Theme and variations
on the concatenation product

Jean-Éric Pin[1][*]

LIAFA, University Paris-Diderot and CNRS, France.

Abstract. The concatenation product is one of the most important operations on regular languages. Its study requires sophisticated tools from algebra, finite model theory and profinite topology. This paper surveys research advances on this topic over the last fifty years.

The concatenation product plays a key role in two of the most important results of automata theory: Kleene's theorem on regular languages [23] and Schützenberger's theorem on star-free languages [60].

This survey article surveys the most important results and tools related to the concatenation product, including connections with algebra, profinite topology and finite model theory. The paper is organised as follows: Section 1 presents some useful algebraic tools for the study of the concatenation product. Section 2 introduces the main definitions on the product and its variants. The classical results are summarized in Section 3. Sections 4 and 5 are devoted to the study of two algebraic tools: Schützenberger products and relational morphisms. Closure properties form the topic of Section 6. Hierarchies and their connection with finite model theory are presented in Sections 7 and 8. Finally, new directions are suggested in Section 9.

1 The instruments

This section is a brief reminder on the algebraic notions needed to study the concatenation product: semigroups and semirings, syntactic ordered monoids, free profinite monoids, equations and identities, varieties and relational morphisms. More information can be found in [1–3, 18, 35, 42, 45].

1.1 Semigroups and semirings

If S is a semigroup, the set $\mathcal{P}(S)$ of subsets of S is also a semiring, with union as addition and multiplication defined, for every $X, Y \in \mathcal{P}(S)$, by

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

In this semiring, the empty set is the zero and for this reason, is denoted by 0. It is also convenient to denote simply by x a singleton $\{x\}$.

If k is a semiring, we denote by $M_n(k)$ be the semiring of square matrices of size n with entries in k.

1.2 Syntactic ordered monoid

Let L be a language of A^*. The *syntactic preorder* of L is the relation $\leqslant_L$ defined on A^* by $u \leqslant_L v$ if and only if, for every $x, y \in A^*$,

$$xvy \in L \Rightarrow xuy \in L$$

The *syntactic congruence* of L is the relation $\sim_L$ defined by $u \sim_L v$ if and only if $u \leqslant_L v$ and $v \leqslant_L u$.
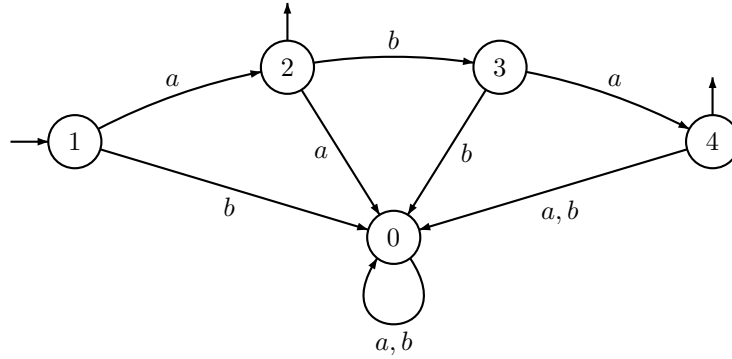
The *syntactic monoid* of L is the quotient $M(L)$ of A^* by $\sim_L$ and the natural morphism $\eta : A^* \rightarrow A^*/\sim_L$ is called the *syntactic morphism* of L. The syntactic preorder $\leqslant_L$ induces an order on the quotient monoid $M(L)$. The resulting ordered monoid is called the *syntactic ordered monoid* of L.

The syntactic ordered monoid can be computed from the minimal automaton as follows. First observe that if $\mathcal{A} = (Q, A, \cdot, q_-, F)$ is a minimal deterministic automaton, the relation $\leqslant$ defined on Q by $p \leqslant q$ if for all $u \in A^*$,

$$q \cdot u \in F \Rightarrow p \cdot u \in F$$

is an order relation, called the *syntactic order* of the automaton. Then the syntactic ordered monoid of a language is the transition monoid of its ordered minimal automaton. The order is defined by $u \leqslant v$ if and only if, for all $q \in Q$, $q \cdot u \leqslant q \cdot v$.

For instance, let L be the language $\{a, aba\}$. Its minimal deterministic automaton is represented below:

The order on the set of states is $2 \leqslant 4$, $1 \leqslant 3$ and $1, 2, 3, 4 \leqslant 0$. Indeed, one has $0 \cdot u = 0$ for all $u \in A^*$ and thus, the formal implication

$$0 \cdot u \in F \Rightarrow q \cdot u \in F$$

holds for any state q. Similarly, $1 \leqslant 3$ since a is the only word such that $3 \cdot a \in F$ and one also has $1 \cdot a \in F$.

The syntactic monoid of L is the monoid $M = \{1, a, b, ab, ba, aba, 0\}$ presented by the relations $a^2 = b^2 = bab = 0$. Its syntactic order is $1 < ab < 0$, $1 < ba < 0$, $a < aba < 0$, $b < 0$.

1.3 Free profinite monoids

We briefly recall the definition of a free profinite monoid. More details can be found in [1, 45]. A finite monoid M *separates* two words u and v of A^* if there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$. We set

$$r(u, v) = \min\{|M| \mid M \text{ is a finite monoid that separates } u \text{ and } v \}$$

and $d(u, v) = 2^{-r(u,v)}$, with the usual conventions $\min \emptyset = +\infty$ and $2^{-\infty} = 0$. Then d is a *metric* on A^* and the completion of A^* for this metric is denoted by $\widehat{A^*}$. The product on A^* can be extended by continuity to $\widehat{A^*}$. This extended product makes $\widehat{A^*}$ a compact topological monoid, called the *free profinite monoid*. Its elements are called *profinite words*.

In a compact monoid, the smallest closed subsemigroup containing a given element s has a unique idempotent, denoted s^ω. This is true in particular in a finite monoid and in the free profinite monoid.

One can show that every morphism φ from A^* into a (discrete) finite monoid M extends uniquely to a a uniformly continuous morphism $\widehat{\varphi}$ from $\widehat{A^*}$ to M. It follows that if x is a profinite word, then $\widehat{\varphi}(x^\omega) = \widehat{\varphi}(x)^\omega$.

1.4 Equations and identities

Let φ be a morphism from A^* into a finite [ordered] monoid M and let x, y be two profinite words of $\widehat{A^*}$. We say that φ *satisfies the profinite equation* $x = y$ $[x \leqslant y]$ if $\widehat{\varphi}(x) = \widehat{\varphi}(y)$ $[\widehat{\varphi}(x) \leqslant \widehat{\varphi}(y)]$.

A regular language of A^* *satisfies a profinite equation* if its syntactic morphism satisfies this equation. More generally, we say that a set of regular languages $\mathcal{L}$ *is defined a set of profinite equations* E if $\mathcal{L}$ is the set of all regular languages satisfying every equation of E.

A *lattice of languages* is a set $\mathcal{L}$ of languages of A^* containing $\emptyset$ and A^* and closed under finite union and finite intersection. It is closed under quotients if, for each $L \in \mathcal{L}$ and $u \in A^*$, the languages $u^{-1}L$ and Lu^{-1} are also in $\mathcal{L}$. It is proved in [19] that a set of regular languages is a lattice [Boolean algebra] closed under quotient if and only if it can be defined by a set of profinite equations of the form $u \leqslant v$ $[u = v]$.

A finite [ordered] monoid M *satisfies the identity* $x = y$ $[x \leqslant y]$ if every morphism from A^* into M satisfies this equation. These notions can be extended to semigroups by considering morphisms from the free semigroup A^+ to a finite semigroup.

1.5 Varieties of monoids

In this paper, we will only consider varieties in Eilenberg's sense. Thus, for us, a *variety of semigroups* is a class of finite semigroups closed under taking subsemigroups, quotients and finite direct products [18]. Varieties of *ordered semigroups*, *monoids* and *ordered monoids* are defined analogously [39].

Given a set E of identities, we denote by $[\![E]\!]$ the class of all finite [ordered] monoids which satisfy all the identities of E. Reiterman's theorem [57] and its extension to ordered structures [53] states that every variety of [ordered] monoids [semigroups] can be defined by a set of identities. For instance, the variety of ordered semigroups $[\![x^\omega y x^\omega \leqslant x^\omega]\!]$ is the variety of ordered semigroups S such that, for each idempotent $e \in S$ and for each $s \in S$, $ese \leqslant e$.

