Polynomial closure and unambiguous product

Jean-Eric Pin and Pascal Weil *

E-mail: `pin@litp.ibp.fr, weil@litp.ibp.fr`

1 Introduction

This paper is a contribution to the algebraic theory of recognizable languages. The main topic of this paper is the *polynomial closure*, an operation that mixes together the operations of union and concatenation. Formally, the *polynomial closure* of a class of languages $\mathcal{L}$ of A^* is the set of languages that are finite unions of marked products of the form $L_0a_1L_1 \cdots a_nL_n$, where the a_i's are letters and the L_i's are elements of $\mathcal{L}$. The *unambiguous polynomial closure* is the closure under disjoint union and unambiguous marked product. One can also define, with a slight modification (see section 4) similar operators for languages of A^+.

Our main result is an algebraic characterization of the polynomial closure of a variety of languages. There are several technical difficulties to achieve this result. First, even if $\mathcal{V}$ is a variety of languages, its polynomial closure is not, in general, a variety of languages. The solution to this problem was given in a recent paper by the first author [18]. If the definition of a variety of languages is slightly modified (instead of all boolean operations, only closure under intersection and union are required in the definition), one still has an Eilenberg type theorem. The new classes of languages are called *positive varieties*, but of course, the algebraic counterpart has to be modified too: they are the varieties of finite ordered semigroups or finite ordered monoids. It turns out that the polynomial closure of a variety of languages is always a positive variety. Now, the next question can be asked: given a variety of monoids $\mathbf{V}$ corresponding to a variety of languages $\mathcal{V}$, describe the variety of ordered monoids corresponding to the polynomial closure of $\mathcal{V}$. Our answer

*LITP/IBP, Université Paris VI et CNRS, Tour 55-65, 4 Place Jussieu, 75252 Paris Cedex 05, France

(statement (1)) fits surprisingly well with two other important results on varieties (statements (2) and (3)):

(1) The algebraic operation corresponding to the polynomial closure is the Mal'cev product $\mathbf{W} \textcircled{\scriptsize M} \mathbf{V}$, where $\mathbf{W}$ is the variety of finite ordered semigroups $(S, \leq)$ in which $ese \leq e$, for each idempotent e and each element s in S.

(2) The algebraic operation corresponding to the unambiguous polynomial closure is the Mal'cev product $\mathbf{LI} \textcircled{\scriptsize M} \mathbf{V}$, where $\mathbf{LI}$ is the variety of semigroups S in which $ese = e$, for each idempotent e and each s in S [20].

(3) The algebraic operation corresponding to the closure under boolean operations and concatenation is the Mal'cev product $\mathbf{A} \textcircled{\scriptsize M} \mathbf{V}$, where $\mathbf{A}$ is the variety of aperiodic semigroups (Straubing [30]).

The proof of our main result is non-trivial and relies on a deep theorem of Simon [27] on factorization forests. Its importance can probably be better understood on its far-reaching consequences. Due to the lack of place, we indicate some of these consequences. Others can be found in the extended version of this article. First, the polynomial closure leads to natural hierarchies among recognizable languages. Define a boolean algebra as a set of languages of A^* (resp. A^+) closed under finite union and complement. Now, start with the trivial boolean algebra of recognizable languages, and call it level 0. Thus the languages of level 0 are the empty language and A^* (resp. A^+). Then define recursively the higher levels as follows: level $n+1/2$ is the polynomial closure of level n and level $n+1$ is the boolean closure of level $n+1/2$. This defines the Straubing (resp. dot-depth) hierarchy. The main open problem is to know whether each level of these hierarchies is decidable.

Levels 0, 1/2 and 1 of the Straubing hierarchy were known to be decidable. Level 3/2 was also known to be decidable but the proof was quite involved and no practical algorithm was known. We give a simple proof of this last result and show that, given a deterministic n-state automaton $\mathcal{A}$ on the alphabet A, one can decide in time polynomial in $2^{|A|}n$ whether the language accepted by $\mathcal{A}$ is of level 3/2 in the Straubing hierarchy. Decidability of level 2 is still an open question, but we make some progress on this problem. First our main result gives a short proof of a result of Cowan [8] characterizing the languages of level 2 whose syntactic monoid is inverse. Second, we formulate a conjecture for the identities of the variety of monoids corresponding to languages of level 2. More generally, we conjecture that the variety of ordered monoids corresponding to the boolean closure of the polynomial closure of $\mathcal{V}$ is the Mal'cev product $\mathbf{B_1} \textcircled{\scriptsize M} \mathbf{V}$, where $\mathbf{B_1}$ is the variety of finite semigroups corresponding to languages of dot-depth one.

