Polynomial closure and unambiguous product

Jean-Eric Pin and Pascal Weil *

pin@litp.ibp.fr, weil@litp.ibp.fr

Abstract

This article is a contribution to the algebraic theory of automata, but it also contains an application to Büchi's sequential calculus. The polynomial closure of a class of languages $\mathcal{C}$ is the set of languages that are finite unions of languages 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 $\mathcal{C}$. Our main result is an algebraic characterization, via the syntactic monoid, of the polynomial closure of a variety of languages. We show that the algebraic operation corresponding to the polynomial closure is a certain Mal'cev product of varieties. This result has several consequences. We first study the concatenation hierarchies similar to the dot-depth hierarchy, obtained by counting the number of alternations between boolean operations and concatenation, For instance, we show that level 3/2 of the Straubing hierarchy is decidable and we give a simplified proof of the partial result of Cowan on level 2. We propose a general conjecture for these hierarchies. We also show that if a language and its complement are in the polynomial closure of a variety of languages, then this language can be written as a disjoint union of marked unambiguous products of languages of the variety. This allows us to extend the results of Thomas on quantifier hierarchies of first-order logic.

1 Introduction

This paper is a contribution to the algebraic theory of recognizable languages, in the spirit of the work of Eilenberg and Schützenberger. Eilenberg's variety theorem gives a bijective correspondence between varieties of languages and varieties of finite semigroups or finite monoids. Varieties of languages are classes of recognizable languages closed under finite boolean

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

operations, inverse morphisms and left and right quotients. Much effort has been devoted in recent years to extend this correspondence to *operations*. That is, given an operation on languages, find the associated operation on the semigroup level, or conversely, given an operation on (finite) semigroups, find the associated operation on the languages level. The most important operations on languages are the boolean operations (union, intersection and complement), the concatenation product and Kleene's star operation [26]. Most classification schemes on recognizable languages proposed in the sixties, like the star-height or the dot-depth are based on these three basic operations.

The main topic of this paper is the *concatenation product*, an operation already widely studied in the literature: Arfi [5, 6], Blanchet-Sadri [10, 11, 13, 14, 12, 15], Brzozowski [17, 18, 19], Cowan [20], Eilenberg [22], Knast [27, 28], Schützenberger[57, 58], Simon [59, 62], Straubing [48, 49, 64, 66, 68, 69, 70], Thérien [49, 69], Thomas [72] and the authors [35, 36, 39, 48, 49, 70, 74, 77].

These former works have shown that, instead of the pure concatenation product, the real fundamental operation 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 introduction of the letters a_i's is a bit surprising and can only be justified by subsequent developments. However, it suffices to say that this operation is more natural in the algebraic perspective we want to stress. It is also much more suitable for the connections with formal logic (see our Section 10). We also consider the *unambiguous polynomial closure*, that is the closure under disjoint union and unambiguous marked product, and the *boolean closure of the polynomial closure*. One can also define, with a slight modification (see Section 5) similar operators for languages of A^+.

The main result of this paper is an algebraic characterization of the polynomial closure. 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 of the first author [44]. 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 $\mathcal{V}$. The solution involves algebraic results on ordered monoids that generalize known results on monoids. For instance, most results about identities defining varieties of monoids carry over for varieties of ordered monoids [51]. The Mal'cev product, one of the most powerful operations on varieties of monoids can also be extended to varieties of ordered monoids [50]. The variety of ordered monoids corresponding to the polynomial closure of $\mathcal{V}$ is precisely a Mal'cev product of the form $\mathbf{W} \circledM \mathbf{V}$, where $\mathbf{W}$ is the variety of finite ordered semigroups $(S, \leqslant)$ in which $ese \leqslant e$, for each idempotent e and each element s in S. The formulation of this result is very close to the algebraic characterization of the unambiguous polynomial closure obtained in [49]: the variety of ordered monoids corresponding to the unambiguous polynomial closure of $\mathcal{V}$ is the Mal'cev product $\mathbf{LI} \circledM \mathbf{V}$, where $\mathbf{LI}$ is the variety of semigroups S in which $ese = e$, for each idempotent e and each s in S.

The proof of our main result is non-trivial and relies on a deep theorem of Simon [60] on factorization forests. Its importance can probably be better understood on its consequences. 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 a given boolean algebra of recognizable languages, and call it level 0. Then define recursively the higher levels as follows: the level $n + 1/2$ is the polynomial closure of the level n and the level $n + 1$ is the boolean closure of the level $n + 1/2$. Note that a set of level m is also a set of level n for every $n \geqslant m$. The main problems concerning these hierarchies is to know whether they are infinite and whether each level is decidable. At least three different hierarchies of this type were proposed in the literature and all three were proved to be infinite: the Straubing hierarchy, whose level 0 are the empty language and A^*, the dot-depth hierarchy, whose level 0 consists of the finite or cofinite languages[1], and the group hierarchy, whose level 0 consists of the group languages. A *group language* is simply a recognizable language accepted by a *permutation automaton*, that is, a complete deterministic finite automaton in which each letter induces a permutation on the set of states.

[1]in this particular case, languages must be considered as subsets of A^+. This is a subtle, but important detail.

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 automaton $\mathcal{A}$ with n states 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 the level 2 is still an open question, but we make some progress on this problem. First we give a short proof of a result of Cowan [20] 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. Several conjectures have been proposed before, but this one is the first that gives explicitly a set of identities for this variety. Actually, our conjecture is a particular case of a more general conjecture on the boolean closure of the polynomial closure. We conjecture that the algebraic counterpart of this operation is also a Mal'cev product. More precisely, 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 $\mathbf{B_1} \textcircled{m} \mathbf{V}$, where $\mathbf{B_1}$ is the variety of finite semigroups corresponding to languages of dot-depth one. We also present an equivalent formulation of this conjecture in terms of ordered monoids (Conjecture 9.1). This last conjecture leads to a promising track. Indeed, the simplest case of our conjecture, obtained by taking for $\mathbf{V}$ the trivial variety, is a nice result of Straubing and Thérien [69] stating that every finite $\mathcal{J}$-trivial monoid is a quotient of an ordered monoid satisfying the identity $x \leqslant 1$. The hope would be to adjust the artful proof of this latter result to some other cases.

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.

Our results on the group hierarchy were announced in [41] in a slightly different form. It is easy to see that level 0 is decidable, but the decidability of level 1 follows from a series of non trivial results in semigroup theory [25, 24]. The languages of level 1/2 were also widely studied. In particular, they are exactly the recognizable open sets of the progroup (or Hall) topology on the free monoid. One of the non-trivial consequences of our main result is that level 1/2 in the group hierarchy is also decidable. Furthermore, our algorithm to decide whether a recognizable language is of level 1/2 gives as a byproduct an algebraic and effective characterization of the recognizable open sets in the progroup topology, a result conjectured in [40].

Our new approach is also related to the Schützenberger product, an algebraic tool studied by several authors [36, 39, 48, 53, 57, 65]. We first observe that the Schützenberger product can be naturally equipped with an order. Thus, given a variety of finite monoids $\mathbf{V}$, the Schützenberger products of members of $\mathbf{V}$ generate a variety of ordered monoids. We show that this variety is precisely the Mal'cev product $\mathbf{W} \circledM \mathbf{V}$ of our main result. This proves the equivalence of the two constructions in the case of monoids. However, our construction still corresponds to the polynomial closure in the case of languages of A^+. This is not the case of the Schützenberger product, contrary to a claim of the first author in [36].

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, it has been known for some time that there are some nice connections between the Straubing hierarchy and formal logic [72, 34, 45]. More precisely, Thomas [72] (see also [34]) 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 [58].

Our paper is organized as follows. Sections 2, 3 and 4 introduce the necessary background. Section 5 contains our main result. The connections with the Schützenberger product are analyzed in Section 6. The results on the unambiguous polynomial closure are discussed in Section 7. Section 8 is devoted to concatenation hierarchies and our conjecture is discussed in Section 9. Section 10 contains the applications to formal logic.

2 Varieties

Our approach in this paper is purely algebraic and relies mainly on the concept of variety. Some very recent developments of the theory of varieties are used in this paper, and thus it seems appropriate to recall these results to keep the paper self-contained. We will review, in order, varieties of semigroups and of ordered semigroups, description by identities in the free profinite semigroup, relational morphisms and the Mal'cev product. If S is a semigroup, S^1 denotes the monoid equal to S if S has an identity and to $S \cup \{1\}$ otherwise.

2.1 Varieties of semigroups and ordered semigroups

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

An ordered semigroup $(S, \leqslant)$ is an *ordered subsemigroup* of $(T, \leqslant)$ if S is a subsemigroup of T and the order on S is the restriction to S of the order on T. An ordered semigroup $(T, \leqslant)$ is an *ordered quotient* of $(S, \leqslant)$ if there exists a surjective morphism of ordered semigroups $\varphi : (S, \leqslant) \to (T, \leqslant)$. For instance, any ordered semigroup $(S, \leqslant)$ is a quotient of $(S, =)$. Given a family $(S_i, \leqslant)_{i \in I}$ of ordered semigroups, the product $\prod_{i \in I}(S_i, \leqslant)$ is the ordered semigroup $\prod_{i \in I} S_i$ equipped with the product order.

Let A be a set and let A^+ be the free semigroup on A. Then $(A^+, =)$ is an ordered semigroup, which is in fact the free ordered semigroup on A.

Recall that a *variety of finite semigroups* (or *pseudovariety*) is a class of finite semigroups closed under the taking of subsemigroups, quotients and finite products. Similarly, a *variety of finite ordered semigroups* is a class of finite ordered semigroups closed under the taking of ordered subsemigroups, ordered quotients and finite products. *Varieties of finite monoids* and *varieties of finite ordered monoids* are defined in the same way. If $\mathbf{V}$ is a variety of finite semigroups, the class of all ordered semigroups of the form $(S, \leqslant)$, where $S \in \mathbf{V}$, is a variety of ordered semigroups, called the variety of ordered semigroups *generated* by $\mathbf{V}$, also denoted $\mathbf{V}$. The context will make clear whether $\mathbf{V}$ is considered as a variety of semigroups or as a variety of ordered semigroups.

Given a variety of finite ordered semigroups, the class of all duals of members of $\mathbf{V}$ form a variety of finite ordered semigroups, called the *dual* of $\mathbf{V}$ and denoted $\check{\mathbf{V}}$. The *join* of two varieties of finite ordered semigroups $\mathbf{V}_1$ and $\mathbf{V}_2$ is the smallest variety of finite ordered semigroups containing $\mathbf{V}_1$ and $\mathbf{V}_2$. The join of a variety and its dual will be of special interest in the sequel.

