Logic, semigroups and automata on words.

Jean-Eric Pin (*)

Abstract This is a survey paper on the connections between formal logic and the theory of automata. The logic we have in mind is the sequential calculus of Buchi, a system which allows to formalize properties of words. In this logic, there is a predicate for each letter and the unique extra non logical predicate is the relation symbol, which is interpreted as the usual order on the integers. Several famous classes have been classified within this logic.

We shall briefly review the main results concerning second order, which covers classes like PH, NP, P, etc. and then study in more detail the results concerning the monadic second order and the first order logic.

1. Introduction

The aim of this paper is to survey the connections between formal logic and the theory of automata. The logic we have in mind is the "sequential calculus" of Büchi, a system which allows to formalize properties of words. A typical formula of this logic looks like

$$\exists x\ \exists y\ \big((x < y) \wedge (R_a x) \wedge (R_b y)\big),$$

and can intuitively be interpreted on a word u by "there exist two integers $x < y$ such that, in u, the letter in position x is an a and the letter in position y is a b".

Thus, if $A = \{a, b\}$ is the alphabet, the set of finite words satisfying our formula is the set of all words containing an occurrence of a followed (but not necessarily immediately) by an occurrence of b, and can be described by the rational expression $A^*aA^*bA^*$. Similarly, the set of infinite (resp. biinfinite) words satisfying the formula is $A^*aA^*bA^\omega$ (resp. $A^{\tilde{\omega}}aA^*bA^\omega$).

This example illustrates the logical point of view to define a set of words, but there are other approaches to do so, including automata, rational expressions and semigroups. As we shall see in this paper, these various points of view complement each other and are, to some extent, equivalent. This leads to a remarkable theory and to numerous problems, some of which are still open.

In this paper, we focus on the logical point of view. We shall classify the most fundamental logical problems that arise in this framework into three categories:

(1) *Descriptive power.* Given a set S of sentences (such as first order sentences, Σ_1 formulæ, etc.,) characterize the sets of words that can be defined by a formula of S.

(2) *Decision problems.* Given a set S of sentences and a rational set of words X, is it decidable whether X can be defined by a sentence of S?

(3) *Elementary equivalence.* Given a set S of sentences, two words are said to be S-equivalent if they satisfy exactly the same sentences of S. The problem is to describe these

(*) Bull Research and Development, rue Jean Jaurès, 78340 Les Clayes-sous-Bois, FRANCE. From 1st Sept 1993, LITP, Université Paris VI, Tour 55-65, 4 Place Jussieu, 75252 Paris Cedex 05, FRANCE. **E-mail**: `pin@litp.ibp.fr`

equivalence relations. The following question, proposed by Parikh, falls into this category: "if two biinfinite words u and v have the same factors, do they satisfy the same first order formulæ of the theory of successor?"

But before we tackle such logical problems, we need to survey the three other approaches: automata, semigroups and rational expressions.

2. Automata, semigroups and rational sets

In this section, we recall the basic definitions of the theory of finite automata needed in this article. Most of them are quite standard, but the reader might not be familiar with some of them, in particular those relative to biinfinite words and to semigroups.

2.1. Words

Let A be a finite set called an *alphabet*, whose elements are *letters*. A *finite word* is a finite sequence of letters, that is, a function u from a finite set of the form $\{0, 1, 2, \ldots, n\}$ into A. If one puts $u(i) = a_i$ for $0 \leq i \leq n$, the word u is usually denoted by $a_0a_1 \cdots a_n$, and the integer $|u| = n + 1$ is the *length* of u. The unique word of length 0 is the empty word, denoted by 1. An *infinite word (on the right)* is a function u from $\mathbb{N}$ into A, usually denoted by $a_0a_1a_2 \cdots$, where $u(i) = a_i$ for all $i \in \mathbb{N}$. An *infinite word on the left* is a function u from the set of non positive integers into A, usually denoted by $\cdots a_{-2}a_{-1}a_0$, where $a_{-i} \in A$ for all $i \in \mathbb{N}$. Finally, a *biinfinite word* is a function u from $\mathbb{Z}$ into A, usually denoted by $\cdots a_{-2}a_{-1}a_0a_1a_2 \cdots$. Two biinfinite words $u : A \to \mathbb{Z}$ and $v : A \to \mathbb{Z}$ are *shift equivalent* if there exists an integer n such that, for all $i \in \mathbb{Z}$, $v(i) = u(i + n)$. A *bilateral word* is an equivalence class for the shift equivalence. It is usually denoted by $\cdots a_{-2}a_{-1}a_0a_1a_2 \cdots$. A *word* is either a finite word, an infinite word on the right, an infinite word on the left or a bilateral word.

Intuitively, the *concatenation* or *product* of two words u and v is the word uv obtained by writing u followed by v. More precisely, if u is finite and v is finite or infinite on the right, then uv is the word defined by

$$(uv)(i) = \begin{cases} u(i) & \text{if } i < |u| \\ v(i - |u|) & \text{if } i \geq |u| \end{cases}$$

if u is infinite on the left and v is finite, then uv is the word defined by

$$(uv)(i) = \begin{cases} u(i) & \text{if } i \leq 0 \\ v(i - 1) & \text{if } i > 0 \end{cases}$$

Finally, if u is infinite on the left and v is infinite on the right, then uv is the shift class of the biinfinite word uv defined by

$$(uv)(i) = \begin{cases} u(i) & \text{if } i \leq 0 \\ v(i - 1) & \text{if } i > 0 \end{cases}$$

We denote respectively by A^*, A^+, $A^{\mathbb{N}}$, $A^{-\mathbb{N}}$ and $A^{\mathbb{Z}}$ the set of all finite words, finite nonempty words, infinite words (on the right), infinite words on the left and bilateral words. A word x is a factor of a word w if there exist two words u and v (possibly empty) such that $w = uxv$. A factor x of a biinfinite word $u = \cdots a_{-2}a_{-1}a_0a_1a_2 \cdots$ occurs "infinitely often on the right" (respectively left) of u if for every n, x is a factor of the infinite word $a_na_{n+1}a_{n+2} \cdots$ (resp. $\cdots a_{n+2}a_{n+1}a_n$). A biinfinite word u is *recurrent* if every factor of u occurs infinitely often on the right and on the left. Since these notions are invariant under shift, they can be extended to bilateral words.

For every finite word $u = a_0a_1 \cdots a_n$, we set $\tilde{u} = a_na_{n-1} \cdots a_0$. Similarly, for every infinite word on the right u, we denote by $\tilde{u}$ the infinite word on the left defined by $\tilde{u}_{-n} = u_n$.

2.2. Rational sets

The *rational operations* are the three operations union, product and star, defined on the set of subsets of A^* as follows

(1) Union : $L_1 \cup L_2 = \{u \mid u \in L_1 \text{ or } u \in L_2\}$
(2) Product : $L_1 L_2 = \{u_1 u_2 \mid u_1 \in L_1 \text{ and } u_2 \in L_2\}$
(3) Star : $L^* = \{u_1 \cdots u_n \mid n \geq 0 \text{ and } u_1, \ldots, u_n \in L\}$

It is also convenient to introduce the operator

$$L^+ = LL^* = \{u_1 \cdots u_n \mid n > 0 \text{ and } u_1, \ldots, u_n \in L\}$$

The set of rational subsets of A^* is the smallest set of subsets of A^* containing the finite sets and closed under finite union, product and star. For instance, $(a \cup ab)^* ab \cup (ba^*b)^*$ denotes a rational set. The rational subsets of A^+ are the rational subsets of A^* that do not contain the empty word.
It is possible to generalize the concept of rational sets to infinite words as follows. First, the product can be extended to $A^* \times A^{\mathbb{N}}$, by setting, for $X \subset A^*$ and $Y \subset A^{\mathbb{N}}$,

$$XY = \{xy \mid x \in X \text{ and } y \in Y\}.$$

Next, we define an infinite iteration ω by setting, for every subset X of A^+

$$X^\omega = \{x_0 x_1 \; \cdots \mid \text{ for all } i \geq 0, x_i \in X\}$$

Equivalently, X^ω is the set of infinite words obtained by concatenating an infinite sequence of words of X. In particular, if $u = a_0 a_1 \cdots a_n$, we set

$$u^\omega = a_0 a_1 \cdots a_n a_0 a_1 \cdots a_n a_0 a_1 \cdots a_n a_0 a_1 \cdots$$

By definition, a subset of $A^{\mathbb{N}}$ is $\mathbb{N}$-*rational* if and only if it can be written as a finite union of subsets of the form XY^ω where X and Y are non-empty rational subsets of A^+.

For biinfinite words, we first extend the product to $A^{-\mathbb{N}} \times A^{\mathbb{N}}$, by setting, for $X \subset A^{-\mathbb{N}}$ and $Y \subset A^{\mathbb{N}}$,

$$XY = \{xy \mid x \in X \text{ and } y \in Y\}.$$

Next, infinite iteration on the left is defined by

$$X^{\tilde{\omega}} = \{u \in A^{-\mathbb{N}} \mid \tilde{u} \in (\tilde{X})^\omega\}$$

where $\tilde{X} = \{\tilde{x} \mid x \in X\}$. By definition, a subset of $A^{\mathbb{Z}}$ is $\mathbb{Z}$-*rational* if and only if it can be written as a finite union of subsets of the form $X^{\tilde{\omega}} Y Z^\omega$ where X, Y and Z are non-empty rational subsets of A^+.

Example 2.1. The set of infinite words on the alphabet $\{a, b\}$ having only a finite number of b's is given by the expression $\{a, b\}^* a^\omega$. The set of biinfinite words on the alphabet $\{a, b\}$ having only a finite number of b's is given by the expression $a^{\tilde{\omega}} \{a, b\}^* a^\omega$.

2.3. Finite automata and recognizable sets

A finite (non deterministic) automaton is a triple $\mathcal{A} = (Q, A, E)$ where Q is a finite set (the set of *states*), A is an alphabet, and E is a subset of $Q \times A \times Q$, called the set of *transitions*. Two transitions (p, a, q) and (p', a', q') are *consecutive* if $q = p'$. A *path* in $\mathcal{A}$ is a finite sequence of consecutive transitions

$$e_0 = (q_0, a_0, q_1), \;\; e_1 = (q_1, a_1, q_2), \;\; \ldots \;\;, \;\; e_n = (q_n, a_n, q_{n+1})$$

also denoted

$$q_0 \xrightarrow{a_0} q_1 \;\; a_1 \!\longrightarrow q_2 \;\; \cdots \;\; q_n \xrightarrow{a_n} q_{n+1}$$

The state q_0 is the *origin* of the path, the state q_{n+1} is its *end*, and the word $x = a_0 a_1 \cdots a_n$ is its *label*.

An $\mathbb{N}$-*path* in $\mathcal{A}$ is a sequence p of consecutive transitions indexed by $\mathbb{N}$,

$$e_0 = (q_0, a_0, q_1), \;\; e_1 = (q_1, a_1, q_2), \;\; \ldots$$

also denoted

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

The state q_0 is the *origin* of the infinite path and the infinite word $a_0 a_1 \cdots$ is its *label*. A state q occurs infinitely often in p if $q_n = q$ for infinitely many n. Similarly, a $\mathbb{Z}$-*path* in $\mathcal{A}$ is a sequence p of consecutive transitions indexed by $\mathbb{Z}$. A state q occurs infinitely often *on the right* (respectively *on the left*) in p if $q_n = q$ for infinitely many positive (respectively negative) n.

Example 2.2. Let $\mathcal{A} = (Q, A, E)$ where

$$Q = \{1, 2\}, A = \{a, b\} \text{ and } E = \{(1, a, 1), (2, b, 1), (1, a, 2), (2, b, 2)\}$$

be the automaton represented below.

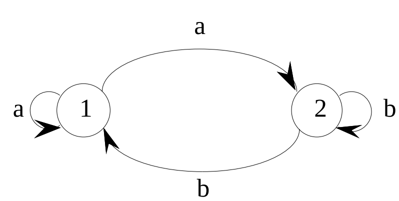

Figure 2.1.

Then $(1, a, 2)(2, b, 2)(2, b, 1)(1, a, 2)(2, b, 2)(2, b, 1)(1, a, 2)(2, b, 2)(2, b, 1) \cdots$ is an infinite path of $\mathcal{A}$.

A finite *Büchi automaton* is a quintuple $\mathcal{A} = (Q, A, E, I, F)$ where

(1) (Q, A, E) is a finite automaton,

(2) I and F are subsets of Q, called the set of *initial* and *final* states, respectively.

A finite path in $\mathcal{A}$ is *successful* if its origin is in I and its end is in F. An $\mathbb{N}$-path p is *successful* if its origin is in I and if some state of F occurs infinitely often in p. A $\mathbb{Z}$-path p is *successful* if some state of I occurs infinitely often on the left in p and if some state of F occurs infinitely often on the right in p.

