Automata and semigroups recognizing infinite words

Olivier Carton[1]
Dominique Perrin[2]
Jean-Éric Pin[1]

[1] LIAFA, CNRS and Université Denis Diderot Paris 7, Case 7014, 75205 Paris Cedex 13, France

[2] Institut Gaspard Monge, Université de Marne-la-Vallée, 5, boulevard Descartes, Champs-sur-Marne, F-77454 Marne-la-Vallée Cedex 2

Olivier.Carton@liafa.jussieu.fr, perrin@univ-mlv.fr,
Jean-Eric.Pin@liafa.jussieu.fr

Abstract

This paper is a survey on the algebraic approach to the theory of automata accepting infinite words. We discuss the various acceptance modes (Büchi automata, Muller automata, transition automata, weak recognition by a finite semigroup, ω-semigroups) and prove their equivalence. We also give two algebraic proofs of McNaughton's theorem on the equivalence between Büchi and Muller automata. Finally, we present some recent work on prophetic automata and discuss its extension to transfinite words.

1 Introduction

Among the many research contributions of Wolfgang Thomas, those regarding automata on infinite words and more generally, on infinite objects, have been highly inspiring to the authors. In particular, we would like to emphasize the historical importance of his early papers [33, 34, 35], his illuminating surveys [36, 37] and the Lecture Notes volume on games and automata [15].

Besides being a source of inspiration, Wolfgang always had nice words for our own research on the *algebraic approach* to automata theory. This survey, which presents this theory for infinite words, owes much to his encouragement.

Büchi has extended the classical theory of languages to infinite words instead of finite ones. Most notions and results known for finite words extend to infinite words, often at the price of more difficult proofs. For example, proving that rational languages are closed under Boolean operations becomes, in the infinite case, a delicate result, the proof of which makes use of Ramsey theorem. In the same way, the determinization of automata, an easy algorithm on finite words, turns to a difficult theorem in the infinite case.

Not surprisingly, the same kind of obstacle occurred in the algebraic approach to automata theory. It was soon recognized that finite automata are closely linked with finite semigroups, thus giving an algebraic counterpart of the definition of recognizability by finite automata. In this setting, every rational language X of A^+ is recognized by a morphism from A^+ onto a finite semigroup. There is also a minimal semigroup recognizing X, called the syntactic semigroup of X. The success of the algebraic approach for studying regular languages was already firmly established by the end of the seventies, but it took another ten years to find the appropriate framework for infinite words. Semigroups are replaced by ω-semigroups, which are, roughly speaking, semigroups equipped with an infinite product. In this new setting, the definitions of recognizable sets of infinite words and of syntactic congruence become natural and most results valid for finite words can be adapted to infinite words. Carrying on the work of Arnold [1], Pécuchet [21, 20] and the second author [22, 23, 24], Wilke [38, 39] has pushed the analogy with the theory for finite words sufficiently far to obtain a counterpart of Eilenberg's variety theorem for finite or infinite words. This theory was further extended by using ordered ω-semigroups [27, 25]. Notwithstanding the importance of the variety theory, we do not cover it in this article but rather choose to present some applications of the algebraic approach to automata theory. The first nontrivial application is the construction of a Muller automaton, given a finite semigroup weakly recognizing a language. The second one is a purely algebraic proof of the theorem of McNaughton stating that any recognizable subset of infinite words is a Boolean combination of deterministic recognizable sets. The third one deals with prophetic automata, a subclass of Büchi automata in which any infinite word is the label of exactly one final path. The main result states that these automata are equivalent to Büchi automata. We show, however, that this result does not extend to words indexed by ordinals.

Our paper has the character of a survey. For the reader's convenience it reproduces some of the material published in the book *Semigroups and automata on infinite words* [26], which owes a debt of gratitude to Wolfgang Thomas. Proofs are often only sketched in the present paper, but complete proofs can be found in [26]. Other surveys on automata and infinite words include [24, 25, 36, 37, 32].

Our article is divided into seven sections. Automata on infinite words are introduced in Section 2. Algebraic recognition modes are discussed in Section 3. The syntactic congruence is defined in Section 4. In Section 5, we show that all recognition modes defined so far are equivalent. Sections 6 and 7 illustrate the power of the algebraic approach. In Section 6, we give an algebraic proof of McNaughton's theorem. Section 7 is devoted to prophetic automata.

2 Automata

Let A be an alphabet. We denote by A^+, A^* and A^ω, respectively, the sets of nonempty finite words, finite words and infinite words on the alphabet A. We also denote by A^∞ the set $A^* \cup A^\omega$ of finite or infinite words on A. By definition, an *ω-rational* subset of A^ω is a finite union of sets of the form XY^ω where X and Y are rational subsets of A^*.

An *automaton* is given by a finite alphabet A, a finite set of states Q and a subset E of $Q \times A \times Q$, called the set of *edges* or *transitions*. Two transitions (p, a, q) and (p', a', q') are called *consecutive* if $q = p'$. An *infinite path* in the automaton $\mathcal{A}$ is an infinite sequence p of consecutive transitions

$$p: \quad q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \quad \cdots$$

The state q_0 is the *origin* of the infinite path and the infinite word $a_0a_1\cdots$ is its *label*. We say that the path p *passes infinitely often* through a state q (or that p *visits* q infinitely often, or yet that q is *infinitely repeated* in p) if there are infinitely many integers n such that $q_n = q$. The set of *infinitely repeated* states in p is denoted by $\mathrm{Inf}(p)$.

An automaton $\mathcal{A} = (Q, A, E)$ is said to have *deterministic transitions*, if, for every state $q \in Q$ and every letter $a \in A$, there is at most one state q' such that (q, a, q') is a transition. It is *deterministic* if it has deterministic transitions and if I is a singleton. Dually, $\mathcal{A}$ has *complete transitions* if, for every state $q \in Q$ and every letter $a \in A$, there is at least one state q' such that (q, a, q') is a transition.

Acceptance modes are usually defined by specifying a set of *successful* finite or infinite paths. This gives rise to different types of automata. We shall only recall here the definition of two classes: the Büchi automata and the Muller automata.

2.1 Büchi automata

In the model introduced by Büchi, one is given a set of *initial states* I and a set of *final states* F. Here are the precise definitions.

Let $\mathcal{A} = (Q, A, E, I, F)$ be a Büchi automaton. We say that an infinite path in $\mathcal{A}$ is *initial* if its origin is in I and *final* if it visits F infinitely often. It is *successful* if it is initial and final. The set of infinite words *recognized* by $\mathcal{A}$ is the set, denoted by $L^\omega(\mathcal{A})$, of labels of infinite successful paths in $\mathcal{A}$. It is also the set of labels of infinite initial paths p in $\mathcal{A}$ and such that $\mathrm{Inf}(p) \cap F \neq \varnothing$.

By definition, a set of infinite words is *recognizable* if it is recognized by some finite Büchi automaton. Büchi has shown that Kleene's theorem on regular languages extends to infinite words.

Theorem 2.1. A set of infinite words is recognizable if and only if it is ω-rational.

The notion of trim automaton can also be adapted to the case of infinite words. A state q is called *accessible* if there is a (possibly empty) finite initial path in $\mathcal{A}$ ending in q. A state q is called *coaccessible* if there exists an infinite final path starting at q. Finally, $\mathcal{A}$ is *trim* if all its states are both accessible and coaccessible.

It is easy to see that every Büchi automaton is equivalent to a trim Büchi automaton. For this reason, we shall assume that all the automata considered in this paper are trim.

So far, extending automata theory to infinite words did not raise any insuperable problems. However, it starts getting harder when it comes to determinism.

The description of the subsets of A^ω recognized by deterministic Büchi automata involves a new operator. For a subset L of A^*, let

$$\overrightarrow{L} = \{u \in A^\omega \mid u \text{ has infinitely many prefixes in } L\}.$$

Example 2.2.

(a) If $L = a^*b$, then $\overrightarrow{L} = \varnothing$.

(b) If $L = (ab)^+$, then $\overrightarrow{L} = (ab)^\omega$.

(c) If $L = (a^*b)^+ = (a+b)^*b$, that is if L is the set of words ending with b, then $\overrightarrow{L} = (a^*b)^\omega$, which is the set of infinite words containing infinitely many occurrences of b.

The following example shows that not every set of words can be written in the form $\overrightarrow{L}$.

Example 2.3. The set $X = (a+b)^*a^\omega$ of words with a *finite* number of occurrences of b is not of the form $\overrightarrow{L}$. Otherwise, the word ba^ω would have a prefix $u_1 = ba^{n_1}$ in L, the word $ba^{n_1}ba^\omega$ would have a prefix $u_2 = ba^{n_1}ba^{n_2}$ in L, etc. and the infinite word $u = ba^{n_1}ba^{n_2}ba^{n_3}\cdots$ would have an infinity of prefixes in L and hence would be in $\overrightarrow{L}$. This is impossible, since u contains infinitely many b's.

A set of infinite words which can be recognized by a deterministic Büchi automaton is called *deterministic*.

Theorem 2.4. A subset X of A^ω is deterministic if and only if there exists a recognizable set L of A^+ such that $X = \overrightarrow{L}$.

2.2 Muller automata

Contrary to the case of finite words, deterministic Büchi automata fail to recognize all recognizable sets of infinite words. This is the motivation for introducing Muller automata which are also deterministic, but have a more powerful acceptance mode. In this model, an infinite path p is final if the set $\mathrm{Inf}(p)$ belongs to a prescribed set $\mathcal{T}$ of sets of states. The definition of initial and successful paths are unchanged.

A *Muller automaton* is a 5-tuple $\mathcal{A} = (Q, A, E, i, \mathcal{T})$ where (Q, A, E) is a deterministic automaton, i is the initial state and $\mathcal{T}$ is a set of subsets of Q, called the *table of states* of the automaton. The set of infinite words recognized by $\mathcal{A}$ is the set, denoted by $L^\omega(\mathcal{A})$, of labels of infinite successful paths in $\mathcal{A}$.

A fundamental result, due to R. McNaughton [18], states that any Büchi automaton is equivalent to a Muller automaton.

Theorem 2.5. Any recognizable set of infinite words can be recognized by a Muller automaton.

This implies in particular that recognizable sets of infinite words are closed under complementation, a result proved for the first time by Büchi in a direct way.

2.3 Transition automata

It is sometimes convenient to use a variant of automata in which a set of final transitions is specified, instead of the usual set of final states. This idea can be applied to all variants of automata.

Formally, a *Büchi transition automaton* is a 5-tuple $\mathcal{A} = (Q, A, E, I, F)$ where (Q, A, E) is an automaton, $I \subseteq Q$ is the set of initial states and $F \subseteq E$ is the set of final transitions. If p is an infinite path, we denote by $\mathrm{Inf}_T(p)$ the set of transitions through which p goes infinitely often. A path p is final if it goes through F infinitely often, that is, if $\mathrm{Inf}_T(p) \cap F \neq \varnothing$.

Similarly, a *transition Muller automaton* is a 5-tuple $\mathcal{A} = (Q, A, E, I, \mathcal{T})$ where (Q, A, E) is a finite deterministic automaton, i is the initial state and $\mathcal{T}$ is a set of subsets of E, called the *table of transitions* of the automaton. A path is final if $\mathrm{Inf}_T(p) \in \mathcal{T}$, that is, if the set of transitions occurring infinitely often in p is an element of the table.

Proposition 2.6.

(1) Büchi automata and transition Büchi automata are equivalent.

(2) Muller automata and transition Muller automata are equivalent.