For the dot-depth hierarchy, only levels 0 and 1 were known to be decidable. We show that level 1/2 is also decidable. There is some evidence that level 3/2 is also decidable, but the proof of this result would require some auxiliary algebraic results that will be studied in a future paper.

Another important consequence of our result is the fact that a language L belongs to the unambiguous polynomial closure of a variety of languages $\mathcal{V}$ if and only if both L and its complement belong to the polynomial closure of $\mathbf{V}$. This result has an interesting consequence in logic. Indeed, Thomas [34] (see also [14, 17]) showed that Straubing's hierarchy is in one-to-one correspondence with a well known hierarchy of first order logic, the Σ_n hierarchy, obtained by counting the alternative use of existential and universal quantifiers in formulas in prenex normal form. We present analogous results for the Δ_n hierarchy of first order logic. We first show that each level of this logical hierarchy defines a variety of languages. Next we give an effective description of the first levels. For the levels 0 and 1, the corresponding variety is trivial. The variety corresponding to level 2 is the smallest variety of languages closed under non-ambiguous product, introduced by Schützenberger [25].

2 Varieties

All semigroups and monoids considered in this paper are finite or free.

2.1 Varieties of semigroups and ordered semigroups

An *ordered semigroup* $(S, \leq)$ is a semigroup S equipped with an order relation $\leq$ on S such that, for every $u, v, x \in S$, $u \leq v$ implies $ux \leq vx$ and $xu \leq xv$. An *order ideal* of $(S, \leq)$ is a subset I of S such that, if $x \leq y$ and $y \in I$, then $x \in I$. A *morphism of ordered semigroups* $\varphi : (S, \leq) \to (T, \leq)$ is a semigroup morphism from S into T such that, for every $x, y \in S$, $x \leq y$ implies $x\varphi \leq y\varphi$. A semigroup S can be considered as an ordered semigroup by taking the equality as order relation. Ordered subsemigroups, quotients and products are defined in the natural way.

A *variety of semigroups* (resp. ordered semigroups) is a class of (ordered) semigroups closed under the taking of (ordered) subsemigroups, quotients and finite products. *Varieties of (ordered) monoids* are defined in the same wa y.

2.2 Identities

Let A be a finite alphabet and let u, v be two words of A^*. A monoid M *separates* u and v if there exists a monoid morphism $\varphi : A^* \to M$ such that $u\varphi \neq v\varphi$. One defines a distance on A^* as follows: if u and v are elements of A^*, let $r(u,v) = \min\{\ |M| \ \big|\ M \text{ separates } u \text{ and } v\ \}$ and $d(u,v) = 2^{-r(u,v)}$. By convention, $\min \emptyset = \infty$ and $2^{-\infty} = 0$. Thus $r(u,v)$ measures the size of the smallest monoid which separates u and v. It is not difficult to verify that d is an ultrametric distance function. For this metric, multiplication in A^* is uniformly continuous, so that A^* is a topological monoid. The completion of the metric space (A^*, d) is a monoid, denoted $\hat{A}^*$.

If we consider each finite monoid M as being equipped with the discrete distance, every monoid morphism from A^* onto M is uniformly continuous and can be extended in a unique way into a continuous morphism from $\hat{A}^*$ onto M. Since $\hat{A}^*$ is a completion of A^*, its elements are limits of sequences of words. An important such limit is the ω-power, which traditionally designates the idempotent power of an element of a finite monoid [9, 16].

Proposition 2.1 *Let $x \in \hat{A}^*$. The sequence $(x^{n!})_{n\geq 0}$ converges in $\hat{A}^*$ to an idempotent denoted x^ω. Furthermore, if $\mu : \hat{A}^* \to M$ is a continuous morphism into a finite monoid, then $x^\omega\mu$ is the unique idempotent power of $x\mu$.*

Another useful example is the following. The set 2^A of subsets of A is a semigroup under union and the function $c : A^* \to 2^A$ defined by $c(a) = \{a\}$ is a semigroup morphism. Thus $c(u)$ is the set of letters occurring in u. Now c extends into a continuous morphism from $\hat{A}^*$ onto 2^A, also denoted c and called the *content* mapping.

