Syntactic semigroups

Jean-Eric Pin

LITP/IBP, CNRS, Université Paris VI,
4 Place Jussieu, 75252 Paris Cedex 05,
FRANCE

1 Introduction

This chapter gives an overview on what is often called the algebraic theory of finite automata. It deals with languages, automata and semigroups, and has connections with model theory in logic, boolean circuits, symbolic dynamics and topology.

Kleene's theorem [70] is usually considered as the foundation of this theory. It shows that the class of *recognizable* languages (i.e. recognized by finite automata), coincides with the class of *rational* languages, which are given by rational expressions. The definition of the *syntactic monoid*, a monoid canonically attached to each language, was first given in an early paper of Rabin and Scott [128], where the notion is credited to Myhill. It was shown in particular that a language is recognizable if and only if its syntactic monoid is finite. However, the first classification results were rather related to automata [89]. A break-through was realized by Schützenberger in the mid sixties [144]. Schützenberger made a non trivial use of the syntactic monoid to characterize an important subclass of the rational languages, the *star-free* languages. Schützenberger's theorem states that a language is star-free if and only if its syntactic monoid is finite and aperiodic.

This theorem had a considerable influence on the theory. Two other important "syntactic" characterizations were obtained in the early seventies: Simon [152] proved that a language is piecewise testable if and only if its syntactic monoid is finite and $\mathcal{J}$-trivial and Brzozowski-Simon [41] and independently, McNaughton [86] characterized the locally testable languages. These successes settled the power of the semigroup approach, but it was Eilenberg who discovered the appropriate framework to formulate this type of results [53].

A variety of finite monoids is a class of monoids closed under taking submonoids, quotients and finite direct product. Eilenberg's theorem states that varieties of finite monoids are in one to one correspondence with certain classes of recognizable languages, the varieties of languages. For instance, the rational languages correspond to the variety of all finite monoids, the star-free languages correspond to the variety of finite aperiodic monoids, and the piecewise testable languages correspond to the variety of finite $\mathcal{J}$-trivial monoids. Numerous similar results have been established during the past fifteen years and, for this reason, the theory of finite automata is now intimately related to the theory of finite monoids.

It is time to mention a sensitive feature of this theory, the role of the empty word. Indeed, there are two possible definitions for a language. A first definition consists in defining a language on the alphabet A as a subset of the free monoid A^*. In this case a language may contain the empty word. In the second definition, a language is defined as a subset of the free semigroup A^+, which excludes the empty word. This subtle distinction has dramatic consequences on the full theory. First, one has to distinguish between $*$-varieties (the first case) and $+$-varieties of languages (the latter case). Next, with the latter definition, monoids have to be replaced by semigroups and Eilenberg's theorem gives a one to one correspondence between varieties of finite semigroups and $+$-varieties of languages. Although it might seem quite annoying to have two such parallel theories, this distinction proved to be necessary. For instance, the locally testable languages form a $+$-variety, which correspond to locally idempotent and commutative semigroups. But no characterization of the locally testable languages is known in terms of syntactic monoids.

An extension of the the notion of syntactic semigroup (or monoid) was recently proposed in [112]. The key idea is to define a partial order on syntactic semigroups, leading to the notion of ordered syntactic semigroups. The resulting extension of Eilenberg's variety theory permits to treat classes of languages that are not necessarily closed under complement, a major difference with the original theory. We have adopted this new point of view throughout this chapter. For this reason, even our definition of recognizable languages may seem unfamiliar to the reader.

The theory has now developed into many directions and has generated a rapidly growing literature. The aim of this chapter is to provide the reader with an overview of the main results. As these results are nowadays intimately related with non commutative algebra, a certain amount of semigroup theory had to be introduced, but we tried to favor the main ideas rather than the technical developments. Some important topics had to be skipped and are briefly mentioned in the last section. Due to the lack of place, no proofs are given, but numerous examples should help the reader to catch the spirit of the more abstract statements. The references listed at the end of the chapter are far from being exhaustive. However, most of the references should be reached by the standard recursive process of tracing the bibliography of the papers cited in the references.

The chapter is organized as follows. The amount of semigroup theory that is necessary to state precisely the results of this chapter is introduced in Section 2. The basic concepts of recognizable set and ordered syntactic semigroup are introduced in Section 3. The variety theorem is stated in Section 4 and examples follow in Section 5. Some algebraic tools are presented in Section 6. Sections 7 and 8 are devoted to the study of the concatenation product and its variants. Connections with the theory of codes are discussed in Section 9. Section 10 gives an overview on the operators on recognizable languages. Various extensions are briefly reviewed in Section 11.

2 Definitions

We review in this section the basic definitions about relations and semigroups needed in this chapter.

2.1 Relations

A relation $\mathcal{R}$ on a set S is *reflexive* if, for every $x \in S$, $x \mathrel{\mathcal{R}} x$, *symmetric*, if, for every $x, y \in S$, $x \mathrel{\mathcal{R}} y$ implies $y \mathrel{\mathcal{R}} x$, and *transitive*, if, for every $x, y, z \in S$, $x \mathrel{\mathcal{R}} y$ and $y \mathrel{\mathcal{R}} z$ implies $x \mathrel{\mathcal{R}} z$. A *quasi-order* is a reflexive and transitive relation. An *equivalence relation* is a reflexive, symmetric and transitive relation. Given a quasi-order $\mathcal{R}$, the relation $\sim$ defined by $x \sim y$ if and only if $x \mathrel{\mathcal{R}} y$ and $y \mathrel{\mathcal{R}} x$ is an equivalence relation, called the *equivalence relation associated with* $\mathcal{R}$. If this equivalence relation is the equality relation, that is, if, for every $x, y \in S$, $x \mathrel{\mathcal{R}} y$ and $y \mathrel{\mathcal{R}} x$ implies $x = y$, then the relation $\mathcal{R}$ is an *order*.

Relations are naturally ordered by inclusion. Let $\mathcal{R}_1$ and $\mathcal{R}_2$ be two relations on a set S. The relation $\mathcal{R}_2$ is *coarser* than $\mathcal{R}_1$ if and only if, for every $s, t \in S$, $s \mathrel{\mathcal{R}_1} t$ implies $s \mathrel{\mathcal{R}_2} t$. In particular, the coarsest relation is the universal relation.

2.2 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 an identity element (usually denoted by 1). If S is a semigroup, S^1 denotes the monoid equal to S if S has an identity element and to $S \cup \{1\}$ otherwise. In the latter case, the multiplication on S is extended by setting $s1 = 1s = s$ for every $s \in S^1$.

A relation on a semigroup S is *stable on the right* (resp. *left*) if, for every $x, y, z \in S$, $x \mathrel{\mathcal{R}} y$ implies $xz \mathrel{\mathcal{R}} yz$ (resp. $zx \mathrel{\mathcal{R}} zy$). A relation is *stable* if it is stable on the right and on the left. A *congruence* is a stable equivalence relation. Thus, an equivalence relation $\sim$ on S is a congruence if and only if, for every $s, t \in S$ and $x, y \in S^1$, $s \sim t$ implies $xsy \sim xty$. If $\sim$ is a congruence on S, then there is a well-defined multiplication on the quotient set $S/{\sim}$, given by

$$[s][t] = [st]$$

where $[s]$ denotes the $\sim$-class of $s \in S$.

An *ordered semigroup* is a semigroup S equipped with a stable order relation $\leq$ on S. Ordered monoids are defined analogously. The notation $(S, \leq)$ will sometimes be used to emphasize the role of the order relation, but most of the time the order will be implicit and the notation S will be used for semigroups as well as for ordered semigroups. If $S = (S, \leq)$ is an ordered semigroup, then $(S, \geq)$ is also an ordered semigroup, called the *dual* of S and denoted $\check{S}$.

A *congruence on an ordered semigroup* S is a stable quasi-order which is coarser than $\leq$. In particular, the order relation $\leq$ is itself a congruence. If $\preceq$ is a congruence on S, then the equivalence relation $\sim$ associated with $\preceq$ is a congruence on S. Furthermore, there is a well-defined stable order on the quotient set $S/{\sim}$, given by

$$[s] \leq [t] \quad \text{if and only if } s \preceq t$$

Thus $(S/{\sim}, \leq)$ is an ordered semigroup, also denoted $S/{\preceq}$.

Given a family $(S_i)_{i \in I}$ of ordered semigroups, the product $\prod_{i \in I} S_i$ is the ordered semigroup defined on the set $\prod_{i \in I} S_i$ by the law

$$(s_i)_{i \in I}(s'_i)_{i \in I} = (s_i s'_i)_{i \in I}$$

and the order given by

$$(s_i)_{i \in I} \leq (s'_i)_{i \in I} \text{ if and only if, for all } i \in I,\ s_i \leq s'_i.$$

Products of semigroups, monoids and ordered monoids are defined similarly.

If M is a monoid, the set $\mathcal{P}(M)$ of the subsets of M is a monoid under the operation

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

2.3 Morphisms

Generally speaking, a morphism between two algebraic structures is a map that preserves the operations and the relations of the structure. This general definition applies in particular to semigroups, monoids, ordered semigroups and ordered monoids. Given two semigroups S and T, a *semigroup morphism* $\varphi : S \to T$ is a map from S into T such that for all $x, y \in S$, $\varphi(xy) = \varphi(x)\varphi(y)$. *Monoid morphisms* are defined analogously, but of course, the condition $\varphi(1) = 1$ is also required. A *morphism of ordered semigroups* $\varphi : S \to T$ is a semigroup morphism from S into T such that, for every $x, y \in S$, $x \leq y$ implies $\varphi(x) \leq \varphi(y)$.

A morphism of semigroups (resp. monoids, ordered semigroups) $\varphi : S \to T$ is an *isomorphism* if there exists a morphism of semigroups (resp. monoids, ordered semigroups) $\psi : T \to S$ such that $\varphi \circ \psi = Id_T$ and $\psi \circ \varphi = Id_S$. It is easy to see that a morphism of semigroups is an isomorphism if and only if it is bijective. This is not true for morphisms of ordered semigroups. In particular, if $(S, \leq)$ is an ordered semigroup, the identity induces a bijective morphism from $(S, =)$ onto $(S, \leq)$ which is not in general an isomorphism. In fact, a morphism of ordered semigroups $\varphi : S \to T$ is an isomorphism if and only if φ is a bijective semigroup morphism and, for every $x, y \in S$, $x \leq y$ is equivalent with $\varphi(x) \leq \varphi(y)$.

For every semigroup morphism $\varphi : S \to T$, the equivalence relation $\sim_\varphi$ defined on S by setting $s \sim_\varphi t$ if and only if $\varphi(s) = \varphi(t)$ is a semigroup congruence. Similarly, for every morphism of ordered semigroups $\varphi : S \to T$, the quasi-order $\preceq_\varphi$ defined on S by setting $s \preceq_\varphi t$ if and only if $\varphi(s) \leq \varphi(t)$ is a congruence of ordered semigroup, called the *nuclear congruence* of φ.

A semigroup (resp. monoid, ordered semigroup) S is a *quotient* of a semigroup (resp. monoid, ordered semigroup) T if there exists a surjective morphism from T onto S. In particular, if $\sim$ is a congruence on a semigroup S, then $S/\sim$ is a quotient of S and the map $\pi : S \to S/\sim$ defined by $\pi(s) = [s]$ is a surjective morphism, called the *quotient morphism* associated with $\sim$. Similarly, let $\preceq$ be a congruence on an ordered semigroup $(S, \leq)$ and let $\sim$ be the equivalence relation associated with $\preceq$. Then $(S/\preceq)$ is a quotient of $(S, \leq)$ and the map $\pi : S \to S/\preceq$ defined by $\pi(s) = [s]$ is a surjective morphism of ordered semigroups.

Let $\sim_1$ and $\sim_2$ be two congruences on a semigroup S and let $\pi_1 : S \to S/\sim_1$ and $\pi_2 : S \to S/\sim_2$ be the quotient morphisms. Then $\sim_2$ is coarser than $\sim_1$ if and only if π_2 *factorizes through* π_1, that is, if there exists a surjective morphism $\pi : S/\sim_1 \to S/\sim_2$ such that $\pi \circ \pi_1 = \pi_2$.

A similar result holds for ordered semigroups. Let $\preceq_1$ and $\preceq_2$ be two congruences on an ordered semigroup S and let $\pi_1 : S \to S/\preceq_1$ and $\pi_2 : S \to S/\preceq_2$ be the quotient morphisms. Then $\preceq_2$ is coarser than $\preceq_1$ if and only if π_2 factorizes through π_1.

Let S be a semigroup (resp. ordered semigroup). A *subsemigroup* (resp. an ordered subsemigroup) of S is a subset T of S such that $t, t' \in T$ implies $tt' \in T$. Subsemigroups are closed under intersection. In particular, given a subset E of S, the smallest subsemigroup of S containing E is called the subsemigroup of S *generated* by G.

A semigroup S *divides* a semigroup T if S is a quotient of a subsemigroup of T. Division is a quasi-order on semigroups. Furthermore, one can show that two finite semigroups divide each other if and only if they are isomorphic.

2.4 Groups

A *group* is a monoid in which every element has an inverse. We briefly recall some standard definitions of group theory. Let p be a prime number. A p-group is a finite group whose order is a power of p. If G is a group, let $G_0 = G$ and $G_{n+1} = [G_n, G]$, the subgroup generated by the commutators $hgh^{-1}g^{-1}$, where $h \in G_n$ and $g \in G$. A finite group is *nilpotent* if and only if $G_n = \{1\}$ for some $n \geq 0$. A finite group is *solvable* if and only if there is a sequence

$$G = G^{(0)}, G^{(1)}, \ldots, G^{(k)} = \{1\}$$

such that, for each $i \geq 0$, $G^{(i+1)}$ is a normal subgroup of $G^{(i)}$ and the quotient $G^{(i)}/G^{(i+1)}$ is commutative. It is a well-known fact that every p-group is nilpotent and every nilpotent group is solvable. See [143]

2.5 Free semigroups

Let A be a finite alphabet. The set of words on A is denoted A^* and the set of non-empty words, A^+. Thus $A^* = A^+ \cup \{1\}$, where 1 is the empty word. The length of a word u is denoted $|u|$. If a is a letter, $|u|_a$ denotes the number of occurrences of a in u. In particular, $|u| = \sum_{a \in A} |u|_a$. A word p is a *prefix* of a word u if $u = pu'$ for some $u' \in A^*$. Symmetrically, a word s is a *suffix* of u if $u = u's$ for some $u' \in A^*$. A word x is a *factor* of u if there exist two words u' and u'' (possibly empty) such that $u = u'xu''$. This notion should not be confused with the notion of subword. A word $a_1 \cdots a_n$ (where the a_i's are letters) is a *subword* of u if $u = u_0 a_1 u_1 \cdots a_n u_n$ for some words $u_0, \ldots, u_n \in A^*$.

The semigroup A^+ is the free semigroup on A and $(A^+, =)$ is the free ordered semigroup on A. Indeed, if $\varphi : A \to S$ is a function from A into an ordered semigroup S, there exists a unique morphism of ordered semigroups $\bar{\varphi} : (A^+, =) \to S$ such that $\varphi(a) = \bar{\varphi}(a)$ for every $a \in A$. Moreover $\bar{\varphi}$ is surjective if and only if $\varphi(A)$ is a generator set of S. It follows that if $\eta : (A^+, =) \to S$ is a morphism of ordered semigroups and $\beta : T \to S$ is a surjective morphism of ordered semigroups, there exists a morphism of ordered semigroups $\varphi : (A^+, =) \to T$ such that $\eta = \beta \circ \varphi$. This property is known as the *universal property* of the free ordered semigroup. Similarly, A^* is the free monoid on A and $(A^*, =)$ is the free ordered monoid on A.

As was explained in the introduction, there are two parallel notions of languages. If we are working with monoids, a *language* on A is a subset of the free monoid A^*. If semigroups are considered, a *language* on A is a subset of the free semigroup A^+.

2.6 Order ideals

An *order ideal* of an ordered semigroup S is a subset I of S such that, if $x \leq y$ and $y \in I$, then $x \in I$. The order ideal *generated by an element* x is the set $\downarrow x$ of all $y \in S$ such that $y \leq x$. The intersection (resp. union) of any family of order ideals is also an order ideal. Furthermore, if I is an order ideal and K is an arbitrary subset of S^1, then the left quotient $K^{-1}I$ and the right quotient IK^{-1} are also order ideals. Recall that, for each subset X of S and for each element s of S^1, the *left (resp. right) quotient* $s^{-1}X$ (resp. Xs^{-1}) of X by s is defined as follows:

$$s^{-1}X = \{t \in S \mid st \in X\} \quad \text{and} \quad Xs^{-1} = \{t \in S \mid ts \in X\}$$

More generally, for any subset K of S^1, the left (resp. right) quotient $K^{-1}X$ (resp. XK^{-1}) of X by K is

$$\begin{aligned} K^{-1}X &= \textstyle\bigcup_{s \in K} s^{-1}X = \{t \in S \mid \text{ there exists } s \in K \text{ such that } st \in X\} \\ XK^{-1} &= \textstyle\bigcup_{s \in K} Xs^{-1} = \{t \in S \mid \text{ there exists } s \in K \text{ such that } ts \in X\} \end{aligned}$$

2.7 Idempotents

An element e of a semigroup S is *idempotent* if $e^2 = e$. A semigroup is *idempotent* if all its elements are idempotent. In this chapter, we will mostly use finite semigroups, in which idempotents play a key role. In particular, the following proposition shows that every non empty finite semigroup contains an idempotent.

Proposition 2.1 *Let s be an element of a finite semigroup. Then the subsemigroup generated by s contains a unique idempotent and a unique maximal subgroup, whose identity is the unique idempotent.*

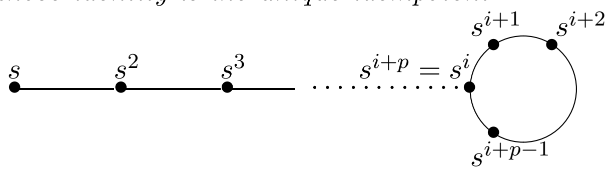

Figure 1: The semigroup generated by s.

If s is an element of a finite semigroup, the unique idempotent power of s is denoted s^ω. If e is an idempotent of a finite semigroup S, the set

$$eSe = \{ese \mid s \in s\}$$

is a subsemigroup of S, called the *local subsemigroup* associated with e. This semigroup is in fact a monoid, since e is an identity in eSe.

A finite semigroup S is said to satisfy *locally* a property $\mathcal{P}$ if every local subsemigroup of S satisfies $\mathcal{P}$. For instance, S is *locally trivial* if, for every idempotent $e \in S$ and every $s \in S$, $ese = e$.

A *zero* is an element 0 such that, for every $s \in S$, $s0 = 0s = 0$. It is a routine exercise to see that there is at most one zero in a semigroup. A non-empty finite semigroup that contains a zero and no other idempotent is called *nilpotent*.

2.8 Green's relations

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

$s \,\mathcal{L}\, t$	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 an $\mathcal{L}$-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 fact that, in finite semigroups, the relations $\mathcal{R}$ and $\mathcal{L}$ commute. The presence of an idempotent in an $\mathcal{H}$-class is indicated 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.

* a_1, a_2	* a_3, a_4	a_5, a_6
b_1, b_2	* b_3, b_4	* b_5, b_6

A $\mathcal{J}$-class.

In this figure, each row is an $\mathcal{R}$-class and each column is an $\mathcal{L}$-class. There are 6 $\mathcal{H}$-classes and 4 idempotents. Each idempotent is the identity of a group of order 2.

A $\mathcal{J}$-class containing an idempotent is called *regular*. One can show that in a regular $\mathcal{J}$-class, every $\mathcal{R}$-class and every $\mathcal{L}$-class contains an idempotent.

A semigroup S is *$\mathcal{L}$-trivial* (resp. *$\mathcal{R}$-trivial, $\mathcal{J}$-trivial, $\mathcal{H}$-trivial*) if two elements of S which are $\mathcal{L}$-equivalent (resp. $\mathcal{R}$-equivalent, $\mathcal{J}$-equivalent, $\mathcal{H}$-equivalent) are equal. See [75, 102] for more details.

2.9 Categories

Some algebraic developments in semigroup theory motivate the introduction of categories as a generalization of monoids. A *category* C is given by

(1) a set $Ob(C)$ of *objects*,

(2) for each pair (u, v) of objects, a set $C(u, v)$ of *arrows*,

(3) for each triple (u, v, w) of objects, a mapping from $C(u, v) \times C(v, w)$ into $C(u, w)$ which associates to each $p \in C(u, v)$ and $q \in C(v, w)$ the composition $pq \in C(u, w)$.

(4) for each object u, an arrow 1_u such that, for each pair (u, v) of objects, for each $p \in C(u, v)$ and $q \in C(v, u)$, $1_u p = p$ and $q 1_u = q$.

Composition is assumed to be associative (when defined).

For each object u, $C(u,u)$ is a monoid, called the *local monoid* of u. In particular a monoid can be considered as a category with exactly one object. A category is said to be *locally idempotent* (resp. *locally commutative*, etc.) if all its local monoids are idempotent (resp. commutative, etc.).

If C and D are categories, a *morphism of categories* $\varphi : C \to D$ is defined by the following data:

(1) an *object map* $\varphi : Ob(C) \to Ob(D)$,

(2) for each pair (u,v) of objects of C, an *arrow map* $\varphi : C(u,v) \to D(\varphi(u), \varphi(v))$ such that

   (a) for each triple (u,v,w) of objects of C, for each $p \in C(u,v)$ and $q \in C(v,w)$, $\varphi(pq) = \varphi(p)\varphi(q)$

   (b) for each object u, $\varphi(1_u) = 1_{\varphi(u)}$.

A category C is a *subcategory* of a category D if there exists a morphism $\varphi : C \to D$ which is injective on objects and on arrows (that is, for each pair of objects (u,v), the arrow map from $C(u,v)$ into $D(\varphi(u), \varphi(v))$ is injective). A category C is a *quotient* of a category D if there exists a morphism $D \to C$ which is bijective on objects and surjective on arrows. Finally C *divides* D if C is a quotient of a subcategory of D.

3 Recognizability

Recognizable languages are usually defined in terms of automata. This is the best definition from an algorithmic point of view, but it is an asymmetric notion. It turns out that to handle the fine structure of recognizable languages, it is more appropriate to use a more abstract definition, using semigroups in place of automata, due to Rabin and Scott [128]. However, we will slightly modify this standard definition by introducing ordered semigroups. As will be shown in the next sections, this order occurs quite naturally and permits to distinguish between a language and its complement. Although these definitions will be mainly used in the context of free semigroups, it is as simple to give them in a more general setting.

3.1 Recognition by ordered semigroups

Let $\varphi : S \to T$ be a surjective morphism of ordered semigroups. A subset Q of S is *recognized* by φ if there exists an order ideal P of T such that

$$Q = \varphi^{-1}(P)$$

This condition implies that Q is an order ideal of S and that $\varphi(Q) = \varphi\varphi^{-1}(P) = P$. By extension, a subset Q of S is said to be *recognized* by an ordered semigroup T if there exists a surjective morphism of ordered semigroups from S onto T that recognizes Q.

It is sometimes convenient to formulate this definition in terms of congruences. Let S be an ordered semigroup and let $\preceq$ a congruence on S. A subset Q of S is said to be *recognized* by $\preceq$ if, for every $q \in Q$, $p \preceq q$ implies $p \in Q$. It is easy to see that a surjective morphism of ordered semigroups φ recognizes Q if and only if the nuclear congruence $\preceq_\varphi$ recognizes Q.

Simple operations on subsets have a natural algebraic counterpart. We now study in this order intersection, union, complement, inverse morphisms and left and right quotients.

Proposition 3.1 *Let $(\eta_i : S \to S_i)_{i \in I}$ be a family of surjective morphisms of ordered semigroups. If each η_i recognizes a subset Q_i of S, then the subsets $\cap_{i \in I} Q_i$ and $\cup_{i \in I} Q_i$ are recognized by an ordered subsemigroup of the product $\prod_{i \in I} S_i$.*

If P is an order ideal of an ordered semigroup S, the set $S \setminus P$ is not, in general, an order ideal of S. However, it is an order ideal of the dual of S.