The following varieties will be used in this paper: the variety $\mathbf{A}$ of *aperiodic* monoids, defined by the identity $x^{\omega+1} = x^\omega$, the variety $\mathbf{R}$ [$\mathbf{L}$] of $\mathcal{R}$-*trivial* [$\mathcal{L}$-*trivial*] monoids, defined by the identity $(xy)^\omega x = x^\omega$ [$y(xy)^\omega = x^\omega$] and the variety $\mathbf{DA}$, which consists of the aperiodic monoids whose regular $\mathcal{J}$-classes are idempotent semigroups. This variety is defined by the identities $x^\omega = x^{\omega+1}$ and $(xy)^\omega (yx)^\omega (xy)^\omega = (xy)^\omega$.

We will also consider two group varieties: the variety $\mathbf{G}_p$ of p-groups (for a prime p) and the variety $\mathbf{Gsol}$ of soluble groups.

Finally, if $\mathbf{V}$ is a variety of monoids, the class of all semigroups S such that, for each idempotent $e \in S$, the "local" monoid eSe belongs to $\mathbf{V}$, form a variety of semigroups, denoted $\mathbf{LV}$. In particular, the variety $\mathbf{LI}$ is the variety of locally trivial semigroups, defined by the identity $x^\omega y x^\omega = x^\omega$.

1.6 Varieties of languages

A *class of languages* $\mathcal{C}$ associates with each alphabet A a set $\mathcal{C}(A^*)$ of regular languages of A^*. A *positive variety of languages* is a class of languages $\mathcal{V}$ such that, for all alphabets A and B,

  (1) $\mathcal{V}(A^*)$ is a lattice of languages closed under quotients,

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

A *variety of languages* is a positive variety $\mathcal{V}$ such that, for each alphabet A, $\mathcal{V}(A^*)$ is closed under complement. We can now state Eilenberg's variety theorem [18] and its counterpart for ordered monoids [39].

Theorem 1.1. *Let $\mathbf{V}$ be a variety of monoids. For each alphabet A, let $\mathcal{V}(A^*)$ be the set of all languages of A^* whose syntactic monoid is in $\mathbf{V}$. Then $\mathcal{V}$ is a variety of languages. Further, the correspondence $\mathbf{V} \to \mathcal{V}$ is a bijection between varieties of monoids and varieties of languages.*

Theorem 1.2. *Let $\mathbf{V}$ be a variety of ordered monoids. For each alphabet A, let $\mathcal{V}(A^*)$ be the set of all languages of A^* whose syntactic ordered monoid is in $\mathbf{V}$. Then $\mathcal{V}$ is a positive variety of languages. Further, the correspondence $\mathbf{V} \to \mathcal{V}$ is a bijection between varieties of ordered monoids and positive varieties of languages.*

A slightly more general definition was introduced by Straubing [71]. Let $\mathcal{C}$ be a class of morphisms between free monoids, closed under composition and containing all length-preserving morphisms. Examples include the classes of all *length-preserving* morphisms, of all *length-multiplying* morphisms (morphisms such that, for some integer k, the image of any letter is a word of length k),

all *non-erasing* morphisms (morphisms for which the image of each letter is a nonempty word), all *length-decreasing* morphisms (morphisms for which the image of each letter is either a letter or the empty word) and all morphisms.

A *positive $\mathcal{C}$-variety of languages* is a class $\mathcal{V}$ of recognisable languages satisfying the first condition defining a positive variety of languages and a second condition

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

A *$\mathcal{C}$-variety of languages* is a positive $\mathcal{C}$-variety of languages closed under complement. When $\mathcal{C}$ is the class of non-erasing morphisms (for which the image of a letter is a nonempty word), we use the term *ne*-variety. These *ne*-varieties are essentially the same thing as Eilenberg's +-varieties (see [49, p. 260–261] for a detailed discussion) and they correspond to varieties of semigroups.

1.7 Relational morphisms

A *relational morphism* between two monoids M and N is a function τ from M into $\mathcal{P}(N)$ such that:

(1) for all $M \in M$, $\tau(m) \neq \emptyset$,

(2) $1 \in \tau(1)$,

(3) for all $m, n \in M$, $\tau(m)\tau(n) \subseteq \tau(mn)$

Let $\mathbf{V}$ be a variety of [ordered] semigroups. A [relational] morphism $\tau : M \to N$ is said to be a [relational] $\mathbf{V}$-*morphism* if for every [ordered] semigroup R of N belonging to $\mathbf{V}$, the [ordered] semigroup $\tau^{-1}(R)$ also belongs to $\mathbf{V}$.

Let me point out an important subtlety. The definition of a [relational] $\mathbf{V}$-morphism adopted in this paper is taken from [44] and differs from the original definition given for instance in [68, 42]. The original definition only requires that, for each idempotent e, the [ordered] semigroup $\tau^{-1}(e)$ also belongs to $\mathbf{V}$. In many cases the two definitions are equivalent: for instance, when $\mathbf{V}$ is one of the varieties $\mathbf{A}$, $[\![x^\omega y x^\omega = x^\omega]\!]$, $[\![x^\omega y = x^\omega]\!]$, $[\![y x^\omega = x^\omega]\!]$ or $\mathbf{LH}$ where $\mathbf{H}$ is a variety of groups. However, the two definitions are *not equivalent* for the variety $[\![x^\omega y x^\omega \leqslant x^\omega]\!]$.

2 Theme and variations: the concatenation product

We now come to the main topic of this article. Just like a piece of classical music, the concatenation product includes theme and variations.

2.1 Main theme

The *product* (or *concatenation product*) of the languages $L_0, L_1, \ldots, L_n$ of A^* is the language

$$L_0 L_1 \cdots L_n = \{u_0 u_1 \cdots u_n \mid u_0 \in L_0, u_1 \in L_1, \cdots, u_n \in L_n\}$$

A language L of A^* is a *marked product* of the languages $L_0, L_1, \ldots, L_n$ if

$$L = L_0 a_1 L_1 \cdots a_n L_n$$

for some letters $a_1, \ldots, a_n$ of A.

2.2 Three variations

Variations include the unambiguous, deterministic, bideterministic and modular products, that are defined below.

Unambiguous product.

A marked product $L = L_0 a_1 L_1 \cdots a_n L_n$ is said to be *unambiguous* if every word of L admits a unique decomposition of the form $u = u_0 a_1 u_1 \cdots a_n u_n$ with $u_0 \in L_0$, $\ldots$, $u_n \in L_n$. For instance, the marked product $\{a, c\}^* a\{1\}b\{b, c\}^*$ is unambiguous.

Deterministic product.

A word x is a *prefix* [*suffix*] of a word u if there is a word v such that $u = xv$ [$u = vx$]. It is a *proper* prefix [suffix] if $x \neq u$. A subset C of A^+ is a *prefix* [*suffix*] *code* if if no element of C is a proper prefix [suffix] of another element of C.

A marked product $L = L_0 a_1 L_1 \cdots a_n L_n$ of n nonempty languages $L_0, L_1, \ldots, L_n$ of A^* is *left* [*right*] *deterministic* if, for $1 \leqslant i \leqslant n$, the set $L_0 a_1 L_1 \cdots L_{i-1} a_i$ [$a_i L_i \cdots a_n L_n$] is a prefix [suffix] code. This means that every word of L has a unique prefix [suffix] in $L_0 a_1 L_1 \cdots L_{i-1} a_i$ [$a_i L_i \cdots a_n L_n$]. It is observed in [9, p. 495] that the marked product $L_0 a_1 L_1 \cdots a_n L_n$ is deterministic if and only if, for $1 \leqslant i \leqslant n$, the language $L_{i-1} a_i$ is a prefix code. Since the product of two prefix codes is a prefix code, any left [right] deterministic product of left [right] deterministic products is left [right] deterministic.

A marked product is said to be *bideterministic* if it is both left and right deterministic.

Modular product of languages.