The set of finite (respectively infinite, bilateral) words *recognized* by $\mathcal{A}$ is the set, denoted $L^+(\mathcal{A})$ (respectively $L^{\mathbb{N}}(\mathcal{A})$, $L^{\mathbb{Z}}(\mathcal{A})$), of the labels of all successful finite (respectively $\mathbb{N}$-, $\mathbb{Z}$-) paths of $\mathcal{A}$. A set of finite (respectively infinite, bilateral) words X is *recognizable* if there exists a finite Büchi automaton $\mathcal{A}$ such that $X = L^+(\mathcal{A})$ (respectively $X = L^{\mathbb{N}}(\mathcal{A})$, $X = L^{\mathbb{Z}}(\mathcal{A})$).

Example 2.3. Let $\mathcal{A}$ be the Büchi automaton obtained from example 2.1 by taking $I = \{1\}$ and $F = \{2\}$. Initial states are represented by an incoming arrow and final states by an arrow going out.

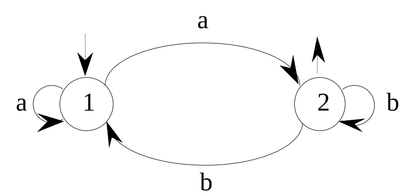

Figure 2.2. A Büchi automaton.

Then $L^+(\mathcal{A}) = a\{a,b\}^*$ is the set of all finite words whose first letter is an a, $L^{\mathbb{N}}(\mathcal{A}) = a(a^*b)^\omega$ is the set of infinite words whose first letter is an a and containing an infinite number of b's and $L^{\mathbb{Z}}(\mathcal{A}) = (ab^*)^{\tilde{\omega}}(a^*b)^\omega$ is the set of infinite words containing infinitely many a's on the left and infinitely many b's on the right.

The relationship between rational and recognizable sets of finite words is given by the famous theorem of Kleene.

Theorem 2.1. *A subset of A^* is rational if and only if it is recognizable.*

The counterpart of Kleene's theorem for infinite words is due to Büchi [12] and for bilateral words to Nivat and Perrin [43].

Theorem 2.2. *A subset of $A^{\mathbb{N}}$ (resp. $A^{\mathbb{Z}}$) is $\mathbb{N}$-rational (resp. $\mathbb{Z}$-rational) if and only if it is recognizable.*

2.4. Semigroups

A *semigroup* is a set equipped with an internal associative operation which is usually written in a multiplicative form. A monoid is a semigroup with identity (usually denoted by 1). If S is a semigroup, S^1 denotes the monoid equal to S if S has an identity and to $S \cup \{1\}$ otherwise. In the latter case, the multiplication on S is extended by setting $s1 = 1s = s$ for every $s \in S^1$. An element e of a semigroup S is *idempotent* if $e^2 = e$. A *zero* is an element 0 such that, for every $s \in S$, $s0 = 0s = 0$. Given two semigroups, S and T, a (semigroup) morphism $\varphi : S \to T$ is an application of S into T such that for all $x, y \in S$, $\varphi(xy) = \varphi(x)\varphi(y)$. A semigroup S is a *quotient* of a semigroup T if there exists a surjective morphism from T onto S. A semigroup S *divides* a semigroup T if S is a quotient of a subsemigroup of T. Division is a quasi-order on finite semigroups (up to an isomorphism).

A *semiring* is a set K equipped with two operations, called respectively addition and multiplication, denoted $(s,t) \to s+t$ and $(s,t) \to s \cdot t$, and an element, denoted 0, such that:

(1) $(K, +, 0)$ is a commutative monoid,
(2) $(K, .)$ is a semigroup,
(3) for all $s, t_1, t_2 \in K$, $s(t_1 + t_2) = st_1 + st_2$ and $(t_1 + t_2)s = t_1 s + t_2 s$,
(4) for all $s \in K$, $0s = s0 = 0$.

Thus the only difference with a ring is that inverses with respect to addition may not exist. We denote by $\mathbb{B}$ the *boolean semiring* defined by the following operations

$+$	0	1
0	0	1
1	1	1

$\cdot$	0	1
0	0	0
1	0	1

Given a semiring K, the set $K^{n\times n}$ of $n \times n$ matrices over K is naturally equipped with a structure of semiring. In particular, $K^{n\times n}$ is a monoid under multiplication defined by

$$(rs)_{i,j} = \sum_{1\leq k\leq n} r_{i,k}s_{k,j}$$

Let Σ be a countable alphabet. Given two words u, v of Σ^+ (resp. Σ^*), a semigroup (resp. monoid) S satisfies the equation $u = v$ if, for every semigroup (monoid) morphism φ from Σ^+ (Σ^*) into S, $\varphi(u) = \varphi(v)$. For instance, a semigroup is commutative if and only if it satisfies the equation $xy = yx$. Let $(u_n = v_n)_{n\in\mathbb{N}}$ be a sequence of equations. A semigroup (resp. monoid) S *ultimately satisfies* the sequence of equations $(u_n = v_n)_{n\in\mathbb{N}}$ if there exists an integer n_S such that, for all $n \geq n_S$, S satisfies the equation $u_n = v_n$. For instance, one can show that a finite monoid is a group if and only if it ultimately satisfies the equations $(u^{n!} = 1)_{n>0}$.

Green's relations on a semigroup S are defined as follows. If s and t are elements of S, we note

$s\,\mathcal{L}\,t$ if there exist $x, y \in S^1$ such that $s = xt$ and $t = ys$,
$s\,\mathcal{R}\,t$ if there exist $x, y \in S^1$ such that $s = tx$ and $t = sy$,
$s\,\mathcal{J}\,t$ if there exist $x, y, u, v \in S^1$ such that $s = xty$ and $t = usv$.
$s\,\mathcal{H}\,t$ if $s\,\mathcal{R}\,t$ and $s\,\mathcal{L}\,t$.

For finite semigroups, these four equivalence relations can be represented as follows. The elements of a given $\mathcal{R}$-class (resp. $\mathcal{L}$-class) are represented in a row (resp. column). The intersection of an $\mathcal{R}$-class and a $\mathcal{L}$-class is an $\mathcal{H}$-class. Each $\mathcal{J}$-class is a union of $\mathcal{R}$-classes (and also of $\mathcal{L}$-classes). It is not obvious to see that this representation is consistent : it relies in particular on the non-trivial fact that, in finite semigroups, the relations $\mathcal{R}$ and $\mathcal{L}$ commute. An idempotent is represented by a star. One can show that each $\mathcal{H}$-class containing an idempotent e is a subsemigroup of S, which is in fact a group with identity e. Furthermore, all $\mathcal{R}$-classes (resp. $\mathcal{L}$-classes) of a given $\mathcal{J}$-class have the same number of elements.

Each column is an $\mathcal{L}$-class

Each row is an $\mathcal{R}$-class

* a_1, a_2	* a_3, a_4	a_5, a_6
a_7, a_8	* a_9, a_{10}	* a_{11}, a_{12}

A $\mathcal{J}$-class.

A semigroup S is $\mathcal{L}$-trivial (resp. $\mathcal{R}$-trivial, $\mathcal{J}$-trivial, $\mathcal{H}$-trivial) if two element of S which are $\mathcal{L}$-equivalent (resp. $\mathcal{R}$-equivalent, $\mathcal{J}$-equivalent, $\mathcal{H}$-equivalent) are equal.

2.5. Semigroups and recognizable sets

In this section, we turn to a more algebraic definition of the recognizable sets, using semigroups in place of automata. Although this definition is more abstract than the definition using automata, it is more suitable to handle the fine structure of recognizable sets. Indeed, as discovered by Eilenberg [21], semigroups provide a powerful and systematic tool to classify recognizable sets. We will see in particular that most of the classes of recognizable sets associated with standard classes of formulæ (first order, existential, etc.) admit some simple algebraic characterizations.

The abstract definition of recognizable sets is based on the following observation. Let $\mathcal{A} = (Q, A, E, I, F)$ be a finite automaton. To each word $u \in A^*$, there corresponds a boolean square matrix of size Card(Q), denoted $\mu(u)$, and defined by

$$\mu(u)_{p,q} = \begin{cases} 1 & \text{if there exists a path from } p \text{ to } q \text{ with label } u \\ 0 & \text{otherwise} \end{cases}$$

Example 2.4. Let $\mathcal{A} = (Q, A, E, I, F)$ be the automaton represented below

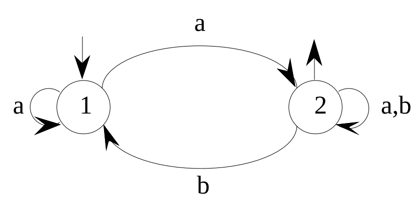

Figure 2.3. A non deterministic automaton.

Then $Q = \{1, 2\}$, $A = \{a, b\}$ and $E = \{(1, a, 1), (1, a, 2), (2, a, 2), (2, b, 1), (2, b, 2)\}$, $I = \{1\}$, $F = \{2\}$, whence

$$\mu(a) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \mu(b) = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \qquad \mu(aa) = \mu(a)$$

$$\mu(ab) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \qquad \mu(ba) = \mu(bb) = \mu(b)$$

It is not difficult to see that μ is a semigroup morphism from A^+ into the multiplicative semigroup of square boolean matrices of size Card(Q). Furthermore, a word u is recognized by $\mathcal{A}$ if and only if there exists a successful path from 1 to 2 with label u, that is, if $\mu(u)_{1,2} = 1$. Therefore, a word is recognized by $\mathcal{A}$ if and only if $\mu(u)_{1,2} = 1$. The semigroup $\mu(A^+) = \{\left(\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)\}$ is called the *transition semigroup* of $\mathcal{A}$.

Let us now give the formal definitions. Let $\varphi : A^+ \to S$ be a semigroup morphism. A subset X of A^+ is *recognized* by φ if there exists a subset P of S such that $X = \varphi^{-1}(P)$. As shown by the previous example, a set recognized by a finite automaton is recognized by the transition semigroup of this automaton. Conversely, given a finite semigroup recognizing a subset X of A^+, one can build a finite automaton recognizing X. Therefore, the two notions of recognizable sets (by finite automata and by finite semigroups) are equivalent:

Theorem 2.3. *A subset of A^+ is recognizable if and only if it is recognized by a finite semigroup.*

Let X be a recognizable set of A^+. Amongst the finite semigroups that recognize X, there is a minimal one (with respect to division). This finite semigroup is called the *syntactic semigroup* of X. It can be defined directly as the quotient of A^+ under the congruence $\sim_X$ defined by $u \sim_X v$ if and only if, for every $x, y \in A^*$, $xuy \in X \Leftrightarrow xvy \in X$. It is also equal to the transition semigroup of the minimal automaton of $\mathcal{A}$. See [54] for more details.

2.6. ω-semigroups

It is possible to extend the previous results to infinite words by replacing semigroups by ω-semigroups, which are, basically, algebras in which infinite products are defined. Although these algebras do not have a finitary signature, standard results on algebras still hold in this case. In particular, A^∞ appears to be the free algebra on the set A and recognizable sets can be defined, as before, as the sets recognized by finite algebras. However, a problem arises since finite algebras have an infinitary signature and thus are not really finite! This problem can be solved by a Ramsey type argument showing that the structure of these finite algebras can be totally determined by only two operations of finite signature. This defines a new type of algebra of finite signature, the Wilke algebras, that suffice to deal with infinite products defined on finite sets. (*)

We now come to the precise definitions. An *ω-semigroup* is an algebra $S = (S_f, S_\omega)$ equipped with the following operations:

- A binary operation defined on S_f and denoted multiplicatively,
- A mapping $S_f \times S_\omega \rightarrow S_\omega$, called *mixed product*, that associates to each couple $(s,t) \in S_f \times S_\omega$ an element of S_ω denoted st,
- A mapping $\pi : S_f^{\mathbb{N}} \rightarrow S_\omega$, called *infinite product*

These three operations should satisfy the following properties :

(1) S_f, equipped with the binary operation, is a semigroup,

(2) for every $s, t \in S_f$ and for every $u \in S_\omega$, $s(tu) = (st)u$,

(3) for every increasing sequence $(k_n)_{n>0}$ and for all $(s_n)_{n\in\mathbb{N}} \in S_f^{\mathbb{N}}$,

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

(4) for every $s \in S_f$ and for every $(s_n)_{n\in\mathbb{N}} \in S_f^{\mathbb{N}}$

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

Conditions (1) and (2) can be thought of as an extension of associativity. Conditions (3) et (4) show that one can replace the notation $\pi(s_0, s_1, s_2, \ldots)$ without ambiguity by the notation $s_0 s_1 s_2 \cdots$. We shall use this simplified notation in the sequel. Intuitively, an ω-semigroup is a sort of semigroup in which infinite products are defined.