Let x, y be elements of $\hat{A}^*$. A monoid (resp. ordered monoid) M *satisfies the identity* $x = y$ (resp. $x \leq y$) if, for every continuous morphism $\varphi : \hat{A}^* \to M$, $x\varphi = y\varphi$ (resp. $x\varphi \leq y\varphi$). Given a set E of identities of the form $x = y$ (resp. $x \leq y$), we denote by $[\![E]\!]$ the class of all monoids (resp. ordered monoids) which satisfy all the identities of E. The following result [22] extends two results of Reiterman [23] and Bloom [6].

Theorem 2.2 *Let E be a set of identities of the form $u = v$ (resp. $u \leq v$). Then the class $[\![E]\!]$ forms a variety of monoids (resp. ordered monoids). Conversely, for each variety of monoids (resp. ordered monoids), there exists a set E of identities such that $\mathbf{V} = [\![E]\!]$.*

A similar theory can be developed for varieties of semigroups, using a distance on the free semigroup A^+ instead of the free monoid A^*. Of particular importance for us is the variety **LI** of locally trivial semigroups, defined by the identity $[\![x^\omega y x^\omega = x^\omega]\!]$. Thus a semigroup S is locally trivial if, for all idempotent e of S and for every $s \in S$, $ese = e$. Similarly, the variety $\mathbf{A} = [\![x^\omega = x^{\omega+1}]\!]$ is the variety of aperiodic monoids [1, 16].

2.3 Relational morphisms and Mal'cev products

A *relational morphism* between semigroups S and T is a relation $\tau : S \to T$ such that:

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

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

If S and T are monoids, a third condition is required

(3) $1 \in 1\tau$

Let $\mathbf{V}$ be a variety of monoids (resp. semigroups) and let $\mathbf{W}$ be a variety of semigroups. The *Mal'cev product* $\mathbf{W} \circledM \mathbf{V}$ is the class of all monoids (resp. semigroups) M such that there exists a relational morphism $\tau : M \to V$ with $V \in \mathbf{V}$ and $e\tau^{-1} \in \mathbf{W}$ for each idempotent e of V. It is easily verified that $\mathbf{W} \circledM \mathbf{V}$ is a variety of monoids (resp. semigroups). The following theorem, obtained by the authors [21], describes a set of identities defining $\mathbf{LI} \circledM \mathbf{V}$.

Theorem 2.3 *Let $\mathbf{V}$ be a variety of monoids. Then $\mathbf{LI} \circledM \mathbf{V}$ is defined by the identities of the form $x^\omega y x^\omega = x^\omega$, where $x, y \in \hat{A}^*$ for some finite set A and $\mathbf{V}$ satisfies $x = y = x^2$.*

These results can be extended to varieties of ordered monoids as follows. Let $\mathbf{V}$ be a variety of monoids and let $\mathbf{W}$ be a variety of ordered semigroups. The *Mal'cev product* $\mathbf{W} \circledM \mathbf{V}$ is the class of all ordered monoids $(M, \leq)$ such that there exists a relational morphism $\tau : M \to V$ with $V \in \mathbf{V}$ and $e\tau^{-1} \in \mathbf{W}$ for each idempotent e of V. It is easily verified that $\mathbf{W} \circledM \mathbf{V}$ is a variety of ordered monoids. A defining set of identities for $[\![x^\omega y x^\omega \leq x^\omega]\!] \circledM \mathbf{V}$ is given in [21].

Theorem 2.4 *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 $x = y = x^2$.*

3 Recognizable languages

Recall that a *variety of languages* is a class of recognizable languages closed under finite union, finite intersection, complement, left and right quotients and inverse morphisms between free semigroups. A *positive variety of languages* is a class of recognizable languages closed under finite union, finite intersection, left and right quotients and inverse morphisms between free semigroups.