Let $L_0, \ldots, L_n$ be languages of A^*, let $a_1, \ldots, a_n$ be letters of A and let r and p be integers such that $0 \leqslant r < p$. We define the *modular product* of the languages $L_0, \ldots, L_n$ with respect to r and p, denoted $(L_0 a_1 L_1 \cdots a_n L_n)_{r,p}$, as the set of all words u in A^* such that the number of factorizations of u in the form $u = u_0 a_1 u_1 \cdots a_n u_n$, with $u_i \in L_i$ for $0 \leqslant i \leqslant n$, is congruent to r modulo p.

A language is a *p-modular product* of the languages $L_0, \ldots, L_n$ if it is of the form $(L_0 a_1 L_1 \cdots a_n L_n)_{r,p}$ for some r.

3 Classical area

The most important results on the concatenation product are due to Schützenberger. They concern the smallest Boolean algebra of languages closed under marked product or one of its variants.

Recall that the set of *star-free languages* is the smallest Boolean algebra of languages of A^* which is closed under marked product.

Theorem 3.1 (Schützenberger [60]). *A regular language is star-free if and only if its syntactic monoid is aperiodic.*

There are essentially two proofs of this result. Schützenberger's original proof [60, 35], slightly simplified in [30], works by induction on the $\mathcal{J}$-depth of the syntactic semigroup. Schützenberger's proof actually gives a stronger result since it shows that the star-free languages form the smallest Boolean algebra of languages of A^* which is closed under marked products of the form $L \to LaA^*$ and A^*aL. In other words, marked products with A^* suffice to generate all star-free languages.

The other proof [17, 28] makes use of a weak form of the Krohn-Rhodes theorem: every aperiodic semigroup divides a wreath product of copies of the monoid $U_2 = \{1, a, b\}$, given by the multiplication table $aa = a$, $ab = b$, $ba = b$ and $bb = b$.

Theorem 3.1 provides an algorithm to decide whether a given regular language is star-free. The complexity of this algorithm is analysed in [16, 65].

Let us define in the same way the set of *unambiguous [right deterministic, left deterministic] star-free languages* as the smallest Boolean algebra of languages of A^* containing the languages of the form B^*, for $B \subseteq A$, which is closed under unambiguous [left deterministic, right deterministic] marked product. The algebraic characterizations of these classes are also known.

Theorem 3.2 (Schützenberger [61]). *A regular language is unambiguous star-free if and only if its syntactic monoid belongs to* **DA***.*

One can show that the set of unambiguous star-free languages of A^* is the smallest set of languages of A^* containing the languages of the form B^*, for $B \subseteq A$, which is closed under finite disjoint union and unambiguous marked product. The languages corresponding to **DA** admit several other nice characterizations: see [72] for a survey.

Deterministic products were also studied in [61]. Alternative descriptions of these languages can be found in [18, 13].

Theorem 3.3 ([18]). *A regular language is left [right] deterministic star-free if and only if its syntactic monoid is $\mathcal{R}$-trivial [$\mathcal{L}$-trivial].*

Similar results are known for the *p*-modular product [18, 66, 73, 76, 29, 78–80].

Theorem 3.4. *Let p be a prime. A language of A^* belongs to the smallest Boolean closed under p-modular product if and only if its syntactic monoid is a p-group.*

Theorem 3.5. *A language of A^* belongs to the smallest Boolean closed under p-modular product for all prime p if and only if its syntactic monoid is a soluble group.*

Finally, one may consider the product and the p-modular products simultaneously.

Theorem 3.6. *A language of A^* belongs to the smallest Boolean closed under product and under p-modular product for all prime p if and only if all the groups in its syntactic monoid are soluble.*

See also [75] for another description of this variety of languages.

4 The ground bass: Schützenberger products

The Schützenberger product is the first algebraic tool used to study the concatenation product. It was first defined by Schützenberger [60] and later generalized by Straubing [67]. An intuitive construction, related to the linear representation of a suitable transducer, was given in [46, 47] and is briefly sketched below. More information on transducers and their linear representations can be found in Sakarovitch's book [59].

4.1 Transducers for the product

The construction given in [46, 47] relies on the following observation. Let τ and τ_a be the transductions from A^* to $A^* \times A^*$ defined by

$$\tau(u) = \{(u_1, u_2) \mid u_1 u_2 = u\}$$
$$\tau_a(u) = \{(u_1, u_2) \mid u_1 a u_2 = u\}$$

It is easy to see that the two transducers pictured below realise these transductions. In these figures, c is a generic letter and the symbol | is a separator between the input letter and the output.
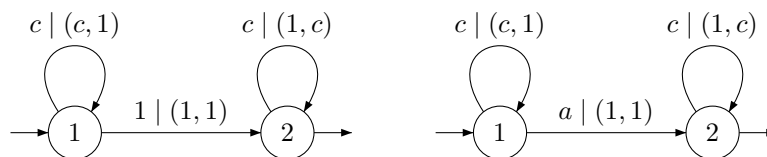
The transducer on the left [right] realizes τ [τ_a]. Now $L_0 L_1 = \tau^{-1}(L_0 \times L_1)$ and $L_0 a L_1 = \tau_a^{-1}(L_0 \times L_1)$ and this equality allows one to compute a monoid recognising $L_0 L_1$ and $L_0 a L_1$, given monoids recognising L_0 and L_1.

This construction can be readily extended to (marked) products of several languages. For instance, given $a_1, \ldots, a_n \in A$, the transduction σ defined by $\sigma(u) = \{(u_0, \cdots, u_n) \in (A^*)^{n+1} \mid u_0 a_1 u_1 \cdots a_n u_n = u\}$ is realised by the transducer

$c \mid (c,1)$ $c \mid (1,c)$ $c \mid (1,c)$ $c \mid (1,c)$

$\boxed{1} \xrightarrow{a_1 \mid (1,1)} \boxed{2}$ $\cdots$ $\boxed{n-1} \xrightarrow{a_n \mid (1,1)} \boxed{n}$

and the marked product $L_0 a_1 L_1 \cdots a_n L_n$ is equal to $\sigma^{-1}(L_0 \times L_1 \times \cdots \times L_n)$. A bit of algebra is now required to make full use of this transduction.

4.2 Linear representations

The R be the semiring $\mathcal{P}(A^* \times A^*)$. Then for each word u in A^*, $\tau_a(u) = \mu(u)_{1,2}$, where $\mu : A^* \to M_2(R)$ is defined by

$$\mu(a) = \begin{pmatrix} (c,1) & (1,1) \\ 0 & (1,c) \end{pmatrix} \quad \text{and} \quad \mu(c) = \begin{pmatrix} (c,1) & 0 \\ 0 & (1,c) \end{pmatrix} \text{ if } c \neq a$$

Indeed, for each $u \in A^*$, one gets

$$\mu(u) = \begin{pmatrix} (u,1) & \{(u_0,u_1) \mid u_0 a u_1 = u\} \\ 0 & (1,u) \end{pmatrix}$$

which gives the result. Let now $\pi_0 : A^* \to M_0$ [$\pi_1 : A^* \to M_1$] be a monoid morphism recognising the language L_0 [L_1] and let $M = M_0 \times M_1$. Let $\pi = \pi_0 \times \pi_1$. Then π is a monoid morphism from $A^* \times A^*$ into M, which can be first extended to a semiring morphism from $A^* \times A^*$ to $\mathcal{P}(M)$ and then to a semiring morphism from $M_2(A^* \times A^*)$ to $M_2(\mathcal{P}(M))$, also denoted by π. It follows that $\pi \circ \mu$ is a morphism from A^* into $M_2(\mathcal{P}(M))$ and it is not difficult to see that this morphism recognises the language $\tau_a^{-1}(L_0 \times L_1)$, that is, $L_0 a L_1$. Further, if u is a word of A^*, the matrix $\pi \circ \mu(u)$ has the form

$$\begin{pmatrix} (m_0,1) & P \\ 0 & (1,m_1) \end{pmatrix}$$

for some $m_0 \in M_0$, $m_1 \in M_1$ and $P \subseteq M_0 \times M_1$. In particular, $L_0 a L_1$ is recognised by the monoid of matrices of this form. This monoid is the *Schützenberger product* of the monoids M_0 and M_1.