Example 2.5. We denote by A^∞ the ω-semigroup $(A^+, A^{\mathbb{N}})$ equipped with the usual concatenation product.

(*) Actually, the chronology is a little bit different. Ramsey type arguments have been used for a long time in semigroup theory [33], the Wilke algebras were introduced by Wilke in [83] under the name of *binoids* to clarify the approach of Arnold [4], Pécuchet [45] and Perrin [46] and the idea of using infinite products on semigroups came last [51].

Given two ω-semigroups $S = (S_f, S_\omega)$ and $T = (T_f, T_\omega)$, a *morphism of ω-semigroup* is a couple $\varphi = (\varphi_f, \varphi_\omega)$ consisting of a semigroup morphism $\varphi_f : S_f \to T_f$ and of a mapping $\varphi_\omega : S_\omega \to T_\omega$ preserving the mixed product and the infinite product: for every sequence $(s_n)_{n \in \mathbb{N}}$ of elements of S_f,

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

and for every $s \in S_f$, $t \in S_\omega$,

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

In the sequel, we shall omit the subscripts, and use the simplified notation φ instead of φ_f and φ_ω.

Algebraic concepts like congruence, ω-subsemigroup, quotient and division are easily adapted to ω-semigroups. The semigroup A^+ is called the *free semigroup* on the set A because it satisfies the following property (which defines free objects in the general setting of category theory): every map from A into a semigroup S can be extended *in a unique way* into a semigroup morphism from A^+ into S. Similarly, it is not difficult to see that the free ω-semigroup on $(A, \emptyset)$ is the ω-semigroup A^∞.

A key result is that when S is finite, the infinite product is totally determined by the elements of the form $s^\omega = sss\cdots$, according to the following result

Theorem 2.4. (Wilke) *Let S_f be a finite semigroup and let S_ω be a finite set. Suppose that there exists a mixed product $S_f \times S_\omega \to S_\omega$ and a map from S_f into S_ω, denoted $s \to s^\omega$, satisfying, for every $s, t \in S_f$, the equations*

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

Then the couple $S = (S_f, S_\omega)$ can be equipped, in a unique way, with a structure of ω-semigroup such that for every $s \in S$, the product $sss\cdots$ is equal to s^ω.

This is a non trivial result, based on a consequence of Ramsey's theorem which is worth mentioning:

Theorem 2.5. *Let $\varphi : A^+ \to S$ be a morphism from A^+ into a finite semigroup S. For every infinite word $u \in A^{\mathbb{N}}$, there exist a couple (s, e) of elements of S such that $se = s$, $e^2 = e$, 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 every $n > 0$.*

A morphism of ω-semigroups $\varphi : A^\infty \to S$ *recognizes a subset* X of $A^{\mathbb{N}}$ if, there exists a subset P of S_ω such that $X = \varphi^{-1}(P)$. By extension, a ω-semigroup S recognizes X if there exists a morphism of ω-semigroup $\varphi : A^\infty \to S$ that recognizes X. As for finite words, the following result holds:

Theorem 2.6. *A subset of $A^{\mathbb{N}}$ is recognizable if and only if it is recognized by a finite ω-semigroup.*

We now give the construction to pass from a finite (Büchi) automaton to a finite ω-semigroup. This construction is much more involved than the corresponding construction for finite words.

Given a finite Büchi automaton $\mathcal{A} = (Q, A, E, I, F)$ recognizing a subset of X of $A^{\mathbb{N}}$, we would like to obtain a finite ω-semigroup recognizing X. Our construction makes use of the semiring $k = \{-\infty, 0, 1\}$ in which addition is the maximum for the ordering $-\infty < 0 < 1$ and multiplication is given in the following table

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

To each letter $a \in A$ is associated a matrix $\mu(a)$ with entries in k defined by

$$\mu(a)_{p,q} = \begin{cases} -\infty & \text{if } (p,a,q) \notin E \\ 0 & \text{if } (p,a,q) \in E \text{ and } p \notin F \text{ and } q \notin F \\ 1 & \text{if } (p,a,q) \in E \text{ and } (p \in F \text{ or } q \in F) \end{cases}$$

We have already used a similar technique to encode automata, but now we discriminate paths that go through a final state. We would like to extend μ into a morphism of ω-semigroup. It is easy to extend μ into a semigroup morphism from A^+ into the multiplicative semigroup of $Q \times Q$-matrices over k. If u is a finite word, one gets

$$\mu(u)_{p,q} = \begin{cases} -\infty & \text{if there exists no path of label } u \text{ from } p \text{ to } q, \\ 1 & \text{if there exists a path from } p \text{ to } q \text{ with label } u \\ & \text{going through a final state,} \\ 0 & \text{if there exists a path from } p \text{ to } q \text{ with label } u \\ & \text{but no such path goes through a final state} \end{cases}$$

However, trouble arises when one tries to equip $k^{Q \times Q}$ with a structure of ω-semigroup. The solution consists in coding infinite paths not by square matrices, but by column matrices, in such a way that each coefficient $\mu(u)_p$ codes the existence of an infinite path of label u starting at p.

Let $S = (S_f, S_\omega)$ where $S_f = k^{Q \times Q}$ is the set of square matrices of size $\mathrm{Card}\, Q$ with entries in k and $S_\omega = k^Q$ is the set of column matrices with entries in $\{-\infty, 1\}$.

In order to define the operation ω on square matrices, we need a convenient definition. If s is a matrix of S_f, we call *infinite s-path starting at p* a sequence $p = p_0, p_1, \ldots$ of elements of Q such that, for $0 \leq i \leq n-1$, $s_{p_i, p_{i+1}} \neq -\infty$.

The s-path is *successful* if $s_{p_i, p_{i+1}} = 1$ for an infinite number of coefficients. Then s^ω is the element of S_ω defined, for every $p \in Q$, by

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

Note that the coefficients of this matrix can be effectively computed. Indeed, computing s^ω_p amounts to check the existence of circuits containing a given edge in a finite graph. Then one can verify that S, equipped with these operations, is a ω-semigroup. Furthermore, we have the following result

Proposition 2.7. *The morphism of ω-semigroup from A^∞ into S induced by μ recognizes the set $L^{\mathbb{N}}(\mathcal{A})$.*

The ω-semigroup $\mu(A^\infty)$ is called the ω-semigroup *associated with $\mathcal{A}$*.

Example 2.6. Let $X = (a\{b,c\}^* \cup \{b\})^\omega$. This set is recognized by the Büchi automaton represented below:

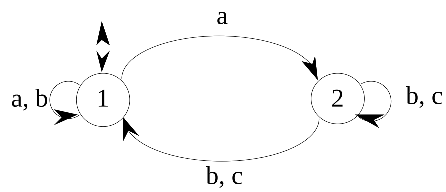

Figure 2.4.

The ω-semigroup associated with this automaton contains 9 elements

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

It is defined by the following relations:

$$\begin{array}{cccccc} a^2 = a & ab = a & ac = a & b^2 = b & bc = c & cb = c \\ c^2 = c & c^\omega = 0 & (ba)^\omega = b^\omega & aa^\omega = a^\omega & ab^\omega = a^\omega & a(ca)^\omega = a^\omega \\ ba^\omega = b^\omega & bb^\omega = b^\omega & b(ca)^\omega = (ca)^\omega & ca^\omega = (ca)^\omega & cb^\omega = (ca)^\omega & c(ca)^\omega = (ca)^\omega \end{array}$$

As in the case of finite words, there exists a minimal finite ω-semigroup (with respect to division) recognizing a given recognizable set X. This ω-semigroup is called the *syntactic ω-semigroup* of X. It can also be defined directly as the quotient of A^∞ under the congruence of ω-semigroup $\sim_X$ defined on A^+ by $u \sim_X v$ if and only if, for every $x, y \in A^*$ and for every $z \in A^+$,

$$\begin{aligned} xuyz^\omega \in X &\iff xvyz^\omega \in X \\ x(uy)^\omega \in X &\iff x(vy)^\omega \in X \end{aligned} \tag{2.1}$$

and on $A^{\mathbb{N}}$ by $u \sim_X v$ if and only if, for every $x \in A^*$,

$$xu \in X \iff xv \in X \tag{2.2}$$

One can also compute the syntactic ω-semigroup of a recognizable set given a finite Büchi automaton $\mathcal{A}$ recognizing X. One first computes the finite ω-semigroup S associated with $\mathcal{A}$ and the image P of X in S. Then the syntactic ω-semigroup is the quotient of S by the congruence $\sim_P$ defined on S_f by $u \sim_P v$ if and only if, for every $r, s \in S_f^1$ and for every $t \in S_f$

$$\begin{aligned} rust^\omega \in P &\iff rvst^\omega \in P \\ r(us)^\omega \in P &\iff r(vs)^\omega \in P \end{aligned}$$

and on S_ω by $u \sim_P v$ if and only if, for every $r \in S_f^1$,

$$ru \in P \iff rv \in P$$

This provides an algorithm to compute the syntactic ω-semigroup of a recognizable set.

For bilateral words, there is a corresponding notion of ζ-semigroup, obtained by considering infinite products indexed by intervals of $\mathbb{Z}$ instead of ω-products. Formally, a ζ-semigroup is a multisorted algebra $S = (S_f, S_\omega, S_{\tilde\omega}, S_\zeta)$ where S_f, S_ω, $S_{\tilde\omega}$, S_ζ are sets intuitively representing the finite products, the infinite products on the right, the infinite product on the left and the biinfinite products, equipped with a product $S_f \times S_f \to S_f$ and three mixed products $S_f \times S_\omega \to S_\omega$, $S_{\tilde\omega} \times S_\omega \to S_\zeta$, and $S_{\tilde\omega} \times S_f \to S_{\tilde\omega}$ satisfying the following axioms:

(1) For all $s, t, u \in S_f$, $(st)u = s(tu)$,
(2) For all $s, t \in S_f$, $u \in S_\omega$, $(st)u = s(tu)$,
(3) For all $s \in S_{\tilde\omega}$, $t, u \in S_f$, $(st)u = s(tu)$,
(4) For all $s \in S_{\tilde\omega}$, $t \in S_f$, $u \in S_\omega$, $(st)u = s(tu)$.

Theorem 2.4 can be adapted to ζ-semigroups as follows

Theorem 2.8. *Let S_f be a finite semigroup and let S_ω, $S_{\tilde\omega}$ and S_ζ be finite sets. Suppose that there exist three mixed products $S_f \times S_\omega \to S_\omega$, $S_{\tilde\omega} \times S_f \to S_{\tilde\omega}$ and $S_{\tilde\omega} \times S_\omega \to S_\zeta$ and two maps $S_f \to S_\omega$, denoted $s \to s^\omega$, and $S_f \to S_{\tilde\omega}$ denoted $s \to s^{\tilde\omega}$ satisfying, for every $s, t \in S_f$, the equations*

$$s(ts)^\omega = (st)^\omega \qquad (ts)^{\tilde\omega} = (st)^{\tilde\omega} s$$
$$(s^n)^\omega = s^\omega \qquad (s^n)^{\tilde\omega} = s^{\tilde\omega} \quad \text{for every } n > 0$$

Then the set $S = (S_f, S_\omega, S_{\tilde\omega}, S_\zeta)$ can be equipped, in a unique way, with a structure of ζ-semigroup such that for every $s \in S$, the product $sss\cdots$ is equal to s^ω and the product $\cdots sss$ is equal to $s^{\tilde\omega}$.

Then one can define the syntactic ζ-semigroup of a recognizable set. As in the case of infinite words, one can effectively compute this finite object, for any given finite Büchi automaton.

3. The sequential calculus

We now come to the major topic of this article, the definition of sets of words by logical formulæ.

3.1. Definitions

For each letter $a \in A$, let R_a denote a unary predicate. In the sequential calculus, a word u is represented as a structure of the form

$$\big(\mathrm{Dom}(u), (R_a)_{a\in A}, S, <\big),$$

where

$$\mathrm{Dom}(u) = \begin{cases} \{0, \ldots, |u| - 1\} & \text{if } u \text{ is a non empty finite word,} \\ \mathbb{N} & \text{if } u \text{ is an infinite word,} \\ \mathbb{Z} & \text{if } u \text{ is a biinfinite word,} \end{cases}$$

and where

$$R_a = \{i \in \mathrm{Dom}(u) \mid u(i) = a\}.$$

Thus, if $u = abbaab$, then $\mathrm{Dom}(u) = \{0, 1, \ldots, 5\}$, $R_a = \{0, 3, 4\}$ and $R_b = \{1, 2, 5\}$. If $u = (aba)^\omega$, then

$$R_a = \{n \in \mathbb{N} \mid n \equiv 0 \bmod 3 \text{ or } n \equiv 2 \bmod 3\} \quad \text{and} \quad R_b = \{n \in \mathbb{N} \mid n \equiv 1 \bmod 3\}.$$

We shall also use two other non-logical symbols, the binary relation symbols $<$ and S, which are interpreted as the usual order and as the successor relation on $\mathrm{Dom}(u)$.