It is a well known fact that varieties of semigroups (in the Birkhoff sense) can be defined by identities. Similarly, a result of Bloom [16] shows that varieties of ordered semigroups can be defined by identities of the form $u \leqslant v$. Analogous results hold for varieties of finite (ordered) semigroups [1, 4, 52, 75], but their statements require some topological preliminaries.

2.2 Profinite completions and identities

Let A be a finite alphabet and let u, v be two words of A^*. A finite 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

$$r(u, v) = \min\{\ |M| \ \bigm|\ M \text{ separates } u \text{ and } v\ \}$$

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 the following, for all $u, v, w \in A^*$:

(1) $d(u, v) = 0$ if and only if $u = v$,
(2) $d(u, v) = d(v, u)$,
(3) $d(u, v) \leqslant \max\big(d(u, w), d(v, w)\big)$,
(4) $d(uw, vw) \leqslant d(u, v)$ and $d(wu, wv) \leqslant d(u, v)$.

That is, 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 $\hat{A}^*$ and called the *free profinite monoid* on A.

We consider each finite monoid M as being equipped with a discrete distance, defined, for every $x, y \in M$ by

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

Then 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 [22, 38].

Proposition 2.1 *Let A be a finite alphabet and let $x \in \hat{A}^*$. The sequence $(x^{n!})_{n\geqslant 0}$ converges in $\hat{A}^*$ to an idempotent denoted x^ω.*

Note that if $\mu : \hat{A}^* \to M$ is a continuous morphism into a finite monoid, then $x^\omega \mu$ is equal to $(x\mu)^\omega$, 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 finite monoid M *satisfies the identity* $x = y$ if, for every continuous morphism $\varphi : \hat{A}^* \to M$, $x\varphi = y\varphi$. Similarly, a finite ordered monoid $(M, \leqslant)$ *satisfies the identity* $x \leqslant y$ if, for every continuous morphism $\varphi : \hat{A}^* \to M$, $x\varphi \leqslant y\varphi$. The context will make clear the sense in which we use the word "identity".

Reiterman's theorem [52] shows that every variety of finite monoids can be defined by a set of identities. The authors have extended this result to varieties of finite ordered monoids [51]. Given a set E of identities, we denote by $[\![E]\!]$ the class of all finite ordered monoids which satisfy all the identities of E.

Proposition 2.2 *Let E be a set of identities. Then the class $[\![E]\!]$ forms a variety of finite ordered monoids. Conversely, for each variety of finite ordered monoids, there exists a set E of identities such that $\mathbf{V} = [\![E]\!]$.*

For instance, the following descriptions of varieties will be used in the sequel. These descriptions make use of the notation ω defined above. The variety $\mathbf{G} = [\![x^\omega = 1]\!]$ is the variety of all finite groups, $\mathbf{A} = [\![x^\omega = x^{\omega+1}]\!]$ is the variety of aperiodic monoids, $\mathbf{J_1} = [\![x^2 = x, xy = yx]\!]$ is the variety of idempotent and commutative monoids and $\mathbf{DA} = [\![x^\omega = x^{\omega+1} \text{ and } (xy)^\omega = (xy)^\omega(yx)^\omega(xy)^\omega]\!]$ is the variety of monoids whose regular $\mathcal{D}$-classes are idempotent semigroups. See Almeida [3] or Pin [38].

If E is a set of identities, we denote by $\breve{E}$ the set of identities of the form $v \leqslant u$ such that the identity $u \leqslant v$ belongs to E. The set $\breve{E}$ is called the *dual* of E. It is intuitively obvious that if E is a set of identities and if $\mathbf{V} = [\![E]\!]$, then $\breve{\mathbf{V}} = [\![\breve{E}]\!]$. In other words, if a variety of finite ordered semigroups is defined by a set E of identities, its dual is defined by the dual of E.

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

2.3 Relational morphisms and Mal'cev products

In this section, we extend the standard notions of relational morphism and Mal'cev product to ordered monoids. One comes across the usual definition when the order relation on the monoids is equality. 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 $s \in S$,
For a relational morphism between two monoids S and T, a third condition is required

(3) $1 \in 1\tau$

Equivalently, τ is a relation whose graph

$$\mathrm{graph}(\tau) = \{\ (s,t) \mid t \in s\tau\ \}$$

is a subsemigroup (resp. submonoid if S and T are monoids) of $S \times T$ that projects onto S.

Let $\mathbf{V}$ be a variety of monoids (resp. semigroups) and let $\mathbf{W}$ be a variety of semigroups. The *Mal'cev product* $\mathbf{W} \textcircled{M} \mathbf{V}$ 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 $\mathbf{W} \textcircled{M} \mathbf{V}$ is a variety of monoids (resp. semigroups).

More generally, if $\mathbf{V}$ be a variety of monoids and $\mathbf{W}$ be a variety of ordered semigroups, the *Mal'cev product* $\mathbf{W} \textcircled{M} \mathbf{V}$ is the class of all ordered monoids $(M, \leqslant)$ 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. One verifies that $\mathbf{W} \textcircled{M} \mathbf{V}$ is a variety of ordered monoids. The following theorem, obtained by the authors [50], describes a set of identities defining $\mathbf{W} \textcircled{M} \mathbf{V}$.

Theorem 2.3 *Let $\mathbf{V}$ be a variety of monoids and let $\mathbf{W}$ be a variety of ordered semigroups. Let E be a set of identities such that $\mathbf{W} = [\![E]\!]$. Then $\mathbf{W} \textcircled{M} \mathbf{V}$ is defined by the identities of the form $x\sigma \leqslant y\sigma$, where $x \leqslant y$ is an identity of E with $x, y \in \hat{B}^*$ for some finite alphabet B and $\sigma : \hat{B}^* \to \hat{A}^*$ is a continuous morphism such that, for all $b, b' \in B$, $\mathbf{V}$ satisfies $b\sigma = b'\sigma = b^2\sigma$.*

We will use in particular the following applications of our result.

Corollary 2.4 *Let $\mathbf{V}$ be a variety of monoids. Then $\mathbf{LI} \textcircled{M} \mathbf{V}$ 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 $x = y = x^2$.*

Corollary 2.5 *Let* $\mathbf{V}$ *be a variety of monoids. Then* $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$ *is defined by the identities of the form* $x^\omega y x^\omega \leqslant x^\omega$*, where* $x, y \in \hat{A}^*$ *for some finite set* A *and* $\mathbf{V}$ *satisfies* $x = y = x^2$*.*

In the case where $\mathbf{V} = \mathbf{G}$, the variety of all finite groups, one can give a much simpler set of identities, but the proof makes use of an important result of Ash [9] (see also the survey [24] for the relevant background). First recall that a submonoid D of a monoid M is *closed under weak conjugacy* if the conditions $sts = s$ and $u \in D$ imply $sut \in D$ and $tus \in D$. Given a monoid M, denote by $D(M)$ the smallest submonoid of M closed under weak conjugacy. The following consequence of Ash's theorem is proved in [50] .

Theorem 2.6 *If an element* x *of* $\hat{A}^*$ *satisfies* $x = 1$ *in* $\mathbf{G}$*, then for each finite monoid* M *and for each morphism* $\varphi : A^* \to M$*,* $x\varphi$ *belongs to* $D(M)$*.*

We are now ready to give the identities of the variety $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{G}$.

Theorem 2.7 *The variety of ordered monoids* $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{G}$ *is defined by the identity* $x^\omega \leqslant 1$*.*

Proof First, by Corollary 2.5, $(M, \leqslant)$ belongs to $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{G}$ if and only if it satisfies the identities $x^\omega y x^\omega \leqslant x^\omega$, for all $x, y \in \hat{A}^*$ such that A is finite and $\mathbf{G}$ satisfies $x = y = 1$ (because $x = x^2$ implies $x = 1$ in a group). This shows in particular, by taking $x = 1$ and $y = u^\omega$, that $(M, \leqslant)$ satisfies the identity $u^\omega \leqslant 1$.

Conversely, assume that $(M, \leqslant)$ satisfies the identity $u^\omega \leqslant 1$. We claim that for every $d \in D(M)$, $d \leqslant 1$. Let $D'(M)$ be the set of all $x \in M$ such that $x \leqslant 1$. Clearly, $D'(M)$ is a submonoid of M. Furthermore, $D'(M)$ is closed under weak conjugacy: indeed, if $sts = s$ and $x \leqslant 1$, then $sxt \leqslant st \leqslant 1$ (since st is idempotent) and similarly, $txs \leqslant 1$. Therefore $D'(M)$ contains $D(M)$, proving the claim. It follows, by Theorem 2.6, that for every $y \in \hat{A}^*$ such that $\mathbf{G}$ satisfies $y = 1$, M satisfies $y \leqslant 1$. This implies in particular that M satisfies $x^\omega y x^\omega \leqslant x^\omega x^\omega = x^\omega$ for every $x, y \in \hat{A}^*$ such that $\mathbf{G}$ satisfies $x = y = 1$. $\square$

3 Recognizable languages

In this section we briefly review the main definitions and results about recognizable languages needed in this paper. In particular, we present the point of view of ordered semigroups recently introduced in [44].

Let $(S, \leqslant)$ be an ordered semigroup and let $\eta : (S, \leqslant) \to (T, \leqslant)$ be a surjective morphism of ordered semigroups. An order ideal Q of S is said to be *recognized* by η if there exists an order ideal P of T such that $Q = P\eta^{-1}$. Notice that this condition implies $Q\eta = P\eta^{-1}\eta = P$. This definition can be applied in particular to languages. A language L of A^+ is recognized by an ordered semigroup $(T, \leqslant)$ if there exists a surjective morphism of ordered semigroups $\eta : (A^+, =) \to (T, \leqslant)$ and an order ideal P of T such that $L = P\eta^{-1}$. A language is *recognizable* if it is recognized by a finite ordered semigroup. This definition is equivalent to the standard definition of a recognizable language: a language L of A^+ is recognizable if and only if there exists a surjective morphism from A^+ onto a finite semigroup T and a subset P of T such that $L = P\eta^{-1}$. Indeed, one may consider simply η as a morphism of ordered semigroups from $(A^+, =)$ onto $(T, =)$, since the condition on orders is trivially satisfied in this case ($x = y$ implies $x\eta = y\eta$) and any subset of $(T, =)$ is an order ideal.

3.1 Syntactic semigroup and syntactic ordered semigroup

Let $(T, \leqslant)$ be an ordered semigroup and let P be an order ideal of T. The *syntactic quasiordering* of P is the relation $\preceq_P$ defined by setting

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

One can show that $\preceq_P$ is a stable quasiorder on T and that the associated equivalence relation $\sim_P$, defined by

$$u \sim_P v \text{ if and only if } u \preceq_P v \text{ and } v \preceq_P u$$

is a semigroup congruence, called the *syntactic congruence* of P. The quotient semigroup $S(P) = T/{\sim_P}$ is called the *syntactic semigroup of* P. The quasiorder $\preceq_P$ on T induces a stable order $\leqslant_P$ on $S(P)$. The ordered semigroup $(S(P), \leqslant_P)$ is called the *syntactic ordered semigroup* of P and the natural morphism $\eta_P : (T, =) \to (S(P), \leqslant_P)$ is called the *syntactic morphism* of P. The universal property of this morphism is given in the next proposition [44].