A similar representation can be given for the transducer σ and this leads to the definition of the Schützenberger product of $n+1$ monoids $M_0, \dots, M_n$. In fact, one can give a slightly more general definition. Let $M = M_0 \times \cdots \times M_n$, let k be a semiring and let $k[M]$ be the monoid algebra of M over k. The *Schützenberger product over k* of the monoids $M_0, \dots, M_n$, is the submonoid of $M_{n+1}(k[M])$ made up of matrices $m = (m_{i,j})$ such that

   (1) $m_{i,j} = 0$, for $i > j$,
   (2) $m_{i,i} = (1, \dots, 1, m_i, 1, \dots, 1)$ for some $m_i \in M_i$,
   (3) $m_{i,j} \in k[1 \times \cdots \times 1 \times M_i \times \cdots \times M_j \times 1 \times \cdots \times 1]$, for $i < j$.

9

This monoid is denoted $k\Diamond(M_0, \ldots, M_n)$. The first condition means that the matrices are uppertriangular, the second one that the entry $m_{i,i}$ can be identified with an element of M_i.

When k is the Boolean semiring, then $k[M]$ is isomorphic to $\mathcal{P}(M)$ and the Schützenberger product is simply denoted $\Diamond(M_0, \ldots, M_n)$. For instance, a matrix of $\Diamond_3(M_1, M_2, M_3)$ will have the form

$$\begin{pmatrix} s_1 & P_{1,2} & P_{1,3} \\ 0 & s_2 & P_{2,3} \\ 0 & 0 & s_3 \end{pmatrix}$$

with $s_i \in M_i$, $P_{1,2} \subseteq M_1 \times M_2$, $P_{1,3} \subseteq M_1 \times M_2 \times M_3$ and $P_{2,3} \subseteq M_2 \times M_3$. The first part of the next proposition is due to Schützenberger [60] for $n = 1$ and to Straubing [67] for the general case.

Proposition 4.1. *Let $L = L_0 a_1 L_1 \cdots a_n L_n$ be a marked product and let M_i be the syntactic monoid of L_i, for $0 \leqslant i \leqslant n$. Then the Schützenberger product $\Diamond_n(M_0, \ldots, M_n)$ recognises L.*

A similar result holds for the p-modular product, for a prime p, by taking $k = \mathbb{F}_p$, the field with p elements [34, 37, 79].

Proposition 4.2. *Let $L = (L_0 a_1 L_1 \cdots a_n L_n)_{r,p}$ be a p-modular product and let M_i be the syntactic monoid of L_i, for $0 \leqslant i \leqslant n$. Then the Schützenberger product $\mathbb{F}_p \Diamond_n(M_0, \ldots, M_n)$ recognises L.*

In view of Proposition 4.1, a natural question arises: what are the languages recognised by a Schützenberger product? In the Boolean case, the answer was first given by Reutenauer [58] for $n = 2$ and by the author [33] in the general case (see also [80, 63]). The case $k = \mathbb{F}_p$ was treated by Weil [79, Theorem 2.2].

Theorem 4.3. *A language is recognised by the Schützenberger product of M_0, $\ldots$, M_n if and only if it belongs to the Boolean algebra generated by the marked products of the form $L_{i_0} a_1 L_{i_1} \cdots a_s L_{i_s}$ where $0 \leqslant i_0 < i_1 < \cdots < i_s \leqslant n$ and L_{i_j} is recognised by M_{i_j} for $0 \leqslant j \leqslant s$.*

Theorem 4.4. *A language is recognised by the monoid $\mathbb{F}_p \Diamond(M_0, \ldots, M_n)$ if and only if it belongs to the Boolean algebra generated by the p-modular products of the form $(L_{i_0} a_1 L_{i_1} \cdots a_s L_{i_s})_{r,p}$ where $0 \leqslant i_0 < i_1 < \cdots < i_s \leqslant n$ and L_{i_j} is recognised by M_{i_j} for $0 \leqslant j \leqslant s$.*

In the Boolean case, it is possible to give an ordered version of Theorem 4.3 [54, 44]. Indeed, the (Boolean) Schützenberger product can be ordered by reverse inclusion: $P \leqslant P'$ if and only if for $1 \leqslant i \leqslant j \leqslant n$, $P_{i,j} \supseteq P'_{i,j}$. The corresponding ordered monoid is denoted $\Diamond_n^+(M_0, \ldots, M_n)$ and is called the *ordered Schützenberger product* of M_1, $\ldots$, M_n.

Theorem 4.5. *A language is recognised by the ordered Schützenberger product of M_0, $\ldots$, M_n if and only if it belongs to the lattice generated by the marked products of the form $L_{i_0} a_1 L_{i_1} \cdots a_s L_{i_s}$ where $0 \leqslant i_0 < i_1 < \cdots < i_s \leqslant n$ and L_{i_j} is recognised by M_{i_j} for $0 \leqslant j \leqslant s$.*

4.3 Algebraic properties of the Schützenberger product

It follows from the definition of the Schützenberger product that the map sending a matrix to its diagonal is a morphism π from $k\Diamond(M_0, \ldots, M_n)$ to M. The properties of this morphism were first analysed by Straubing [67] and by the author [36, 54, 44] in the Boolean case and by Weil [80, Corollary 3.6] when $k = \mathbb{F}_p$. See also [4].

Proposition 4.6. *The morphism* $\pi : \Diamond(M_0, \ldots, M_n) \to M$ *is a* $[\![x^\omega y x^\omega \leqslant x^\omega]\!]$-*morphism.*

Proposition 4.7. *The morphism* $\pi : \mathbb{F}_p\Diamond(M_0, \ldots, M_n) \to M$ *is a* $\mathbf{LG}_p$-*morphism.*

5 Passacaglia: pumping properties

The second method to study the product is to use relational morphisms. This technique was initiated by Straubing [68] and later refined in [10, 8, 36, 44, 50, 54]. We first state the main result under the form of a pumping lemma before turning to a more algebraic formulation.

Let $L = L_0 a_1 L_1 \cdots a_n L_n$ be a marked product of regular languages.

Theorem 5.1. *Let u and v be words of A^* satisfying the following properties:*
  (1) $u^2 \sim_L u$ *and*
  (2) *for each $i \in \{0, \ldots, n\}$, $u^2 \sim_{L_i} u$ and $uvu \leqslant_{L_i} u$.*
Then for all $x, y \in A^*$, the condition $xuy \in L$ implies $xuvuy \in L$.

Another possible formulation of the theorem is to say that, under the assumptions (1) and (2), L is closed under the rewriting system $u \to uvu$.

We now turn to the algebraic version of this statement. For each i, let L_i be a language of A^*, let $\eta_i : A^* \to M(L_i)$ be its syntactic morphism and let

$$\eta : A^* \to M(L_0) \times M(L_1) \times \cdots \times M(L_n)$$

be the morphism defined by $\eta(u) = (\eta_0(u), \eta_1(u), \ldots, \eta_n(u))$. Finally, let $\mu : A^* \to M(L)$ be the syntactic morphism of L. Theorem 5.1 can be reformulated as a property of the relational morphism (see picture below)

$$\tau = \eta \circ \mu^{-1} : M(L) \to M(L_0) \times M(L_1) \times \cdots \times M(L_n)$$

$$
\begin{array}{ccc}
 & A^* & \\
{}^{\mu}\swarrow & & \searrow^{\eta} \\
M(L) \xrightarrow{\ \tau = \eta \circ \mu^{-1}\ } & & M(L_0) \times M(L_1) \times \cdots \times M(L_n)
\end{array}
$$

Theorem 5.2.

   (1) *The relational morphism τ is a relational $[\![x^\omega y x^\omega \leqslant x^\omega]\!]$-morphism.*

   (2) *If the product is unambiguous, it is a relational $[\![x^\omega y x^\omega = x^\omega]\!]$-morphism.*

   (3) *If the product is left deterministic, it is a relational $[\![x^\omega y = x^\omega]\!]$-morphism.*

   (4) *If the product is right deterministic, it is a relational $[\![y x^\omega = x^\omega]\!]$-morphism.*

A similar result holds for the p-modular product.

Proposition 5.3. *Let $L = (L_0 a_1 L_1 \cdots a_n L_n)_{r,p}$ be a p-modular product. The relational morphism $\tau : M(L) \rightarrow M(L_0) \times \cdots \times M(L_n)$ is a relational $\mathbf{LG}_p$-morphism.*

Theorem 5.2 is often used in the following weaker form.