3 Algebraic recognition modes

In this section, we give an historical survey on the various algebraic notions of recognizability that have been considered. The two earlier ones, weak and strong recognition, are now superseded by the notions of ω-semigroupsand Wilke algebras.

Recall that a *semigroup* is a set equipped with an associative operation which does not necessarily admit an identity. If S is a semigroup, S^1 denotes the monoid equal to S if S is a monoid, and to $S \cup \{1\}$ if S is not a monoid. In the latter case, the operation of S is completed by the rules $1s = s1 = s$ for each $s \in S^1$. An element e of S is *idempotent* if $e^2 = e$.

The preorder $\leqslant_{\mathcal{R}}$ is defined on S by setting $s \leqslant_{\mathcal{R}} s'$ if there exists $t \in S^1$ such that $s = s't$. We also write $s \mathrel{\mathcal{R}} s'$ if $s \leqslant_{\mathcal{R}} s'$ and $s' \leqslant_{\mathcal{R}} s$ and $s <_{\mathcal{R}} s'$ if $s \leqslant_{\mathcal{R}} s'$ and $s' \not\leqslant_{\mathcal{R}} s$. The equivalence classes of the relation $\mathcal{R}$ are called the $\mathcal{R}$-*classes* of S.

3.1 Weak recognition

The early attempts aimed at understanding the behaviour of a semigroup morphism from A^+ onto a finite semigroup. The key result is a consequence of Ramsey's theorem in combinatorics, which involves the notion of a linked pair: a *linked pair* of a finite semigroup S is a pair (s, e) of elements of S satisfying $se = s$ and $e^2 = e$.

Theorem 3.1. Let $\varphi : A^+ \to S$ be a morphism from A^+ into a finite semigroup S. For each infinite word $u \in A^\omega$, there exist a linked pair (s, e) of S and a factorization $u = u_0 u_1 \cdots$ of u as a product of words of A^+ such that $\varphi(u_0) = s$ and $\varphi(u_n) = e$ for all $n > 0$.

Theorem 3.1 is frequently used in a slightly different form:

Proposition 3.2. Let $\varphi : A^+ \to S$ be a morphism from A^+ into a finite semigroup S. Let u be an infinite word of A^ω, and let $u = u_0 u_1 \ldots$ be a factorisation of u in words of A^+. Then there exist a linked pair (s, e) of S and a strictly increasing sequence of integers $(k_n)_{n \geqslant 0}$ such that $\varphi(u_0 u_1 \cdots u_{k_0 - 1}) = s$ and $\varphi(u_{k_n} u_{k_n + 1} \cdots u_{k_{n+1} - 1}) = e$ for every $n \geqslant 0$.

Theorem 3.1 lead to the first attempt to extend the notion of recognizable sets. Let us call φ-*simple* a set of infinite words of the form $\varphi^{-1}(s) \left(\varphi^{-1}(e)\right)^\omega$, where (s, e) is a linked pair of S. Then we say that a subset of A^ω is *weakly recognized* by φ if it is a finite union of φ-simple subsets. The following result justifies the term "recognized".

Proposition 3.3. A set of infinite words is recognizable if and only if it is weakly recognized by some morphism onto a finite semigroup.

However, the notion of weak recognition has several drawbacks: there is no natural notion of syntactic semigroup, dealing with complementation is uneasy and more generally, the algebraic tools that were present in the case of finite words are missing.

3.2 Strong recognition

This notion emerged as an attempt to obtain an algebraic proof of the closure of recognizable sets of infinite words under complement.

Let $\varphi : A^+ \to S$ be a morphism from A^+ into a finite semigroup S. Then φ *strongly recognizes* (or *saturates*) a subset X of A^ω if all the φ-simple sets have a trivial intersection with X, that is, for each linked pair (s, e) of S,

$$\varphi^{-1}(s)\left(\varphi^{-1}(e)\right)^\omega \cap X = \varnothing \text{ or } \varphi^{-1}(s)\left(\varphi^{-1}(e)\right)^\omega \subseteq X$$

Theorem 3.1 shows that A^ω is a finite union of φ-simple sets. It follows that if a morphism strongly recognizes a set of infinite words, then it also weakly recognizes it. Furthermore, Proposition 3.3 can be improved.

Proposition 3.4. A set of infinite words is recognizable if and only if it is strongly recognized by some morphism onto a finite semigroup.

The proof relies on a construction which is interesting on its own right. Given a semigroup S, we define a new semigroup

$$T = \{\left(\begin{smallmatrix} s & P \\ 0 & s \end{smallmatrix}\right) \mid \ s \in S,\ P \text{ is a subset of } S \times S\}$$

with multiplication defined by

$$\left(\begin{smallmatrix} s & P \\ 0 & s \end{smallmatrix}\right)\left(\begin{smallmatrix} t & Q \\ 0 & t \end{smallmatrix}\right) = \left(\begin{smallmatrix} st & sQ \cup Pt \\ 0 & st \end{smallmatrix}\right)$$

where $sQ = \{(sq_1, q_2) \mid (q_1, q_2) \in Q\}$ and $Pt = \{(p_1, p_2t) \mid (p_1, p_2) \in P\}$. Let now φ be a morphism from A^+ onto S. Then one can show that the map $\psi : A^+ \to T$ defined by

$$\psi(u) = \begin{pmatrix} \varphi(u) & \tau(u) \\ 0 & \varphi(u) \end{pmatrix} \quad \text{with} \quad \tau(u) = \{(\varphi(u_1), \varphi(u_2)) \mid u = u_1u_2\}$$

is a semigroup morphism and that any set of infinite words weakly recognized by φ is strongly recognized by ψ.

Proposition 3.4 leads to a simple proof of Büchi's complementation theorem.

Corollary 3.5. Recognizable sets of infinite words are closed under complement.

Proof. Indeed, if a morphism strongly recognizes a set of infinite words, it also recognizes its complement. Q.E.D.

3.3 ω-semigroups and Wilke algebras

Although strong recognition constituted an improvement over weak recognition, there were still obstacles to extend to infinite words Eilenberg's variety theorem, which gives a correspondence between recognizable sets and finite semigroups. The solution was found by Wilke [38] and reformulated in slightly different terms by the two last authors in [25]. The idea is to use an algebraic structure, called an ω-semigroup, which is a sort of semigroup in which infinite products are defined. This structure was actually implicit in the original construction of Büchi to recognize the complement [7].

3.3.1 ω-semigroups

An *ω-semigroup* is a two-sorted algebra $S = (S_+, S_\omega)$ equipped with the following operations:

(a) A binary operation defined on S_+ and denoted multiplicatively,
(b) A mapping $S_+ \times S_\omega \to S_\omega$, called *mixed product*, that associates with each pair $(s, t) \in S_+ \times S_\omega$ an element of S_ω denoted st,
(c) A surjective mapping $\pi : S_+^\omega \to S_\omega$, called *infinite product*

These three operations satisfy the following properties:

(1) S_+, equipped with the binary operation, is a semigroup,
(2) for every $s, t \in S_+$ and for every $u \in S_\omega$, $s(tu) = (st)u$,
(3) for every increasing sequence $(k_n)_{n>0}$ and for every sequence $(s_n)_{n\geqslant 0}$ of elements of S_+,

$$\pi(s_0 s_1 \ \cdots \ s_{k_1-1}, s_{k_1} s_{k_1+1} \ \cdots \ s_{k_2-1}, \ \ldots) = \pi(s_0, s_1, s_2, \ldots)$$

(4) for every $s \in S_+$ and for every sequence $(s_n)_{n\geqslant 0}$ of elements of S_+,

$$s\pi(s_0, s_1, s_2, \ldots) = \pi(s, s_0, s_1, s_2, \ldots)$$

These conditions can be thought of as an extension of associativity. In particular, conditions (3) and (4) show that one can replace $\pi(s_0, s_1, s_2, \ldots)$ by $s_0 s_1 s_2 \cdots$ without ambiguity. We shall use this simplified notation in the sequel.

Example 3.6.

(1) We denote by A^∞ the ω-semigroup (A^+, A^ω) equipped with the usual concatenation product. One can show that A^∞ is the free ω-semigroup generated by A.
(2) The *trivial* ω-semigroup is the ω-semigroup $1 = (\{1\}, \{a\})$, obtained by equipping the trivial semigroup $\{1\}$ with an infinite product: the unique way is to declare that every infinite product is equal to a.
(3) Consider the ω-semigroup $S = (\{0, 1\}, \{a\})$ defined as follows: every infinite product is equal to a and every finite product $s_0 s_1 \ldots s_n$ is

equal to 0 except if all the s_i's are equal to 1. In particular, the elements 0 and 1 are idempotents and thus, for all $n > 0$, $1^n \neq 0^n$. Nevertheless $1^\omega = 0^\omega = a$.

These examples, especially the third one, make apparent an algorithmic problem. Even if the sets S_+ and S_∞ are finite, the infinite product is still an operation of infinite arity and it is not clear how to define it as a finite object. The problem was solved by Wilke [38], who proved that finite ω-semigroups are totally determined by only three operations of finite arity. This leads to the notion of Wilke algebras, that we now define.

3.3.2 Wilke Algebras

A *Wilke algebra* is a two-sorted algebra $S = (S_+, S_\omega)$, equipped with the following operations:

(1) an associative product on S_+,

(2) a mixed product, which maps each pair $(s, t) \in S_+ \times S_\omega$ onto an element of S_ω denoted by st, such that, for every $s, t \in S_+$ and for every $u \in S_\omega$, $s(tu) = (st)u$,

(3) a map from S_+ in S_ω, denoted by $s \to s^\omega$ satisfying, for each $s, t \in S_+$,

$$\begin{aligned} s(ts)^\omega &= (st)^\omega \\ (s^n)^\omega &= s^\omega \quad \text{for each } n > 0 \end{aligned}$$

and such that every element of S_ω can be written as st^ω with $s, t \in S_+$.

Wilke's theorem states the equivalence between finite Wilke algebra and finite ω-semigroup. A consequence is that for a finite ω-semigroup, any infinite product is equal to an element of the form st^ω, with $s, t \in S_+$.

Theorem 3.7. Every finite Wilke algebra $S = (S_+, S_\omega)$ can be equipped, in a unique way, with a structure of ω-semigroup that inherits the given mixed product and such that, for each $s \in S_+$, the infinite product $sss\cdots$ is equal to s^ω.

We still need to define morphisms for these algebras. We shall just give the definition for ω-semigroups, but the definition for Wilke algebras would be similar.

3.3.3 Morphisms of ω-semigroups

As ω-semigroups are two-sorted algebras, morphisms are defined as pairs of morphisms. Given two ω-semigroups $S = (S_+, S_\omega)$ and $T = (T_+, T_\omega)$, a *morphism of ω-semigroups* S is a pair $\varphi = (\varphi_+, \varphi_\omega)$ consisting of a semigroup morphism $\varphi_+ : S_+ \to T_+$ and of a mapping $\varphi_\omega : S_\omega \to T_\omega$ preserving the infinite product: for every sequence $(s_n)_{n\in\mathbb{N}}$ of elements of S_+,

$$\varphi_\omega(s_0 s_1 s_2 \cdots) = \varphi_+(s_0)\varphi_+(s_1)\varphi_+(s_2)\cdots$$

It is an easy exercise to verify that these conditions imply that φ also preserves the mixed product, that is, for all $s \in S_+$, and for each $t \in S_\omega$,

$$\varphi_+(s)\varphi_\omega(t) = \varphi_\omega(st)$$

Algebraic concepts like isomorphism, ω-subsemigroup, congruence, quotient, division are easily adapted from semigroups to ω-semigroups. We are now ready for our algebraic version of recognizability.

