RECENT RESULTS ON AUTOMATA AND INFINITE WORDS

Dominique Perrin
L.I.T.P., Université Paris 7
2, Place Jussieu - 75251 Paris Cedex 05 - France

0. INTRODUCTION

The theory of automata on infinite words was initiated by Buchi and McNaughton at the beginning of the sixties, with a flavour of mathematical logic. As for the rest of automata theory, it uses concepts and methods which are familiar to computer scientists. The consideration of infinite words, which could in fact seem at first glance to be highly philosophical, has indeed a natural interpretation when one is wanting to study properties of systems that are not supposed to eventually stop, such as operating systems for instance. The use of automata on ordinary words has now become standard in several areas of algorithm design and I think that the same will happen in the next future with automata on infinite words.

From the mathematical point of view, the study of automata on infinite words is substantially more difficult than the ordinary one. The main problem is that, contrary to the ordinary case, it is not possible to restrict one's attention to deterministic automata. However, a number of intersting results have been obtained during the last years and the aim of this paper is to describe some of them.

The paper is divided into four sections. In the first one, I recall some of the basic definitions and results. Part of them can be found in Chapter XIV of Eilenberg's book [7], whose notation I follow. In the second section, I discuss the matter of star free sets of infinite words. A new result is proved there, according to which one can decide wether a given recognizable set of infinite words is star free. In the third section, an extension of the previous notions to two-sided infinite words is presented. This extension to a higher ordinal has two main features : (1) most results holding in the one sided case still hold in this case, though it is by no means a trivial extension ; (2) extensions to other ordinals such as the theory of infinite trees present quite different features (see [19] for instance, where it is proved that no equivalent of McNaughton's theorem holds on infinite trees). In the last section I present some results on unambiguity.

1. RECOGNIZABILITY

Let A be a (possibly infinite) set called the <u>alphabet</u>. We denote by $A^{\mathbb{N}}$ the set of all one-sided infinite words

$$\alpha = \alpha_0\alpha_1\alpha_2\ldots \qquad (\alpha_i \in A)$$

The set of ordinary (finite) words is denoted as usual by A^*. A morphism

$$\phi : A^* \to M$$

from A^* onto a finite monoid M is said to <u>recognize</u> a subset U of A^N if there exists a set $P \subset M \times M$ such that

$$U = \bigcup_{(m,n)\in P} \phi^{-1}(m)\,[\phi^{-1}(n)]^{\omega}$$

A set $U \subset A^N$ is said to be <u>recognizable</u> if there exists a morphism ϕ from A^* onto a finite monoid that recognizes U.

An equivalent definition of the family $Rec(A^N)$ of recognizable subsets of A^N is the family of sets of the form

$$U = \bigcup_{i=1}^{n} U_i V_i^{\omega}$$

for U_i, $V_i \in Rec(A^*)$.

A morphism $\phi : A^* \to M$ from A^* onto a finite monoid M is said to <u>saturate</u> a subset U of A^N if for any $m,n \in M$ one has either

$$\phi^{-1}(m)\,[\phi^{-1}(n)]^{\omega} \cap U = \emptyset$$

or

$$\phi^{-1}(m)\,[\phi^{-1}(n)]^{\omega} \subset U$$

Obviously, if ϕ saturates U, it also saturates its complement. More precisely, the family of sets saturated by ϕ is the boolean algebra of sets having trivial intersection with the sets $\phi^{-1}(m)\,[\phi^{-1}(n)]^{\omega}$.

In all above definitions, one may replace the pairs $(m,n) \in M\times M$ by a pair of an element m of M and an idempotent e of M since one has $[\phi^{-1}(n)]^{\omega}$

$= [\phi^{-1}(e)]^{\omega}$ when e is the idempotent which is a power of n. The following two propositions relate the notions of recognizability and saturation.

PROPOSITION 1.1 <u>Let ϕ be a morphism from A^* onto a finite monoid which saturates a set $U \subset A^N$. Then ϕ recognizes U and U is a union of sets of the form</u>

$$\phi^{-1}(m)\,[\phi^{-1}(e)]^{\omega}$$

<u>for</u> $m \in M$ <u>and</u> e <u>an idempotent of</u> M.

The above proposition is an easy consequence of the following lemma that we prove directly, although it can be viewed as a particular case of Ramsey's theorem.