Now terms, atomic formulæ, first order formulæ and second order formulæ are formed in the usual way. In first order logic, all variables are element variables, while in second order logic, relation variables are allowed. Monadic second order logic is the restriction of second order logic in which only element variables and set variables are allowed. Weak monadic second order logic is a variant of monadic second order logic in which set variables are restricted to range over *finite* subsets of the domain.

We shall use the notations $F_1(<)$, $MF_2(<)$, $WMF_2(<)$, $F_2(<)$, for the set of first order, monadic second order, weak monadic second order and second order formulæ with signature $\{<, (R_a)_{a\in A}\}$. Similarly, $F_1(S)$, $MF_2(S)$, $WMF_2(S)$, $F_2(S)$, denote the same sets of formulæ with signature $\{S, (R_a)_{a\in A}\}$. If we need to specify the domain, we shall use the notations $F_1(\mathbb{N}, S)$, $F_1(\mathbb{Z}, S)$, etc. In fact, the distinction between the signatures S and $<$ is irrelevant for second order theories according to the following lemma.

Lemma 3.1. *The relation S can be expressed in $F_1(<)$, and the relation $<$ can be expressed in $WMF_2(S)$ and in $MF_2(S)$.*

Proof. We give the proof for the monadic second order theory in $\mathbb{Z}$ and leave the other cases as exercises. First, $S(i, j)$ can be defined by the formula

$$(i < j) \wedge \forall k \left((i < k) \rightarrow ((j = k) \vee (j < k))\right)$$

which says that $j = i + 1$ if i is smaller than j and if there is no element between i and j. Conversely, $i < j$ can be expressed in $MF_2(S)$ as follows:

$$\exists X \left(\left[\forall x \forall y \left(((x \in X) \wedge S(x, y)) \rightarrow (y \in X)\right)\right] \wedge (j \in X) \wedge (i \notin X)\right)$$

which intuitively means there is an interval of the form $[k, +\infty[$ containing j but not i. The relation $<$ can also be expressed in $WMF_2(S)$ (left as an exercise), but not in $F_1(S)$. $\square$

To each sentence φ, one associates the sets of words that satisfy φ:

$$L^+(\varphi) = \{u \in A^+ \mid u \text{ satisfies } \varphi\}$$
$$L^{\mathbb{N}}(\varphi) = \{u \in A^{\mathbb{N}} \mid u \text{ satisfies } \varphi\}$$
$$L^{\mathbb{Z}}(\varphi) = \{u \in A^{\mathbb{Z}} \mid u \text{ satisfies } \varphi\}$$

This last definition requires a justification: normally, one should define the set of infinite words satisfying a given formula. But since this set is always shift invariant, it naturally defines a set of bilateral words.

Example 3.1. Let $\varphi = \exists i\ R_a i$. Then

$$L^+(\varphi) = A^* a A^*, \quad L^{\mathbb{N}}(\varphi) = A^* a A^\omega, \quad L^{\mathbb{Z}}(\varphi) = A^{\tilde{\omega}} a A^\omega.$$

Example 3.2. Let $\varphi = \exists \min\ \forall i\ \min \leq i \wedge\ R_a \min$. Then, according to intuition, min is interpreted as the minimum of the domain and thus

$$L^+(\varphi) = aA^*, \quad L^{\mathbb{N}}(\varphi) = aA^\omega, \quad L^{\mathbb{Z}}(\varphi) = \emptyset.$$

Example 3.3. Let

$$\varphi = \exists \min\ \forall i\ (\min \leq i)\ \wedge\ \exists \max\ \forall i\ (i \leq \max)$$
$$\wedge\ \exists X\ \big(\min \in X \wedge \max \notin X \wedge \forall i \forall j\ (S(i,j) \rightarrow (i \in X \leftrightarrow j \notin X))\big)$$

Again, min and max have their natural interpretation and the set X can be informally described as a set containing the minimum of the domain, not containing the maximum and such that $i \in X$ if and only if $i+1 \notin X$. Thus X is interpreted as the empty set over the domains $\mathbb{N}$ and $\mathbb{Z}$ (because these domains don't have any maximum). On a finite domain, X represents a set of even numbers of the form $\{0, 2, 4, 6 \ldots\}$. Since X does not contain the maximum (that is, $|u| - 1$), this number is odd, and thus $|u|$ is even. It follows that $L^{\mathbb{N}}(\varphi) = \emptyset$, $L^{\mathbb{Z}}(\varphi) = \emptyset$ and $L^+(\varphi)$ is the set of finite words of even length.

Two sentences φ and ψ are said to be *equivalent* (resp. $\mathbb{N}$-*equivalent*, $\mathbb{Z}$-*equivalent*) if $L^+(\varphi) = L^+(\psi)$ (resp. $L^{\mathbb{N}}(\varphi) = L^{\mathbb{N}}(\psi)$, $L^{\mathbb{Z}}(\varphi) = L^{\mathbb{Z}}(\psi)$).

3.2. Full second order

As we have already pointed out, there is no difference between the second order theories of $<$ and of S. The full second order theory defines the polynomial hierarchy. This hierarchy is often defined in terms of Turing machines with oracles, but a direct definition is also possible. Let A be an alphabet. A *polynomial relation* on A^* is a relation which is accepted in polynomial time by a deterministic Turing machine on the alphabet A. For every $k > 0$, we define the set $\Sigma_k^P(A^*)$ as follows. A subset L of A^* belongs to $\Sigma_k^P(A^*)$ if and only if there exist k polynomials $p_1, \ldots, p_k$ and a polynomial $(k+1)$-ary relation R such that

$$\begin{aligned} x \in L \iff & \exists y_1 \in A^* \text{ such that } |y_1| \leq p_1(|x|) \\ & \forall y_2 \in A^* \text{ such that } |y_2| \leq p_2(|x|) \\ & \quad \exists y_3 \in A^* \text{ such that } |y_3| \leq p_3(|x|) \\ & \qquad \ddots \\ & \qquad\quad Q y_k \in A^* \text{ such that } |y_k| \leq p_k(|x|) \\ & \qquad\qquad R(x, y_1, \ldots, y_k) \end{aligned}$$

where Q is an universal quantifier if k is even and an existential quantifier if k is odd. In particular, $\Sigma_0^P = P$, the class of sets recognized by deterministic Turing machines in polynomial time, and $\Sigma_1^P = NP$, the class of sets recognized by non deterministic Turing machines in polynomial time. Finally, the set $\Sigma_*^P = \bigcup_{k \geq 0} \Sigma_k^P$ is the polynomial hierarchy (sometimes also denoted by PH).

Theorem 3.2. (Stockmeyer 1977) *A subset of A^+ is definable in $F_2(<)$ if and only if it belongs to the class PH.*

Restricting to the set $\Sigma_1 F_2(<)$ of existential (second order) formulæ defines the class NP.

Theorem 3.3. (Fagin 1974) *A subset of A^+ is definable in $\Sigma_1 F_2(<)$ if and only if it belongs to the class NP.*

The question whether the class NP is strictly contained in PH is a famous open problem of complexity theory. Therefore it is not known whether existential second order formulæ are less expressive than full second order formulæ. At this point it is tempting to give a logical description of other complexity classes. The four results below give the flavor of this theory. The reader is referred to the original article of Immerman [31] for more information on this topic. Let L (resp. NL, $PSPACE$) be the class of sets accepted by a deterministic Turing machine in logarithmic space (resp. by a non deterministic Turing machine in logarithmic space, by a deterministic Turing machine in polynomial space). The following inclusions are well-known:

$$L \subset NL \subset P \subset NP \subset PH \subset PSPACE$$

Denote by RTC the operator that computes the reflexive and transitive closure of a relation. For instance, if $S(x, y)$ is the successor relation, $RTC[S(x, y)](u, v)$ is the relation $u \leq v$ and $RTC[S(x_1, x'_1) \wedge S(x_2, x'_2)](0, y, x, z)$ is the relation $x + y = z$. Given a set of formulæ F, we denote by $F + RTC$ the set of formulæ expressible using F plus the operator RTC.

Theorem 3.4. (Immerman 1987) *A subset of A^+ is definable in $F_2(<) + RTC$ if and only if it belongs to the class $PSPACE$.*

The characterization of the class NL, also due to Immerman, was originally stated in a slightly more complicated way, but can be simplified since NL is closed under complement [32,72].

Theorem 3.5. (Immerman 1987) *A subset of A^+ is definable in $F_1(S) + RTC$ if and only if it belongs to the class NL.*

To obtain the class L, we need a deterministic version of the operator RTC. Given a first order binary relation R, the deterministic part of R is the relation R_d defined by

$$R_d(x, y) \equiv R(x, y) \wedge (\forall z\ R(x, z) \Longrightarrow z = y)$$

Now define $DRTC(R)$ as the reflexive transitive closure of R_d. For instance, $S = S_d$ on words since every position has at most one successor, so that $<= DRTC[S(x, y)]$. Given a set of formulæ F, we denote by $F + DRTC$ the set of formulæ expressible using F plus the operator $DRTC$.

Theorem 3.6. (Immerman 1987, Vardi) *A subset of A^+ is definable in $F_1(S) + DRTC$ if and only if it belongs to the class L.*

We conclude this section with the class P, for which we introduce the least fixpoint operator LFP. Given a monotone operator φ on relations (that is, such that $R \subset S$ implies $\varphi(R) \subset \varphi(S)$), define

$$LFP(\varphi) = \bigcap_{\varphi(R) = R} R$$

For instance, if $\varphi(R) = (x = y) \vee (\exists z (S(x, z) \wedge R(z, y))$, then $LFP(\varphi)[R]$ is the reflexive and transitive closure of R. In particular, LFP is more powerful than RTC. As for RTC and $DRTC$, given a set of formulæ F, we denote by $F + LFP$ the set of formulæ expressible using F plus the operator LFP.

Theorem 3.7. (Immerman 1987) *A subset of A^+ is definable in $F_1(S) + LFP$ if and only if it belongs to the class P.*

These results are stated when formulæ are interpreted on finite words only. For an extension of these results to finite structures see Immerman [31] and the beautiful survey of Fagin [25].

3.3. Monadic second order

Considering only monadic second order formulæ on words is a much more drastic restriction.

Theorem 3.8. (Büchi [12], Elgot [22]) *Let L be a subset of A^+ (resp. $A^{\mathbb{N}}$, $A^{\mathbb{Z}}$). The following conditions are equivalent:*
(1) L is definable in $WMF_2(<)$,
(2) L is definable in $MF_2(<)$,
(3) L is a rational (resp. $\mathbb{N}$-rational, $\mathbb{Z}$-rational) set.
Furthermore there exists an effective algorithm to pass from a formula to a rational (resp.$\mathbb{N}$-rational, $\mathbb{Z}$-rational) expression and vice-versa.

It is fair to say that this effective algorithm is not very efficient. Indeed, it is shown in [41] that there is no elementary time bounded decision procedure for deciding whether a given sentence of $WMF_2(S)$ is true or not.

Corollary 3.9. *The theories $MF_2(\mathbb{N}, S)$, $MF_2(\mathbb{Z}, S)$, $MF_2(\mathbb{N}, <)$, $MF_2(\mathbb{Z}, <)$, and the corresponding weak monadic second order theories are decidable.*

Proof. We give the proof for $MF_2(\mathbb{N}, <)$, but the other cases are similar. Let φ be a sentence of $MF_2(\mathbb{N}, <)$. Take an empty alphabet. Then theorem 3.8 gives an effective algorithm to compute $L^{\mathbb{N}}(\varphi)$. Then φ is true on $\mathbb{N}$ if and only if $L^{\mathbb{N}}(\varphi) \neq \emptyset$. □

It is amusing to obtain Presburger's result from Büchi's theorem. The nice paper of Hodgson [28] contains other interesting decidability results derived from Büchi's theorem.

Corollary 3.10. (Presburger) *The theory $F_1(\mathbb{N}, +, 0)$ is decidable.*

Proof. The idea is to code the integers by finite subsets of $\mathbb{N}$ and then to interpret the addition within weak monadic second order logic. If

$$n = 2^{i_1} + 2^{i_2} + \cdots + 2^{i_k} \quad \text{with } i_1 < i_2 < \ldots < i_k$$

is the binary expansion of an integer n, then n is coded by the set $\{i_1, i_2, \ \ldots \ , i_k\}$. Thus the coding of an integer n is the set of positions of the bits 1 in the binary expansion of n. For instance, 13, whose binary expansion is 1101, is coded by the set $\{3, 2, 0\}$, and 20, whose binary expansion is 10100 is coded by $\{4, 2\}$.