Corollary 5.4. *The relational morphism $\tau : M(L) \rightarrow M(L_0) \times M(L_1) \times \cdots \times M(L_n)$ is an aperiodic relational morphism.*

6 Chaconne: Closure properties

The results of Section 3 give a description of the smallest Boolean algebra closed under marked product and its variants. The next step would be to characterize all Boolean algebras closed under marked product and its variants. A related problem is to describe the classes of regular languages closed under union and marked product.

Both problems have been solved in the case of a variety of languages, but the description of these results requires an algebraic definition. Let $\mathbf{V}$ be a variety of [ordered] monoids and let $\mathbf{W}$ be a variety of ordered semigroups. The class of all [ordered] monoids M such that there exists a $\mathbf{V}$-relational morphism from M into a monoid of $\mathbf{V}$ is a variety of [ordered] monoids, denoted $\mathbf{W}^{-1}\mathbf{V}$.

6.1 Varieties closed under product

Varieties closed under marked products were described by Straubing [66].

Theorem 6.1. *Let $\mathbf{V}$ be a variety of 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 monoids is $\mathbf{A}^{-1}\mathbf{V}$.*

This important result contains Theorem 3.1 as a particular case, when $\mathbf{V}$ is the trivial variety of monoids. Examples of varieties $\mathbf{V}$ satisfying the equality $\mathbf{A}^{-1}\mathbf{V} = \mathbf{V}$ also include the variety of monoids whose groups belong to a given variety of groups.

Theorem 6.1 has been extended to $\mathcal{C}$-varieties in [15, Theorem 4.1].

6.2 Varieties closed under modular product

Finally, let us mention the results of Weil [80]. A set of languages $\mathcal{L}$ of A^* is *closed under p-modular product* if, for any language $L_0, \ldots, L_n \in \mathcal{L}$, for any letter $a_1, \ldots, a_n \in A$ and for any integer r such that $0 \leqslant r < p$, $(L_0 a_1 L_1 \cdots a_n L_n)_{r,p} \in \mathcal{L}$. A set of languages $\mathcal{L}$ of A^* is *closed under modular product* if it is closed under p-modular product, for each prime p.

Theorem 6.2. *Let p be a prime number, let $\mathbf{V}$ be a variety of 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 p-modular product. Then $\mathcal{W}$ is a variety of languages and the associated variety of monoids is $\mathbf{LG}_p^{-1}\mathbf{V}$.*

Given a variety of groups $\mathbf{H}$, let $\overline{\mathbf{H}}$ be the variety of all monoids whose groups belong to $\mathbf{H}$.

Theorem 6.3. *Let p be a prime number, let $\mathbf{V}$ be a variety of 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 p-modular product. Then $\mathcal{W}$ is a variety of languages and the associated variety of monoids is $\mathbf{L}\overline{\mathbf{G}}_p^{-1}\mathbf{V}$.*

Theorem 6.4. *Let $\mathbf{V}$ be a variety of 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 modular product. Then $\mathcal{W}$ is a variety of languages and the associated variety of monoids is $\mathbf{L}\overline{\mathbf{Gsol}}^{-1}\mathbf{V}$.*

6.3 Polynomial closure

Let $\mathcal{L}$ be a lattice of languages. The *polynomial closure* of $\mathcal{L}$ is the set of languages that are finite unions of marked products of languages of $\mathcal{L}$. It is denoted $\mathrm{Pol}(\mathcal{L})$. Similarly, the *unambiguous polynomial closure* of $\mathcal{L}$ is the set of languages that are finite unions of unambiguous marked products of languages of $\mathcal{L}$. It is denoted $\mathrm{UPol}(\mathcal{L})$. The *left and right deterministic polynomial closure* are defined analogously, by replacing "unambiguous" by "left [right] deterministic". They are denoted $\mathrm{D}^l\mathrm{Pol}(\mathcal{V})$ $[\mathrm{D}^r\mathrm{Pol}(\mathcal{V})]$.

An algebraic characterization of the polynomial closure of a variety of languages was first given in [51, 54]. It was extended to positive varieties in [44].

Theorem 6.5. *Let $\mathbf{V}$ be a variety of [ordered] monoids and let $\mathcal{V}$ be the associated [positive] variety of languages. Then $\mathrm{Pol}(\mathcal{V})$ is a positive variety of languages and the associated variety of ordered monoids is $\llbracket x^\omega y x^\omega \leqslant x^\omega \rrbracket^{-1}\mathbf{V}$.*

Theorem 6.5 has been extended to $\mathcal{C}$-varieties in [49, Theorem 7.2]. For the unambiguous product, one has the following result [32, 50, 4].

Theorem 6.6. *Let* **V** *be a variety of monoids and let* $\mathcal{V}$ *be the associated variety of languages. Then* $UPol(\mathcal{V})$ *is a variety of languages and the associated variety of ordered monoids is* $[\![x^\omega y x^\omega = x^\omega]\!]^{-1}\mathbf{V}$.

For the left (resp. right) deterministic product, similar results hold [32, 50].

Theorem 6.7. *Let* **V** *be a variety of 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* $[\![x^\omega y = x^\omega]\!]^{-1}\mathbf{V}$ *(resp.* $[\![yx^\omega = x^\omega]\!]^{-1}\mathbf{V}$*).*

It is known that the smallest nontrivial variety of aperiodic monoids is the variety $\mathbf{J}_1 = [\![xy = yx, x = x^2]\!]$. One can show that $[\![x^\omega y = x^\omega]\!]^{-1}\mathbf{J}_1$ is equal to the variety $\mathbf{R}$ of all $\mathcal{R}$-trivial monoids, which is also defined by the identity $(xy)^\omega x = (xy)^\omega$. This leads to the following characterization [18, 13].

Corollary 6.8. *For each alphabet* A, $\mathcal{R}(A^*)$ *consists of the languages which are disjoint unions of languages of the form* $A_0^* a_1 A_1^* a_2 \cdots a_n A_n^*$, *where* $n \geqslant 0$, $a_1, \ldots a_n \in A$ *and the* A_i*'s are subsets of* A *such that* $a_i \notin A_{i-1}$, *for* $1 \leqslant i \leqslant n$.

A dual result holds for $\mathcal{L}$-trivial monoids. Finally, $[\![x^\omega y x^\omega = x^\omega]\!]^{-1}\mathbf{J}_1 = \mathbf{DA}$, which leads to the description of the languages of $\mathbf{DA}$ given hereinabove.

6.4 Back to identities

A general result of [52] permits to give identities defining the varieties of the form $\mathbf{V}^{-1}\mathbf{W}$. In particular, we get the following results.

Theorem 6.9. *Let* **V** *be a variety of monoids. Then*
  (1) $\mathbf{A}^{-1}\mathbf{V}$ *is defined by the identities of the form* $x^{\omega+1} = x^\omega$, *where* x *is a profinite word such that* **V** *satisfies the identity* $x = x^2$.
  (2) $[\![x^\omega y x^\omega = x^\omega]\!]^{-1}\mathbf{V}$ *is defined by the identities of the form* $x^\omega y x^\omega = x^\omega$, *where* x, y *are profinite words such that* **V** *satisfies the identity* $x = y = x^2$.
  (3) $[\![x^\omega y x^\omega \leqslant x^\omega]\!]^{-1}\mathbf{V}$ *is defined by the identities of the form* $x^\omega y x^\omega \leqslant x^\omega$, *where* x, y *are profinite words such that* **V** *satisfies the identity* $x = y = x^2$.

7 Hierarchies and bridges

The Boolean algebra $\mathcal{BL}$ generated by a lattice $\mathcal{L}$ is called its *Boolean closure*. In particular, $\mathcal{B}\mathrm{Pol}(\mathcal{L})$ denotes the Boolean closure of $\mathrm{Pol}(\mathcal{L})$.

Concatenation hierarchies are defined by alternating Boolean operations and polynomial operations (union and marked product). More precisely, let $\mathcal{L}$ be a set of regular languages (or more generally, a class of languages). The *concatenation hierarchy* built on $\mathcal{L}$ is the sequence $\mathcal{L}_n$ defined inductively as follows[1]: $\mathcal{L}_0 = \mathcal{L}$ and, for each $n \geqslant 0$:

---

[1] In the literature, concatenation hierarchies are usually indexed by half integers, but it seems simpler to use integers.