3.3.4 Recognition by morphism of ω-semigroups.

In the context of ω-semigroups, it is more natural to define recognizable subsets of A^∞, although we shall mainly use this definition for subsets of A^ω. This global point of view has been confirmed to be the right one in the study of words indexed by ordinals or by linear orders [3, 4, 5, 6, 29]. Thus a subset X of A^∞ is split into two components $X_+ = X \cap A^+$ and $X_\omega = X \cap A^\omega$.

Let $S = (S_+, S_\omega)$ be a finite ω-semigroup, and let $\varphi : A^\infty \to S$ be a morphism. We say that φ *recognizes* a subset X of A^∞ if there exist a pair $P = (P_+, P_\omega)$ with $P_+ \subseteq S_+$ and $P_\omega \subseteq S_\omega$ such that $X_+ = \varphi_+^{-1}(P_+)$ and $X_\omega = \varphi_\omega^{-1}(P_\omega)$. In the sequel, we shall often omit the subscripts and simply write $X = \varphi^{-1}(P)$. It is time again to justify our terminology by a theorem, whose proof will be given in Section 5.

Theorem 3.8. A set of infinite words is recognizable if and only if it is recognized by some morphism onto a finite ω-semigroup.

Example 3.9. Let $A = \{a, b\}$, and consider the ω-semigroup

$$S = (\{1, 0\}, \{1^\omega, 0^\omega\})$$

equipped with the operations $11 = 1$, $10 = 01 = 00 = 0$, $11^\omega = 1^\omega$, $10^\omega = 00^\omega = 01^\omega = 0^\omega$. Let $\varphi : A^\infty \to S$ be the morphism of ω-semigroups defined by $\varphi(a) = 1$ and $\varphi(b) = 0$. We have

$\varphi^{-1}(1) = a^+$ (finite words containing no occurrence of b),

$\varphi^{-1}(0) = A^*bA^*$ (finite words containing at least one occurrence of b),

$\varphi^{-1}(1^\omega) = a^\omega$ (infinite words containing no occurrence of b),

$\varphi^{-1}(0^\omega) = A^\omega \setminus a^\omega$ (infinite words containing at least one occurrence of b),

The morphism φ recognizes each of these sets, as well as any union of these sets.

Example 3.10. Let us take the same ω-semigroup S and consider the morphism of ω-semigroups $\varphi : A^\infty \to S$ defined by $\varphi(a) = s$ for each $a \in A$. We have $\varphi^{-1}(s) = A^+$, $\varphi^{-1}(t) = \varnothing$ and $\varphi^{-1}(u) = A^\omega$. Thus the morphism φ recognizes the empty set and the sets A^+, A^ω and A^∞.

Example 3.11. Let $A = \{a, b\}$, and consider the ω-semigroup

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

equipped with the following operations:

$$\begin{array}{llll} aa = a & ab = a & aa^\omega = a^\omega & ab^\omega = a^\omega \\ ba = b & bb = b & ba^\omega = b^\omega & bb^\omega = b^\omega \end{array}$$

The morphism of ω-semigroups $\varphi : A^\infty \to S$ defined by $\varphi(a) = a$ and $\varphi(b) = b$ recognizes aA^ω since we have $\varphi^{-1}(a^\omega) = aA^\omega$.

Boolean operations can be easily translated in terms of morphisms. Let us start with a result which allows us to treat separately, the subsets of A_+ and those of A_ω.

Proposition 3.12. Let φ be a morphism of ω-semigroups recognizing a subset X of A^∞. Then the subsets X_+, X_ω, $X_+ \cup A^\omega$ and $A^+ \cup X_\omega$ are also recognized by φ.

We now consider the complement.

Proposition 3.13. Let φ be a morphism of ω-semigroups recognizing a subset X of A^∞ (resp. A_+, A_ω). Then φ also recognizes the complement of X in A^∞ (resp. A_+, A_ω).

For union and intersection, we have the following results.

Proposition 3.14. Let $(\varphi_i)_{i\in F} : A^\infty \to S_i$ be a family of surjective morphisms recognizing a subset X_i of A^∞. Then the subsets $\bigcup_{i\in F} X_i$ and $\bigcap_{i\in F} X_i$ are recognized by an ω-subsemigroup of the product $\prod_{i\in F} S_i$.

In the same spirit, the following properties hold:

Proposition 3.15. Let $\alpha : A^\infty \to B^\infty$ be a morphism of ω-semigroups and let φ be a morphism of ω-semigroups recognizing a subset X of B^∞. Then the morphism $\varphi \circ \alpha$ recognizes the set $\alpha^{-1}(X)$.

4 Syntactic congruence

The definition of the syntactic congruence of a recognizable subset of infinite words is due to Arnold [1]. It was then adapted to the context of ω-semigroups. Therefore, this definition can be given for recognizable subsets of A^∞, but we restrict ourself to the case of subsets of A^ω.

The *syntactic congruence* of a recognizable subset of A^ω is defined on A^+ by $u \sim_X v$ if and only if, for each $x, y \in A^*$ and for each $z \in A^+$,

$$\begin{aligned} xuyz^\omega \in X &\iff xvyz^\omega \in X \\ x(uy)^\omega \in X &\iff x(vy)^\omega \in X \end{aligned} \tag{4.1}$$

and on A^ω by $u \sim_X v$ if and only if, for each $x \in A^*$,

$$xu \in X \Longleftrightarrow xv \in X \tag{4.2}$$

The syntactic ω-semigroup of X is the quotient of A^∞ by the syntactic congruence of X.

Example 4.1. Let $A = \{a, b\}$ and $X = \{a^\omega\}$. The syntactic congruence of X divides A^+ into two classes: a^+ and A^*bA^* and A^ω into two classes also: A^*bA^ω and a^ω. The syntactic ω-semigroup of X is the four element ω-semigroup of Example 3.9.

Example 4.2. Let $A = \{a, b\}$ and let $X = aA^\omega$. The syntactic ω-semigroup of X is the ω-semigroup of Example 3.11.

Example 4.3. When X is not recognizable, the equivalence relation $\sim$ defined on A^+ by (4.1) and on A^ω by (4.2) is not in general a congruence. For instance, let $A = \{a, b\}$ and $X = \{ba^1ba^2ba^3b\cdots\}$. We have, for each $n > 0$, $b \sim_X ba^n$, but nevertheless $ba^1ba^2ba^3b\cdots$ is not equivalent to b^ω since $ba^1ba^2ba^3b\cdots \in X$ but $b^\omega \notin X$.

Example 4.4. Let $X = (a\{b, c\}^* \cup \{b\})^\omega$. We shall compute in Example 5.3 an ω-semigroup S recognizing this set. One can show that its syntactic ω-semigroup is $S(X) = (\{a, b, c, ca\}, \{a^\omega, c^\omega, (ca)^\omega\})$, presented by the relations

$$\begin{array}{ccccc}
a^2 = a & ab = a & ac = a & ba = a & b^2 = b \\
bc = c & cb = c & c^2 = c & b^\omega = a^\omega & ba^\omega = a^\omega \\
ac^\omega = c^\omega & ca^\omega = (ca)^\omega & a(ca)^\omega = a^\omega & b(ca)^\omega = (ca)^\omega & c(ca)^\omega = (ca)^\omega
\end{array}$$

The syntactic ω-semigroup is the least ω-semigroup recognizing a recognizable set. More precisely, we have the following statement:

Proposition 4.5. Let X be a recognizable subset of A^∞. An ω-semigroup S recognizes X if and only if the syntactic ω-semigroup of X is a quotient of S.

Note in particular that, if $u \sim_X v$ for two words u, v of A^+, then, for all $x \in A^*$ and $z \in A^\omega$

$$xuz \in X \Longleftrightarrow xvz \in X \tag{4.3}$$

Indeed, if $\varphi : A^\infty \to S$ denotes the syntactic morphism of X, the condition $u \sim_X v$ implies $\varphi(u) = \varphi(v)$. It follows that $\varphi(xuz) = \varphi(xvz)$, which gives (4.3).

5 Conversions from one acceptance mode into one another

In this section, we explain how to convert the various acceptance modes one into one another. We have already seen how to pass from weak to strong recognition by a finite semigroup. We shall now describe, in order, the conversions form weak recognition to Büchi automata, from Büchi automata to ω-semigroups, from strong recognition to ω-semigroups and finally from weak recognition to Muller automata.

5.1 From weak recognition to Büchi automata

Let $\varphi : A^+ \to S$ be a morphism from A^+ onto a finite semigroup S. First observe that, given Büchi automata $\mathcal{A}_1, \ldots, \mathcal{A}_n$, their disjoint union recognizes the set $L^\omega(\mathcal{A}_1) \cup \ldots \cup L^\omega(\mathcal{A}_n)$. Therefore, we may suppose that X is a φ-simple set of infinite words, say $X = \varphi^{-1}(s)(\varphi^{-1}(e))^\omega$ for some linked pair (s, e) of S. We construct a nondeterministic Büchi automaton $\mathcal{A}$ that accepts X as follows. The set Q of states of $\mathcal{A}$ is the set $S^I = S \cup \{f\}$ where f is a new neutral element added to S even if S has already one. The product of S is thus extended to S^I by setting $tf = ft = t$ for any $t \in S^I$. The initial state of $\mathcal{A}$ is s and the unique final state is f. The set of transitions is

$$E = \{\varphi(a)t \xrightarrow{a} t \mid a \in A \text{ and } t \in Q\} \\ \cup \{f \xrightarrow{a} t \mid a \in A, t \in Q \text{ and } \varphi(a)t = e\}.$$

Let $t \in S$. It is easily proved that a word w satisfies $\varphi(w) = t$ if and only if it labels a path from t to f visiting f only at the end. It follows that w labels a path from f to f if and only if $\varphi(w) = e$ and thus $\mathcal{A}$ accepts X.

The previous construction has one main drawback. The transition semigroup of the automaton $\mathcal{A}$ may not belong to the variety of finite semigroups generated by S, as shown by the following example.

Example 5.1. Let S be the semigroup $\{0, 1\}$ endowed with the usual multiplication. Let A be the alphabet $\{a, b\}$ and $\varphi : A^+ \to S$ be the morphism defined by $\varphi(a) = 0$ and $\varphi(b) = 1$. Let (s, e) be the pair $(0, 0)$. The set $\varphi^{-1}(s)(\varphi^{-1}(e))^\omega$ is thus equal to $(b^*a)^\omega$. The automaton $\mathcal{A}$ obtained with the previous construction is pictured in Figure 1. The semigroup S is commutative but the transition semigroup of $\mathcal{A}$ is not. Indeed, there is a path from 1 to 0 labeled by ba but there is no path from 1 to 0 labeled by ab.

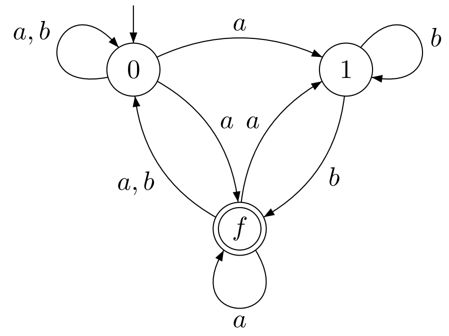


FIGURE 1. The automaton $\mathcal{A}$.