Now, let x, y, and z be three integers, coded by the sets X, Y, and Z, respectively. Then one can code the equality $x + y = z$ by introducing a second order variable R, which codes the positions of carries in the addition. For instance, the addition of the binary expansions of 13 and 20 is represented below.

$$\begin{array}{ll} R = \{5,4,3\} & \quad \overline{1\,1\,1\,0\,0\,0} \\ X = \{4,2\} & \quad \,1\,0\,1\,0\,0 \\ Y = \{3,2,0\} & +\underline{\,0\,1\,1\,0\,1} \\ Z = \{5,0\} & \quad 1\,0\,0\,0\,0\,1 \end{array}$$

Now, i belongs to Z if and only if one or three of the sets X, Y, and R contains i. Similarly, $i+1 \in R$ if and only if i belongs to at least two of the three sets X, Y, and R. This can be easily coded into the following formula:

$$\begin{aligned} &\exists R\ (0 \notin R) \wedge \forall x\ \forall y\ \Big\{ S(x,y) \rightarrow \\ &\quad \Big[(y \in R) \leftrightarrow \big((x \in X) \wedge (x \in Y)\big) \vee \big((x \in X) \wedge (x \in R)\big) \vee \big((x \in R) \wedge (x \in Y)\big)\Big]\Big\} \wedge \\ &\quad \Big\{(x \in Z) \leftrightarrow \Big[\big((x \in R) \wedge (x \in X) \wedge (x \in Y)\big) \vee \big((x \in R) \wedge (x \notin X) \wedge (x \notin Y)\big) \vee \\ &\qquad \big((x \notin R) \wedge (x \in X) \wedge (x \notin Y)\big) \vee \big((x \notin R) \wedge (x \notin X) \wedge (x \in Y)\big)\Big]\Big\} \end{aligned}$$

We now arrive to first order theory. There, the theory branches into two different directions. We first consider the first order theory of the linear ordering and later (section 3.6) the first order theory of successor.

3.4. First order theory of the linear ordering

In this section, we present the results which connect first order logic, star-free sets and aperiodic semigroups. These statements, given the proper definitions, hold for finite, infinite and bilateral words. They summarize a series of deep results of Schützenberger [58], McNaughton and Papert [40], Ladner [34], Thomas [74], Perrin [47] and Perrin-Pin [50]. We first define the key concepts of this statement : star-free sets and aperiodic semigroups.

Boolean operations comprise union, intersection, complementation and set difference. It can be shown that the rational subsets of A^* are closed under finite boolean operations. The set of *star-free* subsets of A^* is the smallest set of subsets of A^* containing the finite sets and closed under finite boolean operations and product. For instance, if X^c denotes the complement of a set X, one has $A^* = \emptyset^c$, showing that the set A^* is star-free, since the empty set is a finite set. More generally, if B is a subset of the alphabet A, the set B^* is also star-free since B^* is the complement of the set of words that contain at least one letter of $B' = A \setminus B$. This leads to the following star-free expression

$$B^* = A^* \setminus A^*(A \setminus B)A^* = (\emptyset^c(A \setminus B)\emptyset^c)^c$$

If $A = \{a, b\}$, the set $(ab)^*$ is star-free, since $(ab)^*$ is the set of words not beginning with b, not finishing by a and containing neither the factor aa, nor the factor bb. This gives the star-free expression

$$(ab)^* = A^* \setminus \big(bA^* \cup A^*a \cup A^*aaA^* \cup A^*bbA^*\big) = \emptyset^c \setminus \big(b\emptyset^c \cup \emptyset^c a \cup \emptyset^c aa\emptyset^c \cup \emptyset^c bb\emptyset^c\big)$$

Readers may convince themselves that the sets $\{ab, ba\}^*$ and $\big(a(ab)^*b\big)^*$ are also star-free but may also wonder whether there exist any non star-free rational sets. In fact, there are some, for instance the sets $(aa)^*$ and $\{b, aba\}^*$, or similar examples that can be derived from the algebraic approach presented below.

The star-free subsets of $A^{\mathbb{N}}$ (resp $A^{\mathbb{Z}}$) are defined in a way similar to the rational subsets of $A^{\mathbb{N}}$ (resp $A^{\mathbb{Z}}$). The set of star-free subsets of $A^{\mathbb{N}}$ is the smallest set $\mathcal{S}$ of subsets of $A^{\mathbb{N}}$ closed under finite boolean operations and such that if X is a star-free subset of A^+ and $Y \in \mathcal{S}$, then $XY \in \mathcal{S}$. The set of star-free subsets of $A^{\mathbb{Z}}$ is the smallest set $\mathcal{S}$ of subsets of $A^{\mathbb{Z}}$ closed under finite boolean operations and such that if Y is a star-free subset of A^+ and X, Z are star-free subsets of $A^{\mathbb{N}}$, then $\tilde{X}YZ \in \mathcal{S}$.

A finite semigroup S is *aperiodic* if and only if it ultimately satisfies the equations $x^n = x^{n+1}$. This notion is in some sense "orthogonal" to the notion of groups. Indeed, one can show that a semigroup is aperiodic if and only if it is $\mathcal{H}$-trivial, or, equivalently, if it contains no non-trivial subgroup. Note that any finite monoid M satisfies an equation of the form $x^{n+p} = x^n$. The two extremal cases are $n = 0$ and $p = 1$. The first case corresponds to groups (if M is a group, then M satisfies the equation $x^{\mathrm{Card}(M)} = 1$), the second case to aperiodic monoids. The connection between aperiodic semigroups and star-free sets was established by Schützenberger [58] for finite words and by Perrin [47,48] for infinite and bilateral words:

Theorem 3.11. *A recognizable subset of A^+ (resp. $A^{\mathbb{N}}$, $A^{\mathbb{Z}}$) is star-free if and only if its syntactic semigroup (resp. ω-semigroup, ζ-semigroup) is aperiodic.*

Example 3.4. Let $A = \{a, b\}$ and consider the set $L = (ab)^+$. Its minimal automaton is represented below:

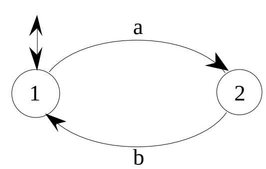

Figure 3.1. The minimal automaton of $(ab)^+$

The transitions and the relations defining the syntactic semigroup S of L are given in the following tables

	1	2
a	2	–
b	–	1
aa	–	–
ab	1	–
ba	–	2

$$a^2 = b^2 = 0$$
$$aba = a$$
$$bab = b$$

Since $a^2 = a^3$, $b^2 = b^3$, $(ab)^2 = ab$ and $(ba)^2 = ba$, S is aperiodic and thus L is star-free. Consider now the set $L' = (aa)^+$. Its minimal automaton is represented below:

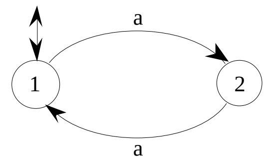

Figure 3.2. The minimal automaton of $(aa)^+$

The transitions and the relations defining the syntactic semigroup S' of L' are given in the following tables

	1	2
a	2	1
b	–	–
aa	1	2

$$a^3 = a$$
$$b = 0$$

Thus S' is not aperiodic and hence L' is not star-free.

The complexity of star-freeness is analyzed in [15,64]. Given a finite automaton, it is a PSPACE-complete problem to know whether this automaton accepts a star-free set of finite words.

Finally, the connection with first order was established by McNaughton and Papert for finite words [40], by Thomas [74] for infinite words and by Perrin and the author [50] for bilateral words.

Theorem 3.12. *Let X be a recognizable subset of A^+ (resp. $A^{\mathbb{N}}$, $A^{\mathbb{Z}}$). Then the following conditions are equivalent:*

(1) X is definable in $F_1(<)$,
(2) X is star-free,
(3) the syntactic semigroup (resp. ω-semigroup, ζ-semigroup) of X is aperiodic.

Since the characterization by syntactic semigroups is effective, one gets the following decidability result.

Corollary 3.13. *It is decidable whether a given sentence of $MF_2(<)$ is equivalent (resp. $\mathbb{N}$-equivalent, $\mathbb{Z}$-equivalent) to some formula of $F_1(<)$.*

3.5. A hierarchy of first order formulæ of the linear ordering

An extension of Theorem 3.12 was discovered by Thomas [76] for finite words and later extended to infinite and bilateral words [50]. It turns out that the hierarchy of star-free sets obtained by counting the number of alternations between concatenation and boolean operations coincides with the hierarchy of first order formulæ in terms of quantifiers alternations. In this section, we give a refined version of this beautiful result and we discuss related problems.

We first define the logical hierarchy. Any first order formula is equivalent to a formula in normal prenex form, that is to say, of the form $\varphi = Q(x_1, \ldots, x_k)\psi$ where $Q(x_1, \ldots, x_k)$ is a sequence of existential or universal quantifiers on the variables $x_1, \ldots x_k$ and where ψ is quantifier free. If $Q(x_1, \ldots, x_k)$ is formed of n blocks of quantifiers such that the first block contains only existential quantifiers (note that this first block may be empty), the second block universal quantifiers, etc., we say that φ is a Σ_n-formula. We denote by Σ_n the set of the Σ_n-formulæ and by $\mathcal{B}\Sigma_n$ the set of boolean combinations(*) of Σ_n-formulæ.

We now turn to the hierarchy of recognizable sets. Let A be an alphabet. The definition of star-free subsets of A^* makes use of two different types of operations: boolean operations and concatenation product. By alternating the use of these two operations, one gets a hierarchy, called the *concatenation hierarchy*, defined as follows.

(1) The sets of level 0 are the empty set $\emptyset$ and A^*,

(2) For every integer $n \geq 0$, the sets of level $n + 1/2$ are the finite unions of sets of the form

$$L_0 a_1 L_1 a_2 \cdots a_k L_k$$

where L_0, L_1, ..., L_k are sets of level n and a_1, ..., a_k are letters

(3) For every integer $n \geq 0$, the sets of level $n + 1$ are the finite boolean combinations of sets of level $n + 1/2$.

Note that a set of level m is also a set of level n for every $n \geq m$. The next result summarizes several results relative to this hierarchy.

Theorem 3.14. (Brzozowski and Knast [10], Perrin and Pin [50])

(1) For every $n \geq 0$, the sets of level n are closed under union, intersection, and complement.

(2) For every $n \geq 0$, the sets of level $n + 1/2$ are closed under union, intersection, and concatenation product.

(3) The hierarchy is strict: if A contains at least two letters, then for every n, there exist some sets of level $n + 1$ that are not of level $n + 1/2$ and some sets of level $n + 1/2$ that are not of level n.

The concatenation hierarchy can be extended to infinite and bilateral words as follows.

(1) The sets of level 0 are the empty set $\emptyset$ and $A^{\mathbb{N}}$ (resp. $A^{\mathbb{Z}}$),

(2) For every integer $n \geq 0$, the sets of level $n + 1/2$ are the finite unions of sets of the form XaY, where X is a set of A^* (resp. $A^{-\mathbb{N}}$) of level $n + 1/2$, Y is a subset of $A^{\mathbb{N}}$ of level n and a is a letter.

(3) For every $n \geq 0$, the sets of level $n + 1$ are the finite boolean combinations of sets of level $n + 1/2$.

Here is the announced connection between the two hierarchies.

(*) boolean operations on formulæ comprise conjunction, disjunction and negation.

Theorem 3.15. (Thomas [76], Perrin and Pin [50]) *A subset of A^+ (resp. $A^{\mathbb{N}}$, $A^{\mathbb{Z}}$) is of level n (resp. $n + 1/2$) if and only if it is $\mathcal{B}\Sigma_n$-definable (resp. Σ_{n+1}-definable).*

We now describe in more details the first levels of this hierarchy. By definition, the sets of level 1/2 are the finite unions of sets of the form $A^*a_1A^*a_2 \cdots a_kA^*$, where the a_i's are letters, and the sets of level 1 are the finite boolean combinations of the same sets. In particular, finite sets are of level 1. The sets of level 3/2 and 2 have a similar description, but this is not a direct consequence of the definition.

Theorem 3.16. (Pin and Straubing [55], Arfi [2,3]) *The sets of level 3/2 of A^* are the finite unions of sets of the form $A_0^*a_1A_1^*a_2 \cdots a_kA_k^*$, where the a_i's are letters and the A_i's are subsets of A. The sets of level 2 are the finite boolean combinations of the same sets.*

The sets of level 1 have a nice algebraic characterization.

Theorem 3.17. (Simon [61]) *A subset of A^+ has level 1 if and only if its syntactic semigroup is $\mathcal{J}$-trivial, or, equivalently, if it ultimately satisfies the equations $x^n = x^{n+1}$ and $(xy)^n = (yx)^n$.*

There exist several proofs of this deep result [1,61,69,63].