LEMMA 1.2 <u>Let ϕ be a morphism from A^* onto a finite monoid M. For each $\alpha \in A^N$ there exists an $m \in M$ and an idempotent e in M such that</u>

$$\alpha \in \phi^{-1}(m)\,[\phi^{-1}(e)]^{\omega}$$

Proof : We proceed by induction on Card(M). Let

$$R = \{r \in M \mid rM = r\}$$

be the set of right zeroes of M. We first consider the case $\alpha \in [\phi^{-1}(R)]^{\omega}$. In this case, we select on $r \in R$ which occurs infinitely often, i.e. such that

$$\alpha = u_0 v_0 u_1 v_1 \ldots$$

with $\phi(v_i) = r$. Then $\phi(v_i u_i) = r$ and therefore $\alpha \in u_0 [\phi^{-1}(r)]^{\omega}$. Since r is an idempotent, the property is proved. Otherwise, one has $\alpha = u\beta$ where β has no left factor in $\phi^{-1}(R)$. Let $n \in M$ be such that β has an infinity of left factors in $\phi^{-1}(n)$, and let N be the monoid $N = \{t \in M \mid nt = n\}$. One has

$$\beta = v_0 v_1 v_2 \ldots$$

with $\phi(v_0 v_1 \ldots v_i) = n$ and therefore $\phi(v_i) \in N$ for all $i \geq 1$. We apply the induction hypothesis to the morphism ψ from B^* (where B is in bijection with the v_i) into N. Since $n \notin R$, one has $\mathrm{Card}(N) < \mathrm{Card}(M)$. Therefore, by the induction hypothesis, there exist $s \in N$ and an idempotent e in N such that $\beta \in \phi^{-1}(s)\, [\phi^{-1}(e)]^{\omega}$ which proves the property.□

The converse of Proposition 1.1 is not true in general as shown in the following example :

EXAMPLE 1.1 Let $\phi : \{a,b\}^* \to \mathbb{Z}/2$ be the morphism defined by

$$\phi(a) = 0, \ \phi(b) = 1$$

then $[\phi^{-1}(0)]^{\omega} = \{\alpha \in \{a,b\}^{N} \mid |\alpha|_b \text{ is even or infinite}\}$

$$\phi^{-1}(1)\, [\phi^{-1}(0)]^{\omega} = \{\alpha \in \{a,b\}^{N} \mid |\alpha|_b \text{ is odd infinite }\}$$

Thus ϕ recognizes the first set but it does not saturate it since the first set has a non trivial intersection with the second one. □

A set of the form $\phi^{-1}(m)\, [\phi^{-1}(e)]^{\omega}$ as in Proposition 1.1 is called a simple recognizable set. Accordingly, any recognizable set is a finite union of simple ones.

To prove a partial converse to Proposition 1.1, we recall the definition of the Schützenberger's product of two monoids M and N. It is the set

$$M \diamond N = \{(m,\rho,n) \mid m \in M,\ \rho \subset M \times N,\ n \in N\}$$

equipped with the product

$$(m,\rho,n)(m',\rho',n') = (mm',\ m\rho' \cup \rho n', n')$$

where $m\rho' = \{(mr',s') \mid (r',s') \in \rho'\}$ and $\rho n' = \{(r,sn') \mid (r,s) \in \rho\}$.

Let $\phi : A^* \to M$ and $\psi : A^* \to N$ be two morphisms. Then one denotes by

$$\phi \diamond \psi : A^* \to M \diamond N$$

the morphism defined for each $w \in A^*$ by

$$\phi \diamond \psi(w) = (\phi(w),\ \rho(w),\ \psi(w))$$

with

$$\rho(w) = \{(\phi(u),\ \psi(v)) \mid w = uv\}$$

The following result is implicitly contained in [16].

PROPOSITION 1.3 <u>Let</u> $\phi : A^* \to M$ <u>be a morphism from</u> A^* <u>into a finite monoid</u> M <u>and let</u> U <u>be a subset of</u> A^N. <u>If</u> ϕ <u>recognizes</u> U, <u>then</u> $\phi \diamond \phi$ <u>saturates</u> U.