Eilenberg's variety theorem can be extended to positive varieties if one replaces varieties of semigroups by varieties of ordered semigroups. Let $(S, \leq)$ be an ordered semigroup and let η be a surjective semigroup morphism from A^+ onto S. A language L of A^+ is said to be *recognized* by η if $L = P\eta^{-1}$ for some order ideal P of S. By extension, L is said to be *recognized* by $(S, \leq)$ if there exists a surjective morphism from A^+ onto S that recognizes L. One defines a stable quasiorder $\preceq_L$ and a congruence relation $\sim_L$ on A^+ by setting $u \preceq_L v$ if and only if, for every $x, y \in A^*$, $xvy \in L$ implies $xuy \in L$ and $u \sim_L v$ if and only if $u \preceq_L v$ and $v \preceq_L u$. The congruence $\sim_L$ is called the *syntactic congruence* of L and the quasiorder $\preceq_L$ induces a stable order $\leq_L$ on $S(L) = A^+/{\sim_L}$. The ordered semigroup $(S(L), \leq_L)$ is called the *syntactic ordered semigroup* of L, the relation $\leq_L$ is called the *syntactic order* of L and the canonical morphism η_L from A^+ onto $S(L)$ is called the *syntactic morphism* of L. The syntactic ordered semigroup is the smallest ordered semigroup that recognizes L. More precisely, an ordered semigroup $(S, \leq)$ recognizes L if and only if $(S(L), \leq_L)$ is a quotient of $(S, \leq)$.

If $\mathbf{V}$ is variety of ordered semigroups and A is a finite alphabet, we denote by $A^+\mathcal{V}$ the set of recognizable languages of A^+ which are recognized by an ordered semigroup of $\mathbf{V}$. Equivalently, $A^+\mathcal{V}$ is the set of recognizable languages of A^+ whose ordered syntactic semigroup belongs to $\mathbf{V}$. It is shown in [26] that the correspondence $\mathbf{V} \to \mathcal{V}$ is a bijective correspondence between varieties of ordered semigroups and positive varieties of languages. A similar result holds if languages are considered as subsets of the free monoid A^*. Then one should consider monoids and varieties of ordered monoids instead of semigroups and varieties of semigroups.

4 Polynomial closure and unambiguous polynomial closure

There are in fact two slightly different notions of polynomial closure, one for $+$-classes and one for $*$-classes.

The *polynomial closure* of a class 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.

The *polynomial closure* of a class of languages $\mathcal{L}$ of A^* is the set of languages that are finite unions of languages of the form $L_0 a_1 L_1 \cdots a_n L_n$, where the a_i's are letters and the L_i's are elements of $\mathcal{L}$.

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, A^+Pol $\mathcal{V}$ (resp. A^*Pol $\mathcal{V}$) is the polynomial closure of $A^+\mathcal{V}$ (resp. $A^*\mathcal{V}$). We also denote by Co-Pol $\mathcal{V}$ the class of languages whose complement is in Pol $\mathcal{V}$ and by BPol $\mathcal{V}$ the boolean closure of Pol $\mathcal{V}$.

Our main result describes the counterpart, on varieties of ordered semigroups, of the operation of polynomial closure on varieties of languages.

Theorem 4.1 *Let* $\mathbf{V}$ *be a variety of semigroups (resp. monoids) and let* $\mathcal{V}$ *be the corresponding* $+$*-variety (resp* $*$*-variety). Then Pol* $\mathcal{V}$ *is a positive* $+$*-variety (resp* $*$*-variety) and the corresponding variety of semigroups (resp. monoids) is the Mal'cev product* $[\![x^\omega y x^\omega \leq x^\omega]\!] \circledM \mathbf{V}$.

In particular, for each alphabet A, A^+Pol $\mathcal{V}$ and A^+Co-Pol $\mathcal{V}$ are closed under finite union and intersection, a result due to Arfi [2, 3].

The marked product $L = u_0 L_1 u_1 \cdots L_n u_n$ (resp. $L_0 a_1 L_1 \cdots a_n L_n$) of n languages $L_1, \ldots, L_n$ of A^+ (resp. A^*) is *unambiguous* if every word u of L admits a unique factorization of the form $u_0 v_1 u_1 \cdots v_n u_n$ (resp. $u_0 a_1 u_1 \cdots a_n u_n$) with $v_1 \in L_1, \ldots, v_n \in L_n$. The *unambiguous polynomial closure* of a class of languages $\mathcal{L}$ of A^+ (resp. A^*) is the set of languages that are finite disjoint unions of (marked) unambiguous products of languages of $\mathcal{L}$.