Proposition 3.2 *Let P be an order ideal of an ordered semigroup $(S, \leq)$. Then $S \setminus P$ is an order ideal of $(S, \geq)$. If P is recognized by a morphism of ordered semigroups $\eta : (S, \leq) \to (T, \leq)$, then $S \setminus P$ is recognized by the morphism of ordered semigroups $\eta : (S, \geq) \to (T, \geq)$.*

Proposition 3.3 *Let $\varphi : R \to S$ and $\eta : S \to T$ be two surjective morphisms of ordered semigroups. If η recognizes a subset Q of S, then $\eta \circ \varphi$ recognizes $\varphi^{-1}(Q)$.*

Proposition 3.4 *Let $\eta : S \to T$ be a surjective morphism of ordered semigroups. If η recognizes a subset Q of S, it also recognizes $K^{-1}Q$ and QK^{-1} for every subset K of S^1.*

3.2 Syntactic order

The syntactic congruence is one of the key notions of this chapter. Roughly speaking, it is the semigroup analog of the notion of minimal automaton. First note that, if S is an ordered semigroup, the congruence $\leq$ recognizes every order ideal of S. The syntactic congruence of an order ideal Q of S is the coarsest congruence among the congruences on S that recognize Q.

Let T be an ordered semigroup and let P be an order ideal of T. Define a relation $\preceq_P$ on T by setting

$$u \preceq_P v \text{ if and only if, for every } x, y \in T^1,\ xvy \in P \Rightarrow xuy \in P$$

One can show that the relation $\preceq_P$ is a congruence of ordered semigroups on T that recognizes P. This congruence is called the *syntactic congruence* of P in T. The equivalence relation associated with $\preceq_P$ is denoted $\sim_P$ and called the *syntactic equivalence* of P in T. Thus $u \sim_P v$ if and only if, for every $x, y \in T^1$,

$$xuy \in P \iff xvy \in P$$

The ordered semigroup $S(P) = T/\preceq_P$ is the *ordered syntactic semigroup* of P, the order relation on $S(P)$ the *syntactic order* of P and the quotient morphism η_P from T onto $S(P)$ the *syntactic morphism* of P. The syntactic congruence is characterized by the following property.

Proposition 3.5 *The syntactic congruence of P is the coarsest congruence that recognizes P. Furthermore, a congruence $\preceq$ recognizes P if and only if $\preceq_P$ is coarser than $\preceq$.*

It is sometimes convenient to state this result in terms of morphisms:

Corollary 3.6 *Let $\varphi : R \to S$ be a surjective morphism of ordered semigroups and let P be an order ideal of R. The following properties hold:*

(1) *The morphism φ recognizes P if and only if η_P factorizes through it.*

(2) *Let $\pi : S \to T$ be a surjective morphism of ordered semigroups. If $\pi \circ \varphi$ recognizes P, then φ recognizes P.*

3.3 Recognizable sets

A subset of an ordered semigroup is *recognizable* if it is recognized by a finite ordered semigroup. Propositions 3.1 and 3.4 show that the recognizable subsets of a given ordered semigroup are closed under finite union, finite intersection and left and right quotients.

In the event where the order relation on S is the equality relation, every subset of S is an order ideal, and the definition of a recognizable set can be slightly simplified.

Proposition 3.7 *Let S be a semigroup. A subset P of S is recognizable if and only if there exists a surjective semigroup morphism φ from S onto a finite semigroup F and a subset Q of F such that $P = \varphi^{-1}(Q)$.*

The case of the free semigroup is of course the most important. In this case, the definition given above is equivalent with the standard definition using finite automata. Recall that a finite (non deterministic) automaton is a quintuple $\mathcal{A} = (Q, A, E, I, F)$, where A denotes the alphabet, Q the set of states, E is the set of transitions (a subset of $Q \times A \times Q$), and I and F are the set of *initial* and *final* states, respectively. An automaton $\mathcal{A} = (Q, A, E, I, F)$ is *deterministic* if I is a singleton and if the conditions $(p, a, q), (p, a, q') \in E$ imply $q = q'$.

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

$$e_0 = (q_0, a_0, q_1), \ \ e_1 = (q_1, a_1, q_2), \ \ \ldots \ \ , \ \ e_{n-1} = (q_{n-1}, a_{n-1}, q_n)$$

also denoted

$$q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \ \cdots \ q_{n-1} \xrightarrow{a_{n-1}} q_n$$

The state q_0 is the *origin* of the path, the state q_n is its *end*, and the word $x = a_0 a_1 \cdots a_{n-1}$ is its *label*. It is convenient to have also, for each state q, an empty path of label 1 from q to q. A path in $\mathcal{A}$ is *successful* if its origin is in I and its end is in F.

The language of A^* *recognized* by $\mathcal{A}$ is the set of the labels of all successful paths of $\mathcal{A}$. In the case of the free semigroup A^+, the definitions are the same, except that we omit the empty paths of label 1.

Automata are conveniently represented by labeled graphs, as in the example below. Incoming arrows indicate initial states and outgoing arrows indicate final states.

Example 3.8 Let $\mathcal{A} = (Q, A, E, I, F)$ be the automaton represented below, with $Q = \{1, 2\}$, $A = \{a, b\}$, $I = \{1\}$, $F = \{2\}$ and

$$E = \{(1, a, 1), (1, a, 2), (2, a, 2), (2, b, 2), (2, b, 1)\}.$$

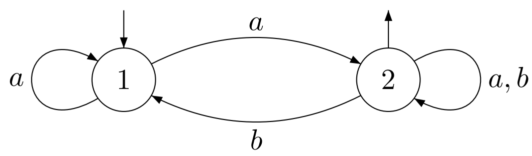

Figure 2: A non deterministic automaton.

The path $(1, a, 1)(1, a, 2)(2, b, 2)$ is a successful path of label aab. The path $(1, a, 1)(1, a, 2)(2, b, 1)$ has the same label but is unsuccessful since its end is 1. The set of words accepted by $\mathcal{A}$ is aA^*, the set of all words whose first letter is a.

The equivalence between automata and semigroups 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 relation on Q, denoted by $\mu(u)$, and defined by $(p, q) \in \mu(u)$ if there exists a path from p to q with label u. It is not difficult to see that μ is a semigroup morphism from A^+ into the semigroup[1] of relations on Q. The semigroup $\mu(A^+)$ is called the *transition semigroup* of $\mathcal{A}$, denoted $S(\mathcal{A})$. For practical computation, it can be conveniently represented as a semigroup of boolean matrices of order $|Q| \times |Q|$. In this case, $\mu(u)$ can be identified with the matrix defined by

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

[1]Given two relations $\mathcal{R}$ and $\mathcal{S}$ on Q, their product is the relation $\mathcal{RS}$ defined by $(p, q) \in \mathcal{RS}$ if and only if there exists r such that $(p, r) \in \mathcal{R}$ and $(r, q) \in \mathcal{S}$

Note that a word u is recognized by $\mathcal{A}$ if and only if $(p, q) \in \mu(u)$ for some initial state p and some final state q. This leads to the next proposition.

Proposition 3.9 *If a finite automaton recognizes a language L, then its transition semigroup recognizes L.*

Example 3.10 If $\mathcal{A} = (Q, A, E, I, F)$ is the automaton of example 3.8, one gets

$$\begin{aligned} \mu(a) &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & \mu(b) &= \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} & \mu(aa) &= \mu(a) \\ \mu(ab) &= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & \mu(ba) &= \mu(bb) = \mu(b) \end{aligned}$$

Thus the transition semigroup of $\mathcal{A}$ is the semigroup of boolean matrices

$$\mu(A^+) = \left\{ \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}.$$

The previous computation can be simplified if $\mathcal{A}$ is deterministic. Indeed, in this case, the transition semigroup of $\mathcal{A}$ is naturally embedded into the semigroup of partial functions on Q under the product $fg = g \circ f$.

Example 3.11 Let $A = \{a, b\}$ and let $\mathcal{A}$ be the deterministic automaton represented below.

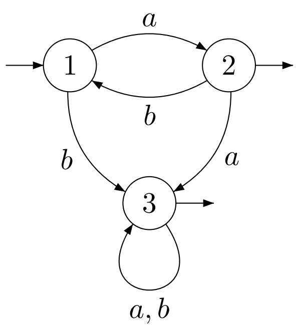

Figure 3: A deterministic automaton.

It is easy to see that $\mathcal{A}$ recognizes the language $A^+ \setminus (ab)^+$. The transition semigroup S of $\mathcal{A}$ contains five elements which correspond to the words a, b, ab, ba and aa. Furthermore aa is a zero of S and thus can be denoted 0. The other relations defining S are $aba = a$, $bab = b$ and $bb = 0$.

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

This semigroup is usually denoted BA_2 in semigroup theory.

Conversely, given a semigroup morphism $\varphi : A^+ \to S$ and a subset P of S, one can build a deterministic automaton recognizing $L = \varphi^{-1}(P)$ as follows. Take the right representation of A on S^1 defined by $s \cdot a = s\varphi(a)$. This defines an automaton $\mathcal{A} = (S^1, A, E, \{1\}, P)$, where $E = \{(s, a, s \cdot a) \mid s \in S^1, a \in A\}$ that recognizes L. We can now conclude.

Proposition 3.12 *A language is recognizable if and only if it is recognized by a finite automaton.*

See [102] for a detailed proof.

3.4 How to compute the syntactic semigroup?

The easiest way to compute the ordered syntactic semigroup of a recognizable language L is to first compute its minimal (deterministic) automaton $\mathcal{A} = (Q, A, \cdot, \{q_0\}, F)$. Then the syntactic semigroup of L is equal to the transition semigroup S of $\mathcal{A}$ and the order on S is given by $s \leq t$ if and only if,

$$\text{for every } x \in S^1, \text{ for every } q \in Q,\ q \cdot tx \in F \Rightarrow q \cdot sx \in F$$

Example 3.13 Let $\mathcal{A}$ be the deterministic automaton of example 3.11. It is the minimal automaton of $L = A^+ \setminus (ab)^+$. The transition semigroup was calculated in the previous section. The syntactic order is given by $0 \leq s$ for every $s \in S$. Indeed, $q \cdot 0 = 3 \in F$ and thus, the formal implication

$$q \cdot sx \in F \Rightarrow q \cdot 0x \in F$$

holds for any $q \in Q$, $s \in S$ and $x \in S^1$. One can verify that there is no other relations among the elements of S. For instance, a and ab are incomparable since $1 \cdot aa = 3$ but $1 \cdot aba = 2 \notin F$ and $1 \cdot abb = 3$ but $1 \cdot ab = 1 \notin F$.

4 Varieties

To each recognizable language is attached a finite ordered semigroup, its ordered syntactic semigroup. It is a natural idea to try to classify recognizable languages according to the algebraic properties of their ordered semigroups. The aim of this section is to introduce the proper framework to formalize this idea.

A *variety of semigroups* is a class of semigroups closed under taking subsemigroups, quotients and direct products. A *variety of finite semigroups*, or *pseudovariety*, is a class of finite semigroups closed under taking subsemigroups, quotients and finite direct products. *Varieties of ordered semigroups* and *varieties of finite ordered semigroups* are defined analogously. Varieties of semigroups or ordered semigroups will be denoted by boldface capital letters, like $\mathbf{V}$. The *join* of two varieties of finite (ordered) semigroups $\mathbf{V}_1$ and $\mathbf{V}_2$ is the smallest variety of finite (ordered) semigroups containing $\mathbf{V}_1$ and $\mathbf{V}_2$.

Given a class C of finite (ordered) semigroups, the variety of finite (ordered) semigroups *generated* by C is the smallest variety of finite (ordered) semigroups containing C. In a more constructive way, the variety of finite (ordered) semigroups *generated* by C is the class of all finite (ordered) semigroups that divide a finite direct product $S_1 \times \cdots \times S_n$, where $S_1, \ldots, S_n \in C$.

4.1 Identities

Varieties are conveniently defined by identities. Let Σ be a denumerable alphabet and let (u, v) be a pair of words of Σ^+. A semigroup S *satisfies the identity* $u = v$ if and only if $\varphi(u) = \varphi(v)$ for every semigroup morphism $\varphi : \Sigma^+ \to S$. Similarly, an ordered semigroup S *satisfies the identity* $u \leq v$ if and only if $\varphi(u) \leq \varphi(v)$ for every morphism of ordered semigroups $\varphi : \Sigma^+ \to S$. If Γ is a set of identities, the class of all semigroups (resp. ordered semigroups) that satisfy all the identities of Γ is a variety of semigroups (resp. ordered semigroups), called the *variety defined by* Γ. The following theorem, due to Birkhoff [22] and to Bloom [33] in the ordered case, shows that the converse also holds.

Theorem 4.1 *A class of semigroups (resp. ordered semigroups) is a variety if and only if it can be defined by a set of identities.*

For instance, the identity $xy = yx$ defines the variety of commutative semigroups and $x = x^2$ defines the variety of idempotent semigroups.

Since we are interested in finite semigroups, it would be interesting to have a similar result for varieties of finite semigroups. The problem was solved by several authors but the most satisfactory answer is due to Reiterman [129] (see also [125] in the ordered case). Reiterman's theorem states that pseudovarieties are also defined by identities. The difference between Birkhoff's and Reiterman's theorem lies in the definition of the identities. For Reiterman, an identity is also a formal equality of the form $u = v$, but u and v are now elements of a certain completion $\widehat{\Sigma}^+$ of the free semigroup Σ^+. Let us make this definition more precise.

A finite semigroup S *separates* two words $u, v \in \Sigma^+$ if $\varphi(u) \neq \varphi(v)$ for some semigroup morphism $\varphi : \Sigma^+ \to S$. Now set, for $u, v \in \Sigma^+$,

$$r(u, v) = \min\{\mathrm{Card}(S) \mid S \text{ is a finite semigroup separating } u \text{ and } v\}$$

and

$$d(u, v) = 2^{-r(u,v)}$$

with the usual conventions $\min \emptyset = +\infty$ and $2^{-\infty} = 0$. One can verify that d is a metric for which two words are close if a large semigroup is required to separate them. For this metric, multiplication in Σ^+ is uniformly continuous, so that Σ^+ is a topological semigroup. The completion of the metric space (Σ^+, d) is a topological semigroup, denoted $\widehat{\Sigma}^+$, in which every element is the limit of some Cauchy sequence of (Σ^+, d). In fact, one can show that $\widehat{\Sigma}^+$ is a compact semigroup.

Some more topology is required before stating Reiterman's theorem. We now consider finite semigroups as metric spaces, endowed with the discrete metric

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$$

Let S be a finite semigroup. Then a map $\varphi : \widehat{\Sigma}^+ \to S$ is *continuous* if and only if, for every converging sequence $(u_n)_{n\geq 0}$ of $\widehat{\Sigma}^+$, the sequence $\varphi(u_n)_{n\geq 0}$ is ultimately constant[2]. A finite semigroup (resp. ordered semigroup) S satisfies an identity $u = v$ (resp. $u \leq v$), where $u, v \in \widehat{\Sigma}^+$, if and only if $\varphi(u) = \varphi(v)$ (resp. $\varphi(u) \leq \varphi(v)$) for every continuous morphism $\varphi : \widehat{\Sigma}^+ \to S$. Note that such a continuous morphism is entirely determined by the values of $\varphi(a)$, for $a \in \Sigma$. Indeed, any map $\varphi : \Sigma \to S$ can be extended in a unique way into a semigroup morphism from Σ^+ into S. Since S is finite, such a morphism is uniformly continuous : if two elements of Σ^+ cannot be separated by S, their images under φ have to be the same. Now, a well known result of topology states that a uniformly continuous map whose domain is a metric space admits a unique continuous extension to the completion of this metric space.

Given a set E of identities of the form $u = v$ (resp. $u \leq v$), where $u, v \in \widehat{\Sigma}^+$, we denote by $[\![E]\!]$ the class of all finite semigroups (resp. ordered semigroups) which satisfy all the identities of E. Reiterman's theorem can now be stated as follows.

Theorem 4.2 *A class of finite semigroups (resp. ordered semigroups) is a variety if and only if it can be defined by a set of identities of $\widehat{\Sigma}^+$.*

Although Theorem 4.2 gives a satisfactory counterpart to Birkhoff's theorem, it is more difficult to understand, because of the rather abstract definition of $\widehat{\Sigma}^+$. Actually, no combinatorial description of $\widehat{\Sigma}^+$ is known and, besides the elements of Σ^+, which are words, the other elements are only defined as limits of sequences of words. An important such limit is the ω-power. Its definition relies on the following proposition.

Proposition 4.3 *For each $x \in \widehat{\Sigma}^+$, the sequence $(x^{n!})_{n\geq 0}$ is a Cauchy sequence. Its limit x^ω is an idempotent of $\widehat{\Sigma}^+$.*

[2]that is, if there exists an integer n_0 such that, for all $n, m \geq n_0$, $\varphi(u_n) = \varphi(u_m)$.

Now, if $\varphi : \widehat{\Sigma}^+ \to S$ is a continuous morphism onto a finite semigroup$S$, then $\varphi(x^\omega)$ is equal to $\varphi(x)^\omega$, the unique idempotent power of $\varphi(x)$, which shows that our notation is consistent. The notation x^ω makes the conversion of algebraic properties into identities very easy. For instance, the variety $[\![yx^\omega = x^\omega]\!]$ is the class of finite semigroups S such that, for every idempotent $e \in S$ and for every $s \in s$, $se = e$. Similarly, a finite semigroup S is locally commutative, if, for every idempotent e, the local monoid eSe is commutative. It follows immediately that finite locally commutative semigroups form a variety, defined by the identity $x^\omega yx^\omega zx^\omega = x^\omega zx^\omega yx^\omega$. More generally, if $\mathbf{V}$ is a variety of finite monoids, $\mathbf{LV}$ denotes the variety of all finite semigroups S, such that, for every idempotent $e \in S$, the local monoid eSe is in $\mathbf{V}$.

Another useful example is the following. The *content* of a word $u \in \Sigma^+$ is the set $c(u)$ of letters of Σ occurring in u. One can show that c is a uniformly continuous morphism from Σ^+ onto the semigroup 2^Σ of subsets of Σ under union. Thus c can be extended in a unique way into a continuous morphism from $\widehat{\Sigma}^+$ onto 2^Σ.

Reiterman's theorem suggests that most standard results on varieties might be extended in some way to pseudovarieties. For instance, it is well known that varieties have free objects. More precisely, if $\mathbf{V}$ is a variety and A is a finite set, there exists an A-generated semigroup $F_A(\mathbf{V})$ of $\mathbf{V}$, such that every A-generated semigroup of $\mathbf{V}$ is a quotient of $F_A(\mathbf{V})$. This semigroup is unique (up to an isomorphism) and is called the *free semigroup* of the variety $\mathbf{V}$. To extend this result to a pseudovariety $\mathbf{V}$, one first relativizes to $\mathbf{V}$ the definition of r and d as follows:

$$r_{\mathbf{V}}(u,v) = \min\bigl\{\mathrm{Card}(S) \bigm| S \in \mathbf{V} \text{ and } S \text{ separates } u \text{ and } v\ \bigr\}$$

and $d_{\mathbf{V}}(u,v) = 2^{-r_{\mathbf{V}}(u,v)}$. The function $d_{\mathbf{V}}(u,v)$ still satisfies the triangular inequality and even the stronger inequality

$$d_{\mathbf{V}}(u,v) \leq \max\ \{\ d_{\mathbf{V}}(u,w), d_{\mathbf{V}}(w,v)\ \}$$

but it is not a metric anymore because one can have $d_{\mathbf{V}}(u,v) = 0$ with $u \neq v$: for instance, if $\mathbf{V}$ is the pseudovariety of commutative finite semigroups, $d_{\mathbf{V}}(xy,yx) = 0$ since xy and yx cannot be separated by a commutative semigroup. However, the relation $\sim_{\mathbf{V}}$ defined on A^+ by $u \sim_{\mathbf{V}} v$ if and only if $d_{\mathbf{V}}(u,v) = 0$ is a congruence and $d_{\mathbf{V}}$ induces a metric on the quotient semigroup $A^+/{\sim_{\mathbf{V}}}$. The completion of this metric space is a topological compact semigroup $\widehat{F}_A(\mathbf{V})$, called the *free pro-$\mathbf{V}$ semigroup*. This semigroup is generated by A as a topological semigroup (this just means that $A^+/{\sim_{\mathbf{V}}}$ is dense in $\widehat{F}_A(\mathbf{V})$) and every A-generated semigroup of $\mathbf{V}$ is a continuous homomorphic image of $\widehat{F}_A(\mathbf{V})$. The combinatorial description of these free objects, for various varieties of finite semigroups, is the object of a very active research [7]. A more detailed presentation of Reiterman's theorem and its consequences can be found in [5, 7, 192].

4.2 The variety theorem

The variety theorem is due to Eilenberg [53]. Eilenberg's original theorem dealt with varieties of finite semigroups. The "ordered" version presented in this section is due to the author [112].

A *class of recognizable languages* is a correspondence $\mathcal{C}$ which associates with each finite alphabet A a set $\mathcal{C}(A^+)$ of recognizable languages of A^+.

If $\mathbf{V}$ is a variety of finite ordered semigroups, we denote by $\mathcal{V}(A^+)$ the set of recognizable languages of A^+ whose ordered syntactic semigroup belongs to $\mathbf{V}$ or, equivalently, which are recognized by an ordered semigroup of $\mathbf{V}$. The correspondence $\mathbf{V} \to \mathcal{V}$ associates with each variety of finite ordered semigroups a class of recognizable languages. The next proposition shows that this correspondence preserves inclusion.

Proposition 4.4 *Let $\mathbf{V}$ and $\mathbf{W}$ be two varieties of finite ordered semigroups. Suppose that $\mathbf{V} \to \mathcal{V}$ and $\mathbf{W} \to \mathcal{W}$. Then $\mathbf{V} \subseteq \mathbf{W}$ if and only if, for every finite alphabet A, $\mathcal{V}(A^+) \subseteq \mathcal{W}(A^+)$. In particular, $\mathbf{V} = \mathbf{W}$ if and only if $\mathcal{V} = \mathcal{W}$.*

It remains to characterize the classes of languages which can be associated with a variety of ordered semigroups. For this purpose, it is convenient to introduce the following definitions. A set of languages of A^+ (resp. A^*) closed under finite intersection and finite union is called a *positive boolean algebra*. Thus a positive boolean algebra always contains the empty language and the full language A^+ (resp. A^*) since $\emptyset = \bigcup_{i\in\emptyset} L_i$ and $A^+ = \bigcap_{i\in\emptyset} L_i$. A positive boolean algebra closed under complementation is a *boolean algebra*.

A *positive variety of languages* is a class of recognizable languages $\mathcal{V}$ such that

(1) for every alphabet A, $\mathcal{V}(A^+)$ is a positive boolean algebra,

(2) if $\varphi : A^+ \to B^+$ is a semigroup morphism, $L \in \mathcal{V}(B^+)$ implies $\varphi^{-1}(L) \in \mathcal{V}(A^+)$,

(3) if $L \in \mathcal{V}(A^+)$ and if $a \in A$, then $a^{-1}L$ and La^{-1} are in $\mathcal{V}(A^+)$.

Proposition 4.5 *Let $\mathbf{V}$ be a variety of finite ordered semigroups. If $\mathbf{V} \to \mathcal{V}$, then $\mathcal{V}$ is a positive variety of languages.*

So far, we have associated a positive variety of languages with each variety of finite ordered semigroups. Conversely, let $\mathcal{V}$ be and let $\mathbf{V}(\mathcal{V})$ be the variety of ordered semigroups generated by the ordered semigroups of the form $S(L)$ where $L \in \mathcal{V}(A^+)$ for a certain alphabet A. This variety is called the variety *associated with* $\mathcal{V}$, in view of the following theorem.

Theorem 4.6 *For every positive variety of languages $\mathcal{V}$, $\mathbf{V}(\mathcal{V}) \to \mathcal{V}$.*

In conclusion, we have the following theorem.

Theorem 4.7 *The correspondence $\mathbf{V} \to \mathcal{V}$ defines a one to one correspondence between the varieties of finite ordered semigroups and the positive varieties of languages.*

A *variety of languages* is a positive variety closed under complement, that is, satisfying

(1′) for every alphabet A, $\mathcal{V}(A^+)$ is a boolean algebra.