Proposition 3.1 *Let $\varphi : (R, \leqslant) \to (S, \leqslant)$ be a surjective morphism of ordered semigroups and let P be an order ideal of $(R, \leqslant)$. Then φ recognizes P if and only if η_P factorizes through φ.*

The previous definitions apply in particular when T is a free semigroup and P is a language. Indeed, if A is a finite alphabet, then $(A^+, =)$ is an ordered semigroup and every subset of A^+ is an order ideal. Furthermore, if $(S, \leqslant)$ is an ordered semigroup, every surjective semigroup morphism $\eta : A^+ \to S$ induces a surjective morphism of ordered semigroups from $(A^+, =)$ onto $(S, \leqslant)$. Therefore, a language $L \subseteq A^+$ is said to be *recognized* by a semigroup morphism $\eta : A^+ \to (S, \leqslant)$ if there exists an order ideal P of S such that $L = P\eta^{-1}$. By extension, given an ordered semigroup $(S, \leqslant)$ and an order ideal P of S, we say that (S, P) recognizes $L \subseteq A^+$ if there exists a surjective semigroup morphism $\eta : A^+ \to S$ such that $L = P\eta^{-1}$.

The syntactic ordered semigroup of the complement of an order ideal is obtained by reversing the order.

Proposition 3.2 *Let P be an order ideal of $(T, \leqslant)$. Then $T \setminus P$ is an order ideal of $(T, \geqslant)$ and the syntactic ordered semigroup of $T \setminus P$ is the dual of the syntactic ordered semigroup of P.*

Proof By definition, $u \preceq_{S \setminus P} v$ if and only if, for all $x, y \in T^1$, $xvy \in S \setminus P$ implies $xuy \in S \setminus P$, which is equivalent to saying that $xuy \in P$ implies $xvy \in P$. Thus $u \preceq_{S \setminus P} v$ if and only if $v \preceq_P u$. $\square$

Corollary 3.3 *Let L be a language of A^+ and let $(S(L), \leqslant_L)$ be its syntactic ordered semigroup. Then the syntactic ordered semigroup of $A^+ \setminus L$ is $(S(L), \geqslant_L)$.*

We have already defined the notion of variety of finite ordered semigroups generated by a variety of finite semigroups. Conversely, we would like to define the variety of semigroups generated by a variety of ordered semigroups $\mathbf{V}$. To be symmetrical, our definition has to give the same result for $\mathbf{V}$ and for its dual. Therefore, it is natural to define the *variety of semigroups generated by* $\mathbf{V}$ to be the class of all semigroups S such that $(S, \leqslant) \in \mathbf{V} \vee \check{\mathbf{V}}$ for some order $\leqslant$ on S. The following result shows that this class is really a variety of semigroups.

Proposition 3.4 *Let S be a finite semigroup and let $\mathbf{V}$ be a variety of ordered semigroups. Then $(S, \leqslant) \in \mathbf{V} \vee \check{\mathbf{V}}$ for some order $\leqslant$ on S if and only if $(S, =) \in \mathbf{V} \vee \check{\mathbf{V}}$.*

Proof If $(S, \leqslant) \in \mathbf{V} \vee \check{\mathbf{V}}$, then $(S, \geqslant) \in \mathbf{V} \vee \check{\mathbf{V}}$ by duality. Now, the diagonal embedding shows that $(S, =)$ is an ordered subsemigroup of $(S, \leqslant) \times (S, \geqslant)$ and thus $(S, =) \in \mathbf{V} \vee \check{\mathbf{V}}$. Furthermore, $(S, \leqslant)$ is a quotient of $(S, =)$. Therefore, if $(S, =) \in \mathbf{V} \vee \check{\mathbf{V}}$, then $(S, \leqslant) \in \mathbf{V} \vee \check{\mathbf{V}}$. $\square$

Here is an equivalent definition.

Proposition 3.5 *Let* $\mathbf{V}$ *be a variety of finite ordered semigroups. A semigroup belongs to the variety of finite semigroups generated by* $\mathbf{V}$ *if and only if it is a quotient of an ordered semigroup of* $\mathbf{V}$.

Proof Let $\mathbf{W}$ be the variety of semigroups generated by $\mathbf{V}$. If T is a quotient of a semigroup S such that $(S, \leqslant) \in \mathbf{V}$ for some order $\leqslant$ on S, then $S \in \mathbf{W}$ by definition and thus $T \in \mathbf{W}$. Conversely, if $T \in \mathbf{W}$, then $(T, =) \in \mathbf{V} \vee \check{\mathbf{V}}$ by Proposition 3.4 and thus there exist two ordered semigroups $(S_1, \leqslant_1) \in \mathbf{V}$ and $(S_2, \leqslant_2) \in \check{\mathbf{V}}$ such that $(T, =)$ is a quotient of an ordered subsemigroup $(S, \leqslant)$ of $(S_1, \leqslant_1) \times (S_2, \leqslant_2)$. It follows that S is a subsemigroup of $S_1 \times S_2$. Now $(S_1, \leqslant_1) \times (S_2, \geqslant_2) \in \mathbf{V}$ and the order $\leqslant'$ induced by $\leqslant_1 \times \geqslant_2$ on S defines an ordered semigroup $(S, \leqslant')$ of $\mathbf{V}$. But T is a quotient of S, concluding the proof. $\square$

3.2 Varieties of languages

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

(1) for every alphabet A, $A^+\mathcal{V}$ is closed under finite union, finite intersection and complement[2],
(2) for every semigroup morphism $\varphi : A^+ \to B^+$, $L \in B^+\mathcal{V}$ implies $L\varphi^{-1} \in A^+\mathcal{V}$,
(3) If $X \in A^+\mathcal{V}$ and $a \in A$, then $a^{-1}L$ and La^{-1} are in $A^+\mathcal{V}$.

Semigroup varieties and $+$-varieties are closely related. To each variety of semigroups $\mathbf{V}$, we associate the $+$-class $\mathcal{V}$ such that, for each alphabet A, $A^+\mathcal{V}$ is the set of recognizable languages of A^+ whose syntactic semigroup belongs to $\mathbf{V}$. One can show that $\mathcal{V}$ is a $+$-variety.

Theorem 3.6 (Eilenberg [22]) *The correspondence* $\mathbf{V} \to \mathcal{V}$ *defines a bijective correspondence between the varieties of finite semigroups and the* $+$-*varieties.*

[2]This includes union and intersection of an empty family of languages. Therefore $\emptyset$ and A^+ are always elements of $A^+\mathcal{V}$.

The variety of finite semigroups corresponding to a given $+$-variety is the variety of semigroups generated by the syntactic semigroups of all the languages $L \in A^+\mathcal{V}$, for every finite alphabet A.

There is a similar statement for varieties of ordered semigroups. A *positive $+$-variety* is a $+$-class of recognizable languages $\mathcal{V}$ such that

(1) for every alphabet A, $A^+\mathcal{V}$ is closed under finite union and finite intersection[3],

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

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

Thus, contrary to a variety, a positive variety need not be closed under complement. To each variety of ordered semigroups $\mathbf{V}$, we associate the $+$-class $\mathcal{V}$ such that, for each alphabet A, $A^+\mathcal{V}$ is the set of recognizable languages of A^+ whose ordered syntactic semigroup belongs to $\mathbf{V}$. One can show that $\mathcal{V}$ is a positive $+$-variety.

Theorem 3.7 [44] *The correspondence $\mathbf{V} \to \mathcal{V}$ defines a bijective correspondence between the varieties of finite ordered semigroups and the positive $+$-varieties.*

Taking the dual of a variety of finite ordered semigroups $\mathbf{V}$ corresponds to complementation at the language level. More precisely, let $\mathcal{V}$ (resp. $\breve{\mathcal{V}}$) be the positive $+$-variety corresponding to $\mathbf{V}$ (resp. to $\breve{\mathbf{V}}$).

Theorem 3.8 *For each alphabet A, $A^+\breve{\mathcal{V}}$ is the class of all complements in A^+ of the languages of $A^+\mathcal{V}$.*

Proof This follows immediately from Corollary 3.3. $\square$

The *join* of two positive $+$-varieties $\mathcal{V}_1$ and $\mathcal{V}_2$ is the smallest positive $+$-variety $\mathcal{V}$ such that, for every alphabet A, $A^+\mathcal{V}$ contains $A^+\mathcal{V}_1$ and $A^+\mathcal{V}_2$. Let $\mathcal{V}$ be a positive $+$-variety and let $\mathbf{V}$ be the corresponding variety of finite ordered semigroups. For each alphabet A, denote by $A^+\mathrm{B}\mathcal{V}$ the boolean algebra generated by $A^+\mathcal{V}$.

Proposition 3.9 *For every positive $+$-variety, $\mathcal{V} \vee \breve{\mathcal{V}} = B\mathcal{V}$. Furthermore, $B\mathcal{V}$ is a $+$-variety and the corresponding variety of semigroups is the variety of finite semigroups generated by $\mathbf{V} \vee \breve{\mathbf{V}}$.*

[3] See the previous footnote.

Proof Let $\mathcal{W}$ be the join of $\mathcal{V}$ and $\check{\mathcal{V}}$ and let A be an alphabet. Then all the languages of $A^+\mathcal{V}$ and their complements are in $A^+\mathcal{W}$. It follows that every language of $A^+\mathrm{B}\mathcal{V}$ is a union of intersections of languages of $A^+\mathcal{W}$ and thus $A^+\mathrm{B}\mathcal{V}$ is contained in $A^+\mathcal{W}$. On the other hand, for each alphabet A, $A^+\mathrm{B}\mathcal{V}$ is a boolean algebra. Furthermore, since boolean operations commute with inverse morphisms and with left and right quotients, $\mathrm{B}\mathcal{V}$ is closed under these operations. Therefore $\mathrm{B}\mathcal{V}$ is a +-variety. Since $\mathcal{W}$ is the smallest positive +-variety containing $\mathcal{V}$ and $\check{\mathcal{V}}$, $\mathcal{W}$ is contained in $\mathrm{B}\mathcal{V}$ and thus $\mathcal{W} = \mathrm{B}\mathcal{V}$. □

Again, there are similar statements for the varieties of finite monoids. In this case, the definitions of a class of languages and of varieties of languages have to be modified by replacing "semigroup" by "monoid" and + by $*$.

A finite semigroup S is *aperiodic* if and only if it satisfies the identity $x^\omega = x^{\omega+1}$. The connection between aperiodic semigroups and star-free sets was established by Schützenberger [57] (see also [29, 33, 38]. Recall that the *star-free* languages of A^* (resp. A^+) form the smallest class of languages containing the finite languages and closed under the boolean operations and the concatenation product.