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, A^+UPol $\mathcal{V}$ (resp. A^*UPol $\mathcal{V}$) is the unambiguous polynomial closure of $A^+\mathcal{V}$ (resp. $A^*\mathcal{V}$). The following result was established in [15, 20] as a generalization of a result of Schützenberger [25].

Theorem 4.2 *Let* $\mathbf{V}$ *be a variety of monoids (resp. semigroups) and let* $\mathcal{V}$ *be the corresponding* $*$*-variety (resp.* $+$*-variety). Then UPol* $\mathcal{V}$ *is a variety of languages, and the associated variety of monoids (resp. semigroups) is* $\mathbf{LI} \circledM \mathbf{V}$.

We give a new characterization of UPol $\mathcal{V}$.

Theorem 4.3 *Let* $\mathcal{V}$ *be a variety of languages. Then Pol* $\mathcal{V} \cap$ *Co-Pol* $\mathcal{V} =$ *UPol* $\mathcal{V}$*. In particular, Pol* $\mathcal{V} \cap$ *Co-Pol* $\mathcal{V}$ *is a variety of languages.*

5 Concatenation hierarchies

Let $\mathcal{V}$ be a variety of languages. The concatenation hierarchy of basis $\mathcal{V}$ is the sequence of varieties $\mathcal{V}_n$ and of positive varieties $\mathcal{V}_{n+1/2}$ defined 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 corresponding varieties of semigroups and ordered semigroups (resp. monoids and ordered monoids) are denoted $\mathbf{V}_n$ and $\mathbf{V}_{n+1/2}$. Theorem 4.1 gives an explicit relation between $\mathbf{V}_n$ and $\mathbf{V}_{n+1/2}$.

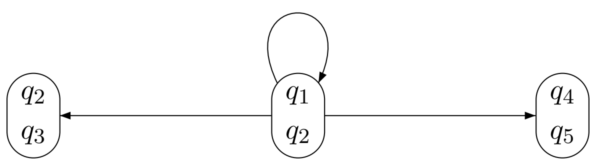
Proposition 5.1 *For every $n \geq 0$, $\mathbf{V}_{n+1/2} = [\![x^\omega y x^\omega \leq x^\omega]\!] \ⓜ \mathbf{V}_n$.*

5.1 Straubing's hierarchy

This is the hierarchy of positive $*$-varieties whose level 0 is the trivial variety. The sets of level 1/2 are the finite unions of sets of the form $A^*a_1A^*a_2 \cdots a_kA^*$, where the a_i's are letters. It is easy to see directly that level 1/2 is decidable (see Arfi [2, 3]). One can also derive this result from our syntactic characterization : a language is of level 1/2 if and only if its ordered syntactic monoid satisfies the identity $x^\omega y x^\omega \leq x^\omega$.

This leads to a polynomial algorithm to decide whether the language accepted by a complete deterministic n-state automaton is of level 1/2. This algorithm relies on the notion of *configuration*. Recall that a *subgraph* of a graph $G = (E, V)$ (where $V \subseteq E \times E$) is a graph $G' = (E', V')$ whose set of edges E' is a subset of E. A graph $G' = (E', V')$ is a *quotient* of G if there exists a map π from G onto G' such that $E' = E\pi$. Finally, a *configuration* of G is a quotient of a subgraph of G.

Theorem 5.2 *Let $\mathcal{A} = (Q, A, E, i, F)$ be a complete deterministic automaton recognizing a language L and let G be the reflexive and transitive closure of the graph of $\mathcal{A} \times \mathcal{A}$. Then L is of level $1/2$ if for every configuration of G of the form*



where the q_i's are states of $\mathcal{A}$, the condition $q_4 \in F$ implies $q_5 \in F$. If $\mathcal{A}$ is minimal, this condition is also sufficient. Therefore, one can decide in polynomial time whether the language accepted by a deterministic n-state automaton is of level $1/2$.

The sets of level 1 are the finite boolean combinations of languages of level 1/2, which were characterized by Simon [26] : a language of A^* is of level 1 if and only if its syntactic monoid satisfies the identities $x^\omega = x^{\omega+1}$ and $(xy)^\omega = (yx)^\omega$. Simon's result yields an algorithm to decide whether a given recognizable set is of level 1. Actually, it was shown by Stern [29] that 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. Indeed, Arfi [2, 3] proved that 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 from our main result the following syntactic charact erization.