Example 3.5. Let $A = \{a, b, c\}$ and let $L = A^*abA^*$. The minimal automaton of L is represented below

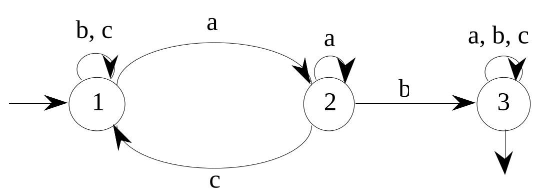

Figure 3.3. The minimal automaton of L.

The transitions and the relations defining the syntactic semigroup S of L are given in the following tables

	1	2	3
a	2	2	3
b	1	3	3
c	1	1	3
ab	3	3	3
ba	2	3	3

$a^2 = a$
$ab = 0$
$ac = c$
$b^2 = b$
$bc = b$
$ca = a$
$cb = c$
$c^2 = c$

The $\mathcal{J}$-class structure of S is represented in the following diagram.

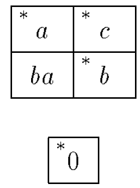

Figure 3.4. The $\mathcal{J}$-classes of S

In particular, $a \mathcal{J} c$ and thus S is not $\mathcal{J}$-trivial. Therefore L has level greater than 1 (in fact 2). Consider now the set $L' = A^*abA^*$ on the alphabet $A = \{a, b\}$. Then the minimal automaton of L' is obtained from that of L by erasing the transitions with label c.

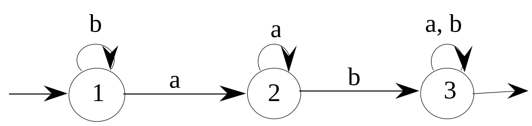

Figure 3.5. The minimal automaton of L'.

The transitions and the relations defining the syntactic semigroup S' of L' are given in the following tables

	1	2	3
a	2	2	3
b	1	3	3
ab	3	3	3
ba	2	3	3

$$a^2 = a$$
$$ab = 0$$
$$b^2 = b$$

The $\mathcal{J}$-class structure of S' is represented in the following diagram.

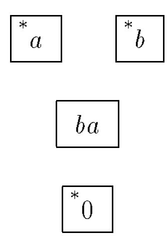

Figure 3.6. The $\mathcal{J}$-classes of S'.

Thus S' is $\mathcal{J}$-trivial and L' has level 1. In fact $L' = A^*aA^*bA^*$.

Theorem 3.17 gives an algorithm to decide whether a given recognizable set (*) is of level 1. The complexity of this algorithm is analyzed in [15,64]. Given a finite automaton, the problem to know whether it recognizes a set of finite words of level 1 is in P and is logspace-complete for NL. Levels 1/2 and 3/2 are also decidable. This is relatively easy to show for the level 1/2, but relies on a deep result of Hashiguchi for the level 3/2.

(*) A recognizable set can be given either by a finite automaton or by a finite semigroup or by a rational expression since there are standard algorithms to pass from one representation to the other.

Theorem 3.18. (Arfi [2,3]) *One can effectively decide whether a given recognizable set of* A^* *is of level* $1/2$, 1 *or* $3/2$.

Corollary 3.19. *It is decidable whether a given sentence of* $MF_2(<)$ *is equivalent to a formula of* Σ_1 *(resp.* $\mathcal{B}\Sigma_1$, Σ_2*).*

The decidability problem for levels 2 and beyond is still open, although much progress has been made on level 2 in the recent years [55,68,81,71,82,19]. Little is known beyond level 2: a semigroup theoretic description of each level of the hierarchy is known [53], but it is not an effective one. In other words, each level admits a description by ultimate equations similar to Theorem 3.17, but these ultimate equations are not known for $n \geq 2$. Furthermore, even if these equations were known, this would not necessarily lead to a decision process for the corresponding variety.

It is not known whether similar results hold for infinite and bilateral words.

3.6. First order theory of successor

We now turn to the first order theory of successor, the characterization of which is more involved than that of the linear ordering. It seems that in this theory, the poorer the logic, the more sophisticated the results!

We need to introduce some new classes of recognizable sets. We first treat the case of finite words. Let $x \in A^+$ and $k \geq 0$. We let

$$F(x,k) = \{u \in A^+ \mid u \text{ contains at least } k \text{ occurrences of } x\}$$

For instance, $F(x,1) = A^*xA^*$ and $F(aba,2) = A^*abaA^*abaA^* \cup A^*ababaA^*$. A set of words of A^+ is *strongly locally testable* (SLT) if it is a boolean combination of sets of the form $F(x,1)$ where $x \in A^+$. It is *locally testable* (LT) if it is a boolean combination of sets of the form uA^*, A^*v or A^*xA^* where $u, v, x \in A^+$. For instance, if $A = \{a,b\}$, the set $(ab)^+$ is locally testable since

$$(ab)^+ = (aA^* \cap A^*b) \setminus (A^*aaA^* \cup A^*bbA^*)$$

More generally, we say that a set of words of A^+ is *strongly locally threshold testable* (SLTT) if it is a boolean combination of sets of the form $F(x,k)$ where $x \in A^+$ and $k > 0$. It is *locally threshold testable* (LTT) if it is a boolean combination of sets of the form uA^*, A^*v or $F(x,k)$ where $u, v, x \in A^+$ and $k > 0$. These families of sets are deeply related to the first order theory of successor.

Theorem 3.20. (Thomas [76]) *A recognizable subset of* A^+ *is definable in* $F_1(S)$ *if and only if it is locally threshold testable.*

Theorem 3.21. (Beauquier and Pin [9]) *A recognizable subset of* A^+ *is definable by a boolean combination of existential formulæ of* $F_1(S)$ *if and only if it is strongly locally threshold testable.*

In fact, these results are particular instances of the general fact that first order formulas can express only local properties [26,78,79].

We now give some effective characterizations of the families of sets introduced above. In order to keep a standard notation in subsequent statements, we shall denote by L a recognizable subset of A^+, by $S(L)$ the syntactic semigroup of L, by $\varphi : A^+ \to S(L)$ the syntactic morphism of L, and by $P(L) = \varphi(L)$ the syntactic image of L. A finite semigroup S is said to be *locally idempotent and commutative* if, for every idempotent e of S, the subsemigroup $eSe = \{ese \mid s \in S\}$ is idempotent and commutative. Equivalently, S is locally idempotent and commutative if, for every $e, s, t \in S$ such that $e = e^2$, $(ese)^2 = (ese)$ and $(ese)(ete) = (ete)(ese)$. We can now state

Theorem 3.22. (Brzozowski and Simon[11], McNaughton[39]) *A recognizable subset L of A^+ is locally testable if and only if $S(L)$ is locally idempotent and commutative.*

Let S be a finite semigroup and let P be a subset of S. We say that P *saturates* the $\mathcal{J}$-classes of S if, for every $\mathcal{J}$-class J of S, $s \in P$ and $s \,\mathcal{J}\, t$ imply $t \in P$.

Theorem 3.23. (Beauquier and Pin [9]) *Let L be a recognizable subset of A^+. Then L is strongly locally testable if and only if $S(L)$ is locally idempotent and commutative and $P(L)$ saturates the $\mathcal{J}$-classes of $S(L)$.*

Example 3.6. Let $A = \{a, b, c\}$, and let $L = c(ab)^* \cup c(ab)^*a$. Then L is recognized by the following automaton.

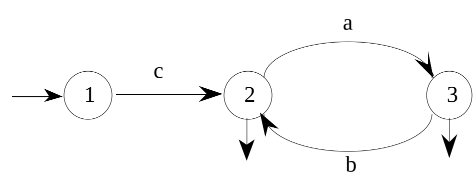

Figure 3.7. An automaton recognizing L.

The transitions and the relations defining the syntactic semigroup S of L are given in the following tables

	1	2	3
a	–	3	–
b	–	–	2
c	2	–	–
aa	–	–	–
ab	–	2	–
ba	–	–	3
ca	3	–	–

$$a^2 = b^2 = c^2 = ac = bc = cb = 0$$
$$aba = a$$
$$bab = b$$
$$cab = c$$

The $\mathcal{J}$-class structure is represented in the following diagram, where the grey box is the image of L.

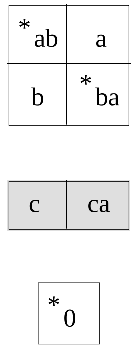

Figure 3.8. The $\mathcal{J}$-class structure.

Thus $P(L)$ saturates the $\mathcal{J}$-classes, and L is SLT. In fact, $L = A^*cA^* \setminus (A^*aaA^* \cup A^*acA^* \cup A^*bbA^* \cup A^*bcA^* \cup A^*cbA^* \cup A^*ccA^*)$.

The syntactic characterization of locally threshold testable sets is more involved and depends on a deep result of Thérien and Weiss [73]. Given a semigroup S, form a graph $G(S)$ as follows: the vertices are the idempotents of S and the edges from e to f are the elements of the form esf.

Theorem 3.24. (Beauquier and Pin [8]) *Let L be a recognizable subset of A^+. Then L is locally threshold testable if and only if $S(L)$ is aperiodic and its graph satisfies the following condition: if p and r are edges from e to f and if q is an edge from f to e, then $pqr = rqp$.*

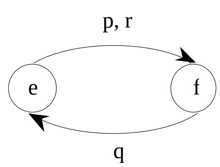

Figure 3.9. The condition $pqr = rqp$.

Example 3.7. Let $A = \{a, b\}$ and let $L = a^*ba^*$. Then L is recognized by the automaton shown in figure 3.7.

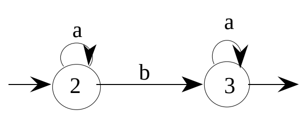

Figure 3.10. The minimal automaton of a^*ba^*.

The transitions and the relations defining the syntactic semigroup S of L are given in the following tables

	1	2
a	1	2
b	2	–
bb	–	–

$$a = 1$$
$$b^2 = 0$$

Thus $S = \{a, b, 0\}$, where $a = 1$ is the identity and $E(S) = \{1, 0\}$. The local semigroups are $0S0 = \{0\}$ and $1S1 = S$. This last local semigroups is not idempotent, since $b^2 \neq b$. Therefore, L is not locally testable. On the other hand, the graph of $S(L)$, represented in Figure 3.11, satisfies the condition $pqr = rqp$.

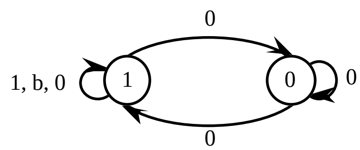

Figure 3.11. The graph of $S(L)$.

Therefore L is locally threshold testable.

We conjecture that an analogous result holds for SLTT subsets of A^+.

Corollary 3.25. *It is decidable whether a given sentence of $MF_2(S)$ is equivalent to a formula of $F_1(S)$.*

We now turn to bilateral words, for which similar classes of recognizable sets can be defined. For each $x \in A^+$ and let $k \geq 0$, we set

$$F^{\mathbb{Z}}(x, k) = \{u \in A^{\mathbb{Z}} \mid u \text{ contains at least } k \text{ occurrences of } x\}$$
$$R^{\mathbb{Z}}(x, \infty) = \{u \in A^{\mathbb{Z}} \mid u \text{ contains infinitely many occurrences of } x \text{ on the right }\}$$
$$L^{\mathbb{Z}}(x, \infty) = \{u \in A^{\mathbb{Z}} \mid u \text{ contains infinitely many occurrences of } x \text{ on the left }\}$$

Note that $F^{\mathbb{Z}}(x, 1) = A^{\tilde{\omega}} x A^{\omega}$. A set of words of $A^{\mathbb{Z}}$ is *strongly locally threshold testable* (SLTT) if it is a boolean combination of sets of the form $F^{\mathbb{Z}}(x, k)$ where $x \in A^+$ and $k > 0$. It is *locally threshold testable* (LTT) if it is a boolean combination of sets of the form $R^{\mathbb{Z}}(x, \infty)$, $L^{\mathbb{Z}}(x, \infty)$ or $F^{\mathbb{Z}}(x, k)$. In the case of bilateral words, the Σ_n hierarchy collapses.

Theorem 3.26. (Beauquier and Pin [9]) *Let X be a recognizable subset of $A^{\mathbb{Z}}$. Then the following conditions are equivalent :*

(1) X is strongly locally threshold testable,
(2) X is definable in $F_1(S)$,
(3) X is definable by a boolean combination of existential formulæ of $F_1(S)$.

Although Theorem 3.22 and 3.23 have been extended to bilateral words (Pécuchet [45], Beauquier and Pin [9]), no effective characterization of the SLTT subsets of $A^{\mathbb{Z}}$ is known.