Theorem 3.10 *A recognizable subset of A^* (resp. A^+) is star-free if and only if its syntactic monoid (resp. semigroup) is aperiodic.*

4 Factorization forests

We review in this section an important combinatorial result of I. Simon on finite semigroups which is a key argument in the proofs of the results of Section 5 below (see in particular Proposition 5.5). A *factorization forest* is a function d that associates to every word x of A^2A^* a factorization $d(x) = (x_1, \ldots, x_n)$ of x such that $n \geqslant 2$ and $x_1, \ldots, x_n \in A^+$. The integer n is the *degree* of the factorization $d(x)$. Thus a factorization forest is just a description of a recursive process to factorize words up to products of letters. To each word x such that $d(x) = (x_1, \ldots, x_n)$ is associated a labeled tree

$t(x)$ defined by $t(x) = (x, (t(x_1), \ldots, t(x_n)))$. For instance, if

$$\begin{aligned} d(x) &= (x_1, a_{12}, x_4) & d(x_1) &= (x_2, a_7, a_8, a_9, a_{10}, a_{11}) \\ d(x_2) &= (a_1, x_3) & d(x_3) &= (a_2, a_3, a_4, a_5, a_6) \\ d(x_4) &= (x_5, x_6, x_9, x_{10}) & d(x_5) &= (a_{13}, a_{14}) \\ d(x_6) &= (a_{15}, x_7, a_{18}, a_{19}, x_8, a_{22}) & d(x_7) &= (a_{16}, a_{17}) \\ d(x_8) &= (a_{20}, a_{21}) & d(x_9) &= (a_{23}, a_{24}) \\ d(x_{10}) &= (a_{25}, a_{26}) \end{aligned}$$

then the tree of x is represented in the figure below.

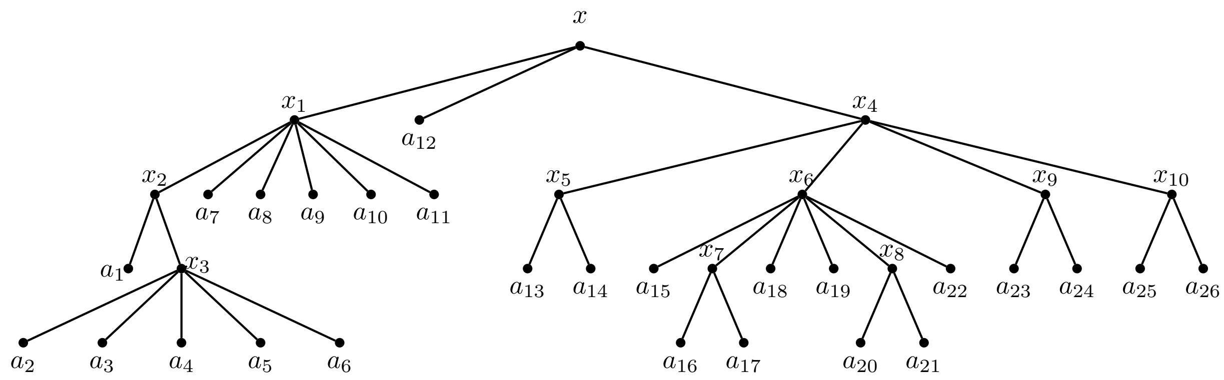


Figure 4.1: The tree $t(x)$.

Given a factorization forest d, the *height function* of d is the function $h : A^+ \to \mathbb{N}$ defined by

$$h(x) = \begin{cases} 0 & \text{if } x \text{ is a letter} \\ 1 + \max \{h(x_i) \mid 1 \leqslant i \leqslant n\} & \text{if } d(x) = (x_1, \ldots, x_n) \end{cases}$$

Thus $h(x)$ is equal to the length of the longest path with origin in x in the tree of x. Finally, the *height* of d is

$$h = \sup \{ \ h(x) \mid x \in A^+ \ \}$$

Let S be a finite semigroup and let $\varphi : A^+ \to S$ be a morphism. A factorization forest d is *Ramseyan modulo* φ if, for every word x of A^2A^*, $d(x)$ is either of degree 2 or there exists an idempotent e of S such that $d(x) = (x_1, \ldots, x_n)$ and $x_1\varphi = x_2\varphi = \ldots = x_n\varphi = e$ for $1 \leqslant i \leqslant n$. The following result is proved in [60, 61].

Theorem 4.1 *Let S be a finite semigroup and let $\varphi : A^+ \to S$ be a morphism. Then there exists a factorization forest of height $\leqslant 9|S|$ which is Ramseyan modulo φ.*

5 Polynomial closure

We now arrive to the main topic of this paper. We describe the counterpart, on varieties of finite ordered monoids, of the operation of polynomial closure on varieties of languages. The terminology *polynomial closure*, first introduced by Schützenberger, comes from the fact that rational languages form a semiring under union as addition and concatenation as multiplication. There are in fact two slightly different notions of polynomial closure, one for $+$-classes and one for $*$-classes.

The *polynomial closure* of a 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 $n \geqslant 0$, 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 $L_0a_1L_1 \cdots a_nL_n$, 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 such that, for every alphabet A, A^+Co-Pol $\mathcal{V}$ (resp. A^*Co-Pol $\mathcal{V}$) is the set of languages L whose complement is in A^+Pol $\mathcal{V}$ (resp. A^*Pol $\mathcal{V}$). Finally, we denote by BPol $\mathcal{V}$ the class of languages such that, for every alphabet A, A^+BPol $\mathcal{V}$ (resp. A^*BPol $\mathcal{V}$) is the closure of A^+Pol $\mathcal{V}$ (resp. A^*Pol $\mathcal{V}$) under finite boolean operations (finite union and complement).

We first establish a simple syntactic property of the concatenation product. For $1 \leqslant i \leqslant n$, let L_i be a recognizable language of A^+, let $\eta_i : A^+ \to S(L_i)$ be its syntactic morphism and let $\eta : A^+ \to S(L_1) \times S(L_2) \times \cdots \times S(L_n)$ be the morphism defined by $u\eta = (u\eta_1, u\eta_2, \ldots, u\eta_n)$. Let u_0, u_1, $\ldots$, u_n be words of A^* and let $L = u_0L_1u_1 \cdots L_nu_n$. Let $\mu : A^+ \to S(L)$ be the syntactic morphism of L. The properties of the relational morphism $\tau = \mu^{-1}\eta : S(L) \to S(L_1) \times S(L_2) \times \cdots \times S(L_n)$ were first studied by Straubing [66] and later by the first author [39]. The next proposition is a more precise version of these results.

Proposition 5.1 *For every idempotent e of $S(L_1) \times S(L_2) \times \cdots \times S(L_n)$, $e\tau^{-1}$ is an ordered semigroup that satisfies the inequality $x^\omega y x^\omega \leqslant x^\omega$.*

Proof Let $e = (e_1, e_2, \ldots, e_n)$ be an idempotent of $S(L_1) \times S(L_2) \times \cdots \times S(L_n)$, and let x and y be words in A^+ such that $x\eta = y\eta = e$. Let k be an integer greater than $n + |u_0 u_1 \cdots u_n|$ such that $x^k\mu$ is idempotent. It suffices to show that for every $u, v \in A^*$, $ux^kv \in L$ implies $ux^kyx^kv \in L$. Since $ux^kv \in L$, there exists a factorization of the form $ux^kv = u_0w_1u_1 \cdots w_nu_n$, where $w_i \in L_i$ for $0 \leqslant i \leqslant n$. By the choice of k, there exist $1 \leqslant h \leqslant n$ and $0 \leqslant j \leqslant k-1$ such that $w_h = w'_h x w''_h$ for some $w'_h, w''_h \in A^*$, $ux^j = u_0w_1 \cdots u_{h-1}w'_h$ and $x^{k-j-1}v = w''_h u_h \cdots w_n u_n$. Now since $x\eta_h = y\eta_h = x^2\eta_h$, the condition $w'_h x w''_h \in L_h$ implies $w'_h x^{k-j} y x^{j+1} w''_h \in L_h$. It follows $ux^kyx^kv \in L$, which concludes the proof. □

There is a similar result for syntactic monoids. Let, for $0 \leqslant i \leqslant n$, L_i be recognizable languages of A^*, let $\eta_i : A^* \to M(L_i)$ be their 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 $u\eta = (u\eta_0, u\eta_1, \ldots, u\eta_n)$. Let $a_1, a_2, \ldots, a_n$ be letters of A and let $L = L_0a_1L_1 \cdots a_nL_n$. Let $\mu : A^* \to M(L)$ be the syntactic morphism of L. Finally, consider the relational morphism $\tau = \mu^{-1}\eta : M(L) \to M(L_0) \times M(L_1) \times \cdots \times M(L_n)$.

Proposition 5.2 *For every idempotent e of $M(L_1) \times M(L_2) \times \cdots \times M(L_n)$, $e\tau^{-1}$ is an ordered semigroup that satisfies the inequality $x^\omega y x^\omega \leqslant x^\omega$.*

Proof Let $e = (e_1, e_2, \ldots, e_n)$ be an idempotent of $M(L_1) \times M(L_2) \times \cdots \times M(L_n)$, and let x and y be words in A^* such that $x\eta = y\eta = e$. Let k be a integer greater that n such that $x^k\mu$ is idempotent. It suffices to show that for every $u, v \in A^*$, $ux^kv \in L$ implies $ux^kyx^kv \in L$. Since $ux^kv \in L$, there exists a factorization of the form $ux^kv = w_0a_1w_1 \cdots a_nw_n$, where $w_i \in L_i$ for $0 \leqslant i \leqslant n$. By the choice of k, there exist $0 \leqslant h \leqslant n$ and $0 \leqslant j \leqslant k-1$ such that $w_h = w'_h x w''_h$ for some $w'_h, w''_h \in A^*$, $ux^j = w_0a_1 \cdots w_{h-1}a_hw'_h$ and $x^{k-j-1}v = w''_h a_{h+1} \cdots a_nw_n$. Now since $x\eta_h = y\eta_h = x^2\eta_h$, the condition $w'_h x w''_h \in L_h$ implies $w'_h x^{k-j} y x^{j+1} w''_h \in L_h$. It follows $ux^kyx^kv \in L$, which concludes the proof. □

There is a subtle difference between the proofs of Propositions 5.1 and 5.2 and that is the reason why Proposition 5.2 is not stated for products of the form $u_0L_1u_1 \cdots L_nu_n$. The difference occurs when x is the empty word in the proof of Proposition 5.2. In this case, if L was equal to $u_0L_1u_1 \cdots L_nu_n$, an occurrence of x^k would not define an occurrence of x in one of the w_i, since x could well occur in the middle of some u_i. But if the u_i's are letters, they do not contain the empty word as a proper factor.