In order to solve this problem, Pécuchet [21] proposed the following construction, which is quite similar to the previous one but has better properties. The set of states of the automaton is still the set $S^I = S \cup \{f\}$. The initial state is s and the unique final state is f. The set E of transitions is modified as follows:

$$E = \{t' \xrightarrow{a} t \mid a \in A, t, t' \in Q \text{ and } (t' = \varphi(a)t \text{ or } t'e = \varphi(a)t)\}$$

The automaton $\mathcal{B}$ obtained with this construction is pictured in Figure 2.

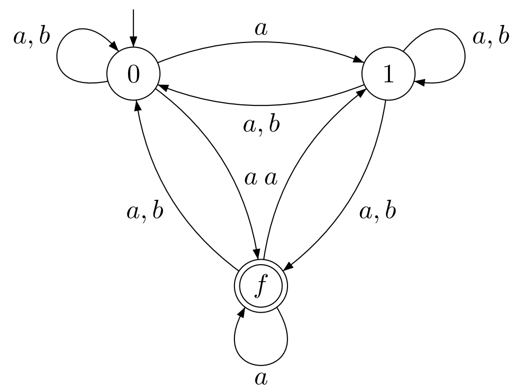


FIGURE 2. The automaton $\mathcal{B}$.

It can be proved that for any states t and t', there is a path from t' to t labeled by w if and only if $t' = \varphi(w)t$ or $t'e = \varphi(w)t$. It follows that if two words w and w' satisfy $\varphi(w) = \varphi(w')$, there is path from t' to t labeled by w if and only if there is path from t' to t labeled by w'. This means that the transition semigroup of the automaton $\mathcal{B}$ divides the semigroup S and hence belongs to the variety of finite semigroups generated by S.

5.2 From Büchi automata to ω-semigroups

Let $\mathcal{A} = (Q, A, E, I, F)$ be a Büchi automaton recognizing a subset X of A^ω. The idea is the following. Given a finite word u and two states p and q, we define a *multiplicity* expressing the following possibilities for the set P of paths from p to q labeled by u:

(1) P is empty,

(2) P is nonempty, but contains no path visiting a final state,

(3) P contains a path visiting a final state.

Our construction makes use of the semiring $\mathbb{K} = \{-\infty, 0, 1\}$ in which addition is the maximum for the ordering $-\infty < 0 < 1$ and multiplication, which extends the Boolean addition, is given in Table 1. Conditions (1), (2) and (3) will be encoded by $-\infty$, 0 and 1, respectively.

	$-\infty$	0	1
$-\infty$	$-\infty$	$-\infty$	$-\infty$
0	$-\infty$	0	1
1	$-\infty$	1	1

Table 1. The multiplication table.

Formally, we associate with each finite word u a $(Q \times Q)$-matrix $\mu(u)$ with entries in $\mathbb{K}$ defined by

$$\mu(u)_{p,q} = \begin{cases} -\infty & \text{in case (1),} \\ 0 & \text{in case (2),} \\ 1 & \text{in case (3)} \end{cases}$$

It is easy to see that μ is a morphism from A^+ into the multiplicative semigroup of $Q \times Q$-matrices with entries in $\mathbb{K}$. Let $S_+ = \mu(A^+)$.

The next step is to complete our structure of Wilke algebra by defining an appropriate set S_ω, an ω-power and a mixed product. The solution consists in coding infinite paths by column matrices of $\mathbb{K}^Q$, in such a way that each coefficient $\mu(u)_p$ codes the existence of an infinite path of label u starting at p.

The usual product of matrices induces a mixed product $\mathbb{K}^{Q\times Q} \times \mathbb{K}^Q \to \mathbb{K}^Q$. In order to define the operation ω on square matrices, we need the following definition. Given a matrix s of S_+, we call *infinite s-path starting at p* a sequence $p = p_0, p_1, \ldots$ of states such that, for all i, $s_{p_i,p_{i+1}} \neq -\infty$.

An s-path is said to be *successful* if $s_{p_i,p_{i+1}} = 1$ for an infinite number of indices i. We define the column matrix s^ω as follows. For every $p \in Q$,

$$s^\omega_p = \begin{cases} 1 & \text{if there exists a successful } s\text{-path of origin } p, \\ -\infty & \text{otherwise} \end{cases}$$

Note that the coefficients of this matrix can be effectively computed. Indeed, computing s^ω_p amounts to checking the existence of circuits containing a given edge in a finite graph.

Finally, S_ω is the set of all column matrices of the form st^ω, with $s, t \in S_+$. One can verify that $S = (S_+, S_\omega)$, equipped with these operations, is a Wilke algebra. Further, the morphism μ can be extended in a unique way as a morphism of ω-semigroups from A^∞ into S which recognizes the set $L^\omega(\mathcal{A})$.

Example 5.2. Let $\mathcal{A}$ be the Büchi automaton represented in Figure 3.

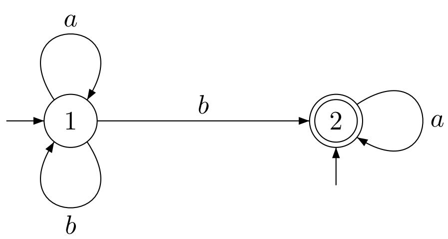


Figure 3. A Büchi automaton.

The morphism $\mu : A^\infty \to S(\mathcal{A})$ is defined by the formula

$$\mu(a) = \begin{pmatrix} 0 & -\infty \\ -\infty & 1 \end{pmatrix} \quad \text{and} \quad \mu(b) = \begin{pmatrix} 0 & 1 \\ -\infty & -\infty \end{pmatrix}$$

The ω-semigroup generated by these matrices contains five elements:

$$a = \begin{pmatrix} 0 & -\infty \\ -\infty & 1 \end{pmatrix} \quad b = \begin{pmatrix} 0 & 1 \\ -\infty & -\infty \end{pmatrix} \quad a^\omega = \begin{pmatrix} -\infty \\ 1 \end{pmatrix} \quad b^\omega = \begin{pmatrix} -\infty \\ -\infty \end{pmatrix} \quad ba^\omega = \begin{pmatrix} 1 \\ -\infty \end{pmatrix}$$

and is presented by the relations:

$$a^2 = a \quad ab = b \quad ba = b \quad b^2 = b \quad aa^\omega = a^\omega \quad ab^\omega = b^\omega \quad bb^\omega = b^\omega$$

Example 5.3. Let $X = (a\{b,c\}^* \cup \{b\})^\omega$. A Büchi automaton recognizing X is represented in Figure 4:

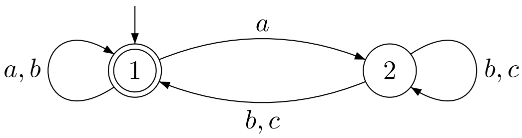


FIGURE 4. An automaton.

For this automaton, the previous computation provides an ω-semigroup with nine elements $S = (\{a, b, c, ba, ca\}, \{a^\omega, b^\omega, c^\omega, (ca)^\omega\})$, where

$$a = \begin{pmatrix} 1 & 1 \\ -\infty & -\infty \end{pmatrix} \qquad b = \begin{pmatrix} 1 & -\infty \\ 1 & 0 \end{pmatrix} \qquad c = \begin{pmatrix} -\infty & -\infty \\ 1 & 0 \end{pmatrix} \qquad ba = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$
$$ca = \begin{pmatrix} -\infty & -\infty \\ 1 & 1 \end{pmatrix}$$
$$a^\omega = \begin{pmatrix} 1 \\ -\infty \end{pmatrix} \qquad b^\omega = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad c^\omega = \begin{pmatrix} -\infty \\ -\infty \end{pmatrix} \qquad (ca)^\omega = \begin{pmatrix} -\infty \\ 1 \end{pmatrix}$$

It is presented by the following relations:

$$\begin{array}{ccccc} a^2 = a & ab = a & ac = a & b^2 = b & bc = c \\ cb = c & c^2 = c & (ba)^\omega = b^\omega & aa^\omega = a^\omega & ab^\omega = a^\omega \\ ac^\omega = c^\omega & a(ca)^\omega = a^\omega & ba^\omega = b^\omega & bb^\omega = b^\omega & bc^\omega = c^\omega \\ b(ca)^\omega = (ca)^\omega & ca^\omega = (ca)^\omega & cb^\omega = (ca)^\omega & cc^\omega = c^\omega & c(ca)^\omega = (ca)^\omega \end{array}$$

Note that the syntactic ω-semigroup $S(X)$ of X is not equal to S. To compute $S(X)$, one should first compute the image of X in S, which is $P = \{a^\omega, b^\omega\}$. Next, one should compute the syntactic congruence $\sim_P$ of P in S, which is defined on S_+ by $u \sim_P v$ if and only if, for every $x, y, z \in S_+$

$$\begin{aligned} xuyz^\omega \in P &\iff xvyz^\omega \in P \\ x(uy)^\omega \in P &\iff x(vy)^\omega \in P \end{aligned} \tag{5.4}$$

and on S^ω by $u \sim_P v$ if and only if, for each $x \in S_+$,

$$xu \in P \iff xv \in P \tag{5.5}$$

Here we get $a \sim_P ba$ and $a^\omega \sim_P b^\omega$ and hence we recovered the semigroup

$$S(X) = (\{a, b, c, ca\}, \{a^\omega, c^\omega, (ca)^\omega\})$$

presented in Example 4.4.

5.3 From strong recognition to ω-semigroups

It is easy to associate a Wilke algebra $\bar{S} = (S, S_\omega)$ to a finite semigroup S.

Let π be the exponent of S, that is, the smallest integer n such that s^n is idempotent for every $s \in S$. Two linked pairs (s, e) and (s', e') of S are said to be *conjugate* if there exist $x, y \in S^1$ such that $e = xy$, $e' = yx$ and $s' = sx$. These equalities also imply $s = s'y$ (since $s'y = sxy = se = s$), showing the symmetry of the definition. One can verify that the conjugacy relation is an equivalence relation on the set of linked pairs of S. We shall denote by $[s, e]$ the conjugacy class of a linked pair (s, e).

We take for S_ω the set of conjugacy classes of the linked pairs of S. One can prove that the set $\bar{S}$ is equipped with a structure of Wilke algebra by setting, for each $[s, e] \in S_\omega$ and $t \in S$,

$$t[s, e] = [ts, e] \quad \text{and} \quad t^\omega = [t^\pi, t^\pi]$$

The definition is consistent since if (s', e') is conjugate to (s, e), then (ts', e') is conjugate to (ts, e). It is now easy to convert strong recognition to recognition by an ω-semigroup.

Proposition 5.4. If a set of infinite words is strongly recognized by a finite semigroup S, then it is recognized by the ω-semigroup $\bar{S}$.

5.4 From weak recognition to Muller automata

The construction given by Le Saec, Pin and Weil [16, 17] permits to convert a semigroup that weakly recognizes a set of infinite words into a transition Muller automaton. It relies, however, on two difficult results of finite semigroup theory. Recall that a semigroup is *idempotent* if all its elements are idempotent and $\mathcal{R}$-*trivial* if the condition $s \mathrel{\mathcal{R}} t$ implies $s = t$.

The first one is a cover theorem also proved in [16, 17]. Recall that the right stabilizer of an element s of a semigroup S is the set of all $t \in S$ such that $st = s$. These stabilizers are themselves semigroups, and reflect rather well the structure of S: if S is a group, every stabilizer is trivial, but if S is has a zero, the stabilizer of the zero is equal to S. Here we consider an intermediate case: the stabilizers are idempotent and $\mathcal{R}$-trivial, which amounts to saying that, for each $s, t, u \in S$, the condition $s = st = su$ implies $t^2 = t$ and $tut = tu$. We can now state the cover theorem precisely.