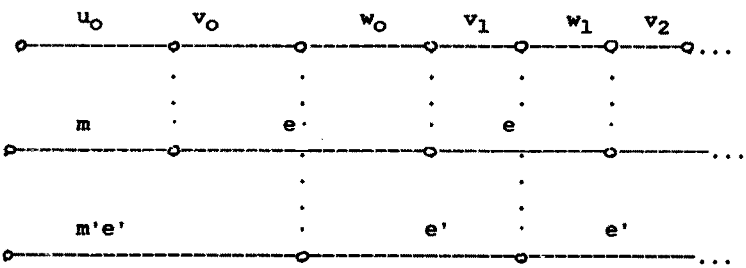
Proof : Let $M' = M \diamond M$ and $\phi' = \phi \diamond \phi$. It is enough to prove that ϕ' saturates each set U of the from $U = \phi^{-1}(m)\ [\phi^{-1}(e)]^\omega$ with $m \in M$ and $e = e^2 \in M$. Suppose that for $m' \in M'$ and $e' = e'^2 \in M'$ there exists a word α such that

$$\alpha \in \phi^{-1}(m)\ [\phi^{-1}(e)]^\omega \cap \phi'^{-1}(m')\ [\phi^{-1}(e')]^\omega$$

Let $m' = (m_1, \rho, m_1)$ and $e' = (e_1, \epsilon, e_1)$. Taking advantage of the fact that e and e' are idempotents, one may write

$$\alpha = u_o v_o w_o v_1 w_1 \ldots$$

with $\phi(u_o) = m$, $\phi(v_i w_i) = e$, $\phi'(u_o v_o) = m'e'$, $\phi'(w_i v_{i+1}) = e'$ (see figure below)



Let now β be any element of the form

$$\beta = r_o s_o s_1 s_2 \ldots$$

with $\phi'(r_o) = m'$, $\phi'(s_i) = e'$. By the definition of ϕ', one has

$$r_o s_o = x_o y_o, \quad s_i = z_i y_{i+1} \qquad (i \geqslant 0)$$

with $\phi(x_o) = \phi(u_o)$, $\phi(y_o) = \phi(v_o)$ and for all $i \geqslant 0$ $\phi(z_i) = \phi(w_i)$, $\phi(y_{i+1}) = \phi(v_{i+1})$. This implies

$$\phi(x_o) = m, \quad \phi(y_i z_i) = e$$

whence $\beta \in \phi^{-1}(m)\ [\phi^{-1}(e)]^\omega$. This proves that ϕ' saturates U. $\square$

EXAMPLE 1.2 Let us come back to the morphism $\phi : \{a,b\}^* \to Z/2$ of Example 1.1 The morphism $\phi' = \phi \diamond \phi$ sends A^* onto a monoid M' with three elements made up with an identity equal to $\phi'(a)$ and a two elements group $\{\phi'(b), \phi'(bb)\}$. In fact, one has

$$\phi'(a) = (0, (0,0),0),\ \phi'(b) = (1, (1,0)+(0,1),1), \phi'(bb) = \{0,(0,0)+(1,1),0\}.$$

The sets recognized by ϕ' are

$\phi'^{-1}(b)\,[\phi'^{-1}(a)]^\omega$: odd number of b's

$\phi'^{-1}(bb)\,[\phi'^{-1}(a)]^\omega$: even number of b's

$[\phi'^{-1}(bb)]^\omega$: infinite number of b's

and ϕ' therefore saturates all sets recognized by ϕ. □

We say that a congruence Θ on A^* saturates a set $U \subset A^N$ if the canonical morphism on the quotient saturates U. If Θ,Θ' are two congruences that saturate U, their upper bound $\Theta \vee \Theta'$ also saturates U. Therefore, for any set $U \subset A^N$ there exists a maximal congruence that saturates U. As for subsets of A^*, it is called the syntactic congruence of U. It has been introduced by A. Arnold in [2] who has proved that for a recognizable subset U of A^N it is characterized in the following way. Define for each word v two sets $\Gamma(v)$ and $\Delta(v)$ by

$$\Gamma(v) = \{(u,w,x) \in A^*xA^*xA^* \mid uvwx^\omega \in U\}$$

$$\Delta(v) = \{(u,x,y) \in A^*xA^*xA^* \mid u(xvy)^\omega \in U\}$$

Then v,v' are equivalent modulo the syntactic congruence iff $\Gamma(v) = \Gamma(v')$ and $\Delta(v) = \Delta(v')$.

The quotient of A^* by the syntactic congruence of U is of course called the syntactic monoid of U.