Proposition 5.1 leads to the following result in terms of varieties.

Corollary 5.3 *Let* $\mathbf{V}$ *be a variety of finite semigroups and let* $\mathcal{V}$ *be the corresponding* $+$*-variety. If* $L \in A^+ Pol\,\mathcal{V}$*, then* $S(L)$ *belongs to the variety of finite ordered semigroups* $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$*.*

Proof Let $\mathbf{W} = [\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$ and let $\mathcal{W}$ be the positive variety corresponding to $\mathbf{W}$. By Theorem 3.7, it suffices to show that L belongs to $A^+\mathcal{W}$. Since $A^+\mathcal{W}$ is closed under finite union, it suffices to prove the result when L is equal to a marked product of the form $u_0 L_1 u_1 \cdots L_n u_n$, where $n \geqslant 0$ and, for $0 \leqslant h \leqslant n$, $u_h \in A^*$ and the L_h are languages in $A^+\mathcal{V}$. But in this case, Proposition 5.1 shows that $S(L) \in \mathbf{W}$. $\square$

Proposition 5.2 leads to an analogous result for $*$-varieties, whose proof is omitted.

Corollary 5.4 *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the corresponding* $*$*-variety. If* $L \in A^* Pol\,\mathcal{V}$*, then* $M(L)$ *belongs to the variety of finite ordered monoids* $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$*.*

We now establish the converse of Corollary 5.3.

Proposition 5.5 *Let* $\mathbf{V}$ *be a variety of finite semigroups and let* $\mathcal{V}$ *be the corresponding* $+$*-variety. Let* L *be a language of* A^+ *and let* $S(L)$ *be its syntactic ordered semigroup. If* $S(L) \in [\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$*, then* $L \in A^+ Pol\,\mathcal{V}$*.*

Proof Let $S = S(L)$ and let $\eta : A^+ \to S$ be the syntactic morphism of L. If $S(L) \in [\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$, there exists a semigroup $V \in \mathbf{V}$ and a relational morphism $\tau : S \to V$ such that, for every idempotent e of V, $e\tau^{-1}$ satisfies the identity $x^\omega y x^\omega \leqslant x^\omega$. Let R be the graph of τ and let $\alpha : R \to S$ and $\beta : R \to V$ be the natural projections. Then α is onto and $\tau = \alpha^{-1}\beta$. By the universal property of A^+, there exists a morphism $\delta : A^+ \to R$ such that $\eta = \delta\alpha$. Let $\mu = \delta\beta$. Then $\eta^{-1}\mu = \alpha^{-1}\delta^{-1}\delta\beta = \alpha^{-1}\beta = \tau$.

$$\begin{tikzcd}
 & A^+ \arrow[ddl, "\eta"'] \arrow[d, "\delta"] \arrow[ddr, "\mu"] & \\
 & R \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\
S \arrow[rr, "\tau"] & & V
\end{tikzcd}$$

Let $K = 2^{9|S||V|}$. We claim that

$$L = \bigcup u_0(e_1\mu^{-1})u_1(e_2\mu^{-1})u_2 \ \cdots \ (e_k\mu^{-1})u_k \qquad (1)$$

where the union is taken over the sequences $(e_1, e_2, \ldots, e_k)$ of idempotents of V such that $k \leqslant K$, $|u_0u_1u_2 \ \cdots \ u_k| \leqslant K$ and

$$u_0(e_1\mu^{-1})u_1(e_2\mu^{-1})u_2 \ \cdots \ (e_k\mu^{-1})u_k \subseteq L.$$

The right hand side of (1) is by construction a subset of L. We now establish the opposite inclusion. By Theorem 4.1, there exists a factorization forest d of height $\leqslant 9|S||V|$ which is Ramseyan modulo δ. We need the following technical lemma.

Lemma 5.6 *Let $x \in A^+$ such that $d(x) = (x_1, \ldots, x_n)$ with $n \geqslant 3$ and let (f, e) be an idempotent of $S \times V$ such that $x_1\delta = \ldots = x_n\delta = (f, e)$. Then, for all $u, v \in A^*$ such that $uxv \in L$, the language $ux_1(e\mu^{-1})x_nv$ is contained in L.*

Proof Since $x = x_1x_2 \ldots x_n$, it follows $x\mu = x_1\mu \cdots x_n\mu = e$ and thus the ordered semigroup $x\eta$ is contained in $e\tau^{-1}$ and satisfies the identity $x^\omega y x^\omega \leqslant x^\omega$. By hypothesis, $x_1, x_n \in e\mu^{-1}$ and hence $x_1\eta = x_n\eta = f \in e\tau^{-1}$. Let now $y \in e\mu^{-1}$. Then $y\eta \in e\tau^{-1}$ and hence $(x_1yx_n)\eta = f(y\eta)f \leqslant f = x\eta$. Therefore, if $u, v \in A^*$, one has

$$(ux_1yx_nv)\eta = u\eta(x_1yx_n)\eta v\eta \leqslant (uxv)\eta$$

Thus, $uxv \in L$ implies $ux_1yx_nv \in L$ since η is the syntactic morphism of L. Therefore $ux_1(e\mu^{-1})x_nv$ is contained in L. $\square$

Now, we associate with every word $x \in A^+$ a language $L(x)$ defined recursively as follows

$$L(x) = \begin{cases} \{x\} & \text{if } x \text{ is a letter} \\ L(x_1)L(x_2) & \text{if } d(x) = (x_1, x_2) \\ L(x_1)e\mu^{-1}L(x_n) & \text{if } d(x) = (x_1, \ldots, x_n) \text{ with } n \geqslant 3 \text{ and} \\ & x_1\delta = \ldots = x_n\delta = (f, e) \end{cases}$$

By induction, $x \in L(x)$ for all x and Lemma 5.6 shows that if $x \in L$, then $L(x)$ is contained in L. Finally, $L(x)$ is of the form

$$u_0(e_1\mu^{-1})u_1(e_2\mu^{-1})u_2 \ \cdots u_k(e_k\mu^{-1})u_{k+1} \tag{2}$$

for some idempotents $e_1, \ldots, e_k$ of V, $k \geqslant 0$, $u_0, u_{k+1} \in A^*$ and u_1, u_2, $\ldots$, u_k in A^+. Furthermore, one can give an upper bound to the length of $u_0u_1u_2 \ \cdots \ u_ku_{k+1}$. Indeed, this word can be obtained by reading the labels of the leaves of the subtree $t'(x)$ of $t(x)$ (see Section 4) obtained by considering the "external" branches only. The tree $t'(x)$ can be defined formally as follows.

$$t'(x) = \begin{cases} x & \text{if } x \text{ is a letter} \\ (x, (t'(x_1), t(x'_n))) & \text{if } d(x) = (x_1, \ldots, x_n) \end{cases}$$

Now $t(x)$ and $t'(x)$ have the same height, but $t'(x)$ is a binary tree. Therefore the number of leaves of $t'(x)$ is bounded by $2^{9|S||V|}$. It follows that $|u_0u_1u_2 \ \cdots \ u_ku_{k+1}| \leqslant K$ and $k \leqslant K$, since $u_1, u_2, \ldots, u_k \in A^+$. This proves formula 1. It follows that $L \in A^+\text{Pol } \mathcal{V}$ since every language $e_i\mu^{-1} \in A^+\mathcal{V}$. $\square$

In the case of $*$-varieties, the previous result also holds with the appropriate definition of polynomial closure.

Proposition 5.7 *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the corresponding* $*$*-variety. Let* L *be a language of* A^* *and let* $M(L)$ *be its syntactic ordered monoid. If* $M(L) \in [\![x^\omega y x^\omega \leqslant x^\omega]\!] \ \text{Ⓜ} \ \mathbf{V}$*, then* $L \in A^*\mathit{Pol}\ \mathcal{V}$.

Proof The beginning of the proof of Proposition 5.5 carries over with the modifications indicated below. Let $M = M(L)$ and let $\eta : A^* \to M$ be the syntactic morphism of L. One obtains as before the following diagram, where $V \in \mathbf{V}$.

$$
\begin{tikzcd}
 & A^* \arrow[ddl, "\eta"'] \arrow[d, "\delta"] \arrow[ddr, "\mu"] & \\
 & R \arrow[dl, "\alpha"] \arrow[dr, "\beta"'] & \\
M \arrow[rr, "\tau"] & & V
\end{tikzcd}
$$

Let $K = 2^{9|M||V|}$. We claim that

$$L = \bigcup (e_0\mu^{-1})a_1(e_1\mu^{-1})a_2 \ \cdots \ a_k(e_k\mu^{-1}) \tag{3}$$

where the union is taken over the sequences $(e_0, e_1, \ldots, e_k)$ of idempotents of V such that $k \leqslant K$ and $(e_0\mu^{-1})a_1(e_1\mu^{-1})a_2 \ \cdots \ a_k(e_k\mu^{-1}) \subseteq L$. Let L' be the right hand side of (3). Then $L' \subseteq L$ by construction. We now establish the opposite inclusion. The proof of Proposition 5.5 shows that, for every $x \in L$, there exists a language $L(x)$, containing x and contained in L, of the form

$$u_0(e_1\mu^{-1})u_1(e_2\mu^{-1})u_2 \ \cdots u_k(e_k\mu^{-1})u_{k+1} \tag{4}$$

where $e_1, \ldots, e_k$ are idempotents of V, $k \geqslant 0$, $u_0, u_{k+1} \in A^*$, $u_1, u_2, \ldots, u_k \in A^+$ and $|u_0u_1u_2 \ \cdots \ u_ku_{k+1}| \leqslant K$. Finally, one can pass from the decomposition given by Formula (4) to a decomposition of the form (3) by inserting languages of the form $1\mu^{-1}$ between the letters of u_0, u_1, $\ldots$, u_{k+1}. Indeed, $1\tau^{-1}$ is a monoid that satisfies $x^\omega y x^\omega \leqslant x^\omega$, and hence $y \leqslant 1$ for each $y \in 1\tau^{-1}$. It follows that, for all $u, v \in A^*$, $uv \in L$ implies $u(1\mu^{-1})v \subseteq L$. Therefore L is contained in L', concluding the proof. $\square$

By combining Corollary 5.4 and Proposition 5.7, we obtain our main result.

Theorem 5.8 *Let* $\mathbf{V}$ *be a variety of finite semigroups and let* $\mathcal{V}$ *be the corresponding* $+$*-variety. Then* $Pol\,\mathcal{V}$ *is a positive* $+$*-variety and the corresponding variety of finite semigroups is the Mal'cev product* $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$.

Theorem 5.9 *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the corresponding* $*$*-variety. Then* $Pol\,\mathcal{V}$ *is a positive* $*$*-variety and the corresponding variety of finite monoids is the Mal'cev product* $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$.