Theorem 5.3 *A language is of level* $3/2$ *if and only if its ordered syntactic monoid satisfies the identity $x^\omega y x^\omega \leq x^\omega$ for every x, y such that $c(x) = c(y)$.*

Arfi [2, 3] proved that level 3/2 is also decidable. But this result relies on a deep result of Hashiguchi, and the corresponding algorithm reduces to a finiteness problem on semigroups of matrices, for which only exponential upper bounds are known. We give below a much more reasonable algorithm. Let $\mathcal{A} = (Q, A, \cdot, i, F)$ be a complete deterministic n-state automaton. Let $\mathcal{B}$ be the automaton that computes the content of a word. Formally, $\mathcal{B} = (2^A, A, \cdot, \emptyset, 2^A)$ where the transition function is defined, for every subset B of A and every letter $a \in A$, by $B \cdot a = B \cup \{a\}$. Consider the product automaton $\mathcal{C} = \mathcal{B} \times \mathcal{A} \times \mathcal{A}$ and let G' be the reflexive and transitive closure of its transition graph.

Theorem 5.4 *Let $\mathcal{A} = (Q, A, E, i, F)$ be a complete automaton recognizing a language L. Then L is of level* $3/2$ *if, for every configuration of G' of the form*

$$\begin{tikzcd} \begin{pmatrix} B \\ q_2 \\ q_3 \end{pmatrix} & \begin{pmatrix} \emptyset \\ q_1 \\ q_2 \end{pmatrix} \arrow[l] \arrow[r] & \begin{pmatrix} B \\ q_1 \\ q_2 \end{pmatrix} \arrow[r] & \begin{pmatrix} B' \\ q_4 \\ q_5 \end{pmatrix} \end{tikzcd}$$

where B and B' are subsets of A and the q_i's are states of $\mathcal{A}$, the condition $q_4 \in F$ implies $q_5 \in F$. This condition is also necessary if $\mathcal{A}$ is minimal. Consequently, 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 $3/2$.

The decidability of level 2 is a challenging open problem, although much progress has been made in recent years [4, 5, 8, 19, 31, 33, 35, 36]. In the case of languages whose syntactic monoid is an inverse monoid, a complete characterization was given by Cowan [8], completing partial results of Straubing and the second author [33, 35, 36]. Our main result gives a much shorter proof of Cowan's result and it is proved in [35, 36] that Cowan's result yields the following important corollary.

Corollary 5.5 *It is decidable whether an inverse monoid belongs to $\mathbf{V}_2$.*

5.2 Dot-depth hierarchy

In this hierarchy, introduced by Brzozowski [7], the level 0 is the trivial +-variety. The languages of level 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 $u_0, \dots, u_k \in A^*$. 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 main result. It also yields to a polynomial algorithm

Proposition 5.6 *A language of A^+ is of dot-depth $1/2$ if and only if its ordered syntactic semigroup satisfies the identity $x^\omega y x^\omega \leq x^\omega$.*

Theorem 5.7 *One can decide in polynomial time whether the language accepted by a deterministic n-state automaton is of dot-depth $1/2$.*

The sets of dot-depth 1 are the finite boolean combinations of languages of dot-depth 1/2. The syntactic characterization of these languages was settled by Knast and relies on the notion of *graph of a semigroup.* Given a semigroup S, form a graph $G(S)$ as follows: the vertices are the idempotents of S and the edges from e to f are the elements of the form esf.

Theorem 5.8 (Knast [10, 11]) *A language of A^+ is of dot-depth 1 if and only if the graph of its syntactic semigroup satisfies the following condition : if e and f are two vertices, p and r edges from e to f, and q and s edges from f to e, then $(pq)^\omega ps(rs)^\omega = (pq)^\omega(rs)^\omega$.*

Theorem 5.9 (Stern [29]) *One can decide in polynomial time whether the language accepted by a deterministic n-state automaton is of dot-depth* 1.

The variety of semigroups satisfying Knast's condition is usually denoted $\mathbf{B_1}$.