Instead of starting with a morphism recognizing a set $U \subset A^N$, one may start with a finite automaton

$$\mathcal{A} = (Q,I,T)$$

and consider the set $U \subset A^N$ of all the infinite words α which are the label of a path starting in I and going infinitely often through the set T of terminal states. This is often called a Büchi automaton and one says that it recognizes U. This is the usual definition of recognizability for subsets of A^N. It is of course equivalent to the definition given previously. One may indeed consider the morphism

$$\phi : A^* \to M$$

where M is the monoid of transition of the automaton $\mathcal{A}$. The image under ϕ of a word $w \in A^*$ is the set of all pairs $(p,q) \in QxQ$ such that there exists a path from p to q with label w. The classical construction performed to recognize the complement of U is the following : one considers the relation $\tau(w)$ defined as the set of pairs $(p,q) \in Q \times Q$ such that there exists a path from p to q with label w and which moreover use at least one element of T. One then considers for a word w the relation $\psi(w)$ over two copies of Q with matrix representation

$$\psi(w) = \begin{bmatrix} \phi(w) & \tau(w) \\ 0 & \phi(w) \end{bmatrix}$$

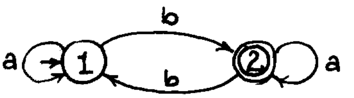
Then ψ is obviously a morphism and it is easy to verify that ψ saturates the set U recognized by the automaton A. In particular, ψ recognizes the complement of U.

It is worthwhile to observe that the morphism ψ can be obtained by representing the monoid $M' = M \diamond M$ by binary relations. In fact, let M be any monoid of relations over a set Q and let T be a subset of Q. For each element $m' = (m,\rho,n)$ of M' define a relation $\Theta(m')$ over two copies of Q defined by its matrix representation as

$$\Theta(m') = \begin{bmatrix} m & r \\ 0 & n \end{bmatrix}$$

where r is the relation over Q defined by $(p,q) \in r$ iff there exists a $t \in T$ and $(u,v) \in \rho$ such that $(p,t) \in u$ and $(t,q) \in v$. It is not difficult to verify that Θ is in fact a morphism.

EXAMPLE 1.3 Coming back again to the example concerning the parity of the number of occurrences of b, we now use the automaton shown below :



with $Q = \{1,2\}$, $I = \{1\}$, $T = \{2\}$. The above construction leads to the representation by binary relations defined by

$$\psi(a) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (b) = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

which is a faithful representation by relations of the monoid defined in Example 1.2. □

An other way of using automata on infinite words involves the notion of a Muller automaton. This is an automaton

$$A = (Q,I,\mathcal{T})$$

with $\mathcal{T}$ a family of subsets of Q. An infinite word $\alpha \in A^{\mathbb{N}}$ is recognized by the automaton if there is a path with label α such that the set of states appearing infinitely often is an element of the family $\mathcal{T}$. It is not difficult to show that this new notion of recognizability coincides with the previous ones.

A fundamental theorem, due to R. McNaughton, says that for each recognizable set $U \subset A^N$ there exists a <u>deterministic</u> Muller automaton recognizing U. This can be rephrased in the following way : for a set $X \subset A^*$, denote by

$$U = \vec{X}$$

the limit of X which is the set of infinite words $\alpha \in A^N$ having an infinity of left factors in X. Then, in an equivalent way, McNaughton's theorem says that the family $\mathrm{Rec}(A^N)$ is equal to the boolean closure of the family of limits of sets in $\mathrm{Rec}(A^*)$:

$$\mathrm{Rec}(A^N) = (\overrightarrow{\mathrm{Rec}(A^*)})^B$$

The link between the two formulations of McNaughton's theorem is the fact that U can be recognized by a deterministic Buchi automaton if and only if it is in $\overrightarrow{\mathrm{Rec}(A^*)}$.

Several proofs of McNaughton's theorem have been given (see [10], [7], [18]), including one by Schutzenberger [15] which uses the following lemma. One starts with a simple recognizable set $U = YZ^\omega$ with

$$Y = \phi^{-1}(m), \quad Z = \phi^{-1}(e)$$

Let $W = \{w \in A^* \mid Z \cap A^* w A^* = \emptyset\}$, $V = W - WA^+$ and $R = A^*V$. Let also $D = Z - ZA^+$.

PROPOSITION 1.4 <u>One has</u>

(1) $Z^\omega = \overrightarrow{ZD}$

(2) $YZ^\omega = \overrightarrow{YZD} - \vec{R}$.

A natural question left open by McNaughton's theorem is how one can decide whether a recognizable set U is a limit. The answer is given by the following result, due to H. Landweber [8].

THEOREM 1.5 <u>Let</u> $\mathcal{A} = (Q, i, \mathcal{T})$ <u>be a deterministic Muller automaton recognizing a set</u> $U \subset A^N$. <u>Then</u>

$$U \in \overrightarrow{\mathrm{Rec}(A^*)}$$

<u>iff</u> U <u>is still recognized by the Muller automaton</u> $(Q, i, \hat{\mathcal{T}})$ <u>where</u> $\hat{\mathcal{T}}$ <u>is the family of subsets of</u> Q <u>containing an element of</u> $\mathcal{T}$.