Finally, we consider the case of infinite words. For each $x \in A^+$ and let $k \geq 0$, we set

$$F^{\mathbb{N}}(x, k) = \{u \in A^{\mathbb{N}} \mid u \text{ contains at least } k \text{ occurrences of } x\}$$
$$R^{\mathbb{N}}(x, \infty) = \{u \in A^{\mathbb{N}} \mid u \text{ contains infinitely many occurrences of } x \text{ on the right }\}$$

A set of words of $A^{\mathbb{N}}$ is *strongly locally threshold testable* (SLTT) if it is a boolean combination of sets of the form $F^{\mathbb{N}}(x, k)$ where $x \in A^+$ and $k > 0$. It is *locally threshold testable* (LTT) if it is a boolean combination of sets of the form $uA^{\mathbb{N}}$, $R^{\mathbb{N}}(x, \infty)$ or $F^{\mathbb{N}}(x, k)$. Finally, it is *left locally threshold testable* (LLTT) if it is a boolean combination of sets of the form uA^{ω} or $F^{\mathbb{N}}(x, k)$. The following results follow from the results of Thomas.

Theorem 3.27. *A recognizable subset of $A^{\mathbb{N}}$ is definable by a boolean combination of existential formulæ of $F_1(S)$ if and only if it is strongly locally threshold testable.*

Theorem 3.28. (Thomas [76]) *A recognizable subset of $A^{\mathbb{N}}$ is definable in $F_1(S)$ if and only if it is left locally threshold testable.*

Again these results are just variations on the theme that first order formulas can express only local properties [26,78,79].

An effective characterization of this last class has been obtained by Wilke. First Theorem 3.24 can be extended to infinite words as follows

Theorem 3.29. (Wilke [84,85]) *A recognizable subset of $A^{\mathbb{N}}$ is locally threshold testable if and only if its syntactic ω-semigroup is aperiodic and its graph satisfies the following condition: if p and r are edges from e to f and if q is an edge from f to e, then $pqr = rqp$.*

The characterization of LLTT sets is more involved, but is effective. It relies on the nice equality

$$LLTT = LTT \cap \Delta_2^0$$

where Δ_2^0 denotes the second level of the Borel hierarchy. The open sets of $A^{\mathbb{N}}$ are the sets of the form $XA^{\mathbb{N}}$ for some $X \subset A^+$. A set is closed if its complement is open. The sets of Δ_2^0 are at the same time countable unions of closed sets and countable intersection of open sets. The recognizable sets of Δ_2^0 are the sets which are accepted by a deterministic Büchi automaton as well as their complement. This class is also decidable, leading to the following conclusion.

Theorem 3.30. (Wilke [84,85]) *One can effectively decide whether a given recognizable subset of $A^{\mathbb{N}}$ is LLTT.*

Corollary 3.31. *It is decidable whether a given sentence of $MF_2(S)$ is $\mathbb{N}$-equivalent to a formula of $F_1(S)$.*

3.7. Modular quantifiers

An interesting extension of the first order theory is obtained by introducing *modular quantifiers* of the form $\exists^{k,r}$. For instance, the sentence $\exists^{5,2} x R_a x$ is interpreted to mean "there are exactly 2 mod 5 positions x such that the letter in position x is an a". Denote by $F_{MOD}(<)$ (resp. $F_{1+MOD}(<)$) the set of formulæ defined by using only modular quantifiers (resp. by using both the ordinary quantifiers and the modular ones). The corresponding classes of languages are decidable, according to the following result.

Theorem 3.32. (Straubing, Thérien and Thomas [70]) *A recognizable subset of A^+ is definable in $F_{MOD}(<)$ (resp. $F_{1+MOD}(<)$) if and only if its syntactic semigroup is a solvable group (resp. a semigroup in which every group is solvable).*

These results can probably be extended in some way to infinite or biinfinite words.

3.8. Elementary equivalence

Given a set of sentences S, we say that two words u and v are S-equivalent if and only if they satisfy exactly the same formulæ of S. This equivalence has been especially studied for bilateral words, for which the next two theorems give a rather exhaustive description.

Theorem 3.33. Perrin and Schupp [52] *Let u, v be two words of $A^{\mathbb{Z}}$. The following conditions are equivalent:*
(1) u and v are either equal or recurrent with the same set of factors,
(2) u and v satisfy the same sentences of $F_1(\mathbb{Z}, <)$,
(3) u and v satisfy the same sentences of $F_2(\mathbb{Z}, <)$.

For the theory of successor, the corresponding result, first conjectured by Parikh, is even simpler.

Theorem 3.34. [8] *Let u, v be two words of $A^{\mathbb{Z}}$. The following conditions are equivalent:*
(1) u and v have the same set of factors,
(2) u and v satisfy the same sentences of $F_1(\mathbb{Z}, S)$,
(3) u and v satisfy the same existential sentences of $F_1(\mathbb{Z}, S)$.

This also is an easy consequence of the results on first-order theory of local structure mentioned above [26,78,79].

4. Temporal logic

An alternative way to define sets of words is to use *propositional linear temporal logic*, a logic used for specifying and verifying correction of computer programs. This logic is intended to represent the structure of time and the basic operators (next time, eventually and until) refer to this intuitive interpretation. As for sequential calculus, to each finite or infinite word is associated a totally ordered set, which represents the structure of time. Thus, in this model, time is always discrete. We follow the notations of [23]. Let A be a finite alphabet. The vocabulary consists of
(1) an atomic proposition p_a for each letter $a \in A$,
(2) connectives $\vee$, $\wedge$ and $\neg$,
(3) temporal operators $\mathbf{X}$ ("next"), $\mathbf{F}$ ("eventually") and $\mathbf{U}$ ("until")
and the formulæ are constructed according to the rules
(1) for every $a \in A$, p_a is a formula,
(2) if φ and ψ are formulæ, so are $\varphi \vee \psi$, $\varphi \wedge \psi$, $\neg\varphi$, $\mathbf{X}\varphi$, $\mathbf{F}\varphi$, $\varphi\mathbf{U}\psi$.
Semantics are defined by induction on the formation rules. Given a word $w \in A^+$, and $n \in \{1, 2, ..., |w|\}$, we define the expression "w satisfies φ at the instant n" (denoted $(w, n) \models \varphi$) as follows
(1) $(w, n) \models p_a$ if the n-th letter of w is an a.
(2) $(w, n) \models \varphi \vee \psi$ (resp. $\varphi \wedge \psi$, $\neg\varphi$) if $(w, n) \models \varphi$ or $(w, n) \models \psi$ (resp. if $(w, n) \models \varphi$ and $(w, n) \models \psi$, if (w, n) does not satisfy φ).
(3) $(w, n) \models \mathbf{X}\varphi$ if $(w, n + 1)$ satisfies φ.
(4) $(w, n) \models \mathbf{F}\varphi$ if there exists m such that $n \leq m \leq |w|$ and $(w, m) \models \varphi$.
(5) $(w, n) \models \varphi\mathbf{U}\psi$ if there exists m such that $n \leq m \leq |w|$, $(w, m) \models \psi$ and, for every k such that $n \leq k < m$, $(w, k) \models \varphi$.
Note that, if $w = w_0 w_1 \cdots w_{|w|}$, $(w, n) \models \varphi$ only depends on the word $w = w_n w_{n+1} \cdots w_{|w|}$. If φ is a temporal formula, we say that w satisfies φ if $(w, 0) \models \varphi$.

Example 4.1. Let $w = abbababcba$. Then $(w,3) \models p_a$ since the fourth letter of w is an a, $(w,3) \models \mathbf{X}p_b$ since the fifth letter of w is a b and $(w,3) \models \mathbf{F}(p_c \wedge \mathbf{X}p_b)$ since cb is a factor of $babcba$.

To each temporal formula φ, one associates the sets of words that satisfy φ:

$$L^+(\varphi) = \{u \in A^+ \mid u \text{ satisfies } \varphi\}$$
$$L^{\mathbb{N}}(\varphi) = \{u \in A^{\mathbb{N}} \mid u \text{ satisfies } \varphi\}$$

In fact, one can show [29,27] that linear temporal logic is equivalent to $F_1(<)$. It follows that one can effectively decide whether a given recognizable set is definable in linear temporal logic. A direct proof of this result is given in [17] for finite words and in [16] for infinite words.

Theorem 4.1. *A subset of A^+ (resp. $A^{\mathbb{N}}$) is definable in linear temporal logic if and only if its (ω)-syntactic semigroup is aperiodic.*

If we omit the "until" operator, we obtain a restricted linear temporal logic that was considered in [27,29]. An effective description of the sets of words definable in this logic is known in the case of finite words.

Proposition 4.2. [17] *Let L be a subset of A^+. The following conditions are equivalent:*
(1) L is definable in restricted linear temporal logic,
(2) L belongs to the smallest boolean algebra of sets containing the languages aA^* and closed under the operations $L \rightarrow A^*L$ and $L \rightarrow aL$ for every $a \in A$,
(3) the syntactic semigroup of L ultimately satisfies the sequence of equations $x^{n+1} = x^n$ and $ux^n(vx^nux^n)^n = (ux^nvx^n)^n$.

There is a simple algebraic interpretation of these equations. Let S be a finite semigroup. Then for every idempotent e, the set $eSe = \{ese \mid s \in S\}$ is a subsemigroup of S, called the *local subsemigroup associated with e*. A semigroup satisfies *locally* a property P if every local subsemigroup satisfies this property. Now a semigroup S satisfies the equations $x^{n+1} = x^n$ and $ux^n(vx^nux^n)^n = (ux^nvx^n)^n$ if and only if it is locally $\mathcal{L}$-trivial.

It remains to find an analogous characterization in the case of infinite words.

5. Conclusion

We have given several description of standard complexity classes in terms of logical formulæ. These results are summarized in the next table.

Formulas	**Complexity classes**
$F_2(<)+$ Reflexive Transitive Closure	PSPACE
$F_2(<)$	PH
$\Sigma_1 F_2(<)$	NP
$F_1(S)+$ Least Fixed Point	P
$F_1(S)+$ Reflexive Transitive Closure	NL
$F_1(S)+$ Deterministic Reflexive Transitive Closure	L

Logical characterizations of some standard complexity classes.

One enters the world of finite automata by considering the monadic second order case.

Formulas	**Finite words**	**Infinite words**	**Bilateral words**
$WMF_2(<)$	Rational	$\mathbb{N}$-Rational	$\mathbb{Z}$-Rational
$MF_2(<)$	Rational	$\mathbb{N}$-Rational	$\mathbb{Z}$-Rational

The monadic second order.

The first order hierarchy of the linear order has been studied intensively for finite words although the decidability of levels 2 and beyond remains open.

Formulas	**Sets of words**	**Algebraic characterization**	**Decidable**
$F_1(<)$	Star-free	aperiodic syntactic semigroup	Yes
$\mathcal{B}\Sigma_0(<)$	$\emptyset, A^+$	trivial syntactic semigroup	Yes
$\Sigma_1(<)$	level 1/2	Yes (technical)	Yes
$\mathcal{B}\Sigma_1(<)$	level 1	$\mathcal{J}$-trivial syntactic semigroup	Yes
$\Sigma_2(<)$	level 3/2	Yes (very difficult)	Yes
$\mathcal{B}\Sigma_2(<)$	level 2	Yes (but non effective so far)	?

The first order hierarchy of $<$.

Definability in the first order theory of the successor is decidable for finite and infinite words, but the case of bilateral words is not yet worked out.

Formulas	**Finite words**	**Infinite words**	**Bilateral words**
$F_1(S)$	LTT (decidable)	LLTT (decidable)	SLTT (?)
$\mathcal{B}\Sigma_1(S)$	SLTT (?)	SLTT (?)	SLTT (?)

The first order hierarchy of S.

We have also discussed the expressivity of linear temporal logic and of restricted linear temporal logic.

Acknowledgements

I would like to thank Jean Goubault, Howard Straubing, Wolfgang Thomas, Pascal Weil and an anonymous referee for comments and suggestions that greatly improved the quality of this article.

6. References

[1] J. Almeida, 1990, Implicit operations on finite $\mathcal{J}$-trivial semigroups and a conjecture of I. Simon, *J. Pure Appl. Algebra* **69**, 205–218.

[2] M. Arfi, 1987, Polynomials operations and relational languages, 4th STACS, Lecture Notes in Computer Science **247**, Springer, 198–206.

[3] M. Arfi, 1991, Opérations polynomiales et hiérarchies de concaténation, *Theoret. Comput. Sci.* **91**, 71–84.

[4] A. Arnold, 1985, A syntactic congruence for rational ω-languages, *Theoret. Comput. Sci.* **39**, 333–335.

[5] D. Beauquier, 1984, Bi-limites de langages reconnaissables, *Theoret. Comput. Sci.* **33**, 335–342.