6 The sequential calculus

Büchi's sequential calculus is built up from a binary relation symbol $<$ and, for each letter $a \in A$, a unary predicate R_a. To each word u is associated a finite structure $\mathfrak{M}_u = \big(\{1, \ldots, |u|\}, (R_a)_{a \in A}, <\big)$ where $R_a = \{i \in \{1, \ldots, |u|\} \mid u(i) = a\}$ is the set of positions of the letter a in u and $<$ is the usual order on $\{1, \ldots, |u|\}$. For instance, if $u = abbaab$, then $R_a = \{1, 4, 5\}$ and $R_b = \{2, 3, 6\}$. Terms, atomic formulæ and first order formulæ are defined in the usual way. A word u satisfies a sentence φ if φ is true when interpreted on the structure $\mathfrak{M}_u$. There is a special convention for the empty word: it satisfies all universal sentences (sentences of the form $\forall x \varphi(x)$) and no existential sentences. To each sentence φ, one associates the sets of words $L(\varphi)$ that satisfy φ. For instance, if $\varphi = \exists i\ R_a i$, then $L(\varphi) = A^*aA^*$. The reader is referred to the survey article [17] for more detail on this logic. The first order definable languages were first characterized by McNaughton and Papert [13] : a recognizable subset of A^* is first-order definable if and only if it is star-free. The correspondence between star-free languages and first order logic is even tighter. Indeed, Thomas has shown that the Straubing hierarchy coincides with the quantifier alternation hierarchy of first order formulæ, defined as follows.

A formula φ is said to be a Σ_n-formula if it is equivalent to a formula of the form $\varphi = Q(x_1, \ldots, x_k)\psi$ where ψ is quantifier free and $Q(x_1, \ldots, x_k)$ is a sequence of n blocks of quantifiers such that the first block contains only existential quantifiers (note that this first block may be empty), the second block universal quantifiers, etc.. Similarly, if $Q(x_1, \ldots, x_k)$ is formed of n alternating blocks of quantifiers beginning with a block of universal quantifiers (which again might be empty), we say that φ is a Π_n-formula.

Denote by Σ_n (resp. Π_n) the class of languages which can be defined by a Σ_n-formula (resp. a Π_n-formula) and by $\mathcal{B}\Sigma_n$ the set of boolean combinations of Σ_n-formulæ. Finally, set, for every $n \geq 0$, $\Delta_n = \Sigma_n \cap \Pi_n$. The connection with Straubing's hierarchy can be stated as follows. Denote by $\mathcal{V}_n$ the class of languages of level n. In particular, $\mathcal{V}_{n+1/2}$ is equal to Pol $\mathcal{V}_n$.

Theorem 6.1 (Thomas [34], Perrin and Pin [14])

(1) *A language is in* $\mathcal{B}\Sigma_n$ *if and only if it is in* $\mathcal{V}_n$
(2) *A language is in* Σ_{n+1} *if and only if it is in Pol* $\mathcal{V}_n$
(3) *A language is in* Π_{n+1} *if and only if it is in Co-Pol* $\mathcal{V}_n$

We now complete this result by giving a characterization of the Δ_n classes, which follows immediately from Theorems 4.3 and 6.1.

Theorem 6.2 *A language of* A^* *is in* Δ_{n+1} *if and only if it is in UPol* $\mathcal{V}_n$.

Finally, our results on logic can be summarized in the following diagrams

$$\begin{tikzcd}[column sep=small]
 & \Sigma_1 \arrow[dr, dash] & & & \Sigma_2 \arrow[dr, dash] & & & \Sigma_3 \cdots \\
\Delta_0 = \Sigma_0 = \Pi_0 = \Delta_1 = \mathcal{B}\Sigma_0 \arrow[ur, dash] \arrow[dr, dash] & & \mathcal{B}\Sigma_1 \arrow[r, dash] & \Delta_2 \arrow[ur, dash] \arrow[dr, dash] & & \mathcal{B}\Sigma_2 \arrow[r, dash] & \Delta_3 \arrow[ur, dash] \arrow[dr, dash] & \\
 & \Pi_1 \arrow[ur, dash] & & & \Pi_2 \arrow[ur, dash] & & & \Pi_3 \cdots
\end{tikzcd}$$