It is not difficult to prove the "only if" part of the theorem. To prove the converse, one considers a deterministic Muller automaton $\mathcal{A} = (Q, i, \mathcal{T})$ with $\mathcal{T} = \hat{\mathcal{T}}$ recognizing a set U. One then builds a new automaton

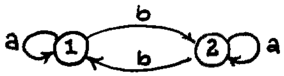
$$\mathcal{B} = (Q \times \mathcal{P}(Q), (i, \emptyset), R)$$

where $R = \{(q,\emptyset) \mid q \in Q\}$ and with the transitions

$$(q,S).a = \begin{cases} (q.a,\ \emptyset) & \text{if} \quad S+q.a \in T \\ (q.a,\ S+q.a) & \text{otherwise} \end{cases}$$

This automaton is deterministic and obviously recognizes (in Büchi's sense) the same set U as the Muller automaton A.

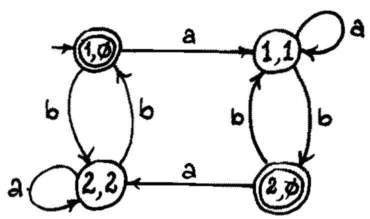
EXAMPLE 1.4 Consider again the automaton of Example 1.3



but this time used as a Muller automaton with $T = \{\{1,2\}\}$. Since $T = \hat{T}$ it recognizes a set which is a limit and it is in fact the set

$$U = (a^*b)^\omega = \overrightarrow{(a+b)^*b}$$

of words having an infinite number of occurrences of b. The previous construction leads to the automaton



giving a strange way of testing an infinite number of b's. □

Two-way automata on infinite words have been considered by J.P. Pécuchet [13]. He has proved that they recognize the same sets as one-way automata, therefore generalizing Shepherdson's result for two-way automata on ordinary words.

2. APERIODIC SETS

The family of <u>star free</u> subsets of A^* denoted by $Sf(A^*)$ is defined as the closure under set-product and boolean operations of the powerset of the alphabet A. Further, a monoid M is said to be <u>aperiodic</u> if there exists an integer $n \geq 0$ such that the identity

$$x^{n+1} = x^n$$

holds for all x in M. According to a classical theorem of Schützenberger (see [7] or [8] or [15]) a recognizable set $X \subset A^*$ is star free if and only if its syntactic monoid is aperiodic. This proves in particular that it is decidable wether a recognizable set is star free. For sets of infinite words, one has the following result, conjectured by R. Ladner and proved by W. Thomas in [17].

THEOREM 2.1 Let U be a recognizable subset of $A^{\mathbb{N}}$. The following conditions are equivalent :

(i) $U = \bigcup_{i=1}^{n} X_i Y_i^{\omega}$ with $X_i, Y_i^* \in Sf(A^*)$

(ii) U is a boolean combination of limits of star free sets :

$$U \in (\overrightarrow{Sf(A^*)})^B$$

(iii) U can be obtained from $\emptyset$ by a finite number of boolean operations and product on the left by an element of $Sf(A^*)$.

We shall denote by $Sf(A^{\mathbb{N}})$ the family of star free subsets of $A^{\mathbb{N}}$ which is defined by the above equivalent conditions.

There is a striking analogy between McNaughton's theorem and the above one. This is emphazized in [14] where a result having both ones as particular cases is proved. The framework used to state the result is that of varieties of semigroups.

The proof of Theorem 2.1 can be essentially handled using Proposition 1.4. Indeed, let F_n be the family of sets satisfying condition (ni). One has trivially

$$F_2 \subset F_1^B$$

Now Proposition 1.4 shows that $F_1 \subset F_2$, whence $F_1 = F_2$. Finally, $F_3 \subset F_1$ is clear since F_1 is closed under boolean operations and $F_2 \subset F_3$ comes from the fact that the complement of a limit can be written as a finite union of sets of the form $Y(A^{\mathbb{N}} - ZA^{\mathbb{N}})$.

We shall now prove the following result which implies in particular that the property of being star free is decidable within $Rec(A^{\mathbb{N}})$.

THEOREM 2.2 A recognizable set $U \subset A^{\mathbb{N}}$ is star free iff its syntactic monoid is aperiodic.

