturmian sequences

Let us end this chapter by evoking an important family of sequences, the so-called Sturmian sequences. Let us start with an example:

The *Fibonacci sequence* is defined over the two-letter alphabet $\{a, b\}$ as the limit of the n-words $\sigma^n(a)$, where the *Fibonacci substitution* σ is defined as $\sigma(a) = ab$ and $\sigma(b) = a$.

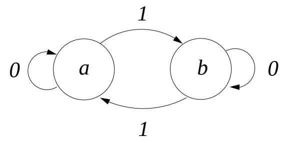

Fig. 1.1. Automaton associated with the Morse sequence

Definition 1.6.1. *A* Sturmian sequence *is defined as a sequence* u *the complexity function* p_u *of which satisfies:*

$$\forall n \in \mathbb{N}, p_u(n) = n + 1.$$

In particular, a Sturmian sequence is defined over a two-letter alphabet. Chapter 6 and 9 shall study in detail the properties of Sturmian sequences. For a proof of the fact that the Fibonacci sequence is Sturmian, see Exercise 3.3.1.

Exercise 1.6.1. For any $s \geq 1$, construct an example of sequence the complexity of which satisfies: $\forall n \in \mathbb{N}, p(n) = s + n$; (*hint: consider non recurrent sequences*).

One thus can define Sturmian sequences over a larger size alphabet as sequences of complexity $n + s$.

One can also consider a generalization of Sturmian sequences over a three-letter alphabet, namely the *Arnoux-Rauzy sequences.* These are recurrent sequences defined over a three-letter alphabet with the following extra combinatorial property [42]: for every n, there is exactly one right special factor and one left special factor of length n (this condition is called the * condition in the seminal paper [42], and more generally in the literature on the subject).

Exercise 1.6.2. Prove that the Arnoux-Rauzy sequences are uniformly recurrent (*hint: prove that the right special factors appear with bounded gaps*).

One shall consider these sequences in Chap. 7 and in Chap. 12.