For varieties of languages, Theorem 4.7 can be modified as follows:

Corollary 4.8 *The correspondence $\mathbf{V} \to \mathcal{V}$ defines a one to one correspondence between the varieties of finite semigroups and the varieties of languages.*

There is an analogous theorem for varieties of ordered monoids. In this case, one defines languages as subsets of a free monoid and the definitions of a class of languages and of a positive variety have to be modified. To distinguish between the two definitions, it is convenient to add the prefixes $+$ or $*$ when necessary: $+$-class or $*$-class, $+$-variety or $*$-variety.

A *$*$-class of recognizable languages* is a correspondence $\mathcal{C}$ which associates with each finite alphabet A a set $\mathcal{C}(A^*)$ of recognizable languages of A^*. A *positive $*$-variety of languages* is a class of recognizable languages $\mathcal{V}$ such that

(1) for every alphabet A, $\mathcal{V}(A^*)$ is a positive boolean algebra,

(2) if $\varphi : A^* \to B^*$ is a monoid morphism, $L \in \mathcal{V}(B^*)$ implies $\varphi^{-1}(L) \in \mathcal{V}(A^*)$,

(3) if $L \in \mathcal{V}(A^*)$ and if $a \in A$, then $a^{-1}L$ and La^{-1} are in $\mathcal{V}(A^*)$.

of course, a *$*$-variety of languages* is a positive $*$-variety closed under complement, that is, satisfying

(1) for every alphabet A, $\mathcal{V}(A^*)$ is a boolean algebra.

The monoid version of Theorem 4.7 can be stated as follows.

Theorem 4.9 *The correspondence $\mathbf{V} \to \mathcal{V}$ defines a one to one correspondence between the varieties of finite ordered monoids and the positive $*$-varieties of languages.*

Corollary 4.10 *The correspondence $\mathbf{V} \to \mathcal{V}$ defines a one to one correspondence between the varieties of finite monoids and the $*$-varieties of languages.*

5 Examples of varieties

In this section, we illustrate the results of the previous section by a few examples. We present, in this order, some standard examples (Kleene's and

Schützenberger's theorems), the commutative varieties and varieties defined by "local" properties. Other examples will be given in the next sections.

Let us start by some general remarks. If $\mathbf{V}$ is a variety of finite ordered semigroups or monoids, the associated positive varieties of languages are denoted by the corresponding cursive letters, like $\mathcal{V}$. There are now several equivalent formulations to state a typical result on varieties, for instance:

> "The variety of finite ordered semigroups $\mathbf{V}$ is associated with the positive variety $\mathcal{V}$."
>
> "A recognizable language belongs to $\mathcal{V}(A^+)$ if and only if it is recognized by an ordered semigroup of $\mathbf{V}$."
>
> "A recognizable language belongs to $\mathcal{V}(A^+)$ if and only if its ordered syntactic semigroup belongs to $\mathbf{V}$."

We shall use mainly statements of the first type, but the last type will be occasionally preferred, especially when there are several equivalent descriptions of the languages. But it should be clear to the reader that all these formulations express exactly the same property.

5.1 Standard examples

The smallest variety of finite monoids is the trivial variety $\mathbf{I}$, defined by the identity $x = 1$. The associated variety of languages is defined, for every alphabet A, by $\mathcal{I}(A^*) = \{\emptyset, A^*\}$.

The largest variety of finite monoids is the variety of all finite monoids $\mathbf{M}$, defined by the empty set of identities. Recall that the set of *rational* languages of A^* is the smallest set of languages containing the languages $\{1\}$ and $\{a\}$ for each letter $a \in A$ and closed under finite union, product and star. Now Kleene's theorem can be reformulated as follows.

Theorem 5.1 *The variety of languages associated with* $\mathbf{M}$ *is the variety of rational languages.*

An important variety of monoids is the variety of aperiodic monoids, defined by the identity $x^\omega = x^{\omega+1}$. Thus, a finite monoid M is *aperiodic* if and only if, for each $x \in M$, there exists $n \geq 0$ such that $x^n = x^{n+1}$. This also means that the cyclic subgroup of the submonoid generated by any element x is trivial (see Proposition 2.1) or that M is $\mathcal{H}$-trivial. It follows that a monoid is aperiodic if and only if it is group-free: every subsemigroup which happens to be a group has to be trivial. Aperiodic monoids form a variety of monoids $\mathbf{A}$.

The associated variety of languages was first described by Schützenberger [144]. Recall that the *star-free* languages of A^* form the smallest boolean algebra containing the languages $\{1\}$ and $\{a\}$ for each letter $a \in A$ and which is closed under product.

Theorem 5.2 *The variety of languages associated with* **A** *is the variety of star-free languages.*

Example 5.3 Let $A = \{a, b\}$ and $L = (ab)^*$. Its minimal (but incomplete) automaton is represented below:

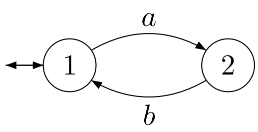

Figure 4: The minimal automaton of $(ab)^*$.

The syntactic monoid M of L is the monoid with zero presented on A by the relations $a^2 = b^2 = 0$, $aba = a$ and $bab = b$. Thus $M = \{1, a, b, ab, ba, 0\}$. Since $1^2 = 1$, $a^3 = a^2$, $b^3 = b^2$, $(ab)^2 = ab$, $(ba)^2 = ba$ and $0^2 = 0$, M is aperiodic and thus $(ab)^*$ is star-free. Indeed, if R^c denotes the complement of a language R, $(ab)^*$ admits the following star-free expression

$$L = \left(b\emptyset^c \cup \emptyset^c a \cup \emptyset^c aa\emptyset^c \cup \emptyset^c bb\emptyset^c\right)^c$$

We shall come back to Schützenberger's theorem in section 8.

When a variety is generated by a single ordered monoid, there is a direct description of the associated variety of languages.

Proposition 5.4 *Let M be a finite ordered monoid, let* **V** *be the variety of ordered monoids generated by M and let $\mathcal{V}$ be the associated positive variety. Then for every alphabet A, $\mathcal{V}(A^*)$ is the positive boolean algebra generated by the languages of the form $\varphi^{-1}(\downarrow m)$, where φ is any monoid morphism from A^* into M and m is any element of M.*

Of course, a similar result holds for varieties of ordered semigroups.

Proposition 5.5 *Let S be a finite ordered semigroup, let* **V** *be the variety of ordered semigroups generated by S and let $\mathcal{V}$ be the associated positive variety. Then for every alphabet A, $\mathcal{V}(A^+)$ is the positive boolean algebra generated by the languages of the form $\varphi^{-1}(\downarrow s)$, where φ is any semigroup morphism from A^+ into S and s is any element of S.*

This result suffices to describe a number of "small" varieties of languages. See for instance propositions 5.6 and 5.7 or [119].

5.2 Commutative varieties

In this section, we will consider only varieties of finite monoids and ordered monoids. A variety of ordered monoids is *commutative* if it satisfies the identity $xy = yx$. The smallest non-trivial variety of aperiodic monoids is the variety $\mathbf{J}_1$ of idempotent and commutative monoids (also called *semilattices*)[3], defined by the identities $xy = yx$ and $x^2 = x$. One can show that $\mathbf{J}_1$ is generated by the monoid $U_1 = \{0, 1\}$, whose multiplication table is given by $0 \cdot 0 = 0 \cdot 1 = 1 \cdot 0 = 0$ and $1 \cdot 1 = 1$. Thus Proposition 5.5 can be applied to get a description of the $*$-variety associated with $\mathbf{J}_1$.

Proposition 5.6 *For every alphabet A, $\mathcal{J}_1(A^*)$ is the boolean algebra generated by the languages of the form A^*aA^* where a is a letter. Equivalently, $\mathcal{J}_1(A^*)$ is the boolean algebra generated by the languages of the form B^* where B is a subset of A.*

Proposition 5.6 can be refined by considering the variety of finite ordered semigroups $\mathbf{J}_1^+$ (resp. $\mathbf{J}_1^-$), defined by the identities $xy = yx$, $x = x^2$ and $x \leq 1$ (resp. $x \geq 1$). One can show that $\mathbf{J}_1^+$ is generated by the ordered monoid $U_1^+ = (U_1, \leq)$ where the order is given by $0 \leq 1$. Then one can apply Proposition 5.4 to find a description of the $*$-variety $\mathcal{J}_1^+$ associated with $\mathbf{J}_1^+$. Let A be an alphabet and B be a subset of A. Denote by $L(B)$ the set of words containing at least one occurrence of every letter of B. Equivalently,

$$L(B) = \bigcap_{a \in B} A^*aA^*$$

Proposition 5.7 *For each alphabet A, $\mathcal{J}_1^+(A^*)$ consists of the finite unions of languages of the form $L(B)$, for some subset B of A.*

Corollary 5.8 *For each alphabet A, $\mathcal{J}_1^-(A^*)$ is the positive boolean algebra generated by the languages of the form B^*, for some subset B of A.*

Another important commutative variety of monoids is the variety of finite commutative groups **Gcom**, generated by the cyclic groups $\mathbb{Z}/n\mathbb{Z}$ $(n > 0)$ and defined by the identities $xy = yx$ and $x^\omega = 1$. For $a \in A$ and $k, n \geq 0$, let

$$F(a, k, n) = \{u \in A^* \mid |u|_a \equiv k \bmod n\}$$

Proposition 5.9 *For every alphabet A, $\mathcal{G}com(A^*)$ is the boolean algebra generated by the languages of the form $F(a, k, n)$, where $a \in A$ and $0 \leq k < n$.*

The largest commutative variety that contains no non-trivial group is the variety **Acom** of aperiodic and commutative monoids, defined by the identities $xy = yx$ and $x^\omega = x^{\omega+1}$. For $a \in A$ and $k \geq 0$, let

$$F(a, k) = \{u \in A^+ \mid |u|_a \geq k\}$$

[3]The notation $\mathbf{J}_1$ indicates that $\mathbf{J}_1$ is the first level of a hierarchy of varieties $\mathbf{J}_n$ that will be defined in section 8.4.

Proposition 5.10 *For every alphabet A, $\mathcal{A}com(A^*)$ is the boolean algebra generated by the languages of the form $F(a,k)$ where $a \in A$ and $k \geq 0$.*

Again, this proposition can be refined by considering ordered monoids. Let $\mathbf{Acom}^+$ be the variety of ordered monoids satisfying the identities $xy = yx$, $x^\omega = x^{\omega+1}$ and $x \leq 1$.

Proposition 5.11 *For every alphabet A, $\mathcal{A}com^+(A^*)$ is the positive boolean algebra generated by the languages of the form $F(a,k)$ where $a \in A$ and $k \geq 0$.*

Finally, the variety **Com** of all finite commutative monoids, defined by the identity $xy = yx$, is the join of the varieties **Gcom** and **Acom**.

Proposition 5.12 *For every alphabet A, $\mathcal{C}om(A^*)$ is the boolean algebra generated by the languages of the form $F(a,k)$ or $F(a,k,n)$ where $a \in A$ and $0 \leq k < n$.*

The "ordered" version is the following. Let $\mathbf{Com}^+$ be the variety of ordered monoids satisfying the identities $xy = yx$ and $x \leq 1$.

Proposition 5.13 *For every alphabet A, $\mathcal{C}om^+(A^*)$ is the positive boolean algebra generated by the languages of the form $F(a,k)$ or $F(a,k,n)$ where $a \in A$ and $0 \leq k < n$.*

5.3 Varieties defined by local properties

Contrary to the previous section, all the varieties considered in this section will be varieties of finite (ordered) semigroups. These varieties are all defined by *local* properties of words. Local properties can be tested by a *scanner*, which is a machine equipped with a finite memory and a sliding window of a fixed size n to scan the input word.

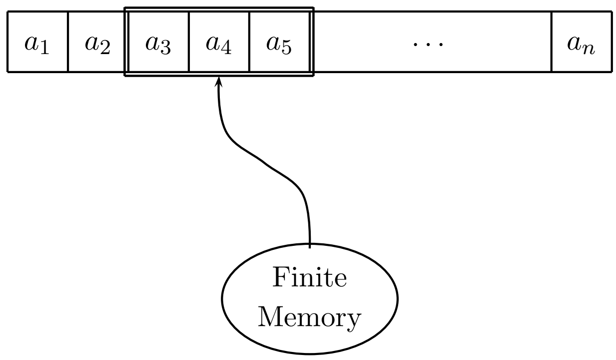

Figure 5: A scanner.

The window can also be moved before the first letter and beyond the last letter of the word in order to read the prefixes and suffixes of length $< n$. For example, if $n = 3$, and if the input word is *abbaaab*, the various positions of the window are represented on the following diagram:

$$\boxed{\,a}\,bbaaab \qquad \boxed{ab}\,baaab \qquad \boxed{abb}\,aaab \qquad a\,\boxed{bba}\,aab \quad \cdots \quad abbaaa\,\boxed{b\,}$$

At the end of the scan, the scanner memorizes the prefixes and suffixes of length $< n$ and the set of factors of length n of the input word. The memory of the scanners contains a table of possible lists of prefixes (resp. suffixes, factors). A word is accepted by the scanner if the list of prefixes (resp. suffixes, factors) obtained after the scan matches one of the lists of the table. Another possibility is to take into account the number of occurrences of the factors of the word.

Local properties can be used to define several varieties of languages. A language is *prefix testable*[4] if it is a boolean combination of languages of the form xA^*, where $x \in A^+$ or, equivalently, if it is of the form $FA^* \cup G$ for some finite languages F and G. Similarly, a language is *suffix testable*[5] if it is a boolean combination of languages of the form A^*x, where $x \in A^+$.

Proposition 5.14 *Prefix (resp. suffix) testable languages form a variety of languages. The associated variety of finite semigroups is defined by the identity $x^\omega y = x^\omega$ (resp. $yx^\omega = x^\omega$).*

Languages that are both prefix and suffix testable form an interesting variety. Recall that a language is *cofinite* if its complement is finite.

Proposition 5.15 *Let L be a recognizable language. The following conditions are equivalent:*

(1) *L is prefix testable and suffix testable,*
(2) *L is finite or cofinite,*
(3) *$S(L)$ satisfies the identities $x^\omega y = x^\omega = yx^\omega$,*
(4) *$S(L)$ is nilpotent*

Proposition 5.15 can be refined as follows

Proposition 5.16 *A language is empty or cofinite if and only if it is recognized by a finite ordered nilpotent semigroup S in which $0 \leq s$ for all $s \in S$.*

Note that the finite ordered nilpotent semigroups S in which 0 is the smallest element form a variety of finite ordered semigroups, defined by the identities $x^\omega y = x^\omega = yx^\omega$ and $x^\omega \leq y$. The dual version of Proposition 5.16 is also of interest.

[4]These languages are called *reverse definite* in the literature.
[5]The suffix testable languages are called *definite* in the literature.

Corollary 5.17 *A language is full or finite if and only if it is recognized by a finite ordered nilpotent semigroup S in which $s \leq 0$ for all $s \in S$.*

A language is *prefix-suffix testable*[6] if it is a boolean combination of languages of the form xA^* or A^*x, where $x \in A^+$.

Proposition 5.18 [90] *Let L be a language. The following conditions are equivalent:*

(1) *L is a prefix-suffix testable language,*

(2) *L is of the form $FA^*G \cup H$ for some finite languages F, G and H,*

(3) *$S(L)$ satisfies the identity $x^\omega y x^\omega = x^\omega$,*

(4) *$S(L)$ is locally trivial.*

Proposition 5.18 shows that prefix-suffix testable languages form a variety of languages. The corresponding variety of semigroups is **LI**.

A language is *positively locally testable* if it is a positive boolean combination of languages of the form $\{x\}$, xA^*, A^*x or A^*xA^* ($x \in A^+$) and it is *locally testable* if it is a boolean combination of the same languages. The syntactic characterization of locally testable languages is relatively simple to state, but its proof, discovered independently by Brzozowski and Simon [41] and by McNaughton [86], requires sophisticated tools that are detailed in section 6.

Theorem 5.19 *Let L be a language. The following conditions are equivalent:*

(1) *L is locally testable,*

(2) *$S(L)$ satisfies the identities $x^\omega y x^\omega z x^\omega = x^\omega z x^\omega y x^\omega$ and $x^\omega y x^\omega y x^\omega = x^\omega y x^\omega$,*

(3) *$S(L)$ is locally idempotent and commutative.*

In the positive case, the identity $x^\omega y x^\omega \leq x^\omega$ must be added. Thus an ordered semigroup S satisfies this identity if and only if, for every idempotent $e \in S$, and for every element $s \in S$, $ese \leq e$. This means that, in the local monoid eSe, the identity e is the maximum element.

Theorem 5.20 *Let L be a language. The following conditions are equivalent:*

(1) *L is positively locally testable,*

(2) *$S(L)$ satisfies the identities $x^\omega y x^\omega z x^\omega = x^\omega z x^\omega y x^\omega$, $x^\omega y x^\omega y x^\omega = x^\omega y x^\omega$ and $x^\omega y x^\omega \leq x^\omega$.*

(3) *$S(L)$ is locally idempotent and commutative and in every local monoid, the identity is the maximum element.*

[6] These languages are called *generalized definite* in the literature.

Example 5.21 Let $A = \{a, b\}$ and let $L = A^+ \setminus (ab)^+$ be the language of Example 3.13. Let S be the ordered syntactic semigroup of L. The idempotents of S are ab, ba and 0 and the local submonoids are $abSab = \{ab, 0\}$, $baSba = \{ba, 0\}$ and $0S0 = \{0\}$. These three monoids are idempotent and commutative and their identity is the maximum element (since $0 \leq s$ for every $s \in S$). Therefore, L is positively locally testable. Indeed, one has

$$L = bA^* \cup A^*a \cup A^*aaA^* \cup A^*bbA^*$$

One can also take into account the number of occurrences of a given word. For each word $x, u \in A^+$, let $\left[\begin{smallmatrix} u \\ x \end{smallmatrix}\right]$ denote the number of occurrences of x as a factor of u. For every integer k, set

$$F(x, k) = \left\{ u \in A^+ \mid \begin{bmatrix} u \\ x \end{bmatrix} \geq k \right\}$$

In particular, $F(x, 1) = A^*xA^*$, the set of words containing at least one occurrence of x. A language is said to be *(positively) threshold locally testable* if it is a (positive) boolean combination of languages of the form $\{x\}$, xA^*, A^*x or $F(x, k)$ ($x \in A^+$, $k \geq 0$).

A new concept is needed to state the syntactic characterization of these languages in a precise way. The *Cauchy category of a semigroup S* is the category C whose objects are the idempotents of S and, given two idempotents e and f, the set of arrows from e to f is the set

$$C(e, f) = \{s \in S \mid es = s = sf\}$$

Given three idempotents e, f, g, the composition of the arrows $p \in C(e, f)$ and $q \in C(f, g)$ is the arrow $pq \in C(e, g)$. The algebraic background for this definition will be given in section 6.2.

Theorem 5.22 *Let L be a recognizable subset of A^+. Then L is threshold locally testable if and only if $S(L)$ is aperiodic and its Cauchy category satisfies the following condition: if p and r are arrows from e to f and if q is an arrow from f to e, then $pqr = rqp$.*

The latter condition on the Cauchy category is called Thérien's condition.

$$\begin{tikzcd} e \arrow[r, bend left, "{p, r}"] & f \arrow[l, bend left, "q"] \end{tikzcd}$$

Figure 6: The condition $pqr = rqp$.

In the positive case, the characterization is the following.

Theorem 5.23 *Let L be a recognizable subset of A^+. Then L is positively threshold locally testable if and only if $S(L)$ is aperiodic, its Cauchy category satisfies Thérien's condition and in every local monoid, the identity is the maximum element.*

Example 5.24 Let $A = \{a, b\}$ and let $L = a^*ba^*$. Then L is recognized by the automaton shown in the figure below.

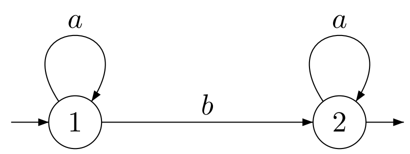

Figure 7: The minimal automaton of a^*ba^*.

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

	a	b	bb
1	1	2	–
2	2	–	–

$$\begin{aligned} a &= 1 \\ b^2 &= 0 \end{aligned}$$

Thus $S = \{1, b, 0\}$ and $E(S) = \{1, 0\}$. The local semigroups are $0S0 = \{0\}$ and $1S1 = S$. The latter is not idempotent, since $b^2 \neq b$. Therefore, L is not locally testable. On the other hand, the Cauchy category of $S(L)$, represented in the figure below, satisfies the condition $pqr = rqp$.

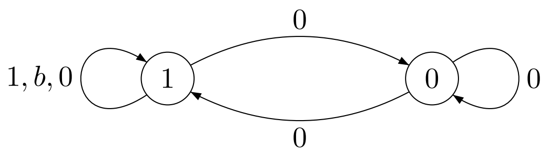

Figure 8: The graph of S.

Therefore L is threshold locally testable and is not positively threshold locally testable since $b \not\leq 1$.

Another way of counting factors is to count modulo n for some integer n. To this purpose, set, for every $x \in A^+$ and for every $k \geq 0$, $n > 0$,

$$F(x, k, n) = \left\{ u \in A^+ \mid \begin{bmatrix} u \\ x \end{bmatrix} \equiv k \bmod n \right\}$$

A language is said to be *modulus locally testable* if it is a boolean combination of languages of the form $\{x\}$, xA^*, A^*x or $F(x, k, n)$ ($x \in A^+$, $k, n \geq 0$).

Theorem 5.25 [171, 169] *Let L be a recognizable subset of A^+. Then L is modulus locally testable if and only if every local monoid of $S(L)$ is a commutative group and the Cauchy category of $S(L)$ satisfies Thérien's condition.*

5.4 Algorithmic problems

Let $\mathbf{V}$ be a variety of finite ordered semigroups and let $\mathcal{V}$ be the associated positive variety of languages. In order to decide whether a recognizable language L of A^+ (given for instance by a finite automaton) belongs to $\mathcal{V}(A^+)$, it suffices to compute the ordered syntactic semigroup S of L and to verify that $S \in \mathbf{V}$. This motivates the following definition: a variety of finite ordered semigroups $\mathbf{V}$ is *decidable* if there is an algorithm to decide whether a given finite ordered semigroup belongs to $\mathbf{V}$. All the varieties that were considered up to now are decidable, but several open problems in this field amount to decide whether a certain variety is decidable or not.

Once it is known that a variety is decidable, it is usually not necessary to compute the ordered syntactic semigroup to decide whether L belongs to $\mathcal{V}(A^+)$. Most of the time, one can obtain a more efficient algorithm by analyzing the minimal automaton of the language [127].

6 Some algebraic tools

The statement of the more advanced results presented in the next sections requires some auxiliary algebraic tools: relational morphisms, Mal'cev products and semidirect products.

6.1 Relational morphisms

Relational morphisms were introduced by Tilson [187]. If S and T are semigroups, a *relational morphism* $\tau : S \to T$ is a relation from S into T, i.e. a mapping from S into $\mathcal{P}(T)$ such that:

(1) $\tau(s)\tau(t) \subseteq \tau(st)$ for all $s, t \in S$,

(2) $\tau(s)$ is non-empty for all $s \in S$,

For a relational morphism between two monoids S and T, a third condition is required

(1) $1 \in \tau(1)$

Equivalently, τ is a relation whose graph

$$\text{graph}(\tau) = \{\ (s, t) \in S \times T \mid t \in \tau(s)\ \}$$

is a subsemigroup (resp. submonoid if S and T are monoids) of $S \times T$, with first-coordinate projection onto S.

It is not necessary to introduce a special notion of relational morphism for ordered semigroups. Indeed, if S and T are ordered, then the graph of τ is naturally ordered as a subsemigroup of $S \times T$ and the projections on S and T are order-preserving.

Semigroup morphisms and the inverses of surjective semigroup morphisms are examples of relational morphisms. This holds even if the semigroups are ordered, and there is no need for the morphisms to be order-preserving. In particular, if $(S, \leq)$ is an ordered semigroup equipped with a non trivial order, then the identity defines a morphism from $(S, =)$ onto $(S, \leq)$ but also a relational morphism from $(S, \leq)$ onto $(S, =)$.