Theorem 5.5. Each finite semigroup is the quotient of a finite semigroup in which the right stabilizers satisfy the identities $x = x^2$ and $xyx = xy$.

The second result we need is a property of path congruences due to I. Simon. A proof of this property can be found in [14]. Given an automaton $\mathcal{A}$, a path congruence is an equivalence relation on the set of finite paths of $\mathcal{A}$ satisfying the following conditions:

(1) any two equivalent paths are coterminal (that is, they have the same origin and the same end),
(2) if p and q are equivalent paths, and if r, p and s are consecutive paths, then rps is equivalent to rqs.

Proposition 5.6 (I. Simon)**.** Let $\sim$ be a path congruence such that, for every pair of loops p, q around the same state, $p^2 \sim p$ and $pq \sim qp$. Then two coterminal paths visiting the same sets of transitions are equivalent.

We are now ready to present our algorithm. Let X be a recognizable subset of A^ω and let $\varphi : A^+ \to S$ be a morphism weakly recognizing X. By Theorem 5.5, we may assume that the stabilizers of S satisfy the identities $x^2 = x$ and $xyx = xy$. Let S^1 be the monoid equal to S if S is a monoid and to $S \cup \{1\}$ if S is not a monoid.

One naturally associates a deterministic automaton $(S^1, A, \cdot)$ to φ by setting, for every $s \in S^1$ and every $a \in A$

$$s \cdot a = s\varphi(a).$$

Let s be a fixed state of S^1. Then every word u is the label of exactly one path with origin s, called the path with origin s defined by u.

Let $\mathcal{A} = (S^1, A, \cdot, 1, \mathcal{T})$ be the transition Muller automaton with 1 as initial state and such that

$$\mathcal{T} = \{\mathrm{Inf}_T(u) \mid u \in X\}.$$

We claim that $\mathcal{A}$ recognizes X. First, if $u \in X$, then $\mathrm{Inf}_T(u) \in \mathcal{T}$ by definition, and thus u is recognized by $\mathcal{A}$. Conversely, let u be an infinite word recognized by $\mathcal{A}$. Then

$$\mathrm{Inf}_T(u) = \mathrm{Inf}_T(v) = T \quad \text{for some } v \in X.$$

Thus, both paths u and v visit only finitely many times transitions out of T. Therefore, after a certain point, every transition of u (resp. v) belongs to T, and every transition of T is visited infinitely often. Consequently, one can find two factorizations $u = u_0u_1u_2\cdots$ and $v = v_0v_1v_2\cdots$ and a state $s \in S$ such that

(1) u_0 and v_0 define paths from 1 to s,

(2) for every $n > 0$, u_n and v_n define loops around s that visit at least once every transition in T and visit no other transition.

The situation is summarized in Figure 5 below

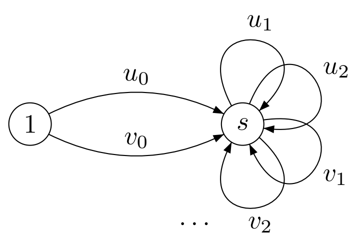


FIGURE 5.

Furthermore, Proposition 3.2 shows that, by grouping the u_i's (resp. v_i's) together, we may assume that

$$\varphi(u_1) = \varphi(u_2) = \varphi(u_3) = \dots \quad \text{and} \quad \varphi(v_1) = \varphi(v_2) = \varphi(v_3) = \dots$$

It follows in particular

$$u_0 v_1^\omega \in X \tag{5.6}$$

since $\varphi(u_0) = \varphi(v_0) = s$, $\varphi(v_1) = \varphi(v_2) = \dots$ and $v_0 v_1 v_2 \cdots \in X$. Furthermore,

$$u \in X \quad \text{if and only if} \quad u_0 u_1^\omega \in X \tag{5.7}$$

To simplify notation, we shall denote by the same letter a path and its label. We define a path equivalence $\sim$ as follows. Two paths p and q are equivalent if p and q are coterminal, and if, for every nonempty path x from 1 to the origin of p, and for every path r from the end of p to its origin, $x(pr)^\omega \in X$ if and only if $x(qr)^\omega \in X$.

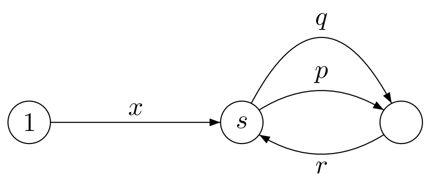


Figure 6.

Lemma 5.7. The equivalence $\sim$ is a path congruence such that, for every pair of loops p, q around the same state, $p^2 \sim p$ and $pq \sim qp$.

Proof. We first verify that $\sim$ is a congruence. Suppose that $p \sim q$ and let u and v be paths such that u, p and v are consecutive. Since $p \sim q$, p and q are coterminal, and thus upv and uqv are also coterminal. Furthermore, if x is a nonempty path from 1 to the origin of upv, and if r is a path from the end of upv to its origin such that $x(upvr)^\omega \in X$, then $(xu)(p(vru))^\omega \in X$, whence $(xu)(q(vru))^\omega \in X$ since $p \sim q$, and thus $x(uqvr)^\omega \in X$. Symmetrically, $x(uqvr)^\omega \in X$ implies $x(upvr)^\omega \in X$, showing that $upv \sim uqv$.

Next we show that if p is a loop around $s \in S$, then $p^2 \sim p$. Let x be a nonempty path from 1 to the origin of p, and let r be a path from the end of p to its origin. Then, since p is a loop, $\varphi(x)\varphi(p) = \varphi(x)$. Now since the stabilisers of S are idempotent semigroups, $\varphi(p) = \varphi(p^2)$ and thus $x(pr)^\omega \in X$ if and only if $x(p^2 r)^\omega \in X$ since φ recognizes X.

Finally, we show that if p and q are loops around the same state s, then $pq \sim qp$. Let, as before, x be a nonempty path from 1 to the origin of p, and let r be a path from the end of p to its origin. Then r is a loop around s. We first observe that

$$x(pq)^\omega \in X \iff x(qp)^\omega \in X \tag{5.8}$$

Indeed $x(pq)^\omega = xp(qp)^\omega$, and since p is a loop, $\varphi(x)\varphi(p) = \varphi(x)$. Thus $xp(qp)^\omega \in X$ if and only if $x(qp)^\omega \in X$, then proving (5.8). Now, we have the following sequence of equivalences

$$\begin{aligned} x(pqr)^\omega \in X &\iff x(pqrq)^\omega \in X \iff x(rqpq)^\omega \in X \\ &\iff x(rqp)^\omega \in X \iff x(qpr)^\omega \in X, \end{aligned}$$

where the second and fourth equivalences follow from (5.8) and the first and third from the identity $xyx = xy$ satisfied by the right stabilizer of $\varphi(x)$. Q.E.D.

We can now conclude the proof of Theorem 2.5. By assumption, the two loops around s defined by u_1 and v_1 visit exactly the same sets of transitions (namely T). Thus, by Lemma 5.7 and by Proposition 5.6, these two paths are equivalent. In particular, since $u_0 v_1^\omega \in X$ by (5.6), we have $u_0 u_1^\omega \in X$, and thus $u \in X$ by (5.7). Therefore $\mathcal{A}$ recognizes X.

6 An algebraic proof of McNaughton's theorem

McNaughton's celebrated theorem states that any recognizable subset of infinite words is a Boolean combination of deterministic recognizable sets. This Boolean combination can be explicitly computed using ω-semigroups. This proof relies on a few useful formulas of independent interest on deterministic sets. Note that McNaughton's theorem can be formulated as the equivalence of Büchi and Muller automata. Thus the construction described in Section 5.4 gives an alternative proof of McNaughton's theorem. Yet another proof is due to Safra [30]. It provides a direct construction leading to a reduced computational complexity.

Let S be a finite ω-semigroup and let $\varphi : A^\infty \to S$ be a surjective morphism recognizing a subset X of A^ω. Set, for each $s \in S_+$, $X_s = \varphi^{-1}(s)$. Finally, we denote by P the image of X in S and by $F(P)$ the set of linked pairs (s, e) of S_+ such that $se^\omega \in P$.

For each $s \in S_+$, the set $P_s = X_s \setminus X_s A^+$ is prefix-free, since a word of P_s cannot be, by definition, prefix of another word of P_s. Put

$$E_s = \{f \in S_+ \mid f^2 \text{ and } sf = s\} = \{f \in S_+ \mid (s, f) \text{ is a linked pair}\},$$

and denote by $\leqslant$ the relation on E_s defined by

$$g \leqslant e \text{ if and only if } eg = g.$$

It is the restriction to the set E_s of the preorder $\leqslant_{\mathcal{R}}$, since, if $g = ex$ then $eg = eex = ex = g$. We shall use the notation $e < g$ if $e \leqslant g$ and if $g \not\leqslant e$. To simplify notation, we shall suppose implicitly that for every expression of the form $X_s X_f^\omega$ or $X_s P_f$, the pair (s, f) is a linked pair of S_+.

Proposition 6.1. For each linked pair (s, e) of S_+, the following formula holds
$$X_s X_e^\omega \subset \overrightarrow{X_s P_e} \subset \bigcup_{f \leqslant e} X_s X_f^\omega. \tag{6.9}$$

Corollary 6.2.
(1) For every idempotent e of S_+, the following formula holds
$$X_e^\omega = \overrightarrow{X_e P_e}. \tag{6.10}$$
(2) For every linked pair (s, e) of S_+, we have
$$\bigcup_{f \leqslant e} X_s X_f^\omega = \bigcup_{f \leqslant e} \overrightarrow{X_s P_f}. \tag{6.11}$$

Proof. Formula (6.10) is obtained by applying (6.9) with $s = e$. Formula (6.11) follows by taking the union of both sides of (6.9) for $f \leqslant e$. Q.E.D.

The previous statement shows that a set of the form X_e^ω, with e idempotent, is always deterministic. This may lead the reader to the conjecture that every subset of the form X^ω, where X is a recognizable subset of A^+, is deterministic. However, this conjecture is ruined by the next example.

Example 6.3. Let $X = (a\{b, c\}^* \cup \{b\})^\omega$. The syntactic ω-semigroup of Y has been computed in Example 4.4. In this ω-semigroup, b is the identity, and all the elements are idempotent. The set X can be split into simple elements as follows:
$$\begin{aligned} X &= \varphi^{-1}(a)\varphi^{-1}(b)^\omega \cup \varphi^{-1}(a)^\omega \\ &= b^* a\{a, b, c\}^* b^\omega \cup (b^* a\{a, b, c\}^*)^\omega. \end{aligned}$$

It is possible to deduce from the previous formulas an explicit Boolean combination of deterministic sets.

Theorem 6.4. The following formula holds
$$X = \bigcup_{(s,e) \in F(P)} \ \bigcup_{f \mathcal{R} e} X_s X_f^\omega \tag{6.12}$$
and, for each $(s, e) \in F(P)$,
$$\bigcup_{f \mathcal{R} e} X_s X_f^\omega = (\overrightarrow{U}_{s,e} \setminus \overrightarrow{V}_{s,e}) \tag{6.13}$$
where $U_{s,e}$ and $V_{s,e}$ are the subsets of A^+ defined by:
$$U_{s,e} = \bigcup_{f \leqslant e} X_s P_f \quad \text{and} \quad V_{s,e} = \bigcup_{f < e} X_s P_f$$
In particular, X is a Boolean combination of deterministic sets.