(1) $\mathcal{L}_{2n+1}$ is the polynomial closure of the level $2n$,

(2) $\mathcal{L}_{2n+2}$ is the Boolean closure of the level $2n + 1$.

The next results summarize the results of [5, 6, 54].

Proposition 7.1. *If $\mathcal{L}$ is a lattice of regular languages, then each even level is a lattice of regular languages and each odd level is a Boolean algebra. Further, if $\mathcal{L}$ is closed under quotients, then every level is closed under quotients.*

Since the polynomial closure of a $\mathcal{C}$-variety of languages is a positive $\mathcal{C}$-variety of languages [49, Theorem 6.2], a similar result holds for $\mathcal{C}$-varieties.

Proposition 7.2. *If $\mathcal{L}$ is a $\mathcal{C}$-variety of languages, then each even level is a positive $\mathcal{C}$-variety of languages and each odd level is a $\mathcal{C}$-variety of languages.*

For instance, the *Straubing-Thérien' hierarchy* $\mathcal{V}_n$ [74, 67, 69] is built on the trivial Boolean algebra $\mathcal{V}_0 = \{\emptyset, A^*\}$. The starting point of Brzozowski's "*dot-depth*" hierarchy $\mathcal{B}_n$ [12] was originally defined as the Boolean algebra of finite and cofinite languages but it was later suggested to start with the Boolean algebra

$$\mathcal{B}_0(A^*) = \{F A^* G \cup H \mid F,\ G,\ H \text{ are finite languages}\}$$

This suggestion was motivated by Theorem 7.4 below.

Another series of concatenation hierarchies is obtained as follows. Let $\mathbf{H}$ be a variety of groups and let $\mathcal{H}$ be the associated variety of languages. The concatenation hierarchy built on $\mathcal{H}$ is denoted by $\mathcal{H}_n$ and these hierarchies are called *group hierarchies*.

It is not immediate to see that all these hierarchies do not collapse. This was first proved by Brzozowski and Knast [14] for the dot-depth hierarchy, but the result also holds for the other hierarchies [26].

Theorem 7.3. *The Straubing-Thérien' hierarchy, the dot-depth hierarchy and the group hierarchies are infinite.*

Let $\mathbf{V}_n$ be the variety of monoids corresponding to $\mathcal{V}_n$ and let $\mathbf{B}_n$ be the variety of semigroups corresponding to $\mathcal{B}_n$. There is a nice algebraic connection between $\mathbf{V}_n$ and $\mathbf{B}_n$, discovered by Straubing [69]. Given a variety of [ordered] monoids $\mathbf{V}$ and a variety of monoids [semigroups] $\mathbf{W}$, let $\mathbf{V} * \mathbf{W}$ be the variety of [ordered] monoids generated by the semidirect products $M * N$ with $M \in \mathbf{V}$ and $N \in \mathbf{W}$.

Theorem 7.4. *The equality $\mathbf{B}_n = \mathbf{V}_n * \mathbf{LI}$ holds for each $n \geqslant 0$.*

There is a similar bridge between $\mathbf{V}_n$ and $\mathbf{H}_n$ for each variety of groups $\mathbf{H}$ [43, 44].

Theorem 7.5. *The equality $\mathbf{H}_n = \mathbf{V}_n * \mathbf{H}$ holds for each $n \geqslant 0$.*

It is still an outstanding open problem to know whether there is an algorithm to compute the concatenation level of a given regular language. Here is a brief summary of the known results. Let us start with the level 1 [26, 39, 41, 56]. Let $\mathbf{G}$ be the variety of all groups.

15

Theorem 7.6. *The following relations hold:* $\mathbf{V}_1 = [\![x \leqslant 1]\!]$, $\mathbf{B}_1 = [\![x^\omega y x^\omega \leqslant x^\omega]\!]$ *and* $\mathbf{G}_1 = [\![x^\omega \leqslant 1]\!]$. *In particular, these varieties are decidable.*

The languages of $\mathcal{G}_1$ are also known to be the open regular sets for the progroup topology [26]. Extensions of this result to the varieties $\mathbf{H}_1$ where $\mathbf{H}$ is a variety of groups is the topic of intensive research. See in particular Steinberg' article [64].

The first decidability result for the level 2 was obtained by Simon [62].

Theorem 7.7. *A language belongs to* $\mathcal{V}_2$ *if and only if its syntactic monoid is* $\mathcal{J}$-*trivial.*

The corresponding result for $\mathcal{B}_2$ is due to Knast [24, 25]

Theorem 7.8. *A language belongs to* $\mathcal{B}_2$ *if and only if its syntactic semigroup satisfies the identity*

$$(x^\omega p y^\omega q x^\omega)^\omega x^\omega p y^\omega s x^\omega (x^\omega r y^\omega s x^\omega)^\omega = (x^\omega p y^\omega q x^\omega)^\omega (x^\omega r y^\omega s x^\omega)^\omega.$$

The corresponding result for $\mathcal{G}_2$ has a long story, related in detail in [38], where several other characterizations can be found.

Theorem 7.9. *A language belongs to* $\mathcal{G}_2$ *if and only if in its syntactic monoid, the submonoid generated by the idempotents is* $\mathcal{J}$-*trivial.*

Theorem 7.9 shows that $\mathcal{G}_2$ is decidable. Again, there is a lot of ongoing work to try to extend this result to varieties of the form $\mathcal{H}_2$. See in particular [7].

Since level 3 is the polynomial closure of level 2, Theorem 6.5 can be applied. One gets in particular the following decidability result [54]. Recall that the *content* of a word is the set of letters occurring in this word.

Theorem 7.10. *A language belongs to* $\mathcal{V}_3$ *if and only if its syntactic ordered monoid satisfies the identities* $x^\omega y x^\omega \leqslant x^\omega$ *for all profinite words* x, y *with the same content.*

The corresponding problem for $\mathcal{B}_3$ is studied in [20, 22, 56]. In fact, Theorem 7.4 can be used to prove the following more general decidability result [56, 69].

Theorem 7.11. *For every integer* n, *the variety* $\mathbf{B}_n$ *is decidable if and only if* $\mathbf{V}_n$ *is decidable.*

It is still an open problem to know whether a similar reduction exists for the hierarchy $\mathbf{G}_n$.

For the level 4, several partial results are known [48, 70] and several conjectures have been formulated and then disproved [54, 64, 55]. Due to the lack of space, we will not detail these results here. Some partial results are also known for the level 5 [21].

8 Harmony with logic

One of the reasons why the decidability problem is particularly appealing is its close connection with finite model theory, first explored by Büchi in the early sixties. Büchi's logic comprises a relation symbol $<$ and, for each letter $a \in A$, a unary predicate symbol $\mathbf{a}$. The set $\mathbf{FO}[<]$ of first order formulas is built in the usual way by using these symbols, the equality symbol, first order variables, Boolean connectives and quantifiers.

A word u is represented as a structure $(\mathrm{Dom}(u), (\mathbf{a})_{a \in A}, <)$ where $\mathrm{Dom}(u) = \{1, \ldots, |u|\}$ and $\mathbf{a} = \{i \in \mathrm{Dom}(u) \mid u(i) = a\}$. The binary relation symbol $<$ is interpreted as the usual order. Thus, if $u = abbaab$, $\mathrm{Dom}(u) = \{1, \ldots, 6\}$, $\mathbf{a} = \{1, 4, 5\}$ and $\mathbf{b} = \{2, 4, 6\}$. Formulas can now be interpreted on words. For instance, the sentence

$$\varphi = \exists x \, \exists y \, \big((x < y) \wedge (\mathbf{a}x) \wedge (\mathbf{b}y)\big)$$

means "there exist two integers $x < y$ such that, in u, the letter in position x is an a and the letter in position y is a b". Therefore, the set of words satisfying φ is $A^*aA^*bA^*$. More generally, the language defined by a sentence φ is the set of words u such that φ satisfies u. The connection with star-free languages was established by McNaughton and Papert [27].

Theorem 8.1. *A language is $\mathbf{FO}[<]$-definable if and only if it is star-free.*

Thomas [77] (see also [31]) refined this result by showing that the concatenation hierarchy of star-free languages corresponds, level by level, to the Σ_n-hierarchy, defined inductively as follows:

(1) Σ_0 consists of the quantifier-free formulas.