Theorems 5.8 and 5.9 lead to a new proof of the following result of Arfi [5, 6].

Corollary 5.10 *For each variety of languages* $\mathcal{V}$*,* $Pol\,\mathcal{V}$ *and* $Co\text{-}Pol\,\mathcal{V}$ *are positive varieties of languages. In particular, for each alphabet* A*,* $A^+ Pol\,\mathcal{V}$ *and* $A^+ Co\text{-}Pol\,\mathcal{V}$ *(resp.* $A^* Pol\,\mathcal{V}$ *and* $A^*\ Co\text{-}Pol\,\mathcal{V}$ *in the case of a* $*$*-variety) are closed under finite union and intersection.*

6 Schützenberger product

One of the most useful tools for studying the concatenation product is the *Schützenberger product* of n monoids, which was originally defined by Schützenberger for two monoids [57], and extended by Straubing [65] for any number of monoids.

Given a monoid M, denote by $\mathcal{P}(M)$ the monoid of subsets of M under the multiplication of subsets, defined, for all $X, Y \subseteq M$ by $XY = \{xy \mid x \in X \text{ and } y \in Y\}$. Then $\mathcal{P}(M)$ is not only a monoid but also a semiring under union as addition and the product of subsets as multiplication. Let $M_1, \ldots, M_n$ be monoids. Denote M the product monoid $M_1 \times \cdots \times M_n$ and $\mathcal{M}_n$ the semiring of square matrices of size n with entries in the semiring $\mathcal{P}(M)$. The *Schützenberger product* of $M_1, \ldots, M_n$, denoted $\diamondsuit_n(M_1, \ldots, M_n)$ is the submonoid of the multiplicative monoid $\mathcal{M}_n$ composed of all the matrices P satisfying the three following conditions:

(1) If $i > j$, $P_{i,j} = 0$

(2) If $1 \leqslant i \leqslant n$, $P_{i,i} = \{(1, \ldots, 1, s_i, 1, \ldots, 1)\}$ for some $s_i \in M_i$

(3) If $1 \leqslant i \leqslant j \leqslant n$, $P_{i,j} \subseteq 1 \times \cdots \times 1 \times M_i \times \cdots \times M_j \times 1 \cdots \times 1$.

Condition (1) shows that the matrices of the Schützenberger product are upper triangular, condition (2) enables us to identify the diagonal coefficient $P_{i,i}$ with an element s_i of M_i and condition (3) shows that if $i < j$, $P_{i,j}$ can be identified with a subset of $M_i \times \cdots \times M_j$. With this convention, a matrix of $\diamondsuit_3(M_1, M_2, M_3)$ 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 Schützenberger product $\diamondsuit_n(M_1, \ldots, M_n)$ is naturally equipped of an order $\leqslant$ defined by

$$P \leqslant P' \text{ if and only if, for } 1 \leqslant i, j \leqslant n,\ P'_{i,j} \subseteq P_{i,j}$$

One could also take the dual order (defined by $P_{i,j} \subseteq P'_{i,j}$), but this one is directly related to the polynomial closure, as we will see below. We first need to verify that this order is stable. Indeed, if $P \leqslant P'$ and if $Q, R \in \diamondsuit_n(M_1, \ldots, M_n)$, then, for $1 \leqslant i, j \leqslant n$,

$$(QP'R)_{i,j} = \sum_{r,s} Q_{i,r}P'_{r,s}R_{s,j} \subseteq \sum_{r,s} Q_{i,r}P_{r,s}R_{s,j} = (QPR)_{i,j}$$

and thus, $QPR \leqslant QP'R$.

The Schützenberger product is closely related to the polynomial closure. We first give a slightly more precise version of Straubing's original result [65].

Proposition 6.1 *Let L_0, ..., L_n be languages of A^* recognized by the monoids M_0, ..., M_n, respectively, and let $a_1, \ldots, a_n$ be letters of A. Then the language $L_0a_1L_1 \cdots a_nL_n$ is recognized by the ordered monoid*

$$(\diamondsuit_{n+1}(M_0, \ldots, M_n), \leqslant).$$

Proof Let, for $0 \leqslant i \leqslant n$, $\eta_i : A^* \to M_i$ be a monoid morphism recognizing L_i. Then there exist subsets R_i of M_i such that $L_i = R_i\eta_i^{-1}$. We let the reader verify that the map $\eta : A^* \to \diamondsuit_{n+1}(M_0, \ldots, M_n)$ defined, for each $u \in A^*$, by

$$(u\eta)_{i,j} = \{(1, \ldots, 1, u_i\eta_i, u_{i+1}\eta_{i+1}, \ldots, u_j\eta_j, 1, \ldots, 1) \mid u_ia_{i+1}u_{i+1} \cdots a_ju_j = u\}$$

is a monoid morphism. Let $N = A^*\eta$. Let Q be the subset of N formed by all matrices $P \in N$ such that for $P_{0,n} \cap R_0 \times \cdots \times R_n \neq \emptyset$. Then Q is an order ideal of N. Indeed, if $P \in Q$ and $P' \leqslant P$, then $P_{0,n} \subseteq P'_{0,n}$ and thus $P'_{0,n} \cap R_0 \times \cdots \times R_n \neq \emptyset$. Let $L = L_0a_1L_1 \cdots a_nL_n$. We claim that $L = Q\eta^{-1}$. Let u be a word of A^* such that $u\eta \in Q$. Then by definition, there exists $(s_0, s_1, \ldots, s_n) \in (u\eta)_{0,n} \cap R_0 \times \cdots \times R_n$. Let $u_0, \ldots, u_n$ be words such that $(u_0\eta, u_1\eta, \ldots, u_n\eta) \in (u\eta)_{0,n} \cap R_0 \times \cdots \times R_n$. Then $u_0 \in L_0$, ..., $u_n \in L_n$ and $u_0a_1u_1 \cdots a_nu_n = u$. Thus $u \in L$. Conversely, if $u \in L$, then $u = u_0 a_1 u_1 \cdots a_n u_n$ for some words $u_0 \in L_0, \ldots, u_n \in L_n$ and thus $(u_0\eta, u_1\eta, \ldots, u_n\eta) \in (u\eta)_{0,n} \cap R_0 \times \cdots \times R_n$, proving the claim. □

This result was extended to varieties by Reutenauer [53] for $n = 1$ and by the author [36] in the general case (see also [76] and Simon [62] for a simpler proof). Here we propose a slightly more precise version of this result. Given a variety of monoids $\mathbf{V}$, $\Diamond\mathbf{V}$ denotes the variety of ordered monoids generated by all Schützenberger products of the form $\Diamond_n(M_1, \ldots, M_n)$ with $n > 0$ and $M_1, \ldots, M_n \in \mathbf{V}$.

Theorem 6.2 *Let $\mathcal{V}$ be the $*$-variety corresponding to $\mathbf{V}$. Then the positive $*$-variety corresponding to $\Diamond\mathbf{V}$ is Pol $\mathcal{V}$.*

Proof The proof relies on the following stronger version of the main result of [39].

Lemma 6.3 *Let $\mathbf{V}$ be a variety of finite monoids and let $M_1, \ldots, M_n$ be monoids of $\mathbf{V}$. Then, for each finite alphabet A, the ordered monoid*

$$(\Diamond_n(M_1, \ldots, M_n), \leqslant)$$

satisfies the identity $x^\omega y x^\omega \leqslant x^\omega$, for each $x, y \in \hat{A}^*$ such that $\mathbf{V}$ satisfies $x = y = x^2$.

Proof Let $\eta : A^* \to \Diamond_n(M_1, \ldots, M_n)$ be a monoid morphism and let $P = x\eta$ and $Q = y\eta$. Since $x = y = x^2$ in $\mathbf{V}$, $P_{i,i} = P^2_{i,i} = Q_{i,i}$. Let ω be the exponent of $\Diamond_n(M_1, \ldots, M_n)$. We may assume that $\omega \geqslant n$. We claim that $P^\omega = P^{\omega+1}$. Indeed,

$$P^\omega_{i,j} = \sum P_{i_0,i_1} P_{i_1,i_2} \cdots P_{i_{\omega-1},i_\omega}$$

where the sum runs over all increasing sequences $i = i_0 \leqslant i_1 \leqslant \ldots \leqslant i_\omega = j$. Now, since $\omega \geqslant n$, there exists in each such sequence an index j such that $i_j = i_{j+1}$. Thus $P_{i_j,i_{j+1}}$ is a diagonal entry and is equal to its square. Therefore one can replace $P_{i_j,i_{j+1}}$ by $P^2_{i_j,i_{j+1}}$ in the product $P_{i_0,i_1} P_{i_1,i_2} \cdots P_{i_{\omega-1},i_\omega}$. It follows that

$$P^\omega_{i,j} = \sum P_{i_0,i_1} P_{i_1,i_2} \cdots P_{i_{\omega-1},i_\omega} \subseteq \sum P_{i_0,i_1} P_{i_1,i_2} \cdots P_{i_\omega,i_{\omega+1}} = P^{\omega+1}_{i,j}$$

Thus $P^\omega \geqslant P^{\omega+1}$ and by induction

$$P^\omega \geqslant P^{\omega+1} \geqslant P^{\omega+2} \geqslant \ldots \geqslant P^{2\omega} = P^\omega$$

which proves the claim. Now,

$$(P^\omega Q P^\omega)_{i,j} = \sum P_{i_0,i_1} P_{i_1,i_2} \cdots P_{i_{\omega-1},i_\omega} Q_{i_\omega,j_0} P_{j_0,j_1} P_{j_1,j_2} \cdots P_{j_{\omega-1},j_\omega}$$

where the sum runs over all increasing sequences $i = i_0 \leqslant i_1 \leqslant \ldots \leqslant i_\omega \leqslant j_0 \leqslant \ldots \leqslant j_\omega = j$. It follows that

$$\begin{aligned}
(P^\omega Q P^\omega)_{i,j} &\supseteq \sum_{i_\omega = j_0} P_{i_0,i_1} P_{i_1,i_2} \cdots P_{i_{\omega-1},i_\omega} Q_{i_\omega,j_0} P_{j_0,j_1} P_{j_1,j_2} \cdots P_{j_{\omega-1},j_\omega} \\
&= \sum_{i_\omega = j_0} P_{i_0,i_1} P_{i_1,i_2} \cdots P_{i_{\omega-1},i_\omega} P_{i_\omega,j_0} P_{j_0,j_1} P_{j_1,j_2} \cdots P_{j_{\omega-1},j_\omega} \\
&= \sum_k P^\omega_{i,k} P_{k,k} P^\omega_{k,j}
\end{aligned}$$