Figure 6.1: The logical hierarchy

$$\begin{tikzcd}[column sep=small]
 & \text{Pol } \mathcal{V}_1 \arrow[dr, dash] & & & \text{Pol } \mathcal{V}_2 \arrow[dr, dash] & & & \text{Pol } \mathcal{V}_3 \cdots \\
\mathcal{V}_0 \arrow[ur, dash] \arrow[dr, dash] & & \mathcal{V}_1 \arrow[r, dash] & \text{UPol } \mathcal{V}_2 \arrow[ur, dash] \arrow[dr, dash] & & \mathcal{V}_2 \arrow[r, dash] & \text{UPol } \mathcal{V}_3 \arrow[ur, dash] \arrow[dr, dash] & \\
 & \text{Co-Pol } \mathcal{V}_1 \arrow[ur, dash] & & & \text{Co-Pol } \mathcal{V}_2 \arrow[ur, dash] & & & \text{Co-Pol } \mathcal{V}_3 \cdots
\end{tikzcd}$$

Figure 6.2: The Straubing-Thérien hierarchy

7 Conclusion and open problems

Let $\mathbf{V}$ be a variety of semigroups and let $\mathcal{V}$ be the corresponding $+$-variety. We have shown that the algebraic counterpart of the operation $\mathcal{V} \to \text{Pol } \mathcal{V}$ 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 $\mathbf{V} \to [\![x^\omega \leq x^\omega y x^\omega]\!] \circledM \mathbf{V}$. We conjecture that the variety of semigroups (resp. monoids) corresponding to BPol $\mathcal{V}$ is $\mathbf{B_1} \circledM \mathbf{V}$. The conjecture was proved to be true if $\mathbf{V}$ is the trivial variety of monoids, the trivial variety of semigroups or the variety of monoids consisting of all groups [32, 10, 12]. Note also that every language of BPol $\mathcal{V}$ is recognized by a semigroup of

$\mathbf{B_1} \circledM \mathbf{V}$. Finally, it is proved in [21] that the identities of $\mathbf{B_1} \circledM \mathbf{V}$ are

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

for all $x, y, p, q, r, s \in \hat{A}^*$ for some finite alphabet A such that $\mathbf{V}$ satisfies $x^2 = x = y = p = q = r = s$.

Acknowledgements

The authors would like to thank Marc Zeitoun for a careful reading of a first version of this article and Jean Goubault for asking a question about the Δ_n hierarchy that led to section 6 of this article.

References

[1] J. Almeida, *Finite Semigroups and Universal Algebra*, World Scientific (Series in Algebra, Volume 3), Singapore, 1995, 511 pp.

[2] M. Arfi, Polynomial operations and rational languages, 4th STACS, *Lecture Notes in Computer Science* **247**, (1987) 198–206.

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

[4] F. Blanchet-Sadri, On dot-depth two, *Informatique Théorique et Applications* **24**, (1990) 521–529.

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

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

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

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

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

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

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

[12] S. W. Margolis and J.E. Pin, Product of group languages, *FCT Conference, Lecture Notes in Computer Science* **199**, (1985) 285–299.

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

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

[15] J.-E. Pin, Propriétés syntactiques du produit non ambigu. *7th ICALP, Lecture Notes in Computer Science* **85**, (1980) 483–499.

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

[17] J.-E. Pin, Logic, Semigroups and Automata on Words, *Annals of Math. and Artificial Intelligence*, to appear.

[18] J.-E. Pin, A variety theorem without complementation, *Izvestija vuzov. Matematika*, to appear.

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

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

[21] J.-E. Pin and P. Weil, Free profinite semigroups, Mal'cev products and identities, to appear.

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

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

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

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

[26] I. Simon, Piecewise testable events, *Proc. 2nd GI Conf., Lecture Notes in Computer Science* **33**, (1975) 214–222.

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

[28] I. Simon, The product of rational languages, *Proceedings of ICALP 1993, Lecture Notes in Computer Science* **700** , (1993), 430–444.

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

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

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

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

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

[34] W. Thomas, Classifying regular events in symbolic logic, *J. Comput. System Sci.* **25**, (1982) 360–375.

[35] P. Weil, Inverse monoids of dot-depth two, *Theoret. Comput. Sci.* **66**, (1989), 233–245.

[36] P. Weil, Some results on the dot-depth hierarchy, *Semigroup Forum* **46** (1993), 352–370.