(2) Σ_n consists of the formulas of the form $\exists^* \, \forall^* \, \exists^* \, \cdots \, \varphi$ with n alternating blocks of quantifiers and φ quantifier-free.

(3) $\mathcal{B}\Sigma_n$ denotes the class of formulas that are Boolean combinations of Σ_n-formulas.

For instance, $\exists x_1 \exists x_2 \forall x_3 \forall x_4 \forall x_5 \exists x_6 \, \varphi$, where φ is quantifier free, is in Σ_3. The next theorem is due to Thomas [77] (see also [31, 40]).

Theorem 8.2.

(1) *A language is $\Sigma_n[<]$-definable if and only if it belongs to $\mathcal{V}_{2n-1}$.*

(2) *A language is $\mathcal{B}\Sigma_n[<]$-definable if and only if it belongs to $\mathcal{V}_{2n}$.*

A slightly expanded logic is required for the dot-depth hierarchy. Let min [max] be a predicate symbol interpreted as the minimum [maximum] of the domain and let P [S] be a relation symbol interpreted as the predecessor [successor] relation. Let $\mathbf{Loc}$ be the signature $\{\mathrm{min}, \mathrm{max}, S, P\} \cup \{(\mathbf{a})_{a \in A}\}$.

Theorem 8.3.

(1) *A language is $\Sigma_n[\mathbf{Loc}]$-definable if and only if it belongs to $\mathcal{B}_{2n-1}$.*

(2) *A language is $\mathcal{B}\Sigma_n[\mathbf{Loc}]$-definable if and only if it belongs to $\mathcal{B}_{2n}$.*

Thus deciding whether a language has level n is equivalent to a very natural problem in finite model theory.

9   Other variations, recent advances

Some specialized topics require even more sophisticated algebraic tools, like the kernel category of a morphism. This is the case for instance for the bideterministic product [9–11] or for the marked product of two languages [4].

Another topic that we did not mention at all, but which is highly interesting, is the extension of these results to infinite words or even to words over ordinals or linear orders.

I would like to conclude with a recent result, which opens a new research direction. We have given in Section 6 various closure properties for varieties or even for $\mathcal{C}$-varieties. The next result of Branco and the author [8] is much more general.

Theorem 9.1. *If $\mathcal{L}$ is a lattice of languages closed under quotients, then $Pol(\mathcal{L})$ is defined by the set of equations of the form $x^{\omega} y x^{\omega} \leqslant x^{\omega}$, where x, y are profinite words such that the equations $x = x^2$ and $y \leqslant x$ are satisfied by $\mathcal{L}$.*

Work is in progress to extend the other results of Section 6 to this more general setting. The difficulty stems from the fact that definitions like $\mathbf{V}^{-1}\mathbf{W}$ are no longer available in this context and one has to work directly on profinite identities.

References

1. J. Almeida, *Finite semigroups and universal algebra*, World Scientific Publishing Co. Inc., River Edge, NJ, 1994. Translated from the 1992 Portuguese original and revised by the author.
2. J. Almeida, Finite semigroups: an introduction to a unified theory of pseudovarieties, in *Semigroups, Algorithms, Automata and Languages*, G. M. S. Gomes, J.-É. Pin and P. Silva (eds.), pp. 3–64, World Scientific, Singapore, 2002.
3. J. Almeida, Profinite semigroups and applications, in *Structural theory of automata, semigroups, and universal algebra*, pp. 1–45, *NATO Sci. Ser. II Math. Phys. Chem.* vol. 207, Lect. Notes Comp. Sci., Dordrecht, 2005.
4. J. Almeida, S. Margolis, B. Steinberg and M. Volkov, Representation theory of finite semigroups, semigroup radicals and formal language theory, *Trans. Amer. Math. Soc.* **361**,3 (2009), 1429–1461.
5. M. Arfi, Polynomial operations on rational languages, in *STACS 87 (Passau, 1987)*, pp. 198–206, *Lect. Notes Comp. Sci.* vol. 247, Lect. Notes Comp. Sci., Berlin, 1987.
6. M. Arfi, Opérations polynomiales et hiérarchies de concaténation, *Theoret. Comput. Sci.* **91**,1 (1991), 71–84.
7. K. Auinger and B. Steinberg, Varieties of finite supersolvable groups with the M. Hall property, *Math. Ann.* **335**,4 (2006), 853–877.
8. M. J. Branco and J.-E. Pin, Equations for the polynomial closure, in *ICALP 2009, Part II*, S. Albers, A. Marchetti-Spaccamela, Y. Matias, S. Nikoletseas and W. Thomas (eds.), Berlin, 2009, pp. 115–126, *Lect. Notes Comp. Sci.* vol. 5556, Springer.

9. M. J. J. BRANCO, The kernel category and variants of the concatenation product, *Internat. J. Algebra Comput.* **7**,4 (1997), 487–509.

10. M. J. J. BRANCO, Two algebraic approaches to variants of the concatenation product, *Theoret. Comput. Sci.* **369**,1-3 (2006), 406–426.

11. M. J. J. BRANCO, Deterministic concatenation product of languages recognized by finite idempotent monoids, *Semigroup Forum* **74**,3 (2007), 379–409.

12. J. A. BRZOZOWSKI, Hierarchies of aperiodic languages, *RAIRO Inform. Théor.* **10**,R–2 (1976), 33–49.

13. J. A. BRZOZOWSKI AND F. E. FICH, Languages of $\mathcal{R}$-trivial monoids, *J. Comput. System Sci.* **20**,1 (1980), 32–49.

14. J. A. BRZOZOWSKI AND R. KNAST, The dot-depth hierarchy of star-free languages is infinite, *J. Comput. System Sci.* **16**,1 (1978), 37–55.

15. L. CHAUBARD, J.-E. PIN AND H. STRAUBING, Actions, Wreath Products of $\mathcal{C}$-varieties and Concatenation Product, *Theoret. Comput. Sci.* **356** (2006), 73–89.

16. S. CHO AND D. T. HUỲNH, Finite-automaton aperiodicity is PSPACE-complete, *Theoret. Comput. Sci.* **88**,1 (1991), 99–116.

17. R. S. COHEN AND J. A. BRZOZOWSKI, On Star-Free Events, in *Proceedings of the Hawaii International Conference on System Sciences Honolulu, HI, January 29-31, 1968*, B. K. Kinariwala and F. F. Kuo (eds.), pp. 1–4, 1968.

18. S. EILENBERG, *Automata, Languages and Machines*, vol. B, Academic Press, New York, 1976.

19. M. GEHRKE, S. GRIGORIEFF AND J.-E. PIN, Duality and equational theory of regular languages, in *ICALP 2008, Part II*, L. Aceto and al. (eds.), Berlin, 2008, pp. 246–257, *Lect. Notes Comp. Sci.* vol. 5126, Springer.

20. C. GLASSER AND H. SCHMITZ, Languages of dot-depth 3/2, in *STACS 2000 (Lille)*, pp. 555–566, *Lecture Notes in Comput. Sci.* vol. 1770, Lect. Notes Comp. Sci., Berlin, 2000.

21. C. GLASSER AND H. SCHMITZ, Level 5/2 of the Straubing-Thérien hierarchy for two-letter alphabets, in *Developments in language theory (Vienna, 2001)*, pp. 251–261, *Lecture Notes in Comput. Sci.* vol. 2295, Lect. Notes Comp. Sci., Berlin, 2002.

22. C. GLASSER AND H. SCHMITZ, Languages of dot-depth 3/2, *Theory Comput. Syst.* **42**,2 (2008), 256–286.

23. S. C. KLEENE, Representation of events in nerve nets and finite automata, in *Automata studies*, pp. 3–41, Princeton University Press, Princeton, N. J., 1956. Annals of mathematics studies, no. 34.

24. R. KNAST, A semigroup characterization of dot-depth one languages, *RAIRO Inform. Théor.* **17**,4 (1983), 321–330.

25. R. KNAST, Some theorems on graph congruences, *RAIRO Inform. Théor.* **17**,4 (1983), 331–342.

26. S. MARGOLIS AND J.-E. PIN, Products of group languages, in *FCT*, Berlin, 1985, pp. 285–299, *Lect. Notes Comp. Sci.* n° 199, Springer.