We claim that this latter product is equal to $P^\omega_{i,j}$. Indeed, for every increasing sequence of $3\omega + 1$ indices $i = i_0 \leqslant i_1 \leqslant \ldots \leqslant i_{3\omega} = j$, there exists an index r such that $\omega \leqslant r < 2\omega$ and $i_r = i_{r+1} = k$. It follows that

$$\begin{aligned}
(P^\omega)_{i,j} &= (P^{3\omega})_{i,j} \\
&\subseteq \sum_k \big((P^\omega)_{i,k} + \ldots + (P^{2\omega-1})_{i,k}\big) P_{k,k} \big((P^\omega)_{k,j} + \ldots + (P^{2\omega-1})_{k,j}\big) \\
&= \sum_k (P^\omega)_{i,k} P_{k,k} (P^\omega)_{k,j} \qquad \text{(since } P^\omega = P^{\omega+1}\text{)} \\
&\subseteq (P^{2\omega+1})_{i,j} = (P^\omega)_{i,j}
\end{aligned}$$

proving the claim. Thus $P^\omega Q P^\omega \leqslant P^\omega$. □

One can now complete the proof of Theorem 6.2. Proposition 6.1 shows that all the languages of Pol $\mathcal{V}$ are recognized by ordered monoids of $\Diamond\mathbf{V}$. Conversely, Lemma 6.3 shows that $\Diamond\mathbf{V}$ is contained in $[\![x^\omega y x^\omega \leqslant x^\omega]\!] ⓜ \mathbf{V}$. Therefore, by Theorem 5.9, the positive $*$-variety corresponding to $\Diamond\mathbf{V}$ contains Pol $\mathcal{V}$. □

Corollary 6.4 *For any variety of finite monoids* $\mathbf{V}$, $\Diamond\mathbf{V} = [\![x^\omega y x^\omega \leqslant x^\omega]\!] ⓜ \mathbf{V}$.

Proof This follows from Theorems 3.7, 5.9 and 6.2. □

Note that the results of this section only hold for positive $*$-varieties and varieties of finite monoids.

7 Unambiguous polynomial closure

The marked product $L = u_0 L_1 u_1 \cdots L_n u_n$ of n languages $L_1, \ldots, L_n$ of A^+ is *unambiguous* if every word u of L admits a unique factorization of the form $u_0 v_1 u_1 \cdots v_n u_n$ 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^+ is the set of languages that are finite disjoint unions of unambiguous products of the form $u_0 L_1 u_1 \cdots L_n u_n$, where the u_i's are words and the L_i's are elements of $\mathcal{L}$.

The marked product $L = L_0 a_1 L_1 \cdots a_n L_n$ of n languages $L_0, L_1, \ldots, L_n$ of A^* is *unambiguous* if every word u of L admits a unique factorization of the form $u_0 a_1 u_1 \cdots a_n u_n$ with $u_0 \in L_0$, $u_1 \in L_1, \ldots, u_n \in L_n$.

The *unambiguous polynomial closure* of a class of languages $\mathcal{L}$ of A^* is the set of languages that are finite disjoint unions of unambiguous products 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 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 [35, 49] as a generalization of an earlier result of Schützenberger [58].

Theorem 7.1 *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the corresponding* $*$*-variety. Then UPol* $\mathcal{V}$ *is a variety of languages, and the associated variety of monoids is* $\mathbf{LI} \circledM \mathbf{V}$.

A similar result holds for varieties of finite semigroups, although this result is not explicitely stated in [49].

Theorem 7.2 *Let* $\mathbf{V}$ *be a variety of finite semigroups and let* $\mathcal{V}$ *be the corresponding* $+$*-variety. Then UPol* $\mathcal{V}$ *is a variety of languages, and the associated variety of semigroups is* $\mathbf{LI} \circledM \mathbf{V}$.

Here is another important characterization of UPol $\mathcal{V}$, which holds for $*$-varieties as well as for $+$-varieties.

Theorem 7.3 *Let* $\mathcal{V}$ *be a variety of languages. Then Pol* $\mathcal{V} \cap$ *Co-Pol* $\mathcal{V} =$ *UPol* $\mathcal{V}$.

Proof We give the proof for $*$-varieties, but the proof would be similar for $+$-varieties. By definition, A^*UPol $\mathcal{V}$ is contained in A^*Pol $\mathcal{V}$. Moreover, by Theorem 7.1, A^*UPol $\mathcal{V}$ is a variety of languages, and hence is closed under complementation. Therefore $A^*\text{UPol}\ \mathcal{V}$ is also contained in $A^*\text{Co-Pol}\ \mathcal{V}$, which proves the inclusion

$$A^*\text{UPol}\ \mathcal{V} \subseteq A^*\text{Pol}\ \mathcal{V} \cap A^*\text{Co-Pol}\ \mathcal{V}$$

Conversely, let L be a language of $A^*\text{Pol}\ \mathcal{V} \cap A^*\text{Co-Pol}\ \mathcal{V}$. By Corollary 5.4, the ordered syntactic monoid $M(L)$ of L belongs to the variety of finite ordered monoids $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$. The identities defining this variety are given in Corollary 2.5. Let B be a finite alphabet, and let x, y be elements of $\hat{B}^*$ such that $\mathbf{V}$ satisfies $x = y = x^\omega$. Then $M(L)$ satisfies $x^\omega y x^\omega \leqslant x^\omega$. Now since $L \in A^*\text{Co-Pol}\ \mathcal{V}$, the complement of L belongs to $A^*\text{Pol}\ \mathcal{V}$ and thus by Proposition 3.3 and Theorem 3.8, $M(L)$ satisfies $x^\omega \leqslant x^\omega y x^\omega$. It follows that $M(L)$ satisfies $x^\omega = x^\omega y x^\omega$. Thus, by Corollary 2.4, $M(L) \in \mathbf{LI} \circledM \mathbf{V}$ and, by Theorem 7.1, $L \in \text{UPol}\ \mathcal{V}$, which concludes the proof. $\square$

Corollary 7.4 *If $\mathcal{V}$ is a variety of languages, then so is Pol $\mathcal{V} \cap$ Co-Pol $\mathcal{V}$.*

8 Concatenation hierarchies

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

(1) level 0 is $\mathcal{V}$
(2) for every integer $n \geqslant 0$, level $n + 1/2$ is the polynomial closure of level n
(3) for every integer $n \geqslant 0$, level $n + 1$ is the boolean closure of level $n + 1/2$.

Theorems 5.8 and 5.9 show that the polynomial closure of a variety of languages is a positive variety of languages and Proposition 3.9 shows that the boolean closure of a positive variety of languages is a varietyof languages. That is, one defines a sequence of varieties $\mathcal{V}_n$ and of positive varieties $\mathcal{V}_{n+1/2}$, where n is an integer, as follows:

(1) $\mathcal{V}_0 = \mathcal{V}$
(2) for every integer $n \geqslant 0$, $\mathcal{V}_{n+1/2} = \text{Pol}\ \mathcal{V}_n$,
(3) for every integer $n \geqslant 0$, $\mathcal{V}_{n+1} = \text{BPol}\ \mathcal{V}_n$.

The corresponding varieties of semigroups and ordered semigroups (resp. monoids and ordered monoids) are denoted $\mathbf{V}_n$ and $\mathbf{V}_{n+1/2}$. Theorems 5.8 and 5.9 give an explicit relation between $\mathbf{V}_n$ and $\mathbf{V}_{n+1/2}$.

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

Three concatenation hierarchies have been considered so far in the literature. The simplest one is the hierarchy of positive $*$-varieties whose basis is the trivial variety. It was first considered by Thérien (implicitly in [71]) and Straubing (explicitly in [67]) and it is called the *Straubing hierarchy*. The hierarchy of positive $+$-varieties whose basis is the trivial variety is the *dot-depth hierarchy*, introduced by Brzozowksi, and it was the first to be studied [17].[4] The third hierarchy to be considered [31] is the hierarchy of positive $*$-varieties whose basis is the variety of group-languages. For the sake of simplicity, we will call it the *group hierarchy*.

The original work of Brzozowski and Knast [18] shows that these three hierarchies are strict: if A contains at least two letters, then for every n, there exist languages of level $n+1$ which are not of level $n+1/2$ and languages of level $n+1/2$ which are not of level n.

The main question is the decidability of each level: given a level n (resp. $n+1/2$) and a recognizable language L, decide whether or not L has level n (resp. $n+1/2$). The language can be given either by a finite automaton, a finite semigroup or a rational expression since there are standard algorithms to pass from one representation to the other. We now describe in more details the first levels of each of these hierarchies. We consider the Straubing hierarchy, the dot-depth hierarchy and the group hierarchy in this order.

8.1 Straubing's hierarchy

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

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

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

By definition, the sets of level 1/2 are the finite unions of languages of the form $A^*a_1A^*a_2 \cdots a_kA^*$, where the a_i's are letters. An alternative description can be given in terms of another operation on languages, the *shuffle*. Recall that the *shuffle* of two words u and v is the set $u \sqcup\!\sqcup v$ of all words w such that $w = u_1v_1u_2v_2 \cdots u_nv_n$ with $u_1, v_1, u_2, v_2, \ldots, u_n, v_n \in A^*$,

[4] Actually, the basis of the original dot-depth hierarchy was the variety of finite or cofinite languages. But this modification does not change the other levels.

$u_1u_2\cdots u_n = u$ and $v_1v_2\cdots v_n = v$. Now a language is a *shuffle ideal* if and only if for every $u \in L$ and $v \in A^*$, $u \mathbin{\text{Ш}} v$ is contained in L.

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

Proof By a well known theorem of Higman (cf [30], chapter 6), every shuffle ideal is a finite union of languages of the form $A^*a_1A^*a_2\cdots a_kA^*$, where the a_i's are letters. Conversely, the languages of this form are clearly shuffle ideals. $\square$

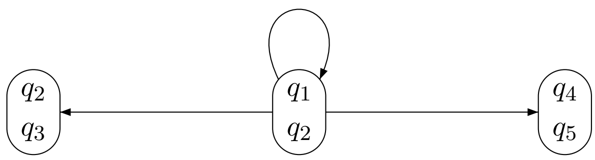
It follows in particular that if a language of A^* and its complement are shuffle ideals, then $L = A^*$ or $L = \emptyset$. It is easy to see directly that level $1/2$ is decidable (see Arfi [5, 6]). One can also derive this result from our syntactic characterization.

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

We shall derive from this result a polynomial algorithm to decide whether the language accepted by a complete deterministic n-state automaton is of level $1/2$. This algorithm, as well as the other algorithms presented in this section, rely on the notion of graph. Recall that a *graph* is a pair $G = (E, V)$, where E is the set of *edges* and $V \subseteq E \times E$ is the set of *vertices*. A *subgraph* of G is a graph $G' = (E', V')$ such that $E' \subseteq E$.