The composition of two relational morphisms is again a relational morphism. In particular, given two surjective semigroup morphisms $\alpha : A^+ \to S$ and $\beta : A^+ \to T$, the relation $\tau = \beta \circ \alpha^{-1}$ is a relational morphism between S and T. We shall consider two examples of this situation in which S and T are syntactic semigroups.

Our first example illustrates a simple, but important property of the concatenation product. Let, for $0 \leq i \leq n$, L_i be recognizable languages of A^*, let $\eta_i : A^* \to M(L_i)$ be their syntactic morphisms and let $\eta : A^* \to M(L_0) \times M(L_1) \times \cdots \times M(L_n)$ be the morphism defined by $\eta(u) = (\eta_0(u), \eta_1(u), \ldots, \eta_n(u))$. Let $a_1, a_2, \ldots, a_n$ be letters of A and let $L = L_0 a_1 L_1 \cdots a_n L_n$. Let $\mu : A^* \to M(L)$ be the syntactic morphism of L. The relational morphism $\tau = \eta \circ \mu^{-1} : M(L) \to M(L_0) \times M(L_1) \times \cdots \times M(L_n)$ has a remarkable property.

Proposition 6.1 *For every idempotent e of $M(L_0) \times M(L_1) \times \cdots \times M(L_n)$, $\tau^{-1}(e)$ is an ordered semigroup that satisfies the identity $x^\omega y x^\omega \leq x^\omega$.*

Proposition 6.1 is a simplified version [126, 127] of an earlier result of Straubing [164].

There is a similar result for syntactic semigroups. In this case, we consider languages of the form $L = u_0 L_1 u_1 \cdots L_n u_n$, where $u_0, u_1, \ldots, u_n$ are words of A^* and $L_1, \ldots, L_n$ are recognizable languages[7] of A^+. Let $\eta_i : A^+ \to S(L_i)$ be the syntactic morphism of L_i and let

$$\eta : A^+ \to S(L_1) \times S(L_2) \times \cdots \times S(L_n)$$

be the morphism defined by $\eta(u) = (\eta_1(u), \eta_2(u), \ldots, \eta_n(u))$. Finally, let $\mu : A^+ \to S(L)$ be the syntactic morphism of L and let $\tau = \eta \circ \mu^{-1}$.

Proposition 6.2 *For every idempotent e of $S(L_1) \times S(L_2) \times \cdots \times S(L_n)$, $\tau^{-1}(e)$ is an ordered semigroup that satisfies the identity $x^\omega y x^\omega \leq x^\omega$.*

[7] The reason of this modification is the following: a product of the form $L_0 a_1 L_1 \cdots a_n L_n$ where $L_1, \ldots, L_n$ are languages of A^* can be written as a finite union of languages of the form $u_0 L_1 u_1 \cdots L_n u_n$, where $u_0, u_1, \ldots, u_n$ are words of A^* and $L_1, \ldots, L_n$ are recognizable languages of A^+.

The product $L = L_0a_1L_1 \cdots a_nL_n$ is *unambiguous* if every word u of L admits a unique factorization of the form $u_0a_1u_1 \cdots a_nu_n$ with $u_0 \in L_0$, $\ldots$, $u_n \in L_n$. It is *left deterministic* (resp. *right deterministic*) if, for $1 \leq i \leq n$, u has a unique prefix (resp. suffix) in $L_0a_1L_1 \cdots L_{i-1}a_i$ (resp. $a_iL_i \cdots a_nL_n$). If the product is unambiguous, left deterministic or right deterministic, Proposition 6.1 can be improved as follows.

Proposition 6.3 *If the product $L_0a_1L_1 \cdots a_nL_n$ is unambiguous (resp. left deterministic, right deterministic), then for every idempotent e of $M(L_0) \times M(L_1) \times \cdots \times M(L_n)$, $\tau^{-1}(e)$ is an ordered semigroup that satisfies the identity $x^\omega y x^\omega = x^\omega$ (resp. $x^\omega y = x^\omega$, $yx^\omega = x^\omega$).*

A similar result holds for languages of A^+ and unambiguous (resp. left deterministic, right deterministic) products of the form

$$L = u_0L_1u_1 \cdots L_nu_n.$$

The product $L = L_0a_1L_1 \cdots a_nL_n$ is *bideterministic* if it is both left and right deterministic. Bideterministic products were introduced by Schützenberger [148] and studied in more detail in [123, 34, 35]. In particular, Branco proved an analog of Proposition 6.3 for the bideterministic product, but the condition bears on the kernel category of the relational morphism τ (see section 6.3).

Our second example, also due to Straubing [164] concerns the star operation. Recall that a language L is *pure* if, for every $u \in L$ and every $n > 0$, $u^n \in L$ implies $u \in L$. Now consider the syntactic morphism $\eta : A^* \to M(L)$ of a recognizable language L and let $\mu : A^* \to M(L^*)$ be the syntactic morphism of L^*. Then consider the relational morphism $\tau = \eta \circ \mu^{-1} : M(L^*) \to M(L)$.

Proposition 6.4 *If L^* is a pure language, then for every idempotent e of $M(L)$, $\tau^{-1}(e)$ is an aperiodic semigroup.*

6.2 Mal'cev product

Let S and T be ordered semigroups and let $\tau : S \to T$ be a relational morphism. Then, for every ordered subsemigroup T' of T, the set

$$\tau^{-1}(e) = \{s \in S \mid e \in \tau(s)\}$$

is an ordered subsemigroup of S. Let $\mathbf{W}$ be a variety of ordered semigroups. A relational morphism $\tau : S \to T$ is called a $\mathbf{W}$*-relational morphism* if, for every idempotent $e \in T$, the ordered semigroup $\tau^{-1}(e)$ belongs to $\mathbf{W}$.

If $\mathbf{V}$ is a variety of semigroups (resp. monoids), the class $\mathbf{W} ⓜ \mathbf{V}$ of all ordered semigroups (resp. monoids) S such that there exists a $\mathbf{W}$-relational morphism from S onto a semigroup (resp. monoid) of $\mathbf{V}$ is a variety of ordered semigroups, called the *Mal'cev product* of **W** and **V**. Let **W** be a variety of ordered semigroups, defined by a set E of identities. The following theorem, proved in [124] describes a set of identities defining $\mathbf{W} \circledM \mathbf{V}$.

Theorem 6.5 *Let* $\mathbf{V}$ *be a variety of monoids and let* $\mathbf{W} = [\![E]\!]$ *be a variety of ordered semigroups. Then* $\mathbf{W} \circledM \mathbf{V}$ *is defined by the identities of the form* $\sigma(x) \leq \sigma(y)$, *where* $x \leq y$ *is an identity of* E *with* $x, y \in \widehat{B}^*$ *for some finite alphabet* B *and* $\sigma : \widehat{B}^* \to \widehat{A}^*$ *is a continuous morphism such that, for all* $b, b' \in B$, $\mathbf{V}$ *satisfies the identity* $\sigma(b) = \sigma(b') = \sigma(b^2)$.

Despite its rather abstract statement, Theorem 6.5 can be used to produce effectively identities of some Mal'cev products [124]. Mal'cev products play an important role in the study of the concatenation product, as will be shown in Section 7.1.

6.3 Semidirect product

Let S and T be semigroups. We write the product in S additively to provide a more transparent notation, but it is not meant to suggest that S is commutative. A *left action* of T on S is a map $(t, s) \to ts$ from $T^1 \times S$ into S such that, for all $s, s_1, s_2 \in S$ and $t, t_1, t_2 \in T$,

(1) $(t_1 t_2)s = t_1(t_2 s)$
(2) $t(s_1 + s_2) = ts_1 + ts_2$
(3) $1s = s$

If S is a monoid with identity 0, the action is *unitary* if it satisfies, for all $t \in T$,

(1) $t0 = 0$

Given such a left action[8], the *semidirect product* of S and T (with respect to this action) is the semigroup $S * T$ defined on the set $S \times T$ by the product

$$(s_1, t_1)(s_2, t_2) = (s_1 + t_1 s_2, t_1 t_2)$$

Given two varieties of finite semigroups $\mathbf{V}$ and $\mathbf{W}$, denote by $\mathbf{V} * \mathbf{W}$ the variety of finite semigroups generated by the semidirect products $S * T$ with $S \in \mathbf{V}$ and $T \in \mathbf{W}$. One can define similarly the semidirect product of two varieties of finite monoids, or of a variety of finite monoid and a variety of finite semigroups. For instance, if $\mathbf{V}$ is a variety of finite monoids and $\mathbf{W}$ is a variety of finite semigroups, $\mathbf{V} * \mathbf{W}$ is the variety of finite semigroups generated by the semidirect products $S * T$ with $S \in \mathbf{V}$ and $T \in \mathbf{W}$ such that the action of T on S is unitary.

[8]We followed Almeida [5] for the definition of a left action. This definition is slightly different from Eilenberg's definition [53], where an action is defined as a map from $T \times S$ into S satisfying (1) and (2).

The *wreath product* is closely related to the semidirect product. The wreath product $S \circ T$ of two semigroups S and T is the semidirect product $S^{T^1} * T$ defined by the action of T on S^{T^1} given by

$$tf(t') = f(tt')$$

for $f : T^1 \to S$ and $t, t' \in T^1$. In particular, the multiplication in $S \circ T$ is given by

$$(f_1, t_1)(f_2, t_2) = (f, t_1t_2) \text{ where } f(t) = f_1(t) + f_2(t_1t) \text{ for all } t \in T^1$$

In a way, the wreath product is the most general semidirect product since every semidirect product $S * T$ is a subsemigroup of $S \circ T$. It follows that $\mathbf{V} * \mathbf{W}$ is generated by all wreath products of the form $S \circ T$, where $S \in \mathbf{V}$ and $T \in \mathbf{W}$. Although the semidirect product is not an associative operation, it become associative at the variety level. That is, if $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$ are varieties of finite semigroups, then $(\mathbf{V}_1 * \mathbf{V}_2) * \mathbf{V}_3 = \mathbf{V}_1 * (\mathbf{V}_2 * \mathbf{V}_3)$.

Wreath products allow to decompose semigroups into smaller pieces. Let U_1 be the monoid $\{1, 0\}$ under usual multiplication and let $U_2 = \{1, a, b\}$ be the monoid defined by the multiplication $aa = ba = a$ and $ab = bb = b$.

Theorem 6.6 *The following decompositions hold:*

(1) *Every solvable group divides a wreath product of commutative groups,*

(2) *Every $\mathcal{R}$-trivial monoid divides a wreath product of copies of U_1,*

(3) *Every aperiodic monoid divides a wreath product of copies of U_2,*

(4) *Every monoid divides a wreath product of groups and copies of U_2,*

Statement (4) is the celebrated Krohn-Rhodes theorem [5, 53, 169]. Wreath product decompositions were first used in language theory to get a new proof of Schützenberger's theorem [44, 88]. This use turns out to be a particular case of Straubing's "wreath product principle" [159, 168], which provides a description of the languages recognized by the wreath product of two finite monoids.

Let M and N be two finite monoids and let $\eta : A^* \to M \circ N$ be a monoid morphism. We denote by $\pi : M \circ N \to N$ the monoid morphism defined by $\pi(f, n) = n$ and we put $\varphi = \pi \circ \eta$. Thus φ is a monoid morphism from A^* into N. Let $B = N \times A$ and $\sigma : A^* \to B^*$ be the map defined by

$$\sigma(a_1a_2 \cdots a_n) = (1, a_1)(\varphi(a_1), a_2) \cdots (\varphi(a_1a_2 \cdots a_{n_1}), a_n)$$

Observe that σ is not a morphism, but a sequential function [19]. Straubing's result can be stated as follows.

Theorem 6.7 *If a language L is recognized by $\eta : A^* \to M \circ N$, then L is a finite boolean combination of languages of the form $X \cap \sigma^{-1}(Y)$, where $Y \subset B^*$ is recognized by M and where $X \subset A^*$ is recognized by N.*

In view of the decomposition results of Theorem 6.6, this principle can be used to describe the variety of languages corresponding to solvable groups (Theorem 7.18 below), $\mathcal{R}$-trivial monoids (Corollaries 7.8 and 7.14) or aperiodic monoids (Theorem 5.2). Theorem 6.6 is also the key result in the proof of Theorems 7.12 and 7.21.

The wreath product principle can be adapted to the case of a wreath product of a monoid by a semigroup. Varieties of the form $\mathbf{V} * \mathbf{LI}$ received special attention. See the examples at the end of this section.

However, all these results yield the following question: if $\mathbf{V}$ and $\mathbf{W}$ are decidable varieties of finite monoids (or semigroups), is $\mathbf{V} * \mathbf{W}$ also decidable ? A negative answer was given in the general case [1], but several positive results are also known. A few more definitions on categories are needed to state these results precisely.

Let M and N be two monoids and let $\tau : M \to N$ be a relational morphism. Let C be the category such that $Ob(C) = N$ and, for all $u, v \in N$,

$$C(u, v) = \{(u, s, v) \in N \times M \times N \mid v \in u\tau(s)\}.$$

Composition is given by $(u, s, v)(v, t, w) = (u, st, w)$. Now the *kernel category* of τ is the quotient of C by the congruence $\sim$ defined by

$$(u, s, v) \sim (u, t, v) \text{ if and only if } ms = mt \text{ for all } m \in \tau^{-1}(u)$$

Thus the kernel category identifies elements with the same action on each fiber[9] $\tau^{-1}(u)$.

The next theorem, due to Tilson [188], relates semidirect products and relational morphisms.

Theorem 6.8 *Let* $\mathbf{V}$ *and* $\mathbf{W}$ *be two variety of finite monoids. A monoid* M *belongs to* $\mathbf{V}*\mathbf{W}$ *if and only if there exists a relational morphism* $\tau : M \to T$*, where* $T \in \mathbf{W}$*, whose kernel category divides a monoid of* $\mathbf{V}$*.*

There is an analogous result (with a few technical modifications) when $\mathbf{W}$ is a variety of finite semigroups. In view of Theorem 6.8, it is important to characterize, given a variety of finite monoids $\mathbf{V}$, the categories that divide a monoid of $\mathbf{V}$. The problem is not solved in general, but one can try to apply one of the following results.

A first idea is to convert a category, which can be considered as a "partial semigroup" under composition of arrows, into a real semigroup. If C be a category, associate with each arrow $p \in C(u, v)$ the triple (u, p, v). Let $S(C)$ be the set of all such triples, along with a new element denoted 0. Next

[9] The action is the multiplication on the right. In some applications [120, 35], it is more appropriate to use a definition of the kernel category that takes also in account the multiplication on the left [138].

define a multiplication on $S(C)$ by setting

$$(u,p,v)(u',p',v') = \begin{cases} (u,pp',v') & \text{if } v = u' \\ 0 & \text{otherwise} \end{cases}$$

It is easy to verify that $S(C)$ is a semigroup. The interest of this construction lies in the following theorem [166, 188]. Recall that BA_2 is the five element semigroup considered in example 3.11.

Theorem 6.9 *Let* $\mathbf{V}$ *be a variety of finite monoids containing the monoid* BA_2^1. *Then a category* C *divides a monoid of* $\mathbf{V}$ *if and only if* $S(C)^1$ *belongs to* $\mathbf{V}$.

Corollary 6.10 *Let* $\mathbf{V}$ *be a decidable variety of finite monoids containing the monoid* BA_2^1. *Then it is decidable whether a given finite category divides a monoid of* $\mathbf{V}$.

Thus the question raised above is solved for varieties of finite monoids that contain BA_2^1, for instance the variety A of finite aperiodic monoids. We are now mainly interested in varieties of finite monoids that do not contain BA_2^1. These varieties are exactly the subvarieties of the variety of finite monoids in which each regular $\mathcal{J}$-class is a semigroup. This variety, denoted $\mathbf{DS}$, is defined by the identities $\big((xy)^\omega(yx)^\omega(xy)^\omega\big)^\omega = (xy)^\omega$.

The problem is also easy to solve for another type of varieties, the *local varieties*. For these varieties, it suffices to check whether the local monoids of the category are in $\mathbf{V}$. More precisely, a variety $\mathbf{V}$ is local if and only if every category whose local monoids are in $\mathbf{V}$ divides a monoid of $\mathbf{V}$. Local varieties were first characterized in [183]. See also [188].

Theorem 6.11 *A non trivial variety* $\mathbf{V}$ *is local if and only if* $\mathbf{V} * \mathbf{LI} = \mathbf{LV}$.

In spite of this theorem, it is not easy in general to know whether a variety of finite monoids is local or not, even for the subvarieties of $\mathbf{DS}$. The next theorem summarizes results of Simon, Thérien, Weiss, Tilson, Jones and Almeida [41, 53, 183, 195, 188, 67, 6]. In this theorem, $\mathbf{DA}$ denotes the intersection of $\mathbf{A}$ and $\mathbf{DS}$. Thus $\mathbf{DA}$ is the variety of finite monoids in which each regular $\mathcal{J}$-class is an idempotent semigroup.

Theorem 6.12 *The following varieties are local: any non trivial variety of finite groups, the varieties* $\mathbf{J}_1$, $\mathbf{DA}$, $\mathbf{DS}$, $[\![x^\omega = x^{\omega+1}, x^\omega y = yx^\omega]\!]$ *and the varieties* $[\![x^n = x]\!]$ *for each* $n > 1$.

Note that if $\mathbf{V}$ is a decidable local variety, one can effectively decide whether a finite category divides a monoid of $\mathbf{V}$. Unfortunately, some varieties, like the trivial variety $\mathbf{I}$, are not local. However, a decidability result can still be obtained in some cases. We just mention the most important of them, which concern four important subvarieties of $\mathbf{DS}$.

Theorem 6.13 [188] *A category C divides a trivial monoid if and only if, for every $u, v \in Ob(C)$, the set $C(u, v)$ has at most one element.*

Theorem 6.14 [72] *A category C divides a finite $\mathcal{J}$-trivial monoid if and only if, for every $u, v \in Ob(C)$, for each $p, r \in C(u, v)$ and each $q, s \in C(v, u)$, $(pq)^\omega ps(rs)^\omega = (pq)^\omega (rs)^\omega$.*

Theorem 6.15 [183] *A category C divides a finite commutative monoid if and only if, for every $u, v \in Ob(C)$, for every $p, r \in C(u, v)$ and every $q \in C(v, u)$, $pqr = rqp$.*

Theorem 6.16 [183] *A category C divides a finite aperiodic commutative monoid if and only if, the local monoids of C are aperiodic and, for every $u, v \in Ob(C)$, for every $p, r \in C(u, v)$ and every $q \in C(v, u)$, $pqr = rqp$.*

Semidirect products of the form $\mathbf{V} * \mathbf{LI}$ deserved special attention. The key result is due to Straubing [166] and was formalized in [188]. It is a generalization of former results of [41, 86, 71].

Theorem 6.17 [166] *Let $\mathbf{V}$ be a variety of finite monoids. A semigroup belongs to $\mathbf{V} * \mathbf{LI}$ if and only if its Cauchy category divides a monoid of $\mathbf{V}$.*

With all these powerful tools in hand, one can now sketch a proof of Theorems 5.19, 5.22 and 5.25. First, one makes use of the wreath product principle to show that the variety of finite semigroups associated with the locally testable (resp. threshold locally testable, modulus locally testable) languages is the variety $\mathbf{J}_1 * \mathbf{LI}$ (resp. $\mathbf{Acom} * \mathbf{LI}$, $\mathbf{Gcom} * \mathbf{LI}$). It follows by Theorem 6.17 that a recognizable language is locally testable (resp. threshold locally testable, modulus locally testable) if and only if the Cauchy category of its syntactic semigroup divides a monoid of $\mathbf{J}_1$ (resp. **Acom**, **Gcom**). It remains to apply Theorem 6.12 (or Theorem 6.16 in the case of **Acom**) to conclude.

Theorem 8.19 below is another application of Theorem 6.17.

6.4 Representable transductions

In this section, we address the following general problem. Let $L_1, \ldots, L_n$ be languages recognized by monoids $M_1, \ldots, M_n$, respectively. Given an operation φ on these languages, find a monoid which recognizes $\varphi(L_1, \ldots, L_n)$. The key idea of our construction is to consider, when it is possible, an operation $\varphi : A^* \times \cdots \times A^* \to A^*$ as the inverse of a transduction $\tau : A^* \to A^* \times \cdots \times A^*$. Then, for a rather large class of transductions, it is possible to solve our problem explicitly. Precise definitions are given below.

Transductions were intensively studied in connection with context-free languages [19]. For our purpose, it suffices to consider transductions τ from a free monoid A^* into an arbitrary monoid M such that, if P is a recognizable subset of M, then $\tau^{-1}(P)$ is a recognizable subset of A^*. It is well known that rational transductions have this property. In this case, τ can be realized by a transducer, which is essentially a non deterministic automaton with output. Now, just as automata can be converted into semigroups, automata with outputs can be converted into matrix representations. In particular, every rational transduction admits a linear representation. It turns out that the important property is to have a matrix representation. Whether this representation is linear or not is actually irrelevant for our purpose. We now give the formal definitions.

Let M be a monoid. A *transduction* $\tau : A^* \to M$ is a relation from A^* into M, i.e. a function from A^* into $\mathcal{P}(M)$. If P is a subset of M, $\tau^{-1}(P)$ is the image of P by the relation $\tau^{-1} : M \to A^*$. Therefore

$$\tau^{-1}(P) = \{u \in A^* \mid \tau(u) \cap P \neq \emptyset\}$$

The definition of a representable transduction requires some preliminaries. The set $\mathcal{P}(M)$ is a semiring under union (as addition) and subset product (as multiplication). Therefore, for each $n > 0$, the set $\mathcal{P}(M)^{n\times n}$ of n by n matrices with entries in $\mathcal{P}(M)$ is again a semiring for addition and multiplication of matrices induced by the operations in $\mathcal{P}(M)$.

Let X be an alphabet and let $M \oplus X^*$ be the free product of M and X^*, that is, the set of the words of the form $m_0x_1m_1x_2m_2\cdots x_km_k$, with $m_0, m_1, \ldots, m_k \in M$ and $x_1, \ldots, x_k \in X$, equipped with the product

$$\begin{aligned}(m_0x_1m_1x_2m_2 &\cdots x_km_k)(m'_0x'_1m'_1x'_2m'_2\cdots x'_km'_k) = \\ &m_0x_1m_1x_2m_2\cdots x_k(m_km'_0)x'_1m'_1x'_2m'_2\cdots x'_km'_k\end{aligned}$$

A series in the non commutative variables X with coefficients in $\mathcal{P}(M)$ is an element of $\mathcal{P}(M \oplus X^*)$, that is, a formal sum of words of the form $m_0x_1m_1x_2m_2\cdots x_km_k$.

A *representation of dimension* n for the transduction τ is a pair (μ, s), where μ is a monoid morphism from A^* into $\mathcal{P}(M)^{n\times n}$ and s is a series in the non commutative variables $\{x_{1,1}, \ldots, x_{n,n}\}$ with coefficients in $\mathcal{P}(M)$ such that, for all $u \in A^*$,

$$\tau(u) = s[\mu(u)]$$

where the expression $s[\mu(u)]$ denotes the subset of M obtained by substituting $(\mu(u)_{1,1}, \ldots, \mu(u)_{n,n})$ for $(x_{1,1}, \ldots, x_{n,n})$ in s. A transduction is *representable* if it admits a representation. The following example should help the reader to understand this rather abstract definition.

Example 6.18 Let $A = \{a, b\}$, $M = A^*$ and let μ be the morphism from A^* into $\mathcal{P}(A^*)^{2\times 2}$ defined by

$$\mu(u) = \begin{pmatrix} u & \emptyset \\ \emptyset & A^{|u|} \end{pmatrix}$$

Then the following transductions from A^* into A^* admit a representation of the form (μ, s) for some series s.

(1) $\tau_1(u) = A^{|u|}uA^{|u|}$

(2) $\tau_2(u) = L_0uL_1u \ldots uL_k$, where $L_0, \ldots, L_k$ are arbitrary languages,

It suffices to take

$$s = x_{2,2}x_{1,1}x_{2,2}$$

in the first case and

$$s = \sum_{u_0 \in L_0, \ldots, u_k \in L_k} u_0x_{1,1}u_1 \cdots x_{1,1}u_k$$

in the second case.