[6] D. Beauquier, 1985, Ensembles reconnaissables de mots bi-infinis, in *Automata on Infinite Words, M. Nivat ed., Lecture Notes in Computer Science 192*, 28–46, Springer.

[7] D. Beauquier and M. Nivat, 1985, About rational sets of factors of a biinfinite word, in *Automata, Languages and Programming*, (W.Brauer, ed.), Lecture Notes in Comput. Sci., **194**, Springer, 33-42.

[8] D. Beauquier and J.E. Pin, 1989, Factors of words, in *Automata, Languages and Programming*, (G. Ausiello, M. Dezani-Ciancaglini and S. Ronchi Della Rocca, eds.), Lecture Notes in Comput. Sci., **372**, Springer, 63–79.

[9] D. Beauquier and J.E. Pin, 1991, Languages and scanners, *Theoret. Comput. Sci.* **84** 3–21.

[10] J.A. Brzozowski and R. Knast, 1978, The dot-depth hierarchy of star-free languages is infinite, *J. Comput. System Sci.* **16** 37–55.

[11] J.A. Brzozowski and I. Simon, 1973, Characterizations of locally testable languages, *Discrete Math.* **4**, 243-271.

[12] J.R. Büchi, 1960, Weak second-order arithmetic and finite automata, *Z. Math. Logik und Grundl. Math.* **6**, 66–92.

[13] J.R. Büchi, 1962, On a decision method in restricted second-order arithmetic, in *Proc. 1960 Int. Congr. for Logic, Methodology and Philosophy of Science*, Stanford Univ. Press, Standford, 1–11.

[14] J.R. Büchi and L.H. Landweber, 1969, Definability in the monadic second-order theory of successors, *J. Symbolic Logic* **34** 166–170.

[15] S. Cho and D.T. Huynh, 1991, Finite automaton aperiodicity is PSPACE-complete, *Theoret. Comput. Sci.* **88** 99–116.

[16] J. Cohen, 1991, On the expressive power of temporal logic for infinite words, *Theoret. Comput. Sci.* **83**, 301–312.

[17] J. Cohen, D. Perrin and J.-E. Pin, 1993, On the expressive power of temporal logic, *J. Comput. System Sci.* **46**, 271–294.

[18] K. Compton, 1983, On rich words, in *Combinatoric Words, Progress and Perspectives*, L. Cummings ed., Academic Press, 39–62.

[19] D. Cowan, 1993, Inverse monoids of dot-depth 2, *Int. Jour. Alg. and Comp.*, to appear.

[20] S. Eilenberg, 1974, *Automata, languages and machines*, Vol. A, Academic Press, New York.

[21] S. Eilenberg, 1976, *Automata, languages and machines*, Vol. B, Academic Press, New York.

[22] C.C. Elgot, 1961, Decision problems of finite automata design and related arithmetics, *Trans. Amer. Math. Soc.* **98**, 21–52.

[23] E.A. Emerson, 1990, Temporal and modal logic, in *Handbook of Theoretical Computer Science, vol B, Formal models and semantics*, Elsevier, 995–1072.

[24] R. Fagin, 1974, Generalized first-order spectra and polynomial-time recognizable sets, *SIAM-AMS Proceedings* **7**, 43–73.

[25] R. Fagin, 1993, Finite-model theory - a personal perspective *Theoret. Comp. Sci.* **116**, 3–31.

[26] H. Gaifman, 1982, On local and non-local properties, in *Proc. of the Herbrandt Symposium*, Logic Colloquium'81 (J. Stern, ed.), Studies in Logic **107**, North-Holland, Amsterdam, 105–135.

[27] D. Gabbay, A. Pnueli, S. Shelah and J. Stavi, 1980, On the temporal analysis of fairness, *Proc. 7th ACM Symp. Princ. Prog. Lang.*, 163–173.

[28] B.R. Hodgson, 1983, Décidabilité par automate fini, *Ann. Sc. Math. Québec* **7** 39–57.

[29] J.A. Kamp, 1968, Tense logic and the theory of linear order, Ph. D. Thesis, Univ. of California, Los Angeles.

[30] S.C. Kleene, 1956, Representation of Events in nerve nets and finite automata, in "*Automata Studies*" C.E. Shannon and J. McCarthy eds.), Princeton University Press, Princeton, New Jersey, 3–42.

[31] N. Immerman, 1987, Languages that capture complexity classes, *SIAM J. Comput.* **16**, 760–778.

[32] N. Immerman, 1988, Nondeterministic space is closed under complement, *SIAM J. Comput.* **17**, 935–938.

[33] J. Justin and G. Pirillo, 1986, On a natural extension of Jacob's rank, *J. Comb. Theory* **43**, 205–218

[34] R. Ladner, 1977, Application of model theoretic games to discrete linear orders and finite automata, *Information and Control*, **33** 281–303.

[35] G. Lallement, 1979, *Semigroups and combinatorial applications*, Wiley, New York.

[36] H. Landweber, 1967, Finite state games - A solvability algorithm for restricted second-order arithmetic, *Notices Amer. Math. Soc.* **14**, 129–130.

[37] H. Landweber, 1969, Decision problems for ω-automata, *Math. Syst. Theor.* **3**, 376–384.

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

[39] R. McNaughton, 1974, Algebraic decision procedures for local testability, *Math. Syst. Theor.* **8**, 60-76.

[40] R. McNaughton and S. Pappert, 1971, *Counter-free Automata,* MIT Press.

[41] A.R. Meyer, 1975, Weak monadic second order theory of successor is not elementary recursive, in Proc; Boston Univ. Logic Colloquium, Lecture Notes in Mathematics 453, Springer-Verlag, Berlin Heidelberg, New-York, 132–154.

[42] C. Michaux and F. Point, 1986, Les ensembles k-reconnaissables sont définissables dans $\langle \mathbb{N}, +, V_k \rangle$, *C. R. Acad. Sc. Paris* **303**, 939–942.

[43] M. Nivat and D. Perrin, 1986, Ensembles reconnaissables de mots biinfinis, *Canad. J. Math.* XXXVIII, 513–537.

[44] J.P. Pécuchet, 1986, Variétés de semigroupes et mots infinis, in B. Monien and G. Vidal-Naquet eds., *STACS 86, Lecture Notes in Comput. Sci.* **210**, Springer, 180–191.

[45] J.P. Pécuchet, 1986, Etude syntaxique des parties reconnaissables de mots infinis, in *Proc. 13th ICALP, (L. Kott ed.) Lecture Notes in Comput. Sci.* **226**, 294–303.

[46] D. Perrin, 1982, Variétés de semigroupes et mots infinis, *C.R. Acad. Sci. Paris* **295**, 595–598.

[47] D. Perrin, 1984, Recent results on automata and infinite words, in *Mathematical Foundations of Computer Science*, Lecture Notes in Comput. Sci. **176** Springer-Verlag, New-York/Berlin, 134–148.

[48] D. Perrin, 1984, An introduction to automata on infinite words, in *Automata on Infinite Words* (Nivat, M. ed.), Lecture Notes in Comput. Sci. **192**, Springer, 2–17.

[49] D. Perrin, 1990, *Automata*, Chapter 1 in Handbook of Theoretical Computer Science (Van Leeuwen, J. ed.), Vol B : Formal Models and Semantics, Elsevier.

[50] D. Perrin and J.E. Pin, First order logic and star-free sets, *J. Comput. System Sci.* **32**, 1986, 393–406.

[51] D. Perrin and J.-E. Pin, *Mots infinis*, to appear (LITP report 93–40).

[52] D. Perrin and P.E. Schupp, 1986, Automata on the integers, recurrence distinguishability and the equivalence and decidability of monadic theories, *Proc. 1st IEEE Symp. on Logic in Computer Science*, 301–304.

[53] J.-E. Pin, 1984, Hiérarchies de concaténation, *RAIRO Informatique Théorique* **18** 23–46.

[54] J.-E. Pin, 1984 *Variétés de langages formels*, Masson, Paris; English translation: 1986, *Varieties of formal languages*, Plenum, New-York.

[55] J.-E. Pin and H. Straubing, 1981, Monoids of upper triangular matrices, *Colloquia Mathematica Societatis Janos Bolyai 39, Semigroups, Szeged*, 259–272.

[56] W.V. Quine, 1946, Concatenation as a basis for arithmetic, *J. Symbolic Logic* **11**, 105–114.

[57] F.D. Ramsey, 1929, On a problem of formal logic, *Proc. of the London Math. Soc.* **30**, 338–384.

[58] M.P. Schützenberger, 1965, On finite monoids having only trivial subgroups, *Information and Control* **8**, 190–194.

[59] A.L. Semenov, 1980, On certain extensions of the arithmetic of addition of natural numbers, *Math. USSR Izvestiya* **15** 401–418.

[60] D. Siefkes, 1970, Büchi's monadic second order successor arithmetic, Lecture Notes in Math. **120**, Springer, Berlin.

[61] I. Simon, 1975, Piecewise testable events, *Proc. 2nd GI Conf., Lect. Notes in Comp. Sci.* **33**, Springer, Berlin, 214–222.

[62] I. Simon, 1990, Factorization forests of finite height, *Theoret. Comput. Sci.* **72**, 65–94.

[63] J. Stern, 1985, Characterization of some classes of regular events, *Theoret. Comp. Sci.* **35**, 17–42.

[64] J. Stern, 1985, Complexity of some problems from the theory of automata, *Inform. and Control* **66**, 163–176.

[65] L. Stockmeyer, 1977, The polynomial time hierarchy, *Theoret. Comp. Sci.* **3**, 1–22.

[66] L.J. Stockmeyer and A.R. Meyer, 1973, Word problems requiring exponential time: preliminary report, in *Proc. 5th Ann. ACM Symp. on the Theory of Computing*, 1–9.

[67] H. Straubing, 1979, Aperiodic homomorphisms and the concatenation product of recognizable sets, *Journal of Pure and Applied Algebra* **15**, 319–327.

[68] H. Straubing, 1988, Semigroups and languages of dot-depth two, *Theoret. Comp. Sci.* **58**, 361–378.

[69] H. Straubing and D. Thérien, 1985, Partially ordered finite monoids and a theorem of I. Simon, *J. of Algebra* **119**, 393–399.

[70] H. Straubing, D. Thérien and W. Thomas, 1988, Regular Languages Defined with Generalized Quantifiers, in *Proc. 15th ICALP*, Springer Lecture Notes in Computer Science **317**, 561–575.

[71] H. Straubing and P. Weil, 1992, On a conjecture concerning dot-depth two languages, *Theoret. Comp. Sci.* **104**, 161–183.

[72] R. Szelepcsényi, 1988, The method of foreced enumeration for nondeterministic automata, *Acta Informatica* **26**, 279–284.

[73] D. Thérien and A. Weiss, 1985, Graph congruences and wreath products, *J. Pure Applied Algebra* **35**, 205-215.

[74] W. Thomas, 1979, Star-free regular sets of ω-sequences, *Information and Control* **42** 148–156.

[75] W. Thomas, 1981, A combinatorial approach to the theory of ω-automata, *Information and Control* **48**, 261–283.

[76] W. Thomas, 1982, Classifying regular events in symbolic logic, *J. Comput. Syst. Sci* **25**, 360–375.

[77] W. Thomas, 1990, Automata on infinite objects, in *Handbook of Theoretical Computer Science, vol B, Formal models and semantics*, Elsevier, 135–191.

[78] W. Thomas, 1991, On logics, tilings, and automata, *Proc. 18th ICALP, Madrid, (J. Leach Albert et al., eds.), Lect. Notes in Comp. Sci.* **510**, Springer, Berlin, 441–454.

[79] W. Thomas, 1993, On the Ehrenfeucht-Fraïssé Game in Theoretical Computer Science, *TAPSOFT'93, M.C. Gaudel, J.P. Jouannaud (Eds.), Lect. Notes in Comp. Sci.* **668**, Springer, Berlin, 559–568.

[80] P. Weil, 1988, *Inverse monoids and the dot-depth hierarchy*, Ph. D. Dissertation, University of Nebraska, Lincoln.

[81] P. Weil, 1989, Inverse monoids of dot-depth two, *Theoret. Comp. Sci.* **66**, 233–245.

[82] P. Weil, Some results on the dot-depth hierarchy, to appear in *Semigroup Forum*.

[83] Wilke, T., 1991, An Eilenberg theorem for ∞-languages, Automata, Languages and Programming, Proc. of 18th ICALP Conference, LNCS **510**, Springer, Berlin, 588–599.

[84] Th. Wilke, 1992, An algebraic theory for regular languages of finite and infinite words, *International Journal of Algebra and Computation*, to appear.

[85] Th. Wilke, 1993, Locally threshold testable languages of infinite words, in *STACS 93, P. Enjalbert, A. Finkel, K.W. Wagner (Eds.), Lect. Notes in Comp. Sci.* **665**, Springer, Berlin, 607–616.