For a proof, see [26, p. 120]. One can also obtain a characterization of the deterministic subsets.

Theorem 6.5. The set X is deterministic if and only if, for each linked pairs (s, e) and (s, f) of S_+ such that $f \leqslant e$, the condition $se^\omega \in P$ implies $sf^\omega \in P$. In this case

$$X = \bigcup_{(s,e) \in F(P)} \overrightarrow{X_s P_e} \tag{6.14}$$

For a proof, see [26, Theorem 9.4, p. 121].

Example 6.6. We return to Example 6.3. The image of X in its syntactic ω-semigroup is the set $P = \{a^\omega\}$. Now, the pairs (a, b) and (a, c) are linked pairs of S_+ since $ab = ac = a$ and we have $c \leqslant b$ since $bc = c$. But $ab^\omega = a^\omega \in P$, and $ac^\omega = c^\omega \notin P$. Therefore X is not deterministic.

The proof of McNaughton's theorem described above is due to Schützenberger [31]. It is related to the proof given by Rabin [28] and improved by Choueka [12]. See [26, p. 72] for more details.

7 Prophetic automata

In this section, we introduce a new type of automata, called prophetic, because in some sense, all the information concerning the future is encoded in the initial state. We first need to make precise a few notions on Büchi automata.

7.1 More on Büchi automata

There are two competing versions for the notions of determinism and co-determinism for a trim automaton. In the first version, the notions are purely local and are defined by a property of the transitions set. They give rise to the notions of automaton with deterministic or co-deterministic transitions introduced in Section 2. The second version is global: a trim automaton is *deterministic* if it has exactly one initial state and if every word is the label of at most one initial path. Similarly, a trim automaton is *co-deterministic* if every word is the label of at most one final path.

The local and global notions of determinism are equivalent. The local and global notions of co-determinism are also equivalent for finite words. However, for infinite words, the global version is strictly stronger than the local one.

Lemma 7.1. A trim Büchi automata is deterministic if and only if it has exactly one initial state and if its transitions are deterministic. Further, if a trim Büchi automata is co-deterministic, then its transitions are co-deterministic.

The notions of complete and co-complete Büchi automata are also global notions. A trim Büchi automata is *complete* if every word is the label of at least one initial path. It is *co-complete* if every word is the label of at least one final path.

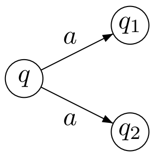
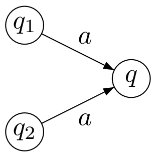
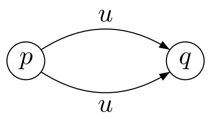
<table>
<tr><th>Det. transitions</th><th>Co-det. transitions</th><th>Unambiguous</th></tr>
<tr><td>Forbidden configuration:

where a is a letter.</td><td>Forbidden configuration:

where a is a letter.</td><td>Forbidden configuration:

where u is a word.</td></tr>
<tr><th>Deterministic</th><th>Co-deterministic</th><th>Unambiguous</th></tr>
<tr><td>Two initial paths with the same label are equal + exactly one initial state</td><td>Two final paths with the same label are equal</td><td rowspan="3">Two successful paths with the same label are equal</td></tr>
<tr><th>Complete</th><th>Co-complete</th></tr>
<tr><td>Every word is the label of some initial path</td><td>Every word is the label of some final path</td></tr>
</table>

Table 2. Summary of the definitions.

Unambiguity is another global notion. A Büchi automaton $\mathcal{A}$ is said to be *ω-unambiguous* if every infinite word in is the label of at most one successful path. It is clear that any deterministic or co-deterministic Büchi automaton is ω-unambiguous, but the converse is not true. The various terms are summarized in Table 2.

7.2 Prophetic automata

By definition, a *prophetic automaton* is a co-deterministic, co-complete Büchi automaton. Equivalently, a Büchi automaton is prophetic if every word is the label of exactly one final path. Therefore, a word is accepted if the unique final path it defines is also initial. The main result of this section shows that prophetic and Büchi automata are equivalent.

Theorem 7.2. Any recognizable set of infinite words can be recognized by a prophetic automaton.

It was already proved independently in [19] and [2] that any recognizable set of infinite words is recognized by a codeterministic automaton, but the construction given in [2] does not provide unambiguous automata.

Prophetic automata recognize infinite words, but the construction can be adapted to biinfinite words. Two unambiguous automata on infinite words can be merged to make an unambiguous automaton on biinfinite words. This leads to an extension of McNaughton's theorem to the case of biinfinite words. See [26, Section 9.5] for more details.

Theorem 7.2 was originally formulated by Michel in the eighties but remained unpublished for a long time. Another proof was found by the first author and the two proofs were finally published in [10, 11]. Our presentation follows the proof which is based on ω-semigroups.

We start with a simple characterization.

Proposition 7.3. Let $\mathcal{A} = (Q, A, E, I, F)$ be a Büchi (resp. transition Büchi) automaton and let, for each $q \in Q$, $L_q = L^\omega(Q, A, E, q, F)$.

(1) $\mathcal{A}$ is co-deterministic if and only if the L_q's are pairwise disjoint.

(2) $\mathcal{A}$ is co-complete if and only if $\cup_{q\in Q} L_q = A^\omega$.

Proof. (1) If $\mathcal{A}$ is co-deterministic, the L_q's are clearly pairwise disjoint. Suppose that the L_q's are pairwise disjoint and let $p_0 \xrightarrow{a_0} p_1 \xrightarrow{a_1} p_2 \ \cdots$ and $q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \ \cdots$ be two infinite paths with the same label $u = a_0 a_1 \cdots$. Then, for each $i \geqslant 0$, $a_i a_{i+1} \cdots \in L(p_i) \cap L(q_i)$, and thus $p_i = q_i$. Thus $\mathcal{A}$ is co-deterministic.

(2) follows immediately from the definition of co-complete automata.

Q.E.D.

Example 7.4. A prophetic automaton is presented in Figure 7. The corresponding partition of A^ω is the following:

$L_0 = A^* b a^\omega$	(at least one, but finitely many b)
$L_1 = a^\omega$	(no b)
$L_2 = a(A^* b)^\omega$	(first letter a, infinitely many b)
$L_3 = b(A^* b)^\omega$	(first letter b, infinitely many b)

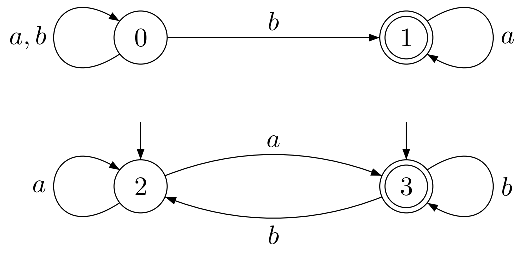


Figure 7. A prophetic automaton.

Example 7.5. Another example, recognizing the set $A^*(ab)^\omega$, is presented in Figure 8.

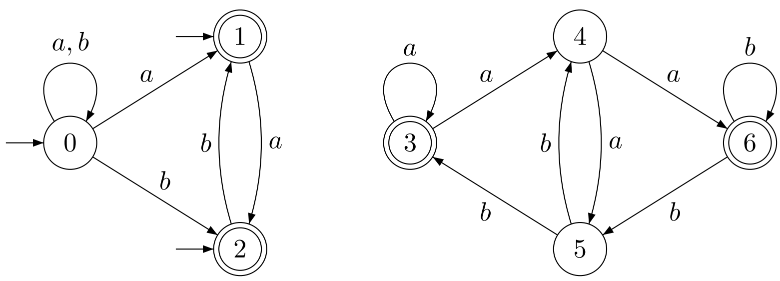


Figure 8. A prophetic automaton recognizing $A^*(ab)^\omega$.

Complementation becomes easy with prophetic automata.

Proposition 7.6. Let $\mathcal{A} = (Q, A, E, I, F)$ be a prophetic automaton recognizing a subset X of A^ω. Then the Büchi automaton $(Q, A, E, Q \setminus I, F)$ recognizes the complement of X.

It is easier to prove Theorem 7.2 for a variant of prophetic automata that we now define. A *prophetic transition automaton* is a co-deterministic, co-complete, transition automaton. Proposition 2.6 states that Büchi automata and transition Büchi automata are equivalent. It is not difficult to adapt this result to prophetic automata [26, Proposition I.8.1].

Proposition 7.7. Prophetic and transition prophetic automata are equivalent.

Thus Theorem 7.2 can be reformulated as follows.

Theorem 7.8. Any recognizable set of infinite words can be recognized by a prophetic transition automaton.

Proof. Let X be a recognizable subset of A^ω, let $\varphi : A^\infty \to S$ be the syntactic morphism of X and let $P = \varphi(X)$. Our construction strongly relies on the properties of $\geqslant_{\mathcal{R}}$-chains of the semigroup S_+ and requires a few preliminaries.

We shall denote by R the set of all nonempty $>_{\mathcal{R}}$-chains of S_+:

$$R = \{(s_0, s_1, \ldots, s_n) \mid n \geqslant 0,\ s_0, \ldots, s_n \in S \text{ and } s_0 >_{\mathcal{R}} s_1 >_{\mathcal{R}} \cdots >_{\mathcal{R}} s_n\}$$

In order to convert a $\geqslant_{\mathcal{R}}$-chain into a strict $>_{\mathcal{R}}$-chain, we introduce the reduction ρ, defined inductively as follows

$$\rho(s) = (s)$$

$$\rho(s_1, \ldots, s_n) = \begin{cases} \rho(s_1, \ldots, s_{n-1}) & \text{if } s_n \mathrel{\mathcal{R}} s_{n-1} \\ (\rho(s_1, \ldots, s_{n-1}), s_n) & \text{if } s_{n-1} >_{\mathcal{R}} s_n \end{cases}$$

In particular, for each finite word $u = a_0 a_1 \cdots a_n$ (where the a_i's are letters), let $\hat\varphi(u)$ be the $>_{\mathcal{R}}$-chain $\rho(s_0, s_1, \ldots, s_n)$, where $s_i = \varphi(a_0 a_1 \cdots a_i)$ for $0 \leqslant i \leqslant n$. The definition of $\hat\varphi$ can be extended to infinite words. Indeed, if $u = a_0 a_1 \cdots$ is an infinite word,

$$s_0 \geqslant_{\mathcal{R}} s_1 \geqslant_{\mathcal{R}} s_2 \ldots$$

and since S_+ is finite, there exists an integer n, such that, for all $i, j \geqslant n$, $s_i \mathrel{\mathcal{R}} s_j$. Then we set $\hat\varphi(u) = \hat\varphi(a_0 \ldots a_n)$.

Define a map from $A \times S_+^1$ into S_+^1 by setting, for each $a \in A$ and $s \in S_+^1$,

$$a \cdot s = \varphi(a)s$$

We extend this map to a map from $A \times R$ into R by setting, for each $a \in A$ and $(s_1, \ldots, s_n) \in R$,

$$a \cdot (s_1, \ldots, s_n) = \rho(a \cdot 1, a \cdot s_1, \ldots, a \cdot s_n)$$

To extend this map to A^+, it suffices to apply the following induction rule, where $u \in A^+$ and $a \in A$

$$(ua) \cdot (s_1, \ldots, s_n) = u \cdot (a \cdot (s_1, \ldots, s_n))$$

This defines an action of the semigroup A^+ on the set R in the sense that, for all $u, v \in A^*$ and $r \in R$,

$$(uv) \cdot r = u(v \cdot r)$$

The connections between this action, φ and $\hat\varphi$ are summarized in the next lemma.