Suppose that a transduction $\tau : A^* \to M$ admits a representation (μ, s), where $\mu : A^* \to \mathcal{P}(M)^{n\times n}$. Then every monoid morphism $\varphi : M \to N$ induces a monoid morphism $\varphi : \mathcal{P}(M)^{n\times n} \to \mathcal{P}(N)^{n\times n}$. The main property of representable transductions can now be stated.

Theorem 6.19 [115, 116] *Let (μ, s) be a representation for a transduction $\tau : A^* \to M$. If P is a subset of M recognized by φ, then the language $\tau^{-1}(P)$ is recognized by the monoid $(\varphi \circ \mu)(A^*)$.*

Corollary 6.20 *Let $\tau : A^* \to M$ be a representable transduction. Then for every recognizable subset P of M, the language $\tau^{-1}(P)$ is recognizable.*

The precise description of the monoid $(\varphi\circ\mu)(A^*)$ is the key to understand several operations on languages.

Example 6.21 This is a continuation of Example 6.18. Let

$$\tau_1(u) = A^{|u|}uA^{|u|}.$$

Then, for every language L of A^*,

$$\tau_1^{-1}(L) = \{u \in A^* \mid \text{ there exist } u_0,\ u_1 \text{ with } |u_0| = |u| = |u_1| \\ \text{and } u_0uu_1 \in L\}$$

Thus $\tau_1^{-1}(L)$ is the set of "middle thirds" of words of L.

Let $\tau_2(u) = u^2$. Then, for every language L of A^*,

$$\tau_2^{-1}(L) = \{u \in A^* \mid u^2 \in L\}$$

Thus $\tau_2^{-1}(L)$ is the "square root" of L. In both cases, the transduction has a representation of the form (μ, s), where $\mu : A^* \to \mathcal{P}(A^*)^{2\times 2}$ is defined by

$$\mu(u) = \begin{pmatrix} u & \emptyset \\ \emptyset & A^{|u|} \end{pmatrix}$$

Thus if $\varphi : A^* \to N$ is a monoid morphism and if $\varphi(A) = X$, $(\varphi \circ \mu)(A^*)$ is a monoid of matrices of $\mathcal{P}(N)^{2\times 2}$ of the form

$$\begin{pmatrix} \{x\} & \emptyset \\ \emptyset & X^k \end{pmatrix}$$

with $x \in N$ and $k \geq 0$. Thus this monoid can be identified with a submonoid of $N \times C$, where C is the submonoid of $\mathcal{P}(N)$ generated by X. In particular, C is commutative.

We now give some examples of application of Theorem 6.19.

Inverse substitutions.

Recall that a substitution from A^* into M is a monoid morphism from A^* into $\mathcal{P}(M)$. Therefore a substitution $\sigma : A^* \to M$ has a representation of dimension 1. Thus if L is a subset of M recognized by a monoid N, then $\sigma^{-1}(L)$ is recognized by a submonoid of $\mathcal{P}(N)$.

Length preserving morphisms.

Let $\varphi : A^* \to B^*$ be a length preserving morphism. Then the transduction $\varphi^{-1} : B^* \to A^*$ is a substitution. Thus, if L is a subset of A^* recognized by a monoid N, then $\varphi(L)$ is recognized by a submonoid of $\mathcal{P}(N)$.

Shuffle product.

Recall that the *shuffle* of n words $u_1, \ldots, u_n$ is the set $u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_n$ of all words of the form

$$u_{1,1}u_{2,1} \cdots u_{n,1}u_{1,2}u_{2,2} \cdots u_{n,2} \cdots u_{1,k}u_{2,k} \cdots u_{n,k}$$

with $k \geq 0$, $u_{i,j} \in A^*$, such that $u_{i,1}u_{i,2} \cdots u_{i,k} = u_i$ for $1 \leq i \leq n$. The shuffle of k languages $L_1, \ldots, L_k$ is the language

$$L_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup L_k = \bigcup_{u_1 \in L_1, \ldots, u_k \in L_k} u_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup u_k$$

Let $\tau : A^* \to A^* \times \cdots \times A^*$ be the transduction defined by

$$\tau(u) = \{(u_1, \cdots, u_k) \in A^* \times \cdots \times A^* \mid u \in u_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup u_k\}$$

Then $\tau^{-1}(L_1 \times \cdots \times L_k) = L_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup L_k$. Furthermore τ is a substitution defined, for every $a \in A$, by

$$\tau(a) = \{(a, 1, \ldots, 1), (1, a, 1, \ldots, 1), \ldots, (1, 1, \ldots, 1, a)\}$$

Thus, if $L_1, \ldots, L_k$ are languages recognized by monoids $M_1, \ldots, M_k$, respectively, then $L_1 \text{ш} \cdots \text{ш} L_k$ is recognized by a submonoid of $\mathcal{P}(M_1 \times \cdots \times M_k)$.

Other examples include the concatenation product — which leads to a construction on monoids called the *Schützenberger product* [163] — and the inverse of a sequential function, or more generally of rational function, which is intimately related to the wreath product. See [115, 116, 105] for the details.

7 The concatenation product

The concatenation product is certainly the most studied operation on languages. As was the case in section 6.1, we shall actually consider products of the form $L_0a_1L_1 \cdots a_nL_n$, where the a_i's are letters or of the form $u_0L_1u_1 \cdots L_nu_n$, where the u_i's are words.

7.1 Polynomial closure

Polynomial operations comprise finite union and concatenation product. This terminology, first introduced by Schützenberger, comes from the fact that languages form a semiring under union as addition and concatenation as multiplication. There are in fact two slightly different notions of polynomial closure, one for $+$-classes and one for $*$-classes.

The *polynomial closure* of a set of languages $\mathcal{L}$ of A^* is the set of languages of A^* that are finite unions of languages of the form

$$L_0a_1L_1 \cdots a_nL_n$$

where $n \geq 0$, the a_i's are letters and the L_i's are elements of $\mathcal{L}$.

The *polynomial closure* of a set of languages $\mathcal{L}$ of A^+ is the set of languages of A^+ that are finite unions of languages of the form

$$u_0L_1u_1 \cdots L_nu_n$$

where $n \geq 0$, the u_i's are words of A^* and the L_i's are elements of $\mathcal{L}$. If $n = 0$, one requires of course that u_0 is not the empty word.

By extension, if $\mathcal{V}$ is a $*$-variety (resp. $+$-variety), we denote by Pol $\mathcal{V}$ the class of languages such that, for every alphabet A, Pol $\mathcal{V}(A^*)$ (resp. Pol $\mathcal{V}(A^+)$) is the polynomial closure of $\mathcal{V}(A^*)$ (resp. $\mathcal{V}(A^+)$). Symmetrically, we denote by Co-Pol $\mathcal{V}$ the class of languages such that, for every alphabet A, Co-Pol $\mathcal{V}(A^*)$ (resp. Co-Pol $\mathcal{V}(A^+)$) is the set of languages L whose complement is in Pol $\mathcal{V}(A^*)$ (resp. Pol $\mathcal{V}(A^+)$). Finally, we denote by BPol $\mathcal{V}$ the class of languages such that, for every alphabet A, BPol $\mathcal{V}(A^*)$ (resp. BPol $\mathcal{V}(A^+)$) is the closure of Pol $\mathcal{V}(A^*)$ (resp. Pol $\mathcal{V}(A^+)$) under finite boolean operations (finite union and complement).

Proposition 6.1 above is the first step in the proof of the following algebraic characterization of the polynomial closure [126, 127], which makes use of a deep combinatorial result of semigroup theory [153, 154, 155].

Theorem 7.1 *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the associated variety of languages. Then Pol* $\mathcal{V}$ *is a positive variety and the associated variety of finite ordered monoids is the Mal'cev product* $[\![x^\omega y x^\omega \leq x^\omega]\!] \circledM \mathbf{V}$.

In the case of +-varieties, the previous result also holds with the appropriate definition of polynomial closure. The following consequence was first proved by Arfi [11, 12].

Corollary 7.2 *For each variety of languages* $\mathcal{V}$*, Pol* $\mathcal{V}$ *and Co-Pol* $\mathcal{V}$ *are positive varieties of languages. In particular, for each alphabet* A*, Pol* $\mathcal{V}(A^*)$ *and Co-Pol* $\mathcal{V}(A^*)$ *(resp. Pol* $\mathcal{V}(A^+)$ *and Co-Pol* $\mathcal{V}(A^+)$ *in the case of a +-variety) are closed under finite union and intersection.*

7.2 Unambiguous and deterministic polynomial closure

The *unambiguous polynomial closure* of a set of languages $\mathcal{L}$ of A^* is the set of languages of A^* that are finite unions of unambiguous products of the form $L_0a_1L_1 \cdots a_nL_n$, where $n \geq 0$, the a_i's are letters and the L_i's are elements of $\mathcal{L}$. Similarly, the unambiguous polynomial closure of a set of languages $\mathcal{L}$ of A^+ is the set of languages of A^+ that are finite unions of unambiguous products of the form

$$u_0L_1u_1 \cdots L_nu_n$$

where $n \geq 0$, the u_i's are words of A^* and the L_i's are elements of $\mathcal{L}$. If $n = 0$, one requires that u_0 is not the empty word.

The *left and right deterministic polynomial closure* are defined analogously, by replacing "unambiguous" by "left (resp. right) deterministic".

By extension, if $\mathcal{V}$ is a variety of languages, we denote by UPol $\mathcal{V}$ the class of languages such that, for every alphabet A, UPol $\mathcal{V}(A^*)$ (resp. UPol $\mathcal{V}(A^+)$) is the unambiguous polynomial closure of $\mathcal{V}(A^*)$ (resp. $\mathcal{V}(A^+)$). Similarly, the left (resp. right) deterministic polynomial closure of $\mathcal{V}$ is denoted D^ℓPol $\mathcal{V}$ (resp. D^rPol $\mathcal{V}$). The algebraic counterpart of the unambiguous polynomial closure is given in the following theorems [97, 120].

Theorem 7.3 *Let* $\mathbf{V}$ *be a variety of finite monoids (resp. semigroups) and let* $\mathcal{V}$ *be the associated variety of languages. Then UPol* $\mathcal{V}$ *is a variety of languages, and the associated variety of monoids (resp. semigroups) is* $[\![x^\omega y x^\omega = x^\omega]\!] \circledM \mathbf{V}$.

Theorem 6.5 leads to the following description of a set of identities defining the variety $[\![x^\omega y x^\omega \leq x^\omega]\!] \circledM \mathbf{V}$.

Proposition 7.4 *Let* $\mathbf{V}$ *be a variety of monoids. Then* $[\![x^\omega y x^\omega \leq x^\omega]\!] \circledM \mathbf{V}$ *is defined by the identities of the form* $x^\omega y x^\omega \leq x^\omega$*, where* $x, y \in \hat{A}^*$ *for some finite set* A *and* $\mathbf{V}$ *satisfies* $x = y = x^2$.

Recall that the variety $[\![x^\omega y x^\omega = x^\omega]\!]$ is the variety $\mathbf{LI}$ of locally trivial semigroups. Thus, the $+$-variety associated with $\mathbf{LI}$, described in Proposition 5.18, is also the smallest $+$-variety closed under unambiguous product. This is a consequence of Theorem 7.3, applied with $\mathbf{V} = \mathbf{I}$. For $\mathbf{V} = \mathbf{J}_1$, one can show that $\mathbf{LI} \circledM \mathbf{J}_1$ is equal to $\mathbf{DA}$, the variety of finite monoids in which each regular $\mathcal{J}$-class is an idempotent semigroup. This variety is defined by the identities $(xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega$ and $x^\omega = x^{\omega+1}$ [148].

Corollary 7.5 *For each alphabet* A*,* $\mathcal{DA}(A^*)$ *is the smallest set of languages of* A^* *containing the languages of the form* B^**, with* $B \subseteq A$*, and closed under disjoint union and unambiguous product. Equivalently,* $\mathcal{DA}(A^*)$ *is the set of languages that are disjoint unions of unambiguous products of the form* $A_0^* a_1 A_1^* a_2 \cdots a_k A_k^*$*, where the* a_i*'s are letters and the* A_i*'s are subsets of* A.

An interesting consequence of the conjunction of Theorems 7.1 and 7.3 is the following characterization of UPol $\mathcal{V}$, which holds for $*$-varieties as well as for $+$-varieties.

Theorem 7.6 *Let* $\mathcal{V}$ *be a variety of languages. Then Pol* $\mathcal{V} \cap$ *Co-Pol* $\mathcal{V} =$ *UPol* $\mathcal{V}$.

For the left (resp. right) deterministic product, similar results hold [97, 98]. We just state the result for $*$-varieties.

Theorem 7.7 *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the associated variety of languages. Then* D^ℓ*Pol* $\mathcal{V}$ *(resp.* D^r*Pol* $\mathcal{V}$*) is a variety of languages, and the associated variety of monoids is* $[\![x^\omega y = x^\omega]\!] \circledM \mathbf{V}$ *(resp.* $[\![y x^\omega = x^\omega]\!] \circledM \mathbf{V}$*).*

One can show that $[\![x^\omega y = x^\omega]\!] \circledM \mathbf{J}_1$ is equal to the variety $\mathbf{R}$ of all finite $\mathcal{R}$-trivial monoids, which is also defined by the identity $(xy)^\omega x = (xy)^\omega$. This leads to the following characterization [53, 39]

Corollary 7.8 *For each alphabet* A*,* $\mathcal{R}(A^*)$ *consists of the languages which are disjoint unions of languages of the form* $A_0^* a_1 A_1^* a_2 \cdots a_k A_k^*$*, where* $k \geq 0$*,* $a_1, \ldots a_n \in A$ *and the* A_i*'s are subsets of* A *such that* $a_i \notin A_{i-1}$*, for* $1 \leq i \leq k$.

A dual result holds for $\mathcal{L}$-trivial monoids.

7.3 Varieties closed under product

A set of languages $\mathcal{L}$ is *closed under product*, if, for each $L_0, \ldots, L_n \in \mathcal{L}$ and $a_1, \ldots, a_n \in A$, $L_0 a_1 L_1 \cdots a_n L_n \in \mathcal{L}$. A $*$-variety of languages $\mathcal{C}$ is *closed under product*, if, for each alphabet A, $\mathcal{V}(A^*)$ is closed under product. The next theorem, due to Straubing [160], shows, in essence, that closure under product also corresponds to a Mal'cev product.

Theorem 7.9 *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the associated variety of languages. For each alphabet* A*, let* $\mathcal{W}(A^*)$ *be the smallest boolean algebra containing* $\mathcal{V}(A^*)$ *and closed under product. Then* $\mathcal{W}$ *is a* $*$*-variety and the associated variety of finite monoids is* $\mathbf{A} \circledM \mathbf{V}$*.*

This important result contains Theorem 5.2 as a particular case, when $\mathbf{V}$ is the trivial variety of monoids. Examples of varieties of finite monoids $\mathbf{V}$ satisfying the equality $\mathbf{A} \circledM \mathbf{V} = \mathbf{V}$ include varieties of finite monoids defined by properties of their groups. A group in a monoid M is a subsemigroup of M containing an identity e, which can be distinct from the identity of M. As was mentioned in section 2.8, the maximal groups in a monoid are exactly the $\mathcal{H}$-classes containing an idempotent. Given a variety of finite groups $\mathbf{H}$, the class of finite monoids whose groups belong to $\mathbf{H}$ form a variety of finite monoids, denoted $\overline{\mathbf{H}}$. In particular, if $\mathbf{H}$ is the trivial variety of finite groups, then $\overline{\mathbf{H}} = \mathbf{A}$, since a monoid is aperiodic if and only if it contains no non trivial group.

Theorem 7.10 *For any variety of finite groups* $\mathbf{H}$*,* $\mathbf{A} \circledM \overline{\mathbf{H}} = \overline{\mathbf{H}}$*.*

It follows, by Theorem 7.9, that the $*$-variety associated with a variety of monoids of the form $\overline{\mathbf{H}}$ is closed under product. Varieties of this type will be considered in Theorems 7.19 and 9.4 below.

7.4 The operations $L \to LaA^*$ and $L \to A^*aL$

A slightly stronger version of Theorem 7.9 can be given [98, 191, 193].

Theorem 7.11 *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the associated variety of languages. Let* $\mathcal{W}$ *be the variety of languages associated with* $\mathbf{A} \circledM \mathbf{V}$*. Then, for each alphabet* A*,* $\mathcal{W}(A^*)$ *is the smallest boolean algebra of languages containing* $\mathcal{V}(A^*)$ *and closed under the operations* $L \to LaA^*$ *and* $L \to A^*aL$*, where* $a \in A$*.*

In view of this result, it is natural to look at the operation $L \to LaA^*$ [98, 191, 193].

Theorem 7.12 *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the associated variety of languages. For each alphabet* A*, let* $\mathcal{W}(A^*)$ *be the boolean* *algebra generated by the languages* L *or* LaA^**, where* $a \in A$ *and* $L \in \mathcal{V}(A^*)$*. Then* $\mathcal{W}$ *is a variety of languages and the associated variety of monoids is* $\mathbf{J}_1 * \mathbf{V}$*.*

Since the variety $\mathbf{R}$ of $\mathcal{R}$-trivial monoids is the smallest variety closed under semidirect product and containing the commutative and idempotent monoids, one gets immediately

Corollary 7.13 *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the associated variety of languages. For each alphabet* A*, let* $\mathcal{W}(A^*)$ *be the smallest boolean algebra of languages containing* $\mathcal{V}(A^*)$ *and closed under the operations* $L \to LaA^*$*, for all* $a \in A$*. Then* $\mathcal{W}$ *is a variety of languages and the associated variety of monoids is* $\mathbf{R} * \mathbf{V}$*.*

In particular, this leads to another description of the languages associated with $\mathbf{R}$ (compare with Corollary 7.8).

Corollary 7.14 *For each alphabet* A*,* $\mathcal{R}(A^*)$ *is the smallest boolean algebra of languages closed under the operations* $L \to LaA^*$*, for all* $a \in A$*.*

Operations of the form $L \to LaA^*$ and $L \to A^*aL$ were also used by Thérien [176] to describe the languages whose syntactic monoid is idempotent (see also [38, 53]). This characterization is somewhat unusual, since it is given by induction on the size of the alphabet.

Theorem 7.15 *Let* $\mathbf{V}$ *be the variety of finite idempotent monoids and let* $\mathcal{V}$ *be the associated variety of languages. Then* $\mathcal{V}(\emptyset^*) = \{\emptyset, \{1\}\}$*, and for each non empty alphabet* A*,* $\mathcal{V}(A^*)$ *is the smallest boolean algebra of languages containing the languages of the form* A^*aA^**,* LaA^* *and* A^*aL*, where* $a \in A$*,* $L \in \mathcal{V}((A \setminus \{a\})^*$ *and* $L \in \mathcal{V}(A \setminus \{a\})^*$*.*

The languages associated with subvarieties of the variety of finite idempotent monoids are studied in [151].

7.5 Product with counters

Let $L_0, \ldots, L_k$ be languages of A^*, let $a_1, \ldots, a_k$ be letters of A and let r and p be integers such that $0 \le r < p$. We define $(L_0a_1L_1 \cdots a_kL_k)_{r,p}$ to be the set of all words u in A^* such that the number of factorizations of u in the form

$$u = u_0a_1u_1 \cdots a_ku_k$$

with $u_i \in L_i$ for $0 \le i \le k$, is congruent to r modulo p. This *product with counter* is especially useful for the study of group languages. A recognizable language is called a *group language* if its syntactic monoid is a group. Since equality is the only stable order on a finite group, the ordered syntactic monoid is useless in the case of a group language.

A frequently asked question is whether there is some "nice" combinatorial description of the group languages. No such description is known for the variety of all group languages, but there are simple descriptions for some subvarieties. We have already described the variety of languages corresponding to commutative groups. We will now consider the varieties of p-groups, nilpotent groups and solvable groups.

The p-groups form a variety of finite monoids $\mathbf{G}_p$. The associated variety of languages $\mathcal{G}_p$ is given in [53], where the result is credited to Schützenberger. For $u, v \in A^*$, denote by $\binom{v}{u}$ the number of distinct factorizations of the form $v = v_0a_1v_1 \cdots a_nv_n$ such that $v_0, \cdots, v_n \in A^*$, $a_1, \cdots, a_n \in A$ and $a_1 \cdots a_n = u$. In other words $\binom{v}{u}$ is the number of distinct ways to write u as a subword of v. For example

$$\binom{abab}{ab} = 3 \qquad \binom{aabbaa}{aba} = 8$$

Theorem 7.16 *For each alphabet A, $\mathcal{G}_p(A^*)$ is the boolean algebra generated by the languages*

$$S(u, r, p) = \left\{ v \in A^* \mid \binom{v}{u} \equiv r \bmod p \right\}$$

for $0 \leq r < p$ and $u \in A^*$. It is also the boolean algebra generated by the languages $(A^*a_1A^* \cdots a_kA^*)_{r,p}$ where $0 \leq r < p$ and $k \geq 0$.

Nilpotent groups form a variety of finite monoids **Gnil**. A standard result in group theory states that a finite group is nilpotent if and only if it is isomorphic to a direct product $G_1 \times \cdots \times G_n$, where each G_i is a p_i-group for some prime p_i. This result leads to the following description of the variety of languages $\mathcal{G}nil$ associated with **Gnil** [53, 173].

Theorem 7.17 *For each alphabet A, $\mathcal{G}nil(A^*)$ is the boolean algebra generated by the languages $S(u, r, p)$, where p is a prime number, $0 \leq r < p$ and $u \in A^*$. It is also the boolean algebra generated by the languages $(A^*a_1A^* \cdots a_kA^*)_{r,p}$, where $a_1, \ldots, a_k \in A$, $0 \leq r < p$, p is a prime number and $k \geq 0$.*

Solvable groups also form a variety of finite monoids **Gsol**. The associated variety of languages $\mathcal{G}sol$ was described by Straubing [159].

Theorem 7.18 *For each alphabet A, $\mathcal{G}sol(A^*)$ is the smallest boolean algebra of languages closed under the operations $L \to (LaA^*)_{r,p}$, where $a \in A$, p is a prime number and $0 \leq r < p$.*

The variety of languages associated with $\overline{\mathbf{Gsol}}$ was first described by Straubing [159]. See also [178, 169]. The formulation given below is due to Weil [191, 193].

Theorem 7.19 *Let $\mathcal{V}$ be the $*$-variety associated with $\overline{\mathbf{Gsol}}$. For each alphabet A, $\mathcal{V}(A^*)$ is the smallest boolean algebra of languages closed under the operations $L \to LaA^*$ and $L \to (LaA^*)_{r,p}$, where $a \in A$, p is a prime number and $0 \leq r < p$.*

The variety of languages associated with the variety of finite monoids in which every $\mathcal{H}$-class is a solvable group is described in the next theorem, which mixes the ideas of Theorems 7.19 and 7.15.

Theorem 7.20 Thérien [176] *Let $\mathbf{V}$ be a variety of finite monoids in which every $\mathcal{H}$-class is a solvable group and let $\mathcal{V}$ be the associated variety of languages. Then $\mathcal{V}(\emptyset^*) = \{\emptyset, \{1\}\}$, and for each non empty alphabet A, $\mathcal{V}(A^*)$ is the smallest boolean algebra of languages closed under the operations $L \to (LaA^*)_{r,n}$, where $a \in A$ and $0 \leq r < n$, and containing the languages of the form A^*aA^*, LaA^* and A^*aL, where $a \in A$, $L \in \mathcal{V}((A \setminus \{a\})^*$.*

7.6 Varieties closed under product with counter

Finally, let us mention the results of Weil [193]. Let n be an integer. A set of languages $\mathcal{L}$ of A^* is *closed under product with n-counters* if, for any language $L_0, \ldots, L_k \in \mathcal{L}$, for any letter $a_1, \ldots, a_k \in A$ and for any integer r such that $0 \leq r < n$, $(L_0a_1L_1 \cdots a_kL_k)_{r,n} \in \mathcal{L}$. A set of languages $\mathcal{L}$ of A^* is *closed under product with counters* if it is closed under product with n-counters, for arbitrary n.