Proof : Suppose first that U is star free. By condition (i) there exists a morphism ϕ from A^* onto an aperiodic monoid M recognizing U. Now, by Proposition 1.3, the morphism $\phi \diamond \phi$ saturates U. But it is well known (see for instance [7]) that $M \diamond M$ is aperiodic if M is. Therefore the syntactic monoid of U is aperiodic.

Conversely, if U is recognized by a morphism onto an aperiodic monoid, it satisfies certainely condition (i). $\square$

Star free subsets of $A^{\mathbb{N}}$ have an interesting characterization in terms of logic. Let indeed $L(A)$ be the set of first order formulas on $\mathbb{N}$ with a relation $<$ and a set of predicates $\Pi_a(i)$ for $a \in A$. An infinite word $\alpha \in A^{\mathbb{N}}$ satisfies the formula $\phi \in L(A)$ written $\alpha \models \phi$, if the formula is true when the variables are interpreted in $\mathbb{N}$, the relation $<$ as the usual ordering on the integers and the predicate $\Pi_a(i)$ as : the i-th letter of α is an a.

For a formula $\phi \in L(A)$, we denote by

$$M(\phi) = \{\alpha \in A^N \mid \alpha \models \phi\}$$

the set of words satisfying the formula. We say that the set $M(\phi)$ is <u>defined</u> by formula ϕ. The following result has been proved by W. Thomas [17].

THEOREM 2.3 <u>A set $U \subset A^N$ is star free iff it can be defined by a first order formula</u> $\phi \in L(A)$.

An analogous result for finite words also holds and it was one of the motivations that lead R. Mc Naughton to consider star free sets.

EXAMPLE 2.1 Let $A = \{a,b\}$ and

$$U = (a+b)^* a^\omega$$

then U is star free according to condition (i) of Theorem 2.1. An expression in the form of condition (ii) is

$$U = (a+b)^\omega - a^* b (a+b)^\omega$$

The syntactic monoid of U is composed of three elements $1, \alpha = \phi(a^+)$, $\beta = \phi(A^* bA^*)$. This gives the following decomposition of U into simple recognizable sets :

$$U = (a+b)^* ba^\omega + a^\omega$$

A simple formula that defines U is

$$\exists x\, (\forall y\, ((x<y) \to \Pi_a(y))). \qquad \square$$

3. TWO-SIDED INFINITE WORDS

We consider now two-sided infinite words, that is to say elements $\alpha \in A^Z$ written

$$\alpha = \ldots \alpha_{-1}\alpha_0\alpha_1\alpha_2 \ldots$$

with $\alpha_i \in A$. Let $\mathcal{A} = (Q,I,T)$ be a finite automaton. THe set of words $\alpha \in A^Z$ <u>recognized</u> by $\mathcal{A}$ is, by definition, formed by the labels of two-sided infinite paths in $\mathcal{A}$ going infinitely often through I on the left and infinitely often through T on the right. A set $W \subset A^Z$ is said to be recognizable, written $W \in \mathrm{Rec}(A^Z)$, if it can be recognized by a finite automaton.

Observe an essential feature of this definition : a recognizable set is invariant under the shift. This is the point which makes both the interest of the objects and also the difficulty in their study since it adds a new phenomenon which was not present (nor possible to define) with one-sided infinite words. To see the relationship with the one sided case, we introduce the notation $[U,V]$ for $U,V \subset A^N$ to represent the set of all $\alpha \in A^Z$ such that

$$\alpha_0\alpha_{-1}\alpha_{-2}\ldots \in U,\ \alpha_0\alpha_1\alpha_2\ldots \in V$$

whereas $\widetilde{UV}$ is the closure of $[U,V]$ under the shift. We also note ${}^{\omega}X$ for the reversal of X^{ω}. Then it is not difficult to prove that the three following conditions are equivalent :

(i) $W \in \mathrm{Rec}(A^{\mathbb{Z}})$

(ii) W is shift invariant and is a finite union of sets $[U,V]$ for $U,V \in \mathrm{Rec}(A^{\mathbb{N}})$

(iii) W is a finite union of sets ${}^{\omega}XYZ^{\omega}$ for $X,Y,Z \in \mathrm{Rec}(A^*)$.

Condition (ii) can be used to prove easily that the family $\mathrm{Rec}(A^{\mathbb{Z}})$ is closed under complementation.

The notion of a morphism $\phi : A^* \to M$ that recognizes or saturates a set $U \subset A^{\mathbb{Z}}$ extends without difficulty to the two-sided case. The main problem is the extension of McNaughton's theorem.