Lemma 7.9. The following formulas hold:

(1) For each $u \in A^+$ and $v \in A^\omega$, $u \cdot \varphi(v) = \varphi(uv)$

(2) For each $u, v \in A^+$, $u \cdot \hat{\varphi}(v) = \hat{\varphi}(uv)$

Proof. (1) follows directly from the definition of the action and it suffices to establish (2) when u reduces to a single letter a. Let $v = a_0 a_1 \dots a_n$, where the a_i's are letters and let, for $0 \leqslant i \leqslant n$, $s_i = \varphi(a_0 a_1 \dots a_i)$. Then, by definition, $\hat{\varphi}(v) = \rho(s_0, \dots, s_n)$ and since, the relation $\geqslant_{\mathcal{R}}$ is stable on the left,

$$a \cdot \hat{\varphi}(v) = \rho(a \cdot 1, a \cdot s_0, a \cdot s_1, \dots, a \cdot s_n) = \hat{\varphi}(av)$$

which gives (2). Q.E.D.

We now define a transition Büchi automaton $\mathcal{A} = (Q, A, E, I, F)$ by setting

$$Q = \Big\{ \big((s_1, \dots, s_n), se^\omega\big) \mid (s_1, \dots, s_n) \in R, \\ (s, e) \text{ is a linked pair of } S_+ \text{ and } s_n \mathrel{\mathcal{R}} s \Big\}$$

$$I = \Big\{ \big((s_1, \dots, s_n), se^\omega\big) \in Q \mid se^\omega \in P \Big\}$$

$$E = \Big\{ \Big(\big(a \cdot (s_1, \dots, s_n), a \cdot se^\omega\big), a, \big((s_1, \dots, s_n), se^\omega\big) \Big) \\ \mid a \in A \text{ and } \big((s_1, \dots, s_n), se^\omega\big) \in Q \Big\}$$

A transition $\Big(\big(a \cdot (s_1, \dots, s_n), a \cdot se^\omega\big), a, \big((s_1, \dots, s_n), se^\omega\big) \Big)$ is said to be *cutting* if the last two elements of the $\geqslant_{\mathcal{R}}$-chain $(a \cdot 1, a \cdot s_1, \dots, a \cdot s_n)$ are $\mathcal{R}$-equivalent.

We choose for F the set of cutting transitions of the form

$$\Big(\big(a \cdot (s_1, \dots, s_n), a \cdot e^\omega\big), a, \big((s_1, \dots, s_n), e^\omega\big) \Big)$$

where e is an idempotent of S_+ such that $s_n \mathrel{\mathcal{R}} e$.

Note that $\mathcal{A}$ has co-deterministic transitions. A typical transition is shown in Figure 9.

$$\big(a \cdot (s_1, \dots, s_n), a \cdot se^\omega\big) \xrightarrow{\quad a \quad} \big((s_1, \dots, s_n), se^\omega\big)$$

FIGURE 9. A transition of $\mathcal{A}$.

The first part of the proof consists in proving that every infinite word is the label of a final path. Let $u = a_0 a_1 \cdots$ be an infinite word, and let, for each

$i \geqslant 0$, $x_i = a_i a_{i+1} \cdots$ and $q_i = \big(\hat\varphi(x_i), \varphi(x_i)\big)$. Each q_i is a state of Q, and Lemma 7.9 shows that

$$p = q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \;\; \cdots$$

is a path of $\mathcal{A}$.

Lemma 7.10. The path p is final.

Proof. Let $(u_i)_{i\geqslant 0}$ be a factorization of u associated with the linked pair (s, e). Then for each $i > 0$, $\varphi(u_i u_{i+1} \cdots) = e^\omega$. Fix some $i > 0$ and let $n_i = |u_0 u_1 \cdots u_i|$. Then $q_{n_i} = \big((s_1, \ldots, s_n), e^\omega\big)$ with $(s_1, \ldots, s_n) = \hat\varphi(u_{i+1} u_{i+2} \cdots)$. In particular, $s_n \mathrel{\mathcal{R}} e$ and hence $es_n = s_n$. Suppose first that $n \geqslant 2$. Then $\varphi(u_i)s_{n-1} = es_{n-1} \leqslant_{\mathcal{R}} e$ and $\varphi(u_i)s_n = es_n = s_n \mathrel{\mathcal{R}} e$. Therefore the relation $\varphi(u_i)s_{n-1} >_{\mathcal{R}} \varphi(u_i)s_n$ does not hold. If $n = 1$, the same argument works by replacing s_{n-1} by 1. It follows that in the path of label u_i from $q_{n_{i-1}}$ to q_{n_i}, at least one of the transitions is cutting. Thus p contains infinitely many cutting transitions and one can select one, say (q, a, q'), that occurs infinitely often. This gives a factorization of the form

$$p = q_0 \xrightarrow{x_0} q \xrightarrow{a} q' \xrightarrow{x_1} q \xrightarrow{a} q' \xrightarrow{x_2} \cdots$$

Up to taking a superfactorization, we can assume, by Proposition 3.2, that for some idempotent f, $\varphi(x_i a) = f$ for every $i > 0$. It follows that the second component of q' is $\varphi(x_i a x_{i+1} a \cdots) = f^\omega$ and thus the transition (q, a, q') is final, which proves the lemma. Q.E.D.

Furthermore, p is successful if and only if $\varphi(u) \in P$, or, equivalently, if $u \in X$. Thus $\mathcal{A}$ recognizes X and is co-complete. It just remains to prove that $\mathcal{A}$ is co-deterministic, which, by Proposition 7.3, will be a consequence of the following lemma.

Lemma 7.11. Any final path of label u starts at state $\big(\hat\varphi(u), \varphi(u)\big)$.

Proof. Let p be a final path of label u. Then some final transition, say (q, a, q'), occurs infinitely often in p. Highlighting this transition yields a factorization of p

$$q_0 \xrightarrow{v_0} q \xrightarrow{ⓐ} q' \xrightarrow{v_1} q \xrightarrow{ⓐ} q' \xrightarrow{v_2} \cdots$$

Let $q' = \big((s_1, \ldots, s_n), e^\omega\big)$, and consider a factor of the path p labelled by a word of the form $v = v_i a v_{i+1} a \cdots v_j a$, with $i > 0$ and $j - i \geqslant n$. By the choice of v, $q' = v \cdot q'$, and the first component of q' is obtained by reducing the $\geqslant_{\mathcal{R}}$-chain

$$\big(\varphi(v[0, 0]), \varphi(v[0, 1]), \ldots, \varphi(v), \varphi(v)s_1, \ldots, \varphi(v)s_n\big)$$

Now, since the cutting transition (q, a, q') occurs $n+1$ times in this factor, the last $n+1$ elements of this chain are $\mathcal{R}$-equivalent. It follows that the first component of q' is simply equal to $\hat{\varphi}(v)$.

Consider now a superfactorization $u = w_0 w_1 w_2 \cdots$ obtained by grouping the factors $v_i a$

$$u = \underbrace{(v_0 a \cdots v_{i_0 - 1} a)}_{w_0} \underbrace{(v_{i_0} a \cdots v_{i_1 - 1} a)}_{w_1} \underbrace{(v_{i_1} a \cdots v_{i_2 - 1} a)}_{w_2}$$

in such a way that, for some idempotent f, $\varphi(w_1) = \varphi(w_2) = \cdots = f$. We may also assume that $i_0 > 0$ and $i_1 - i_0 \geqslant n+1$. Thus $q' = w_1 \cdot q' = w_1 w_2 \cdot q' = \cdots$, and

$$(s_1, \cdots, s_n) = \hat{\varphi}(w_1) = \hat{\varphi}(w_1 w_2) = \cdots = \hat{\varphi}(w_1 w_2 \cdots)$$

It follows in particular $s_n \mathcal{R} \varphi(w_1) = f$. Furthermore, $s_n \mathcal{R} e$ since (q, a, q') is a final transition and thus $e \mathcal{R} f$. Therefore $e^\omega = f^\omega = \varphi(w_1 w_2 \cdots)$. Thus $q' = \big(\hat{\varphi}(w_1 w_2 \cdots), \varphi(w_1 w_2 \cdots)\big)$ and it follows from Lemma 7.9 that $q_0 = w_0 \cdot q' = \big(\hat{\varphi}(u), \varphi(u)\big)$. Q.E.D.

Q.E.D.

The construction given in the proof of Theorem 7.2 is illustrated in the following examples.

Example 7.12. Let $A = \{a, b\}$ and let $X = aA^\omega$. The syntactic ω-semigroup S of X, already computed in Example 4.2 is $S = (S_+, S_\infty)$ where $S_+ = \{a, b\}$, $S_\omega = \{a^\omega, b^\omega\}$, submitted to the following relations

$$\begin{array}{llll} aa = a & ab = a & aa^\omega = a^\omega & ab^\omega = a^\omega \\ ba = b & bb = b & ba^\omega = b^\omega & bb^\omega = b^\omega \end{array}$$

The syntactic morphism φ of X is defined by $\varphi(a) = a$ and $\varphi(b) = b$. The transition Büchi automaton associated with φ is shown in Figure 10. The final transitions are circled.

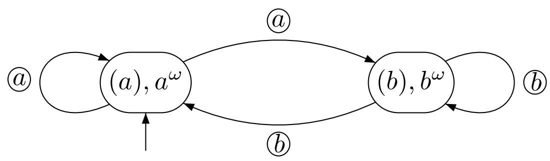


FIGURE 10. The transition Büchi automaton associated with φ.

Example 7.13. Let $A = \{a, b\}$ and let $X = (A^*a)^\omega$. The syntactic ω-semigroup S of X is $S = (S_+, S_\infty)$ where $S_+ = \{0, 1\}$, $S_\omega = \{0^\omega, 1^\omega\}$, submitted to the following relations

$$\begin{array}{llll} 1 \cdot 1 = 1 & 1 \cdot 0 = 0 & 10^\omega = 0^\omega & 11^\omega = 1^\omega \\ 0 \cdot 1 = 0 & 0 \cdot 0 = 0 & 00^\omega = 0^\omega & 01^\omega = 1^\omega \end{array}$$

The syntactic morphism φ of X is defined by $\varphi(a) = 0$ and $\varphi(b) = 1$. The transition Büchi automaton associated with φ is shown in Figure 11.

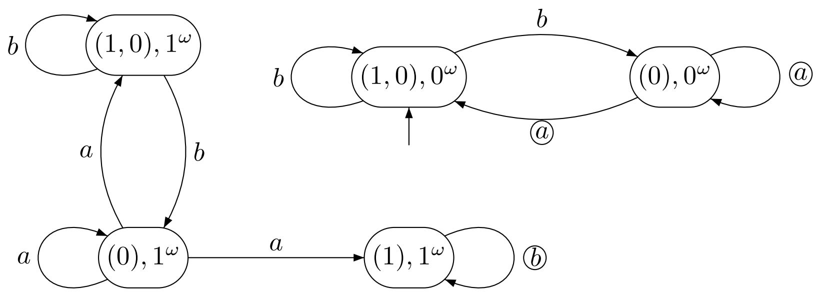


FIGURE 11. The transition Büchi automaton associated with φ.