Theorem 7.21 *Let p be a prime number, let $\mathbf{V}$ be a variety of finite monoids and let $\mathcal{V}$ be the associated variety of languages. For each alphabet A, let $\mathcal{W}(A^*)$ be the smallest boolean algebra containing $\mathcal{V}(A^*)$ and closed under product with p-counters. Then $\mathcal{W}$ is a variety of languages and the associated variety of monoids is $\mathbf{LG}_p \circledM \mathbf{V}$.*

Theorem 7.22 *Let p be a prime number, let $\mathbf{V}$ be a variety of finite monoids and let $\mathcal{V}$ be the associated variety of languages. For each alphabet A, let $\mathcal{W}(A^*)$ be the smallest boolean algebra containing $\mathcal{V}(A^*)$ and closed under product and product with p-counters. Then $\mathcal{W}$ is a variety of languages and the associated variety of monoids is $\mathbf{L}\overline{\mathbf{G}}_p \circledM \mathbf{V}$.*

Theorem 7.23 *Let $\mathbf{V}$ be a variety of finite monoids and let $\mathcal{V}$ be the associated variety of languages. For each alphabet A, let $\mathcal{W}(A^*)$ be the boolean algebra containing $\mathcal{V}(A^*)$ and closed under product with counters. Then $\mathcal{W}$ is a variety of languages and the associated variety of monoids is $\mathbf{LGsol} \circledM \mathbf{V}$.*

For instance, if $\mathbf{V} = \mathbf{J}_1$ it is known that $\mathbf{LGsol} \circledM \mathbf{J}_1$ is the variety of monoids whose regular $\mathcal{D}$-classes are unions of solvable groups.

Theorem 7.24 *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the associated variety of languages. For each alphabet* A*, let* $\mathcal{W}(A^*)$ *be the smallest boolean algebra containing* $\mathcal{V}(A^*)$ *and closed under product and product with counters. Then* $\mathcal{W}$ *is a variety of languages and the associated variety of monoids is* $\mathbf{L\overline{Gsol}} \circledM \mathbf{V}$*.*

8 Concatenation hierarchies

By alternating the use of the polynomial closure and of the boolean closure one can obtain hierarchies of recognizable languages. Let $\mathcal{V}$ be a variety of languages. The concatenation hierarchy of basis $\mathcal{V}$ is the hierarchy of classes of languages defined as follows.

(1) level 0 is $\mathcal{V}$

(2) for every integer $n \geq 0$, level $n + 1/2$ is the polynomial closure of level n

(3) for every integer $n \geq 0$, level $n + 1$ is the boolean closure of level $n + 1/2$.

Theorem 7.1 shows that the polynomial closure of a variety of languages is a positive variety of languages. Furthermore the boolean closure of a positive variety of languages is a variety of languages. Therefore, one defines a sequence of varieties $\mathcal{V}_n$ and of positive varieties $\mathcal{V}_{n+1/2}$, where n is an integer, as follows:

(1) $\mathcal{V}_0 = \mathcal{V}$

(2) for every integer $n \geq 0$, $\mathcal{V}_{n+1/2} = \text{Pol } \mathcal{V}_n$,

(3) for every integer $n \geq 0$, $\mathcal{V}_{n+1} = \text{BPol } \mathcal{V}_n$.

The associated varieties of semigroups and ordered semigroups (resp. monoids and ordered monoids) are denoted $\mathbf{V}_n$ and $\mathbf{V}_{n+1/2}$. Theorem 7.1 gives an explicit relation between $\mathbf{V}_n$ and $\mathbf{V}_{n+1/2}$.

Proposition 8.1 *For every integer* $n \geq 0$*,* $\mathbf{V}_{n+1/2} = [\![x^\omega y x^\omega \leq x^\omega]\!] \circledM \mathbf{V}_n$*.*

Three concatenation hierarchies have been considered so far in the literature. The first one, introduced by Brzozowski [36] and called the *dot-depth hierarchy*, is the hierarchy of positive $+$-varieties whose basis is the trivial variety. The second one, first considered implicitly in [174] and explicitly in [163, 166] is called the *Straubing-Thérien hierarchy*: it is the hierarchy of positive $*$-varieties whose basis is the trivial variety. The third one, introduced in [83], is the hierarchy of positive $*$-varieties whose basis is the variety of group-languages. It is called the *group hierarchy*.

It can be shown that these three hierarchies are strict: if A contains at least two letters, then for every n, there exist languages of level $n + 1$ which are not of level $n + 1/2$ and languages of level $n + 1/2$ which are not of level n. This was first proved for the dot-depth hierarchy in [40].

The main question is the decidability of each level: given an integer n (resp. $n + 1/2$) and a recognizable language L, decide whether or not L is of level n (resp. $n + 1/2$). The language can be given either by a finite automaton, by a finite semigroup or by a rational expression since there are standard algorithms to pass from one representation to the other.

There is a wide literature on the concatenation product: Arfi [11, 12], Blanchet-Sadri [24, 25, 26, 27, 28, 29, 30, 31, 32], Brzozowski [36, 40, 41], Cowan [47], Eilenberg [53], Knast [71, 72, 40], Schützenberger[144, 148], Simon [41, 152, 156], Straubing [118, 120, 160, 164, 167, 170, 172], Thérien [120, 170], Thomas [184], Weil [190, 172, 194, 126, 127] and the author [97, 100, 104, 118, 120, 126, 127]. The reader is referred to the survey articles [110, 111] for more details on these results.

We now describe in more details the first levels of each of these hierarchies. We consider the Straubing-Thérien hierarchy, the dot-depth hierarchy and the group hierarchy, in this order.

8.1 Straubing-Thérien's hierarchy

Level 0 is the trivial $*$-variety. Therefore a language of A^* is of level 0 if and only if it is empty or equal to A^*. This condition is easily characterized.

Proposition 8.2 *A language is of level* 0 *if and only if its syntactic monoid is trivial.*

It is also well known that one can decide in polynomial time whether the language of A^* accepted by a deterministic n-state automaton is empty or equal to A^* (that is, of level 0).

By definition, the sets of level 1/2 are the finite unions of languages of the form $A^*a_1A^*a_2 \cdots a_kA^*$, where the a_i's are letters. An alternative description can be given in terms of shuffle. A language is a *shuffle ideal* if and only if for every $u \in L$ and $v \in A^*$, $u \text{ Ш } v$ is contained in L.

Proposition 8.3 *A language is of level* $1/2$ *if and only if it is a shuffle ideal.*

It follows from Theorem 7.6 that the only shuffle ideals whose complement is also a shuffle ideal are the full language and the empty language. It is easy to see directly that level 1/2 is decidable. One can also derive this result from our syntactic characterization.

Proposition 8.4 *A language is of level* $1/2$ *if and only if its ordered syntactic monoid satisfies the identity* $x \leq 1$*.*

One can derive from this result a polynomial algorithm to decide whether the language accepted by a complete deterministic n-state automaton is of level 1/2. See [126, 127] for details.

Corollary 8.5 *One can decide in polynomial time whether the language accepted by a complete deterministic n-state automaton is of level $1/2$.*

The sets of level 1 are the finite boolean combinations of languages of the form $A^*a_1A^*a_2 \cdots a_kA^*$, where the a_i's are letters. In particular, all finite sets are of level 1. The sets of level 1 have a nice algebraic characterization, due to Simon [152]. There exist now several proofs of this deep result [3, 170, 157, 64].

Theorem 8.6 *A language of A^* is of level 1 if and only if its syntactic monoid is $\mathcal{J}$-trivial, or, equivalently, if and only if it satisfies the identities $x^\omega = x^{\omega+1}$ and $(xy)^\omega = (yx)^\omega$.*

Thus $\mathbf{V}_1 = \mathbf{J}$, the variety of finite $\mathcal{J}$-trivial monoids. Theorem 8.6 yields an algorithm to decide whether a given recognizable set is of level 1. See [157].

Corollary 8.7 *One can decide in polynomial time whether the language accepted by a deterministic n-state automaton is of level 1.*

The sets of level 3/2 also have a simple description, although this is not a direct consequence of the definition [118].

Theorem 8.8 *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.*

We derive the following syntactic characterization [126, 127].

Theorem 8.9 *A language is of level $3/2$ if and only if its ordered syntactic monoid satisfies the identity $x^\omega yx^\omega \leq x^\omega$ for every x, y such that $c(x) = c(y)$.*

Corollary 8.10 *There is an algorithm, in time polynomial in $2^{|A|}n$, for testing whether the language of A^* accepted by a deterministic n-state automaton is of level $3/2$.*

We now arrive to the level 2. Theorem 8.8 gives a combinatorial description of the languages of level 2 [118].

Theorem 8.11 *The languages of level 2 of A^* are the finite boolean combinations of the languages 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 next theorem [118] gives a non-trivial (but unfortunately non effective) algebraic characterization of level 2. Given a variety of monoids $\mathbf{V}$, denote by $\mathbf{PV}$ the variety generated by all monoids of the form $\mathcal{P}(M)$, where $M \in \mathbf{V}$.

Theorem 8.12 *A language is of level* 2 *in the Straubing-Thérien hierarchy if and only if its syntactic monoid belongs to* **PJ**.

Let $\mathbf{V}_2$ be the variety of finite monoids associated with the languages of level 2. Theorem 8.12 shows that $\mathbf{V}_2 = \mathbf{PJ}$. Unfortunately, no algorithm is known to decide whether a finite monoid divides the power monoid of a $\mathcal{J}$-trivial monoid. In other words, the decidability problem for level 2 is still open, although much progress has been made in recent years [25, 29, 47, 118, 167, 172, 190, 194]. This problem is a particular case of a more general question discussed in section 8.5.

In the case of languages whose syntactic monoid is an inverse monoid, a complete characterization can be given [47, 126, 127]. An *inverse automaton* is a deterministic automaton $\mathcal{A} = (Q, A \cup \bar{A}, i, F)$ over a symmetrized alphabet $A \cup \bar{A}$, which satisfies, for all $a \in A$, $q, q' \in Q$

$$q \cdot a = q' \text{ if and only if } q' \cdot \bar{a} = q$$

Note however that this automaton is not required to be complete. In other words, in an inverse automaton, each letter defines a partial injective map from Q to Q and the letters a and $\bar{a}$ define mutually reciprocal transitions.

Theorem 8.13 *The language recognized by an inverse automaton* $\mathcal{A} = (Q, A \cup \bar{A}, i, F)$ *is of level* 2 *in the Straubing-Thérien hierarchy if and only if, for all* $q, q' \in Q$, $u, v \in (A \cup \bar{A})^*$, *such that* $q \cdot u$ *and* $q' \cdot u$ *are defined,* $q \cdot v = q'$ *and* $c(v) \subseteq c(u)$ *imply* $q = q'$.

Actually, one can show [126, 127] that each language recognized by an inverse automaton $\mathcal{A}$ is the difference of two languages of level 3/2 recognized by the completion of $\mathcal{A}$. It is proved in [190, 194] that Theorem 8.13 yields the following important corollarylary.

Corollary 8.14 *It is decidable whether an inverse monoid belongs to* $\mathbf{V}_2$.

Example 8.15 Let $A = \{a, b\}$ and let $L = (ab)^*$ be the language of example 5.3. Its minimal automaton satisfies the conditions of Theorem 8.13 and thus it has level 2. In fact, by observing that $\emptyset^* = \{1\}$, L can be written in the form

$$(\emptyset^* \cup \emptyset^* aA^* \cup A^* b\emptyset^*) \setminus (\emptyset^* bA^* \cup \emptyset^* A^* a \cup A^* a\emptyset^* aA^* \cup A^* b\emptyset^* bA^*)$$

Theorem 8.11 describes the languages of level 2 of A^* as finite boolean combinations of the languages of the form $A_0^* a_1 A_1^* a_2 \cdots a_k A_k^*$, where the a_i's are letters and the A_i's are subsets of A. Several subvarieties of languages can be obtained by imposing various conditions on the A_i's. This was the case, for instance for Corollaries 7.5 and 7.8. The known results are summarized in the table below.

Conditions	Variety	Algebraic description	Ref.
Unambiguous product	**DA**	Regular $\mathcal{J}$-classes are idempotent semigroups	[148]
Left det. product	**R**	$\mathcal{R}$-trivial	[97, 98]
Right det. product	**L**	$\mathcal{L}$-trivial	[97, 98]
Bidet. product	$\mathbf{J} \cap \mathbf{Ecom}$	$\mathcal{J}$-trivial and idempotents commute	[123]
$a_i \notin A_{i-1}$	**R**	$\mathcal{R}$-trivial	[53, 39]
$a_i \notin A_i$	**L**	$\mathcal{L}$-trivial	[53, 39]
$a_i \notin A_{i-1} \cup A_i$	$\mathbf{J} \cap \mathbf{Ecom}$	$\mathcal{J}$-trivial and idempotents commute	[17]
$a_i \notin A_{i-1}$ and $A_{i-1} \subseteq A_i$	$\mathbf{L}_1$	Idempotent and $\mathcal{L}$-trivial	[119]
$a_i \notin A_{i-1}$ and $A_i \subseteq A_{i-1}$	$\mathbf{R}_1$	Idempotent and $\mathcal{R}$-trivial	[119]
$A_i \cap A_j = \emptyset$ for $i \neq j$ and $a_i \notin A_{i-1} \cup A_i$	$\mathbf{J} \cap \mathbf{LJ}_1 \cap \mathbf{Ecom}$	The syntactic semigroup is $\mathcal{J}$-trivial, locally idempotent and commutative and its idempotents commute	[150]

The variety of monoids associated with the condition $A_0 \subseteq \cdots \subseteq A_n$ is also described in [5], pp. 234–239, but this description, which requires an infinite number of identities, is too technical to be reproduced here.

Little is known beyond level 2: a semigroup theoretic description of each level of the hierarchy is known, but it is not effective. Each level of the hierarchy is a variety or a positive variety and the associated variety of (ordered) semigroups admits a description by identities, but these identities are not known for $n \geq 2$. Furthermore, even if these identities were known, this would not necessarily lead to a decision process for the corresponding variety. See also the conjecture discussed in section 8.5.

8.2 Dot-depth hierarchy

Level 0 is the trivial $+$-variety. Therefore a language of A^+ is of dot-depth 0 if and only if it is empty or equal to A^+ and one can decide in polynomial time whether the language of A^+ accepted by a deterministic n-state automaton is of level 0.

Proposition 8.16 *A language is of dot-depth* 0 *if and only if its syntactic semigroup is trivial.*

The languages of dot-depth 1/2 are by definition finite unions of languages of the form $u_0A^+u_1A^+ \cdots u_{k-1}A^+u_k$, where $k \geq 0$ and $u_0, \ldots, u_k \in A^*$. But since $A^* = A^+ \cup \{1\}$, these languages can also be expressed as finite unions of languages of the form

$$u_0A^*u_1A^* \cdots u_{k-1}A^*u_k$$

The syntactic characterization is a simple application of our Theorem 7.1.

Proposition 8.17 *A language of A^+ is of dot-depth* 1/2 *if and only if its ordered syntactic semigroup satisfies the identity $x^\omega y x^\omega \leq x^\omega$.*

Corollary 8.18 *One can decide in polynomial time whether the language accepted by a deterministic n-state automaton is of dot-depth* 1/2.

The languages of dot-depth 1 are the finite boolean combinations of languages of dot-depth 1/2. The syntactic characterization of these languages, due to Knast [71, 72], makes use of the Cauchy category.

Theorem 8.19 *A language of A^+ is of dot-depth* 1 *if and only if the Cauchy category of its syntactic semigroup satisfies the following condition (K): if p and r are arrows from e to f and if q and s are arrows from f to e, then $(pq)^\omega ps(rs)^\omega = (pq)^\omega(rs)^\omega$.*

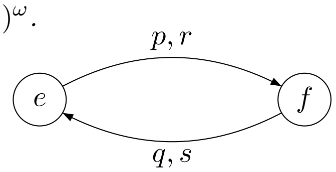

Figure 9: The condition (K).

The variety of finite semigroups satisfying condition (K) is usually denoted $\mathbf{B}_1$ ($\mathbf{B}$ refers to Brzozowski and 1 to level 1). Thus $\mathbf{B}_1$ is defined by the identity

$$(x^\omega py^\omega qx^\omega)^\omega x^\omega py^\omega sx^\omega(x^\omega ry^\omega sx^\omega)^\omega = (x^\omega py^\omega qx^\omega)^\omega(x^\omega ry^\omega sx^\omega)^\omega$$

Theorem 8.19 can be proved in the same way as Theorems 5.19, 5.22 and 5.25. First, one makes use of the wreath product principle to show that a language is of dot-depth 1 if and only if its syntactic semigroup belongs to $\mathbf{J} * \mathbf{LI}$. Then Theorems 6.17 and 6.14 can be applied. See also [5] for a different proof.

The algorithm corresponding to Theorem 8.19 was analyzed by Stern [158].

Corollary 8.20 *One can decide in polynomial time whether the language accepted by a deterministic n-state automaton is of dot-depth* 1.

Straubing [166] discovered an important connection between the Straubing-Thérien and the dot-depth hierarchies. Let $\mathbf{B}_n$ be the variety of finite semigroups corresponding to the languages of dot-depth n and let $\mathbf{V}_n$ be the variety of finite monoids corresponding to the languages of level n in the Straubing-Thérien hierarchy.

Theorem 8.21 *For every integer $n > 0$, $\mathbf{B}_n = \mathbf{V}_n * \mathbf{LI}$.*

Now Theorems 6.17, 6.11 and 6.14 reduce decidability questions about the dot-depth hierarchy to the corresponding questions about the Straubing-Thérien hierarchy.

Theorem 8.22 *For every integer $n \geq 0$, $\mathbf{B}_n$ is decidable if and only if $\mathbf{V}_n$ is decidable.*

It is very likely that Theorem 8.22 can be extended to take in account the half levels, but this is not yet formally proved. In particular, it is not yet known whether level 3/2 of the dot-depth hierarchy is decidable.

8.3 The group hierarchy

We consider in this section the concatenation hierarchy based on the group languages, or group hierarchy. By definition, a language of A^* is of level 0 in this hierarchy if and only if its syntactic monoid is a finite group. This can be easily checked on any deterministic automaton recognizing the language [126, 127].

Proposition 8.23 *One can decide in polynomial time whether the language accepted by a deterministic n-state automaton is a group language.*

The languages of level 1/2 are by definition finite unions of languages of the form $L_0 a_1 L_1 \cdots a_k L_k$ where the a_i's are letters and the L_i's are group languages. By Theorem 7.1, a language is of level 1/2 if and only if its ordered syntactic monoid belongs to the variety $[\![x^\omega y x^\omega \leq x^\omega]\!] \textcircled{\scriptsize M} \mathbf{G}$. The identities of this variety are given by Theorem 6.5.

Proposition 8.24 *A language is of level* 1/2 *in the group hierarchy if and only if its syntactic ordered monoid satisfies the identity $x^\omega \leq 1$.*

Proposition 8.24 can be converted to an algorithm on automata.

Corollary 8.25 *There is a polynomial time algorithm for testing whether the language accepted by a deterministic n-state automaton is of level* 1/2 *in the group hierarchy.*

Level 1/2 is also related with the topology of the free monoid defined by the distance $d_{\mathbf{G}}$. This topology is called the *pro-group* topology. An example of a converging sequence is given by the following proposition [135].

Proposition 8.26 *For every word $u \in A^*$, $\lim_{n\to\infty} u^{n!} = 1$.*

As the multiplication is continuous and a closed set contains the limit of any converging sequence of its elements, it follows that if L is a closed set in the pro-group topology, and if $xu^ny \in L$ for all $n > 0$, then $xy \in L$. The converse is also true if L is recognizable.

Theorem 8.27 *A recognizable set L of A^* is closed in the pro-group topology if and only if for every $u, x, y \in A^*$, $xu^+y \subseteq L$ implies $xy \in L$.*

Since an open set is the complement of a closed set, one can also state:

Theorem 8.28 *A recognizable set L of A^* is open in the pro-group topology if and only if for every $u, x, y \in A^*$, $xy \in L$ implies $xu^+y \cap L \neq \emptyset$.*

These conditions can be easily converted in terms of syntactic monoids.

Theorem 8.29 *Let L be a recognizable language of A^*, let M be its syntactic monoid and let P be its syntactic image.*

(1) *L is closed in the pro-group topology if and only if for every $s, t \in M$ and $e \in E(M)$, $set \in P$ implies $st \in P$.*
(2) *L is open in the pro-group topology if and only if for every $s, t \in M$ and $e \in E(M)$, $st \in P$ implies $set \in P$.*
(3) *L is clopen[10] in the pro-group topology if and only if M is a group.*

Finally, condition (1) states exactly that the syntactic ordered monoid of L satisfies the identity $1 \leq x^\omega$ and condition (2) states that the syntactic ordered monoid of L satisfies the identity $x^\omega \leq 1$. In particular, we get the following result.

Corollary 8.30 *Let L be a recognizable language of A^*. The following conditions are equivalent.*

(1) *L is open in the pro-group topology,*
(2) *L is of level 1/2 in the group hierarchy,*
(3) *the syntactic ordered monoid of L satisfies the identity $x^\omega \leq 1$.*

To finish with these topological properties, let us mention a last result.

Theorem 8.31 *Let L be a language of A^* and let $\bar{L}$ be its closure in the pro-group topology. If L is recognizable, then $\bar{L}$ is recognizable. If L is open and recognizable, then $\bar{L}$ is clopen.*

[10] clopen is a common abbreviation for "closed and open"

A few more definitions are needed to state the algebraic characterization of the languages of level 1. A *block group* is a monoid such that every $\mathcal{R}$-class (resp. $\mathcal{L}$-class) contains at most one idempotent. Block groups form a variety of monoids, denoted **BG**, which is defined by the identity $(x^\omega y^\omega)^\omega = (y^\omega x^\omega)^\omega$. Thus **BG** is a decidable variety.

Theorem 8.32 *A language is of level* 1 *in the group hierarchy if and only if its syntactic monoid belongs to* **BG**.

Corollary 8.33 *There is a polynomial time algorithm for testing whether the language accepted by a deterministic n-state automaton is of level* 1 *in the group hierarchy.*

Several other descriptions of **BG** are known.

Theorem 8.34 *The following equalities holds:* $\mathbf{BG} = \mathbf{PG} = \mathbf{J} * \mathbf{G} = \mathbf{J} \circledM \mathbf{G}$. *Furthermore, a finite monoid belongs to* **BG** *if and only if its set of idempotents generates a $\mathcal{J}$-trivial monoid.*

The results of this section are difficult to prove. They rely on the one hand on a topological conjecture of Reutenauer and the author [114], recently proved by Ribes and Zalesskii [140], and on the other hand on the solution by Ash [15, 16] of a famous open problem in semigroup theory, the Rhodes "type II" conjecture. Actually, as was shown in [106], the topological and algebraic aspects are intimately related. See the survey [62] for more references and details.

Theorem 8.31 is proved in [113]. The second part of the statement was suggested by Daniel Lascar. The study of the languages of level 1 in the group hierarchy started in 1985 [83] and was completed in [63] (see also [62]). The reader is referred to the survey article [111] for a more detailed discussion. See also [101, 106, 114, 62, 108].

8.4 Subhierarchies

Several subhierarchies were considered in the literature. One of the most studied is the subhierarchy inside level 1 of a concatenation hierarchy. Recall that if $\mathcal{V}$ is a $*$-variety of languages, then BPol$\mathcal{V}$ is the level one of the concatenation hierarchy built on $\mathcal{V}$. For each alphabet A, let $\mathcal{V}_{1,n}(A^*)$ be the boolean algebra generated by languages of the form

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

where the a_i's are letters, $L_i \in \mathcal{V}(A^*)$ and $0 \leq k \leq n$. Then the sequence $\mathcal{V}_{1,n}$ is an increasing sequence of $*$-varieties whose union is BPol$\mathcal{V}$.

For the Straubing-Thérien's hierarchy, $\mathcal{V}_{1,n}(A^*)$ is the boolean algebra generated by the languages of the form $A^*a_1A^*a_2 \cdots a_kA^*$, where the a_i's are letters, and $0 \leq k \leq n$. The corresponding variety of finite monoids is denoted $\mathbf{J}_n$. In particular, $\mathcal{J}_1$ is the $*$-variety already encountered in Proposition 5.6. The hierarchy $\mathbf{J}_n$ is defined in [36]: Simon proved that $\mathbf{J}_2$ is defined by the identities $(xy)^2 = (yx)^2$ and $xyxzx = xyzx$: the variety $\mathbf{J}_3$ is defined by the identities $xzyxvxwy = xzxyxvxwy$ [24], $ywxvxyzx = ywxvxyxzx$ and $(xy)^3 = (yx)^3$ but there are no finite basis of identities for $\mathbf{J}_n$ when $n > 3$ [30].