For this, we first introduce the notion of <u>two-sided limit</u> : for $X \subset A^+$, the two-sided limit of X, denoted by $\overleftrightarrow{X}$ is the set of $\gamma \in A^{\mathbb{Z}}$ such that

$$\gamma_{-n}\gamma_{n+1}\cdots\gamma_{m-1}\gamma_m \in X$$

for infinitely many $n \geq 0$ and infinitely many $m \geq 0$. The following result, proved in [12] by M. Nivat and myself, generalizes McNaughton's theorem to the two-sided case.

THEOREM 3.1 <u>The family</u> $\mathrm{Rec}(A^{\mathbb{Z}})$ <u>is the boolean closure of the two-sided limits of elements of</u> $\mathrm{Rec}(A^*)$:

$$\mathrm{Rec}(A^{\mathbb{Z}}) = (\overleftrightarrow{\mathrm{Rec}(A^*)})^B$$

The proof of this result relies on a generalization to the two-sided case of Proposition 1.4. One starts with a simple recognizable set

$$W = {}^{\omega}XYZ^{\omega}$$

with $Y = \phi^{-1}(m)$ and X,Z the minimal generating sets of $\phi^{-1}(e)$, $\phi^{-1}(f)$ for two idempotents $e,f \in \phi(A^*) = M$. One then considers

$$G = X - A^+X, \quad D = Z - ZA^+$$

and two sets $L = KA^*$, $R = A^*J$ where K,J are the basis of the two sided ideals which are the complements of the set of factors of X^*, Z^* respectively.

PROPOSITION 3.2 <u>One has the equality</u>

$${}^{\omega}XYZ^{\omega} = \overleftrightarrow{GXYZD} - \overleftrightarrow{L} - \overleftrightarrow{R}.$$

The relationship with deterministic automata goes through the definition of a new type of two-sided automaton or, equivalently, through the following notion : define the family $Det(A^{\mathbb{Z}})$ of deterministic sets as that of shift-invariant subsets of $A^{\mathbb{Z}}$ which are finite unions of sets of the form :

$$[U,V]$$

with $U,V \in Det(A^{\mathbb{N}}) = Rec(A^*)$. One then has the following result proved by D. Beauquier [3].

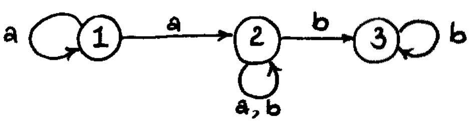
PROPOSITION 3.3 The family $Det(A^{\mathbb{Z}})$ is equal to that of two sided limits of recognizable sets of A^* :

$$Det(A^{\mathbb{Z}}) = \overleftrightarrow{Rec(A^*)}$$

EXAMPLE 3.1 Let $A = \{a,b\}$ and consider

$$W = {}^{\omega}a(a+b)^*b^{\omega}$$

which can be recognized by the automaton



An expression according to Theorem 3.1 is

$$W = \overleftrightarrow{A^*} - \overleftrightarrow{A^*a} - \overleftrightarrow{bA^*}. \qquad \square$$

An interesting result on two-sided infinite words has been obtained by L. Compton [5]. To state it, say that $\alpha \in A^{\mathbb{Z}}$ is generic if each word $w \in A^*$ appears an infinity of times on the right and on the left in α. Then, one has the following result [5] :

THEOREM 3.4 A recognizable subset W of $A^{\mathbb{Z}}$ either contains all generic words or none of them.

This result says, in other terms, that it is not possible to distinguish two generic words using a finite automaton.

4. UNAMBIGUITY

Since it is not always possible to restrict one's attention to deterministic Büchi automata, the problem of unambiguity in the recognition of infinite words is more difficult than in the case of finite ones. However deterministic Muller automata can be used to obtain unambiguous representations as follows.

We say that a family (X_i,Y_i), $1 \leq i \leq n$ of sets in A^* is unambiguous if for each infinite word $\alpha \in A^{\mathbb{N}}$ there is at most one decomposition in the form

$$\alpha = xy_1y_2\ldots$$

with $x \in X_i$, $y_j \in Y_i$ and $1 \leq i \leq n$. This means that the set

$$U = \bigcup_{i=1}^{n} X_i Y_i^{\omega}$$

is described in this way by an unambiguous expression. The following result is due to A. Arnold [1] :

THEOREM 4.1 For each recognizable set $U \subset A^{N}$ there exists an unambiguous family (X_i, Y_i) defining U, that is such that $U = \bigcup_{i=1}^{n} X_i Y_i^{\omega}$.