27. R. MCNAUGHTON AND S. PAPERT, *Counter-free automata*, The M.I.T. Press, Cambridge, Mass.-London, 1971. With an appendix by William Henneman, M.I.T. Research Monograph, No. 65.

28. A. R. MEYER, A note on star-free events, *J. Assoc. Comput. Mach.* **16** (1969), 220–225.

29. P. PÉLADEAU, Sur le produit avec compteur modulo un nombre premier, *RAIRO Inform. Théor. Appl.* **26**,6 (1992), 553–564.

30. D. PERRIN, Finite automata, in *Handbook of theoretical computer science, Vol. B*, pp. 1–57, Elsevier, Amsterdam, 1990.

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

32. J.-E. Pin, Propriétés syntactiques du produit non ambigu, in *7th ICALP*, Berlin, 1980, pp. 483–499, *Lect. Notes Comp. Sci.* n° 85, Springer.

33. J.-E. Pin, Hiérarchies de concaténation, *RAIRO Informatique Théorique* **18** (1984), 23–46.

34. J.-E. Pin, Finite group topology and *p*-adic topology for free monoids, in *12th ICALP*, Berlin, 1985, pp. 445–455, *Lect. Notes Comp. Sci.* n° 194, Springer.

35. J.-E. Pin, *Varieties of formal languages*, North Oxford, London and Plenum, New-York, 1986. (Traduction de Variétés de langages formels).

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

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

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

39. J.-E. Pin, A variety theorem without complementation, *Russian Mathematics (Iz. VUZ)* **39** (1995), 80–90.

40. J.-E. Pin, Logic, Semigroups and Automata on Words, *Annals of Mathematics and Artificial Intelligence* **16** (1996), 343–384.

41. J.-E. Pin, Polynomial closure of group languages and open sets of the Hall topology, *Theoret. Comput. Sci.* **169** (1996), 185–200. Journal version of the article of ICALP 1994.

42. J.-E. Pin, Syntactic semigroups, in *Handbook of formal languages*, G. Rozenberg and A. Salomaa (eds.), vol. 1, ch. 10, pp. 679–746, Springer, 1997.

43. J.-E. Pin, Bridges for concatenation hierarchies, in *25th ICALP*, Berlin, 1998, pp. 431–442, *Lect. Notes Comp. Sci.* n° 1443, Springer.

44. J.-E. Pin, Algebraic tools for the concatenation product, *Theoret. Comput. Sci.* **292** (2003), 317–342.

45. J.-E. Pin, Profinite methods in automata theory, in *26th International Symposium on Theoretical Aspects of Computer Science (STACS 2009)*, S. Albers and J.-Y. Marion (eds.), pp. 31–50, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2009.

46. J.-E. Pin and J. Sakarovitch, Some operations and transductions that preserve rationality, in *6th GI Conference*, Berlin, 1983, pp. 277–288, *Lect. Notes Comp. Sci.* n° 145, Springer.

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

48. J.-E. Pin and H. Straubing, Monoids of upper triangular boolean matrices, in *Semigroups (Szeged, 1981)*, Amsterdam, 1985, pp. 259–272, *Colloq. Math. Soc. János Bolyai* vol. 39, North-Holland.

49. J.-E. Pin and H. Straubing, Some results on $\mathcal{C}$-varieties, *Theoret. Informatics Appl.* **39** (2005), 239–262.

50. J.-E. Pin, H. Straubing and D. Thérien, Locally trivial categories and unambiguous concatenation, *J. of Pure and Applied Algebra* **52** (1988), 297–311.

51. J.-E. Pin and P. Weil, Polynomial closure and unambiguous product, in *22th ICALP*, Berlin, 1995, pp. 348–359, *Lect. Notes Comp. Sci.* n° 944, Springer.

52. J.-E. Pin and P. Weil, Profinite semigroups, Mal'cev products and identities, *J. of Algebra* **182** (1996), 604–626.

53. J.-E. Pin and P. Weil, A Reiterman theorem for pseudovarieties of finite first-order structures, *Algebra Universalis* **35** (1996), 577–595.

54. J.-E. Pin and P. Weil, Polynomial closure and unambiguous product, *Theory Comput. Systems* **30** (1997), 383–422.

55. J.-E. Pin and P. Weil, A conjecture on the concatenation product, *Theoret. Informatics Appl.* **35** (2001), 597–618.

56. J.-E. Pin and P. Weil, The wreath product principle for ordered semigroups, *Communications in Algebra* **30** (2002), 5677–5713.

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

58. C. Reutenauer, Sur les variétés de langages et de monoïdes, in *Theoretical computer science (Fourth GI Conf., Aachen, 1979)*, pp. 260–265, *Lecture Notes in Comput. Sci.* vol. 67, Lect. Notes Comp. Sci., Berlin, 1979.

59. J. Sakarovitch, *Elements of automata theory*, Cambridge University Press, Cambridge, 2009. Translated from the 2003 French original by Reuben Thomas.

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

61. M.-P. Schützenberger, Sur le produit de concaténation non ambigu, *Semigroup Forum* **18** (1976), 331–340.

62. I. Simon, Piecewise testable events, in *Proc. 2nd GI Conf.*, H. Brackage (ed.), pp. 214–222, *Lecture Notes in Comp. Sci.* vol. 33, Springer Verlag, Berlin, Heidelberg, New York, 1975.

63. I. Simon, The product of rational languages, in *Automata, languages and programming (Lund, 1993)*, pp. 430–444, *Lecture Notes in Comput. Sci.* vol. 700, Lect. Notes Comp. Sci., Berlin, 1993.

64. B. Steinberg, Polynomial closure and topology, *Internat. J. Algebra Comput.* **10**,5 (2000), 603–624.

65. J. Stern, Characterizations of some classes of regular events, *Theoret. Comput. Sci.* **35**,1 (1985), 17–42.

66. H. Straubing, Families of recognizable sets corresponding to certain varieties of finite monoids, *J. Pure Appl. Algebra* **15**,3 (1979), 305–318.

67. H. Straubing, A generalization of the Schützenberger product of finite monoids, *Theoret. Comput. Sci.* **13**,2 (1981), 137–150.

68. H. Straubing, Relational morphisms and operations on recognizable sets, *RAIRO Inform. Théor.* **15**,2 (1981), 149–159.

69. H. Straubing, Finite semigroup varieties of the form $\mathbf{V} * \mathbf{D}$, *J. Pure Appl. Algebra* **36**,1 (1985), 53–94.

70. H. Straubing, Semigroups and languages of dot-depth two, *Theoret. Comput. Sci.* **58**,1-3 (1988), 361–378. Thirteenth International Colloquium on Automata, Languages and Programming (Rennes, 1986).

71. H. Straubing, On logical descriptions of regular languages, in *LATIN 2002*, Berlin, 2002, pp. 528–538, *Lect. Notes Comp. Sci.* n° 2286, Springer.

72. P. Tesson and D. Therien, Diamonds Are Forever: The Variety DA, in *Semigroups, Algorithms, Automata and Languages, Coimbra (Portugal) 2001*, pp. 475–500, World Scientific, 2002.

73. D. Thérien, Languages of Nilpotent and Solvable Groups (Extended Abstract), in *6th ICALP*, H. A. Maurer (ed.), pp. 616–632, *Lect. Notes Comp. Sci.* vol. 71, Springer, 1979.

74. D. Thérien, Classification of finite monoids: the language approach, *Theoret. Comput. Sci.* **14**,2 ang. (1981), 195–208.

75. D. Thérien, Sur les monoïdes dont tous les groupes sont résolubles, *Semigroup Forum* **26**,1-2 (1983), 89–101.

76. D. Thérien, A language theoretic interpretation of the Schützenberger representations with applications to certain varieties of languages, *Semigroup Forum* **28**,1-3 (1984), 235–248.
77. W. Thomas, Classifying regular events in symbolic logic, *J. Comput. System Sci.* **25**,3 (1982), 360–376.
78. P. Weil, An extension of the Schützenberger product, in *Lattices, semigroups, and universal algebra (Lisbon, 1988)*, pp. 315–321, Plenum, New York, 1990.
79. P. Weil, Products of languages with counter, *Theoret. Comput. Sci.* **76** (1990), 251–260.
80. P. Weil, Closure of varieties of languages under products with counter, *J. Comput. System Sci.* **45** (1992), 316–339.