For the group hierarchy, the variety $\mathcal{V}_{1,1}$ admits several equivalent descriptions [85], which follow in part from Theorem 7.12.

Theorem 8.35 *Let* $\mathbf{A}$ *be an alphabet and let* K *be a recognizable language of* A^**. The following conditions are equivalent:*

(1) *K is in the boolean algebra generated by the languages of the form L or LaA^*, where L is a group language and $a \in A$,*
(2) *K is in the boolean algebra generated by the languages of the form L or A^*aL, where L is a group language and $a \in A$,*
(3) *K is in the boolean algebra generated by the languages of the form L or LaL', where L and L' are group languages and $a \in A$,*
(4) *idempotents commute in the syntactic monoid of K.*

Thus the corresponding variety of finite monoids is the variety of monoids with commuting idempotents, defined by the identity $x^\omega y^\omega = y^\omega x^\omega$. It is equal to the variety $\mathbf{Inv}$ generated by finite inverse monoids. Furthermore, the following non trivial equalities hold [14, 82, 85]

$$\mathbf{Inv} = \mathbf{J}_1 * \mathbf{G} = \mathbf{J}_1 \circledM \mathbf{G}$$

More generally, one can define hierarchies indexed by trees [100]. These hierarchies were studied in connection with game theory by Blanchet-Sadri [24, 26].

8.5 Boolean-polynomial closure

Let $\mathbf{V}$ be a variety of finite monoids and let $\mathcal{V}$ be the associated $*$-variety. We have shown that the algebraic counterpart of the operation $\mathcal{V} \to \mathrm{Pol}\ \mathcal{V}$ on varieties of languages is the operation $\mathbf{V} \to [\![x^\omega y x^\omega \leq x^\omega]\!] \circledM \mathbf{V}$. Similarly, the algebraic counterpart of the operation $\mathcal{V} \to \text{Co-Pol}\ \mathcal{V}$ is the operation $\mathbf{V} \to [\![x^\omega \leq x^\omega y x^\omega]\!] \circledM \mathbf{V}$. It is tempting to guess that the algebraic counterpart of the operation $\mathcal{V} \to \mathrm{BPol}\ \mathcal{V}$ is also of the form $\mathbf{V} \to \mathbf{W} \circledM \mathbf{V}$ for some variety $\mathbf{W}$. A natural candidate for $\mathbf{W}$ is the join of the two varieties $[\![x^\omega y x^\omega \leq x^\omega]\!]$ and $[\![x^\omega \leq x^\omega y x^\omega]\!]$, which is equal to the variety $\mathbf{B}_1$ defined in section 8.2 [126, 127]. We can thus formulate our conjecture as follows:

Conjecture 8.1 Let $\mathcal{V}$ be a variety of languages and let $\mathbf{V}$ be the associated variety of finite semigroups (resp. monoids). Then the variety of finite semigroups (resp. monoids) associated with BPol $\mathcal{V}$ is $\mathbf{B}_1 \circledM \mathbf{V}$.

One inclusion in the conjecture is certainly true.

Proposition 8.36 *The variety of finite semigroups (resp. monoids) associated with BPol $\mathcal{V}$ is contained in $\mathbf{B}_1 \circledM \mathbf{V}$.*

Now, by Theorem 6.5, the identities of $\mathbf{B}_1 \circledM \mathbf{V}$ are

$$(x^\omega py^\omega qx^\omega)^\omega x^\omega py^\omega sx^\omega (x^\omega ry^\omega sx^\omega)^\omega \quad = \quad (x^\omega py^\omega qx^\omega)^\omega (x^\omega ry^\omega sx^\omega)^\omega \quad (1)$$

for all $x, y, p, q, r, s \in \widehat{A}^*$ for some finite alphabet A such that $\mathbf{V}$ satisfies $x^2 = x = y = p = q = r = s$. These identities lead to another equivalent statement for our conjecture.

Proposition 8.37 *The conjecture is true if and only if every finite semigroup (resp. monoid) satisfying the identities (1) is a quotient of an ordered semigroup (resp. ordered monoid) of the variety $[\![x^\omega yx^\omega \leq x^\omega]\!] \circledM \mathbf{V}$.*

Conjecture 8.1 was proved to be true in a few particular cases. First, if $\mathbf{V}$ is the trivial variety of monoids, then $\mathbf{B}_1 \circledM \mathbf{V} = \mathbf{J}$. In this case, the second form of the conjecture is known to be true. This is in fact a consequence of Theorem 8.6 and it was also proved directly by Straubing and Thérien [170].

Theorem 8.38 *Every $\mathcal{J}$-trivial monoid is a quotient of an ordered monoid satisfying the identity $x \leq 1$.*

Second, if $\mathbf{V}$ is the trivial variety of semigroups, then $\mathbf{B}_1 \circledM \mathbf{V} = \mathbf{B}_1$ is, by Theorem 8.19, the variety of finite semigroups associated with the languages of dot-depth 1. Therefore, the conjecture is true in this case, leading to the following corollarylary.

Corollary 8.39 *Every monoid of $\mathbf{B}_1$ is a quotient of an ordered monoid satisfying the identity $x^\omega yx^\omega \leq x^\omega$.*

Third, if $\mathbf{V} = \mathbf{G}$, the variety of monoids consisting of all finite groups, $\mathbf{B}_1 \circledM \mathbf{G} = \mathbf{J} \circledM \mathbf{G} = \mathbf{PG} = \mathbf{BG}$ is the variety associated with the level 1 of the group hierarchy. Therefore, the conjecture is also true in this case.

Corollary 8.40 *Every monoid of* $\mathbf{BG}$ *is a quotient of an ordered monoid satisfying the identity $x^\omega \leq 1$.*

The level 2 of the Straubing hierarchy corresponds to the case $\mathbf{V} = \mathbf{J}_1$. Therefore, one can formulate the following conjecture for this level:

Conjecture 8.2 A recognizable language is of level 2 in the Straubing hierarchy if and only if its syntactic monoid satisfies the identities

$$(x^\omega py^\omega qx^\omega)^\omega x^\omega py^\omega sx^\omega (x^\omega ry^\omega sx^\omega)^\omega \quad = \quad (x^\omega py^\omega qx^\omega)^\omega (x^\omega ry^\omega sx^\omega)^\omega$$

for all $x, y, p, q, r, s \in \widehat{A}^*$ for some finite alphabet A such that $c(x) = c(y) = c(p) = c(q) = c(r) = c(s)$.

If this conjecture was true, it would imply the decidability of the levels 2 of the Straubing hierarchy and of the dot-depth. It is known that Conjecture 8.2 holds true for languages recognized by an inverse monoid [172, 190, 194] and for languages on a two letter alphabet [167].

More generally, the conjecture $\mathbf{V}_{n+1} = \mathbf{B}_1 \circledM \mathbf{V}_n$ would reduce the decidability of the Straubing hierarchy to a problem on the Mal'cev products of the form $\mathbf{B}_1 \circledM \mathbf{V}$. However, except for a few exceptions (including $\mathbf{G}$, $\mathbf{J}$ and the finitely generated varieties, like the trivial variety or $\mathbf{J}_1$), it is not known whether the decidability of $\mathbf{V}$ implies that of $\mathbf{B}_1 \circledM \mathbf{V}$.

9 Codes and varieties

In this section, we will try to follow the terminology of the book of Berstel and Perrin [20], which is by far the best reference on the theory of codes. Recall that a subset P of A^+ is a *prefix code* if no element of P is a proper prefix of another element of P, that is, if $u, uv \in P$ implies $v = 1$. The next statement shows that the syntactic monoids of finite codes give, in some sense, a good approximation of any finite monoid.

Theorem 9.1 [84] *For every finite monoid M, there exists a finite prefix code P such that*

(1) *M divides $M(P^*)$*
(2) *there exists a relational morphism $\tau : M(P^*) \to M$, such that, for every idempotent e of M, $\tau^{-1}(e)$ is an aperiodic semigroup.*

If $\mathcal{V}$ is a $*$-variety of languages, denote by $\mathcal{V}'$ the least $*$-variety such that, for each alphabet A, $\mathcal{V}'(A^*)$ contains the languages of $\mathcal{V}(A^*)$ of the form P^*, where P is a finite prefix code. Similarly, if $\mathcal{V}$ be a $+$-variety of languages, let $\mathcal{V}'$ be the least $+$-variety such that, for each alphabet A, $\mathcal{V}'(A^+)$ contains the languages of $\mathcal{V}(A^+)$ of the form P^+, where P is a finite prefix code. By construction, $\mathcal{V}'$ is contained in $\mathcal{V}$, but the inclusion is proper in general. A variety of languages $\mathcal{V}$ is said to be *described by its finite prefix codes* if $\mathcal{V} = \mathcal{V}'$. The next theorem summarizes the results of [84] and [79].

Theorem 9.2

(1) *Every $*$-variety closed under product is described by its finite prefix codes.*
(2) *The $+$-variety of locally testable is described by its finite prefix codes.*
(3) *The $*$-variety associated with the variety $[\![x^\omega y^\omega = y^\omega x^\omega]\!]$ is described by its finite prefix codes.*

(4) *Let $\mathbf{H}_1, \ldots, \mathbf{H}_n$ be varieties of finite groups. Then the $*$-variety associated with the variety $\mathbf{A} * \mathbf{H}_1 * \mathbf{A} \cdots * \mathbf{H}_n * \mathbf{A}$ is described by its finite prefix codes.*

In contrast, the varieties of languages corresponding to the varieties $\mathbf{R}$, $\mathbf{J}$, $\mathbf{DA}$ or $\mathbf{B}_1$ are not described by their finite prefix codes. Schützenberger [148] conjectured that the $+$-variety of languages corresponding to $\mathbf{LG}$ is described by its finite prefix codes. This conjecture is still open.

Prefix codes were also used to impose a restriction on the star operation. Let us say that a set $\mathcal{L}$ of languages of A^* is *closed under polynomial operations* if, for $L_0, L_1, \ldots, L_k \in \mathcal{L}$ and $a_1, \ldots, a_k \in A$, $L_0 a_1 L_1 \ldots a_k L_k \in \mathcal{L}$. Then Kleene's theorem can be stated as follows: "the rational languages of A^* form the smallest set of languages of A^* closed under polynomial operations and star". Schützenberger has obtained a similar statement for the star-free languages, in which the star operation is restricted to a special class of languages. A language L of A^* is *uniformly synchronous* if there exists an integer $d \geq 0$ such that, for all $x, y \in A^*$ and for all $u, v \in L^d$,

$$xuvy \in L^* \text{ implies } xu, vy \in L^*$$

The theorem of Schützenberger can now be stated. The paradoxical aspect of this theorem is that it gives a characterization of the star-free languages that makes use of the star operation!

Theorem 9.3 [146, 147] *The class of star-free languages of A^* is the smallest class of languages of A^* closed under polynomial operations and star operation restricted to uniformly synchronous prefix codes.*

Schützenberger also characterized the languages associated with the variety of finite monoids $\overline{\mathbf{Gcom}}$ [146].

Theorem 9.4 *Let $\mathcal{V}$ be the $*$-variety of languages associated with $\overline{\mathbf{Gcom}}$. For each alphabet A, $\mathcal{V}(A^*)$ is the least boolean algebra $\mathcal{C}$ of languages of A^* closed under polynomial operations and under the star operation restricted to languages of the form P^n, where $n > 0$ and P is a uniformly synchronous prefix code of $\mathcal{C}$.*

Another series of results concern the circular codes. A prefix code P is *circular* if, for every $u, v \in A^+$, $uv, vu \in C^+$ implies $u, v \in C^+$. Circular codes are intimately related to locally testable languages [49, 50, 51]

Theorem 9.5 *A finite prefix code[11] P is circular if and only if P^+ is a locally testable language.*

[11] This is a "light" version of the theorem, which holds for a larger class of (non necessarily prefix) codes.

Theorem 9.5 leads to the following characterization of the +-variety of locally testable languages [99].

Theorem 9.6 *Let +-variety of locally testable languages is the smallest +-variety containing the finite languages and closed under the operation $P \to P^+$ where P is a finite prefix circular code.*

Similar results hold for pure codes and star-free languages. The definition of a pure language was given in section 6.1

Theorem 9.7 [133] *A finite prefix code P is pure if and only if P^* is a star-free language.*

Theorem 9.8 [99] *Let $*$-variety of star-free languages is the smallest $*$-variety containing the finite languages and closed under the operation $P \to P^*$ where P is a finite prefix pure code.*

However, Theorems 9.6 and 9.8 are less satisfactory than Theorems 9.3 and 9.4, because they lead to a description of the languages involving all the operations of a $*$-variety, including complement, left and right quotient and inverse morphisms.

10 Operators on languages and varieties

In view of Eilenberg's theorem, one may expect some relationship between the operators on languages (of combinatorial nature) and the operators on semigroups (of algebraic nature). The following table summarizes the results of this type related to the concatenation product.

Closure of $\mathcal{V}$ under the operations ...	$\mathbf{V}$
Product and union	$[\![x^\omega y x^\omega \leq x^\omega]\!] ⓜ \mathbf{V}$
Unambiguous product and union	$[\![x^\omega y x^\omega = x^\omega]\!] ⓜ \mathbf{V}$
Left deterministic product and union	$[\![x^\omega y = x^\omega]\!] ⓜ \mathbf{V}$
Right deterministic product and union	$[\![y x^\omega = x^\omega]\!] ⓜ \mathbf{V}$
Product, boolean operations	$\mathbf{A} ⓜ \mathbf{V}$
Product with p-counters, boolean operations	$\mathbf{LG}_p ⓜ \mathbf{V}$
Product with counters, boolean operations	$\mathbf{LGsol} ⓜ \mathbf{V}$
Product, product with p-counters, boolean op.	$\mathbf{L\overline{G}}_p ⓜ \mathbf{V}$
Product, product with counters, boolean op.	$\mathbf{L\overline{Gsol}} ⓜ \mathbf{V}$
BPol $\mathcal{V}$	$\mathbf{B}_1 ⓜ \mathbf{V}$ (Conjecture)

Other standard operations on languages have been studied. A $*$-variety is *closed under star* if, for every alphabet A, $L \in \mathcal{V}(A^*)$ implies $L^* \in \mathcal{V}(A^*)$.

Theorem 10.1 [94] *The only $*$-variety of languages closed under star is the variety of rational languages.*

Similarly, a $+$-variety is closed under the operation $L \to L^+$ if, for every alphabet A, $L \in \mathcal{V}(A^+)$ implies $L^+ \in \mathcal{V}(A^+)$.

Theorem 10.2 *The only $+$-varieties of languages closed under the operation $L \to L^+$ are the variety of languages associated with the trivial variety, the variety of finite semigroups and the varieties $[\![xy = x]\!]$ and $[\![yx = x]\!]$.*[12]

A $*$-variety is *closed under shuffle* if, for every alphabet A, $L_1, L_2 \in \mathcal{V}(A^*)$ implies $L_1 ш L_2 \in \mathcal{V}(A^*)$. The classification of the varieties of languages closed under shuffle was initiated in [96] and was recently completed by Esik and Simon [54]. If $\mathbf{H}$ is a variety of commutative groups, denote by $\mathbf{Com_H}$ the variety of commutative monoids whose groups belong to $\mathbf{H}$.

Theorem 10.3 *The only $*$-varieties of languages closed under shuffle are the variety of rational languages and the varieties of languages associated with the varieties of the form $\mathbf{Com_H}$, for some variety of commutative groups $\mathbf{H}$.*

The definition of the operator $\mathbf{V} \to \mathbf{PV}$ was given in section 8.1. The next theorem shows that length-preserving morphisms or inverse substitutions form the corresponding operator on languages [134, 161, 96]. Recall that a *substitution* $\sigma : A^* \to B^*$ is a monoid morphism from A^* into $\mathcal{P}(B^*)$.

Theorem 10.4 *Let $\mathbf{V}$ be a variety of monoids and let $\mathcal{V}$ and $\mathcal{W}$ be the varieties of languages corresponding respectively to $\mathbf{V}$ and $\mathbf{PV}$. Then for every alphabet A,*

(1) *$\mathcal{W}(A^*)$ is the boolean algebra generated by the languages of the form $\varphi(L)$, where $\varphi : B^* \to A^*$ is a length-preserving morphism and $L \in \mathcal{W}(B^*)$*

(2) *$\mathcal{W}(A^*)$ is the boolean algebra generated by the languages of the form $\sigma^{-1}(L)$, where $\sigma : A^* \to B^*$ is a substitution and $L \in \mathcal{W}(B^*)$.*

Theorem 10.4 motivated the systematic study of the varieties of the form $\mathbf{PV}$. At present, this classification is not yet complete, although many results are known. Almeida's book [5] gives a complete overview of this topic along with the relevant references.

[12] A language of A^+ is recognized by a semigroup of $[\![xy = x]\!]$ (resp. $[\![yx = x]\!]$) if and only if it is of the form BA^* (resp. A^*B) where $B \subseteq A$.

11 Further topics

The notion of syntactic semigroup and the correspondence between languages and semigroups has been generalized to other algebraic systems. For instance, Rhodes and Weil [139] partially extend the algebraic theory presented in this chapter to torsion semigroups and torsion languages. Reutenauer [136] introduces a notion of *syntactic algebra* for formal power series in non commutative variables and proves an analog of the variety theorem.

Büchi was the first to use finite automata to define sets of infinite words. Although this involves a non trivial generalization of the semigroup structure, a theory similar to the one for finite words can be developed. In particular, a notion of syntactic ω-semigroup can be defined and a variety theorem also holds in this case. See [13, 93, 197, 198] for more details.

There are several important topics that could not be covered in this chapter. The first one is the algebraic study of varieties. In particular, there is a lot of literature on the description of the free profinite semigroups of a given variety of finite semigroups, the computation of the identities for the join, the semidirect product or the Mal'cev product of two varieties. See the book of Almeida [5] for an overview and references. See also the survey articles [7, 62] and the thematic issue **39** of *Russian Mathematics (Izvestiya VUZ Matematika)*, (1995), devoted to pseudovarieties.

The second one is the important connections between varieties and formal logic. The main results [42, 87, 184] relate monadic second order to rational languages, first order to star-free languages and the Σ_n hierarchy of first order formulas to the concatenation hierarchies. There are also some results about the expressive power of the linear temporal logic [68, 56, 46]. See the chapter of W. Thomas in this Handbook or the survey articles [107, 109].

The third one is the connection with boolean circuits initiated by Barrington and Thérien. The recent book of Straubing [169] is an excellent introduction to this field. It contains also several results about the connections with formal logic.

The fourth one is the theory of recognizable and rational sets in arbitrary monoids. In particular, the following particular cases have been studied to some extent: product of free monoids (relations and transductions), free groups, free inverse monoids, commutative monoids, partially commutative monoids (traces). The survey paper [21] is an excellent reference.

The fifth one is the star-height problem for rational languages. We refer the reader to the survey articles [37, 110] for more details. The star-height of a rational expression counts the number of nested uses of the star operation. The *star-height of a language* is the minimum of the star-heights of the rational expressions representing the language. The *star-height problem* is to find an algorithm to compute the star-height. There is a similar problem, called the *extended star-height problem*, if extended rational expressions are allowed. Extended rational expressions allow complement in addition to the usual operations union, product and star.

It was shown by Dejean and Schützenberger [48] that there exists a language of star-height n for each $n \geq 0$. It is easy to see that the languages of star-height 0 are the finite languages, but the effective characterization of the other levels was left open for several years until Hashiguchi first settled the problem for star-height 1 [59] and a few years later for the general case [60].

The languages of extended star-height 0 are the star-free languages. Therefore, there are languages of extended star-height 1, such as $(aa)^*$ on the alphabet $\{a\}$, but, as surprising as it may seem, nobody has been able so far to prove the existence of a language of extended star-height greater than 1, although the general feeling is that such languages do exist. In the opposite direction, our knowledge of the languages proven to be of extended star-height ≤ 1 is rather poor (see [120, 141, 142] for recent advances on this topic).

Acknowledgements

I would like to thank Jorge Almeida, Mario Branco and Pascal Weil for many useful suggestions on a first version of this chapter.

References

[1] D. Albert, R. Baldinger and J. Rhodes, The identity problem for finite semigroups, *J. Symbolic Logic* **57**, (1992), 179–192.

[2] J. Almeida, Equations for pseudovarieties, *Formal properties of finite automata and applications, J.-E. Pin (ed.) Lect. Notes in Comp. Sci.* **386**, Springer Verlag, Berlin, Heidelberg, New York, (1989), 148–164.

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

[4] J. Almeida, On pseudovarieties, varieties of languages, filters of congruences, pseudoidentities and related topics, *Algebra Universalis* **27**, (1990), 333–350.

[5] J. Almeida, *Finite semigroups and universal algebra*, Series in Algebra Vol 3, World Scientific, Singapore, 1994.

[6] J. Almeida, *A syntactical proof of the locality of* **DA**, *International Journal of Algebra and Computation*, to appear.

[7] J. Almeida and P. Weil, Relatively free profinite monoids: an introduction and examples, in NATO Advanced Study Institute *Semigroups, Formal Languages and Groups, J. Fountain (ed.)*, Kluwer, (1995), 73-117.

[8] J. Almeida and P. Weil, Free profinite semigroups over semidirect products, with J. Almeida, *Izvestiya VUZ Matematika* **39** (1995) 3–31. English version, *Russian Mathem. (Iz. VUZ)* **39** (1995) 1–27.

[9] J. Almeida and P. Weil, Reduced factorization in free profinite groups and join decompositions of pseudo-varieties, *International Journal of Algebra and Computation* **4**, (1994), 375–403.

[10] J. Almeida and P. Weil, Profinite categories and semidirect products, to appear.

[11] M. Arfi, Polynomial operations and rational languages, *4th STACS, Lect. Notes in Comp. Sci.* **247**, Springer Verlag, Berlin, Heidelberg, New York, (1987), 198–206.

[12] M. Arfi, Opérations polynomiales et hiérarchies de concaténation, *Theor. Comput. Sci.* **91**, (1991), 71–84.

[13] A. Arnold, A syntactic congruence for rational ω-languages, *Theor. Comput. Sci.* **39** (1985), 333–335.

[14] C. J. Ash, Finite semigroups with commuting idempotents, in *J. Austral. Math. Soc. Ser. A* **43**, (1987), 81–90.

[15] C. J. Ash, Inevitable sequences and a proof of the type II conjecture, in *Proceedings of the Monash Conference on Semigroup Theory*, World Scientific, Singapore, (1991), 31–42.

[16] C. J. Ash, Inevitable Graphs: A proof of the type II conjecture and some related decision procedures, *Int. Jour. Alg. and Comp.* **1**, (1991), 127–146.

[17] C. J. Ash, T.E. Hall and J.E. Pin, On the varieties of languages associated to some varieties of finite monoids with commuting idempotents, *Information and Computation* **86**, (1990), 32–42.

[18] Y. Bar-Hillel and E. Shamir, Finite-State Languages, *Bull. Res. Council Israel.* **8F**, (1960), 155–166.

[19] J. Berstel, *Transductions and context-free languages*, Teubner, Stuttgart (1979).

[20] J. Berstel and D. Perrin, *Theory of codes*, Academic Press, London (1985).

[21] J. Berstel and J. Sakarovitch, Recent results in the theory of rational sets, *Proceedings of MFCS'86, Lect. Notes in Comp. Sci.* **233**, (1986), 15–28.

[22] G. Birkhoff, On the structure of abstract algebras, *Proc. Cambridge Phil. Soc.* **31**, (1935), 433–454.

[23] P. Blanchard, Morphismes et comptages sur les langages rationnels, *J. Inform. Process. Cybernet.* **23**, (1987), 3–11.

[24] F. Blanchet-Sadri, Games, equations and the dot-depth hierarchy, *Comput. Math. Appl.* **18**, (1989), 809–822.

[25] F. Blanchet-Sadri, On dot-depth two, *Informatique Théorique et Applications* **24**, (1990), 521–530.