An other kind of unambiguity result was suggested to me by considering two-sided infinite words. The basic idea is to use automata on infinite words that start reading "at infinity". More precisely, let us say that an automaton

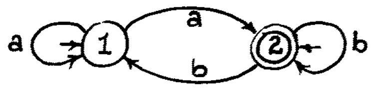
$$\mathcal{A} = (Q, I, T)$$

is codeterministic if for any two arrows $(p,a,q),(p',a,q')$ labeled by the same letter, $q = q'$ implies $p = p'$. The following result was obtained by D. Beauquier and myself [4]. It was in fact, previous to our publication, announced by A. Mostowski in [11] but we have not been able to understand his proof.

THEOREM 4.2 Any recognizable set in A^{N} can be recognized by a codeterministic automaton.

The construction used to build a codeterministic automaton is the following : let $U = YZ^{\omega}$ be a simple recognizable set with $Y = \phi^{-1}(m)$, $Z = \phi^{-1}(e)$ and e an idempotent. Consider a codeterministic automaton $\mathcal{A}$ that simultane ously recognizes YZ and GZ with $G = Z - A^{+}Z$. Let I be the set of initial states for YZ and J that of GZ. Then one can prove that the codeterministic automaton $\mathcal{A}$ having I as initial states and J as terminal states recognizes YZ^{ω}.

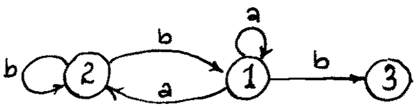
EXAMPLE 4.1 Let $A = \{a,b\}$ and consider $U = (a^{*}b)^{\omega}$. Then the following codeterministic automaton recognizes U :



The behaviour of this automaton can be understood as follows : if, in state 1, the input is an a, there it has to make a transition to state 2 if the next input is going to be a b ; otherwise it would stop on the next b. In terms of the above construction, one has

$$Y = 1, \quad Z = A^{*}b, \quad G = b$$

A codeterministic automaton recognizing YZ and GZ is shown below



With the above notations, one has $I = \{1,2\}$, $J = \{1\}$. The automaton obtained by using I as initial states and J as terminal states is almost identical to the above one, up to removal of the useless state 3. □

BIBLIOGRAPHY

[1] Arnold, A., 1983, Rational ω-languages are non-ambiguous, Theoret. Comput. Sci., 26, 221-224.

[2] Arnold, A., 1984, A syntactic congruence for rational ω-languages, to appear in Theoret. Comput. Sci.

[3] Beauquier, D., 1984, Bilimites de langages reconnaissables, to appear in Theoret. Comput. Sci.

[4] Beauquier, D., Perrin, D., 1984, Automates codeterministes sur les mots infinis, to appear.

[5] Compton, L., 1984, in Progress in Combinatorics on Words, Academic Press.

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

[7] Eilenberg, S., 1974, Automata, Languages and Machines, Vol. A, Academic Press, New York, Vol. B, 1976.

[8] Lallement, G., 1979, Semigroups and Combinatorial Applications, Wiley.

[9] Landweber, L.H., 1969, Decision problems for ω-automata, Math. Syst. Theory, 3, 376-384.

[10] McNaughton, R., 1966, Testing and generating infinite sequences by a finite automaton, Information and Control, 9, 521-530.

[11] Mostowski, A., 1982, Determinancy of sinking automata on infinite trees and inequalities between various Rabin's pair indices, Information Processing Letters, 15, 159-163.

[12] Nivat, M., Perrin D., 1982, Ensembles reconnaissables de mots biinfinis, Proc. 14th ACM Symp. on Theory of Computing, 47-59.

[13] Pécuchet, J.P., 1983, Automates boustrophédons et mots infinis, à paraitre dans Theoret. Comput. Sci.

[14] Perrin D., Variétés de langages et mots infinis, C.R. Acad. Sci. Paris, 295, 595-598.

[15] Pin, J.E., Variétés de langages formels, Masson, 1984.

[16] Schützenberger, M.P., 1972, Sur les relations rationnelles fonctionnelles, in Automata, Languages and Programming (M. Nivat ed.) North Holland, 103-114.

[17] Thomas, W., 1979, Star free regular sets of ω-sequences, Inform. Control, 42, 148-156.

[18] Thomas, W., 1981, A combinatorial approach to the theory of ω-automata, Inform. Control, 48, 261-283.

[19] Thomas, W., 1982, A hierarchy of sets of infinite trees, in Theoretical Computer Science, Springer Lecture Notes on Comput. Sci., 145.