7.3 Transfinite words

A natural extension to finite and infinite words is to consider words indexed by an ordinal, also called *transfinite word*. Automata on ordinals were introduced by Büchi [8, 9]. This leads to the notion of recognizable set of transfinite words. Subsequent work [3, 4, 5, 12, 13, 40] has shown that a number of results on infinite words can be extended to transfinite words (and even to words on linear orders [6, 29]).

An extension of the notion of ω-semigroup to countable ordinals was given in [3, 4, 5]. A further extension to countable linear orders is given in [6].

It is not difficult to extend the notion of prophetic automata to transfinite words. We show however that prophetic automata do not accept all recognizable sets of transfinite words.

First recall that an automaton on transfinite words is given by a finite set Q of states, sets I and F of initial and final states and a set E of transitions. Each transition is either a triple (p, a, q) where p and q are states and a is a letter or a pair (q, P) where where q is a state and P a subset of states. The former ones are called *successor transitions* and the latter ones *limit transitions*.

Let α be an ordinal. A *path* labeled by a word $x = (a_\beta)_{\beta<\alpha}$ of length α is a sequence $c = (q_\beta)_{\beta\leqslant\alpha}$ of states of length $\alpha + 1$ with the following properties.

(1) for each $\beta < \alpha$, the triple $(q_\beta, a_\beta, q_{\beta+1})$ is a successor transition of $\mathcal{A}$.

(2) for each limit ordinal $\beta \leqslant \alpha$, the pair $(\lim_\beta(c), c_\beta)$ is a limit transition of $\mathcal{A}$, where $\lim_\beta(c)$ is the set of states q such that, for each ordinal $\gamma < \beta$, there is an ordinal η such that $\gamma < \eta < \beta$ and $q = q_\eta$.

Note that since Q is finite, the set $\lim_\beta(c)$ is nonempty for each limit ordinal $\beta \leqslant \alpha$. A path $c = (q_\beta)_{\beta\leqslant\alpha}$ is initial if its first state q_0 is initial and it is final if its last state q_α is final. It is *accepting* if it is both initial and final. A word x is accepted if it is the label of an accepting path.

The notion of prophetic automaton can be readily adapted to transfinite words: an automaton is *prophetic* if any transfinite word is the label of exactly one final path. However, the next result shows that not every automaton is equivalent to a prophetic one.

Proposition 7.14. The set A^{ω^2} of words of length ω^2 cannot be accepted by a prophetic automaton.

Proof. Suppose there is a prophetic automaton $\mathcal{A}$ accepting the set A^{ω^2}. Since the word a^{ω^2} is accepted by $\mathcal{A}$, there is a unique successful path $c = (q_\beta)_{\beta\leqslant\omega^2}$ labeled by a^{ω^2}. In particular, q_0 is an initial state and q_{ω^2} is a final state. We claim that the word a^ω is also accepted by $\mathcal{A}$.

We first prove that $q_\beta = q_0$ for any $\beta < \omega^2$. The path $(q_\beta)_{1\leqslant\beta\leqslant\omega^2}$ is also a final path labeled by a^{ω^2}. It must therefore be equal to c. This shows that $q_n = q_0$ for any $n < \omega$. Similarly, the path $(q_\beta)_{\omega\leqslant\beta\leqslant\omega^2}$ is a final path labeled by a^{ω^2} and hence $q_\beta = q$ for any $\beta < \omega^2$. Since the set $\lim_{\omega^2}(c)$ is equal to $\{q_0\}$, the pair $(\{q_0\}, q_{\omega_2})$ must be a limit transition of $\mathcal{A}$. Thus the path $c' = (q'_\beta)_{\beta<\omega}$ defined by $q'_\beta = q_0$ if $\beta < \omega$ and $q'_\omega = q_{\omega^2}$ is a successful path labeled by a^ω. Q.E.D.

References

[1] A. Arnold, A syntactic congruence for rational ω-languages, *Theoret. Comput. Sci.* **39**,2-3 (1985), 333–335.

[2] D. Beauquier and D. Perrin, Codeterministic automata on infinite words, *Inform. Process. Lett.* **20**,2 (1985), 95–98.

[3] N. Bedon, Finite automata and ordinals, *Theoret. Comput. Sci.* **156**,1-2 (1996), 119–144.

[4] N. Bedon, Automata, semigroups and recognizability of words on ordinals, *Internat. J. Algebra Comput.* **8**,1 (1998), 1–21.

[5] N. Bedon and O. Carton, An Eilenberg theorem for words on countable ordinals, in *LATIN'98: theoretical informatics (Campinas, 1998)*, pp. 53–64, Springer, Berlin, 1998.

[6] V. Bruyère and O. Carton, Automata on linear orderings, in *MFCS'2001*, J. Sgall, A. Pultr and P. Kolman (ed.), pp. 236–247, *Lecture Notes in Comput. Sci.* vol. 2136, Springer Verlag, Berlin, Heidelberg, New York, 2001.

[7] J. R. Büchi, On a decision method in restricted second order arithmetic, in *Logic, Methodology and Philosophy of Science (Proc. 1960 Internat. Congr .)*, pp. 1–11, Stanford Univ. Press, Stanford, Calif., 1962.

[8] J. R. Büchi, Transfinite automata recursions and weak second order theory of ordinals, in *Logic, Methodology and Philos. Sci. (Proc. 1964 Internat. Congr.)*, pp. 3–23, North-Holland, Amsterdam, 1965.

[9] J. R. Büchi, The monadic second-order theory of ω_1, in *The Monadic Second-Order Theory of All Countable Ordinals*, J. R. Büchi and D. Siefkes (ed.), pp. 1–127, *Lecture Notes in Math.* vol. 328, Springer Verlag, Berlin, Heidelberg, New York, 1973.

[10] O. Carton and M. Michel, Unambiguous Büchi automata, in *LATIN'2000*, G. Gonnet, D. Panario and A. Viola (ed.), Berlin, 2000, pp. 407–416, *Lecture Notes in Comput. Sci.* vol. 1776, Springer.

[11] O. Carton and M. Michel, Unambiguous Büchi automata, *Theoret. Comput. Sci.* **297** (2003), 37–81.

[12] Y. Choueka, Theories of automata on ω-tapes: a simplified approach, *J. Comput. System Sci.* **8** (1974), 117–141.

[13] Y. Choueka, Finite automata, definable sets, and regular expressions over ω^n-tapes, *J. Comput. System Sci.* **17**,1 (1978), 81–97.

[14] S. Eilenberg, *Automata, languages, and machines. Vol. B*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976. With two chapters ("Depth decomposition theorem" and "Complexity of semigroups and morphisms") by Bret Tilson, Pure and Applied Mathematics, Vol. 59.

[15] E. Grädel, W. Thomas and T. Wilke (ed.), *Automata, logics, and infinite games*, *Lecture Notes in Computer Science* vol. 2500, Springer-Verlag, Berlin, 2002. A guide to current research.

[16] B. Le Saëc, J.-E. Pin and P. Weil, A purely algebraic proof of McNaughton's theorem on infinite words, in *Foundations of software technology and theoretical computer science (New Delhi, 1991)*, pp. 141–151, Springer, Berlin, 1991.

[17] B. Le Saëc, J.-E. Pin and P. Weil, Semigroups with idempotent stabilizers and applications to automata theory, *Internat. J. Algebra Comput.* **1**,3 (1991), 291–314.

[18] R. McNaughton, Testing and generating infinite sequences by a finite automaton, *Information and Control* **9** (1966), 521–530.

[19] A. W. Mostowski, Determinancy of sinking automata on infinite trees and inequalities between various Rabin's pair indices, *Inform. Process. Lett.* **15**,4 (1982), 159–163.

[20] J.-P. Pécuchet, Étude syntaxique des parties reconnaissables de mots infinis, in *Automata, languages and programming (Rennes, 1986)*, pp. 294–303, Springer, Berlin, 1986.

[21] J.-P. Pécuchet, Variétés de semigroupes et mots infinis, in *STACS 86*, B. Monien and G. Vidal-Naquet (ed.), pp. 180–191, *Lecture Notes in Comput. Sci.* vol. 210, Springer Verlag, Berlin, Heidelberg, New York, 1986.

[22] D. Perrin, Variétés de semigroupes et mots infinis, *C. R. Acad. Sci. Paris Sér. I Math.* **295**,10 (1982), 595–598.

[23] D. Perrin, Recent results on automata and infinite words, in *Mathematical foundations of computer science, 1984 (Prague, 1984)*, pp. 134–148, Springer, Berlin, 1984.

[24] D. Perrin, An introduction to finite automata on infinite words, in *Automata on infinite words (Le Mont-Dore, 1984)*, pp. 2–17, Springer, Berlin, 1985.

[25] D. Perrin and J.-E. Pin, Semigroups and automata on infinite words, in *Semigroups, formal languages and groups (York, 1993)*, pp. 49–72, Kluwer Acad. Publ., Dordrecht, 1995.

[26] D. Perrin and J.-E. Pin, *Infinite Words*, *Pure and Applied Mathematics* vol. 141, Elsevier, 2004. ISBN 0-12-532111-2.

[27] J.-E. Pin, Positive varieties and infinite words, in *Latin'98*, C. Lucchesi and A. Moura (ed.), pp. 76–87, *Lecture Notes in Comput. Sci.* vol. 1380, Springer Verlag, Berlin, Heidelberg, New York, 1998.

[28] M. O. Rabin, Decidability of second-order theories and automata on infinite trees., *Trans. Amer. Math. Soc.* **141** (1969), 1–35.

[29] C. Rispal and O. Carton, Complementation of rational sets on countable scattered linear orderings, *Internat. J. Found. Comput. Sci.* **16**,4 (2005), 767–786.

[30] S. Safra, On the complexity of the ω-automata, in *Proc. 29th Ann. IEEE Symp. on Foundations of Computer Science*, pp. 319–327, IEEE, 1988.

[31] M. Schützenberger, À propos des relations rationnelles fonctionnelles, in *Automata, languages and programming (Proc. Sympos., Rocquencourt, 1972)*, pp. 103–114, North Holland, Amsterdam, 1973.

[32] L. Staiger, ω-languages, in *Handbook of formal languages, Vol. 3*, pp. 339–387, Springer, Berlin, 1997.

[33] W. Thomas, Star-free regular sets of ω-sequences, *Inform. and Control* **42**,2 (1979), 148–156.

[34] W. Thomas, A combinatorial approach to the theory of ω-automata, *Inform. and Control* **48**,3 (1981), 261–283.

[35] W. Thomas, Classifying regular events in symbolic logic, *J. Comput. System Sci.* **25**,3 (1982), 360–376.

[36] W. Thomas, Automata on infinite objects, in *Handbook of Theoretical Computer Science*, J. van Leeuwen (ed.), vol. B, Formal models and semantics, pp. 135–191, Elsevier, 1990.

[37] W. Thomas, Languages, automata, and logic, in *Handbook of formal languages, Vol. 3*, pp. 389–455, Springer, Berlin, 1997.

[38] T. Wilke, An Eilenberg theorem for ∞-languages, in *Automata, Languages and Programming*, pp. 588–599, *Lecture Notes in Computer Sci.* vol. 510, Springer Verlag, Berlin, Heidelberg, New York, 1991.

[39] T. Wilke, An algebraic theory for regular languages of finite and infinite words, *Int. J. Alg. Comput.* **3** (1993), 447–489.

[40] J. Wojciechowski, The ordinals less than ω^ω are definable by finite automata, in *Algebra, combinatorics and logic in computer science, Vol. I, II (Győr, 1983)*, pp. 871–887, North-Holland, Amsterdam, 1986.