[26] F. Blanchet-Sadri, Games, equations and dot-depth two monoids, *Discrete Appl. Math.* **39**, (1992), 99–111.

[27] F. Blanchet-Sadri, The dot-depth of a generating class of aperiodic monoids is computable, *International Journal of Foundations of Computer Science* **3**, (1992), 419–442.

[28] F. Blanchet-Sadri, Equations and dot-depth one, *Semigroup Forum* **47**, (1993), 305–317.

[29] F. Blanchet-Sadri, On a complete set of generators for dot-depth two, *Discrete Appl. Math.*, **50**, (1994), 1–25.

[30] F. Blanchet-Sadri, Equations and monoids varieties of dot-depth one and two, *Theor. Comput. Sci.* **123**, (1994), 239–258.

[31] F. Blanchet-Sadri, Equations on the semidirect product of a finite semilattice by a $\mathcal{J}$-trivial monoid of height k, *Informatique Théorique et Applications* **29**, (1995), 157–170.

[32] F. Blanchet-Sadri, Some logical characterizations of the dot-depth hierarchy and applications, *J. Comput. Syst. Sci.* **51**, (1995), 324–337.

[33] S. L. Bloom, Varieties of ordered algebras, *J. Comput. System Sci.* **13**, (1976), 200–212.

[34] M. J. J. Branco, On the Pin-Thérien expansion of idempotent monoids, *Semigroup Forum* **49**, (1994), 329–334.

[35] M. J. J. Branco, Some variants of the concatenation product and the kernel category, (1995), to appear.

[36] J. A. Brzozowski, Hierarchies of aperiodic languages, *RAIRO Inform. Théor.* **10**, (1976), 33–49.

[37] J. A. Brzozowski, Open problems about regular languages, *Formal language theory, perspectives and open problems, R.V. Book (ed.)*, Academic Press, New York, (1980), 23–47.

[38] J. A. Brzozowski, K. Čulik II and A. Gabrielian, Classification of non-counting events, *J. Comp. Syst. Sci.* **5**, (1971), 41–53.

[39] J. A. Brzozowski and F. E. Fich, Languages of $\mathcal{R}$-trivial monoids, *J. Comput. System Sci.* **20** (1980), 32–49.

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

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

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

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

[44] R. S. Cohen and J. A. Brzozowski, On star-free events, *Proc. Hawaii Int. Conf. on System Science*, Honolulu, (1968), 1–4.

[45] R. S. Cohen and J. A. Brzozowski, Dot-depth of star-free events, *J. Comput. System Sci.* **5**, (1971), 1–15.

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

[47] D. Cowan, Inverse monoids of dot-depth 2, *Int. Jour. Alg. and Comp.* **3**, (1993), 411–424.

[48] F. Dejean and M. P. Schützenberger, On a question of Eggan, *Information and Control* **9**, (1966), 23–25.

[49] A. De Luca, On some properties of the syntactic semigroup of very pure subsemigroups, *RAIRO Inf. Théor.* **14**, (1980), 39–56.

[50] A. De Luca, D. Perrin, A. Restivo and S. Termini, Synchronization and simplification, *Discrete Math.* **27**, (1979), 297–308.

[51] A. De Luca and A. Restivo, A characterization of strictly locally testable languages and its application to subsemigroups of a free semigroup, *Inf. and Control* **44**, (1980), 300–319.

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

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

[54] Z. Esik and I. Simon, Modeling literal morphisms by shuffle, *Semigroup Forum*, to appear.

[55] F. E. Fich and J. A. Brzozowski, A characterization of a dot-depth two analogue of generalized definite languages, *Proc. 6th ICALP, Lect. Notes in Comp. Sci.* **71**, Springer Verlag, Berlin, Heidelberg, New York, (1979), 230–244.

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

[57] A. Ginzburg, About some properties of definite, reverse definite and related automata, *IEEE Trans. Electron. Comput.* **15**, (1966), 806–810.

[58] M. Hall Jr., A topology for free groups and related groups, *Ann. of Maths*, **52**, (1950), 127–139.

[59] K. Hashiguchi, Regular languages of star height one, *Information and Control* **53** (1982) 199–210.

[60] K. Hashiguchi, Representation theorems on regular languages, *J. Comput. System Sci.* **27**, (1983) 101–115.

[61] K. Hashiguchi (1988), Algorithms for Determining Relative Star Height and Star Height, *Information and Computation* **78** 124–169.

[62] K. Henckell, S. W. Margolis, J.-E. Pin and J. Rhodes, Ash's Type II Theorem, Profinite Topology and Malcev Products, *Int. Jour. Alg. and Comp.* **1**, (1991), 411–436.

[63] K. Henckell and J. Rhodes, The theorem of Knast, the $PG = BG$ and Type II Conjectures, in *Monoids and Semigroups with Applications, J. Rhodes (ed.)*, Word Scientific, (1991), 453–463.

[64] P. Higgins, A prrof of Simon's Theorem on piecewise testable languages, submitted.

[65] J. Howie, *An Introduction to Semigroup Theory*, Academic Press, London, 1976.

[66] J. Howie, *Automata and Languages*, Oxford Science Publications, Clarendon Press, Oxford, 1991.

[67] P. Jones, Monoid varieties defined by $x^{n+1} = x$ are local, *Semigroup Forum* **47**, (1993), 318–326.

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

[69] M. Keenan and G. Lallement, On certain codes admitting inverse semigroups as syntactic monoids, *Semigroup Forum* **8**, (1974), 312–331.

[70] S.C. Kleene, 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, (1956), 3–42.

[71] R. Knast, A semigroup characterization of dot-depth one languages, *RAIRO Inform. Théor.* **17**, (1983), 321–330.

[72] R. Knast, Some theorems on graph congruences, *RAIRO Inform. Théor.* **17**, (1983), 331–342.

[73] G. Lallement, Regular semigroups with $\mathcal{D} = \mathcal{R}$ as syntactic monoids of finite prefix codes, *Theor. Comput. Sci.* **3** (1977), 35–49.

[74] G. Lallement, Cyclotomic polynomials and unions of groups, *Discrete Math.* **24** (1978), 19–36.

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

[76] G. Lallement and E. Milito, Recognizable languages and finite semilattices of groups, *Semigroup Forum* **11**, (1975), 181–185.

[77] E. Le Rest and M. Le Rest, Sur le calcul du monoïde syntactique d'un sous-monoïde finiment engendré, *Semigroup Forum* **21**, (1980), 173–185.

[78] M. Lothaire, *Combinatorics on Words*, Cambridge University Press, 1982.

[79] S. W. Margolis, On the syntactic transformation semigroup of the language generated by a finite biprefix code, *Theor. Comput. Sci.*, **21**, (1982), 225–230.

[80] S. W. Margolis, Kernel and expansions: an historical and technical perspective, in *Monoids and semigroups with applications, J. Rhodes (ed.)*, World Scientific, (1989), 3–30.

[81] S. W. Margolis, On maximal pseudo-varieties of finite monoids and semigroups, *Izvestiya VUZ Matematika* **39** (1995) 65–70. English version, *Russian Mathem. (Iz. VUZ)* **39** (1995) 60–64.

[82] S. W. Margolis and J.E. Pin, Languages and inverse semigroups, *11th ICALP, Lect. Notes in Comp. Sci.* **172**, Springer, Berlin, (1984), 337–346.

[83] S. W. Margolis and J.E. Pin, Product of group languages, *FCT Conference, Lect. Notes in Comp. Sci.* **199**, Springer Verlag, Berlin, Heidelberg, New York, (1985), 285–299.

[84] S. W. Margolis and J.E. Pin, On varieties of rational languages and variable-length code II, *J. Pure Appl. Algebra* **41**, (1986), 233–253.

[85] S. W. Margolis and J.E. Pin, Inverse semigroups and varieties of finite semigroups, *Journal of Algebra* **110** (1987), 306–323.

[86] R. McNaughton, Algebraic decision procedures for local testability, *Math. Syst. Theor.* **8**, (1974), 60–76.

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

[88] A. R. Meyer, A note on star-free events, *J. Assoc. Comput. Mach.* **16**, (1969), 220–225.

[89] M. Perles, M.O. Rabin and E. Shamir, The theory of definite automata, *IEEE. Trans. Electron. Comput.* **12**, (1963), 233–243.

[90] D. Perrin, Sur certains semigroupes syntactiques, *Séminaires de l'IRIA, Logiques et Automates, Paris* (1971), 169–177.

[91] D. Perrin and J.-F. Perrot, A propos des groupes dans certains monoïdes syntactiques, *Lecture Notes in Mathematics* **855**, Springer, Berlin (1980), 82–91.

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

[93] D. Perrin and J.-E. Pin, Semigroups and automata on infinite words, in NATO Advanced Study Institute *Semigroups, Formal Languages and Groups, J. Fountain (ed.)*, Kluwer academic publishers, (1995), 49–72.

[94] J. F. Perrot, Variétés de langages et opérations, *Theor. Comput. Sci.* **7**, (1978), 197–210.

[95] J.-E. Pin, Sur le monoïde de L^* lorsque L est un langage fini. *Theor. Comput. Sci.* **7** (1978), 211–215.

[96] J.-E. Pin, Variétés de langages et monoïde des parties, *Semigroup Forum* **20** (1980), 11–47.

[97] J.-E. Pin, Propriétés syntactiques du produit non ambigu. *7th ICALP, Lect. Notes in Comp. Sci.* **85**, Springer Verlag, Berlin, Heidelberg, New York, (1980), 483–499.

[98] J.-E. Pin, Variétés de langages et variétés de semigroupes, Thèse d'état, Université Paris VI, (1981)

[99] J.-E. Pin, Langages reconnaissables et codage préfixe pur, Langages reconnaissables et codage préfixe pur, in *8th ICALP, Lect. Notes in Comp. Sci.* **115**, pringer Verlag, Berlin, Heidelberg, New York, (1981), 184–192.

[100] J.-E. Pin, Hiérarchies de concaténation, *RAIRO Inform. Théor.* **18**, (1984), 23–46.

[101] J.-E. Pin, Finite group topology and p-adic topology for free monoids, in *12th ICALP, Lect. Notes in Comp. Sci.* **194**, Springer Verlag, Berlin, Heidelberg, New York, (1985), 445–455.

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

[103] J.-E. Pin, Power semigroups and related varieties of finite semigroups, *Semigroups and Their Applications, S. M. Goberstein and P. M. Higgins (ed.)*, D. Reidel (1986), 139–152.

[104] J.-E. Pin, A property of the Schützenberger product, *Semigroup Forum* **35**, (1987) 53–62.

[105] J.-E. Pin, Relational morphisms, transductions and operations on languages, in *Formal properties of finite automata and applications, J.-E. Pin (ed.), Lect. Notes in Comp. Sci.* **386**, Springer Verlag, Berlin, Heidelberg, New York, (1989), 120–137.

[106] J.-E. Pin, Topologies for the free monoid, *Journal of Algebra* **137**, (1991), 297–337.

[107] J.-E. Pin, Logic, semigroups and automata on words, *Theor. Comput. Sci., Mathematics and Artificial Intelligence*, to appear.

[108] J.-E. Pin, Polynomial closure of group languages and open sets of the Hall topology, *ICALP 1994, Lect. Notes in Comp. Sci.* **820**, Springer Verlag, Berlin, Heidelberg, New York, (1994), 424–435.

[109] J.-E. Pin, Logic on words, *Bulletin of the European Association of Theoretical Computer Science* **54**, (1994), 145–165.

[110] J.-E. Pin, Finite semigroups and recognizable languages: an introduction, in NATO Advanced Study Institute *Semigroups, Formal Languages and Groups, J. Fountain (ed.)*, Kluwer academic publishers, (1995), 1–32.

[111] J.-E. Pin, **BG** = **PG**, a success story, in NATO Advanced Study Institute *Semigroups, Formal Languages and Groups, J. Fountain (ed.)*, Kluwer academic publishers, (1995), 33–47.

[112] J.-E. Pin, A variety theorem without complementation, *Izvestiya VUZ Matematika* **39** (1995) 80–90. English version, *Russian Mathem. (Iz. VUZ)* **39** (1995) 74–83.

[113] J.-E. Pin, Polynomial closure of group languages and open sets of the Hall topology, extended version of [108], to appear in Theor. Comput. Sci.

[114] J.-E. Pin and C. Reutenauer, A conjecture on the Hall topology for the free group, *Notices of the London Math. Society* **23**, (1991), 356–362.

[115] J.-E. Pin and J. Sakarovitch, Operations and transductions that preserve rationality, *6th GI Conf., Lect. Notes in Comp. Sci.* **145**, Springer, Berlin, (1983), 277–288.

[116] J.-E. Pin and J. Sakarovitch, Une application de la représentation matricielle des transductions, *Theor. Comput. Sci.* **35**, (1985), 271–293.

[117] J.-E. Pin and H. Straubing, Remarques sur le dénombrement des variétés de monoïdes finis, *C. R. Acad. Sci. Paris, Sér. A* **292**, (1981), 111–113.

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

[119] J.-E. Pin, H. Straubing and D. Thérien, Small varieties of finite semigroups and extensions, *J. Austral. Math. Soc.* **37**, (1984), 269–281.

[120] J.-E. Pin, H. Straubing and D. Thérien, Locally trivial categories and unambiguous concatenation, *J. Pure Appl. Algebra* **52**, (1988), 297–311.

[121] J.-E. Pin, H. Straubing and D. Thérien, New results on the generalized star height problem, *STACS 89, Lect. Notes in Comp. Sci.* **349**, (1989), 458–467.

[122] J.-E. Pin, H. Straubing and D. Thérien, Some results on the generalized star height problem, *Information and Computation*, **101**, (1992), 219–250.

[123] J.-E. Pin and D. Thérien, The bideterministic concatenation product, *Int. Jour. Alg. and Comp.* **3**, (1993), 535–555.

[124] J.-E. Pin and P. Weil, Profinite semigroups, Malcev products and identities, *J. Algebra*, to appear.

[125] J.-E. Pin and P. Weil, A Reiterman theorem for pseudovarieties of finite first-order structures, *Algebra Universalis*, to appear.

[126] J.-E. Pin and P. Weil, Polynomial closure and unambiguous product, *ICALP 1995, Lect. Notes in Comp. Sci.* **944**, Springer Verlag, Berlin, Heidelberg, New York, (1995), 348–359.

[127] J.-E. Pin and P. Weil, Polynomial closure and unambiguous product (full version), submitted.

[128] M. O. Rabin and D. Scott, Finite automata and their decision problems, *IBM J. Res. and Develop.* **3**, (1959), 114–125. Reprinted in *Sequential Machines, E. F. Moore (ed.)*, Addison-Wesley, Reading, Massachussetts, (1964), 63–91.

[129] J. Reiterman, The Birkhoff theorem for finite algebras, *Algebra Universalis* **14**, (1982), 1–10.

[130] A. Restivo, Codes and aperiodic languages, *Lect. Notes in Comp. Sci.* **2**, Springer Verlag, Berlin, Heidelberg, New York, (1973), 175–181.

[131] A. Restivo, On a question of McNaughton and Papert, *Inf. Control* **25**, (1974), 93–101.

[132] A. Restivo, A combinatorial property of codes having finite synchronization delay, *Theor. Comput. Sci.* **1**, (1975), 95–101.

[133] A. Restivo, Codes and automata, in *Formal properties of finite automata and applications, J.-E. Pin (ed.), Lect. Notes in Comp. Sci.* **386**, Springer Verlag, Berlin, Heidelberg, New York, (1989), 186–198.

[134] C. Reutenauer, Sur les variétés de langages et de monoïdes, *Lect. Notes in Comp. Sci.* **67**, Springer Verlag, Berlin, Heidelberg, New York, (1979), 260–265.

[135] C. Reutenauer, Une topologie du monoïde libre, *Semigroup Forum* **18**, (1979), 33–49.

[136] C. Reutenauer, Séries formelles et algèbres syntactiques, *J. Algebra* **66**, (1980), 448–483.

[137] C. Reutenauer, Sur mon article “Une topologie du monoïde libre”, *Semigroup Forum* **22**, (1981), 93–95.

[138] J. Rhodes and B. Tilson, The kernel of monoid morphisms, *J. Pure Appl. Algebra* **62**, (1989), 227–268.

[139] J. Rhodes and P. Weil, Algebraic and topological theory of languages, *Informatique Théorique et Applications* **29**, (1995), 1–44.

[140] L. Ribes and P.A. Zalesskii, On the profinite topology on a free group, *Bull. London Math. Soc.* **25**, (1993), 37–43.

[141] M. Robson, Some Languages of Generalised Star Height One, *LITP Technical Report* **89-62**, 1989, 9 pages.

[142] More Languages of Generalised Star Height 1, *Theor. Comput. Sci.* **106**, (1992) 327–335.

[143] J. J. Rotman, *The theory of groups, An introduction*, Allyn and Bacon, Boston, 1973.

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

[145] M.P. Schützenberger, Sur certaines variétés de monoïdes finis, in *Automata Theory, E.R. Caianiello (ed.)*, Academic Press, New-York, (1966), 314–319.

[146] M. P. Schützenberger, Sur les monoïdes finis dont les groupes sont commutatifs, *RAIRO Inf. Theor.* **1**, (1974), 55–61.

[147] M. P. Schützenberger, Sur certaines opérations de fermeture dans les langages rationnels, *Ist. Naz. Alta Math. Symp. Math.* **15** (1975), 245–253.

[148] M. P. Schützenberger, Sur le produit de concaténation non ambigu, *Semigroup Forum* **13** (1976), 47–75.

[149] C. Selmi, Over testable languages, *Theor. Comput. Sci.*, to appear.

[150] C. Selmi, Strongly locally testable semigroups with commuting idempotents and associated languages, *Informatique Théorique et Applications*, to appear.

[151] H. Sezinando, The varieties of languages corresponding to the varieties of finite band monoids, *Semigroup Forum* **44**, (1992), 283–305.

[152] I. Simon, Piecewise testable events, *Proc. 2nd GI Conf., Lect. Notes in Comp. Sci.* **33**, Springer Verlag, Berlin, Heidelberg, New York, (1975), 214–222.

[153] I. Simon, Properties of factorization forests, in *Formal properties of finite automata and applications, J.-E. Pin (ed.), Lect. Notes in Comp. Sci.* **386**, Springer Verlag, Berlin, Heidelberg, New York, (1989), 34–55.

[154] I. Simon, Factorization forests of finite height, *Theor. Comput. Sci.* **72**, (1990), 65–94.

[155] I. Simon, A short proof of the factorization forest theorem, in *Tree Automata and Languages, M. Nivat and A. Podelski eds.*, Elsevier Science Publ., Amsterdam, (1992) 433–438.

[156] I. Simon, The product of rational languages, *Proceedings of ICALP 1993, Lect. Notes in Comp. Sci.* **700**, Springer Verlag, Berlin, Heidelberg, New York, (1993), 430–444.

[157] J. Stern, Characterization of some classes of regular events, *Theor. Comput. Sci.* **35**, (1985), 17–42.

[158] J. Stern, Complexity of some problems from the theory of automata, *Inform. and Control* **66**, (1985), 63–176.

[159] H. Straubing, Families of recognizable sets corresponding to certain varieties of finite monoids, *J. Pure Appl. Algebra*, **15** (1979) 305–318.

[160] H. Straubing, Aperiodic homomorphisms and the concatenation product of recognizable sets, *J. Pure Appl. Algebra* **15**, (1979), 319–327.

[161] H. Straubing, Recognizable sets and power sets of finite semigroups, *Semigroup Forum*, **18**, (1979), 331–340.

[162] H. Straubing, On finite $\mathcal{J}$-trivial monoids, *Semigroup Forum* **19**, (1980), 107–110.

[163] H. Straubing, A generalization of the Schützenberger product of finite monoids, *Theor. Comput. Sci.* **13**, (1981), 137–150.

[164] H. Straubing, Relational morphisms and operations on recognizable sets, *RAIRO Inf. Theor.* **15**, (1981), 149–159.

[165] H. Straubing, The variety generated by finite nilpotent monoids, *Semigroup Forum* **24**, (1982), 25–38.

[166] H. Straubing, Finite semigroups varieties of the form $\mathbf{V} * \mathbf{D}$, *J. Pure Appl. Algebra* **36**, (1985), 53–94.

[167] H. Straubing, Semigroups and languages of dot-depth two, *Theor. Comput. Sci.* **58**, (1988), 361–378.

[168] H. Straubing, The wreath product and its application, in *Formal properties of finite automata and applications, J.-E. Pin (ed.), Lect. Notes in Comp. Sci.* **386**, Springer Verlag, Berlin, Heidelberg, New York, (1989), 15–24.

[169] H. Straubing, *Finite automata, formal logic and circuit complexity*, Birkhäuser, Boston, (1994).

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

[171] H. Straubing, D. Thérien and W. Thomas, Regular languages defined with generalized quantifiers, in *Proc. 15th ICALP*, Springer Lecture Notes in Computer Science **317**, (1988), 561–575.

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

[173] D. Thérien, Languages of nilpotent and solvable groups, *Proceedings of ICALP 1979, Lect. Notes in Comp. Sci.* **71**, Springer Verlag, Berlin, Heidelberg, New York, (1979), 616–632.

[174] D. Thérien, Classification of finite monoids: the language approach, *Theoret. Comp. Sci.* **14**, (1981), 195–208.

[175] D. Thérien, Recognizable languages and congruences, *Semigroup Forum* **23**, (1981), 371–373.

[176] D. Thérien, Languages recognized by unions of groups, Technical Report SOCS-82.13, McGill University, (1982).

[177] D. Thérien, Subword counting and nilpotent groups, in *Combinatorics on Words: Progress and Perspectives, L.J. Cummings (ed.)*, Academic Press, New York, (1983), 297–306.

[178] D. Thérien, Sur les monoïdes dont les groupes sont résolubles, *Semigroup Forum* **26**, (1983), 89–101.

[179] D. Thérien, A language theoretic interpretation of Schützenberger representations, *Semigroup Forum* **28**, (1984), 235–248.

[180] D. Thérien, Catégories et langages de dot-depth un, *RAIRO Inf. Théor. et Applications* **22**, (1988), 437–445.

[181] D. Thérien, On the equation $x^t = x^{t+r}$ in categories, *Semigroup Forum* **37**, (1988), 265–271.

[182] D. Thérien, Two-sided wreath products of categories, *J. Pure Appl. Algebra* **74**, (1991), 307–315.

[183] D. Thérien and A. Weiss, Graph congruences and wreath products, *J. Pure Appl. Algebra* **36**, (1985), 205–215.

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

[185] W. Thomas, An application of the Ehrenfeucht-Fraïssé game in formal language theory, *Bull. Soc. Math. de France* **16**, (1984), 11–21.

[186] W. Thomas, A concatenation game and the dot-depth hierarchy, in *Computation Theory and Logic, E. Böger (ed.), Lect. Notes in Comp. Sci.* **270**, Springer, New-York, (1987), 415–426.

[187] B. Tilson, Chapters 11 and 12 of [53].

[188] B. Tilson, Categories as algebras, *J. Pure Appl. Algebra* **48**, (1987), 83–198.

[189] P. Weil, Groups in the syntactic monoid of a composed code, *J. Pure Appl. Algebra* **42**, (1986), 297–319.

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

[191] P. Weil, Products of languages with counter, *Theoret. Comp. Sci.* **76**, (1990), 251–260.

[192] P. Weil, Implicit operations on pseudovarieties: an introduction, in *Monoids and semigroups with applications, J. Rhodes (ed.)*, World Scientific, Singapore, (1991), 89-104.

[193] P. Weil, Closure of varieties of languages under products with counter, *J. Comput. System Sci.* **45**, (1992), 316–339.

[194] P. Weil, Some results on the dot-depth hierarchy, *Semigroup Forum* **46**, (1993), 352–370.

[195] A. Weiss, The local and global varieties induced by nilpotent monoids, *Informatique Théorique et Applications* **20**, (1986), 339–355.

[196] A. Weiss and D. Thérien, Varieties of finite categories, *Informatique Théorique et Applications* **20**, (1986), 357–366.

[197] T. Wilke, An Eilenberg theorem for ∞-languages, in *Automata, Languages and Programming, Lect. Notes in Comp. Sci.* **510**, Springer Verlag, Berlin, Heidelbe rg, New York, (1991), 588–599.

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

Index