Let $\mathcal{A} = (Q, A, \cdot, i, F)$ be an n-state complete deterministic automaton and let $\mathcal{C} = \mathcal{A} \times \mathcal{A}$. Thus $\mathcal{C} = (Q \times Q, A, \cdot, (i, i), F \times F)$ and the transition function is defined by $(q, q') \cdot a = (q \cdot a, q' \cdot a)$. Let $G_2(\mathcal{A})$ be the graph whose vertices are the states of $\mathcal{C}$ and the edges are the pairs $((q_1, q_2), (q_1', q_2'))$ such that there is a word $u \in A^*$ such that $q_1 \cdot u = q_1'$ and $q_2 \cdot u = q_2'$ in $\mathcal{A}$. In other words, $G_2(\mathcal{A})$ is the reflexive and transitive closure of the graph of $\mathcal{C}$.

Theorem 8.5 *Let* $\mathcal{A} = (Q, A, E, i, F)$ *be a complete deterministic automaton recognizing a language* L. *Then* L *is of level* $1/2$ *if for every subgraph of* $G_2(\mathcal{A})$ *of the form*



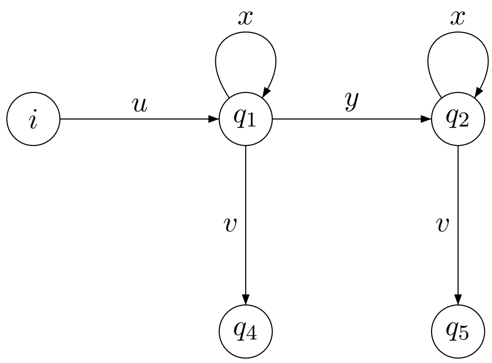
where the q_i*'s are states of* $\mathcal{A}$, *the condition* $q_4 \in F$ *implies* $q_5 \in F$. *If* $\mathcal{A}$ *is minimal, this condition is also sufficient.*

Proof Let ω be the exponent of the transition monoid of $\mathcal{A}$. Suppose that $\mathcal{A}$ satisfies the condition stated in the theorem. Let x, y, u and v be words such that $ux^\omega v \in L$. Set $q_1 = i \cdot ux^\omega$, $q_2 = q_1 \cdot yx^\omega$, $q_3 = q_2 \cdot yx^\omega$, $q_4 = q_1 \cdot v$ and $q_5 = q_2 \cdot v$. Then $q_1 \cdot x^\omega = q_1$ and $q_2 \cdot x^\omega = q_2$. It follows that, in $\mathcal{C}$,

$$\begin{aligned}(q_1, q_2) \cdot x^\omega &= (q_1, q_2)\\ (q_1, q_2) \cdot yx^\omega &= (q_2, q_3)\\ (q_1, q_2) \cdot v &= (q_4, q_5)\end{aligned}$$

Thus we have found a subgraph of the required type. Now $q_4 = i \cdot ux^\omega v$ and hence $q_4 \in F$. Therefore $q_5 \in F$ and since $q_5 = q_2 \cdot v = i \cdot ux^\omega yx^\omega v$, it follows $ux^\omega yx^\omega v \in L$. Thus the syntactic ordered monoid of L satisfies the identity $x^\omega yx^\omega \leqslant x^\omega$.

Conversely, suppose $\mathcal{A}$ is minimal and that the syntactic ordered monoid $(M, \leqslant)$ of L satisfies the identity $x^\omega yx^\omega \leqslant x^\omega$. If one has a subgraph of the type above, there exist three words x, y and u such that $q_1 \cdot x = q_1$, $q_2 \cdot x = q_2$, $q_1 \cdot y = q_2$, $q_2 \cdot y = q_3$, $q_1 \cdot v = q_4$ and $q_2 \cdot v = q_5$. Let u be a word such that $i \cdot u = q_1$.



Then $i \cdot ux^\omega v = q_4$. Therefore, if $q_4 \in F$, $ux^\omega v \in L$ and thus $ux^\omega yx^\omega v \in L$. But $i \cdot ux^\omega yx^\omega v = q_5$ and thus $q_5 \in F$. □

Corollary 8.6 *One can decide in polynomial time whether the language accepted by a deterministic n-state automaton is of level* $1/2$.

Proof First, one can minimize a given deterministic automaton in polynomial time and thus we may assume that $\mathcal{A}$ is minimal. Now $\mathcal{C}$ has n^2 states and therefore $G_2(\mathcal{A})$ can be computed in polynomial time. The condition of Theorem 8.5 can then be tested in polynomial time also. □

The sets of level 1 are the finite boolean combinations of languages of level 1/2. In particular, all finite sets are of level 1. The sets of level 1 have a nice algebraic characterization [59], which yields a polynomial time algorithm to decide whether the language accepted by a deterministic n-state automaton is of level 1 [63]. See also [2, 59, 69, 63, 45, 42] for more details on these results.

It is shown in Arfi [5, 6] that the sets of level 3/2 of A^* are the finite unions of sets of the form $A_0^* a_1 A_1^* a_2 \cdots a_k A_k^*$, where the a_i's are letters and the A_i's are subsets of A.

We derive the following syntactic characterization. Denote by $\mathcal{J}_1$ and $\mathcal{DA}$, respectively, the variety of languages corresponding to the variety of monoids $\mathbf{J_1}$ (idempotent and commutative monoids) and $\mathbf{DA}$. Recall that $A^*\mathcal{J}_1$ is the boolean algebra generated by the languages of the form B^*, where $B \subseteq A$ ([38], page 40) and that $A^*\mathcal{DA}$ is the smallest class of languages of A^* containing languages of the form B^*, with $B \subseteq A$, and closed under disjoint union and unambiguous product ([38], page 110).

Theorem 8.7 *A language is of level* 3/2 *if and only if its ordered syntactic monoid satisfies the identity* $x^\omega y x^\omega \leqslant x^\omega$ *for every* x, y *such that* $c(x) = c(y)$.

Proof From the obvious inclusions $\mathbf{J_1} \subset \mathbf{J} \subset \mathbf{DA}$, it follows, for each alphabet A, the inclusions

$$\text{Pol } \{B^* \mid B \subseteq A\} \subseteq \text{Pol } A^*\mathcal{J}_1 \subseteq \text{Pol } A^*\mathcal{J} \subseteq \text{Pol } A^*\mathcal{DA}$$

Now since $A^*\mathcal{J}$ is the set of languages of level 1, Pol $A^*\mathcal{J}$ is the set of languages of level 3/2. On the other hand the description of $A^*\mathcal{DA}$ recalled above shows that Pol $A^*\mathcal{DA} =$ Pol $\{B^* \mid B \subseteq A\}$. It follows that Pol $\mathcal{J}_1 =$ Pol $\mathcal{J}$ and thus, by Theorem 5.9 the variety corresponding to the languages of level 3/2 is $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{J_1}$. Now, by Theorem 2.3, the identities of this variety are precisely $x^\omega y x^\omega \leqslant x^\omega$ for every x, y such that $c(x) = c(y)$. $\square$

Relying on a difficult result of Hashiguchi, Arfi [5, 6] proved that level 3/2 is also decidable.

Theorem 8.8 (Arfi [5, 6]) *One can effectively decide whether a given recognizable set of* A^* *is of level* 3/2.

The complexity of this algorithm was never explicitly evaluated but was certainly exponential, due to the huge bounds occurring in the proof of Hashiguchi's result. We give below a much more reasonable algorithm, which is a modification of the algorithm presented for the level 1/2.

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 $G'(\mathcal{A})$ be the reflexive and transitive closure of the graph of $\mathcal{C}$

Theorem 8.9 *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 subgraph of $G'(\mathcal{A})$ of the form*

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

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.

Proof Let ω be the exponent of the transition monoid of $\mathcal{A}$. Suppose that $\mathcal{A}$ satisfies the condition stated in the theorem. Let x and y be words with the same content B and let $u, v \in A^*$ be such that $ux^\omega v \in L$. Set $q_1 = i \cdot ux^\omega$, $q_2 = q_1 \cdot yx^\omega$, $q_3 = q_2 \cdot yx^\omega$, $q_4 = q_1 v$, $q_5 = q_2 \cdot v$ and $B' = B \cup c(v)$. Then $q_1 \cdot x^\omega = q_1$ and $q_2 \cdot x^\omega = q_2$. It follows that, in $\mathcal{C}$,

$$\begin{aligned} (\emptyset, q_1, q_2) \cdot x^\omega &= (B, q_1, q_2) \\ (\emptyset, q_1, q_2) \cdot yx^\omega &= (B, q_2, q_3) \\ (B, q_1, q_2) \cdot v &= (B', q_4, q_5) \end{aligned}$$

Consequently, we have found a subgraph of the required type. Now $q_4 = i \cdot ux^\omega v$ and thus $q_4 \in F$. Therefore $q_5 \in F$ and since $q_5 = q_2 \cdot v = i \cdot ux^\omega yx^\omega v$, it follows $ux^\omega yx^\omega v \in L$. It follows that the syntactic ordered monoid of L satisfies the identity $x^\omega yx^\omega \leqslant x^\omega$ for all words such that $c(x) = c(y)$. Thus, by Theorem 8.7, L is of level 3/2.

Conversely, assume that $\mathcal{A}$ is minimal, and let $(M, \leqslant)$ be the syntactic ordered monoid of L. Suppose that M satisfies the identity $x^\omega yx^\omega \leqslant x^\omega$ for $c(x) = c(y)$. If one has a subgraph of the type above, there exist words x, y and v such that $c(x) = c(y) = B$, $q_1 \cdot x = q_1$, $q_2 \cdot x = q_2$, $q_1 \cdot y = q_2$,

$q_1 \cdot v = q_4$ and $q_2 \cdot v = q_5$. Let u be a word such that $i \cdot u = q_1$. Then $i \cdot ux^\omega v = q_4$. Therefore, if $q_4 \in F$, $ux^\omega v \in L$ and thus $ux^\omega yx^\omega v \in L$. Now $i \cdot ux^\omega yx^\omega v = q_1 \cdot yx^\omega v = q_2 \cdot x^\omega v = q_2 \cdot v = q_5$. Therefore $q_5 \in F$. $\square$

Corollary 8.10 *There is a 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* $3/2$.

Proof First, one can minimize a given deterministic automaton in polynomial time and thus we may assume that $\mathcal{A}$ is minimal. Now $\mathcal{C}$ has $2^{|A|}n^2$ states and thus $G'(\mathcal{A})$ can be computed in time polynomial in $2^{|A|}n$. The condition of Theorem 8.9 can then be tested in polynomial time also. $\square$

We arrive now to the level 2. It is shown in [48] that the languages of level 2 of A^* are the finite boolean combinations of the languages of the form $A_0^* a_1 A_1^* a_2 \cdots a_k A_k^*$, where the a_i's are letters and the A_i's are subsets of A. Let $\mathbf{V}_2$ be the variety of finite monoids corresponding to the languages of level 2. A non-trivial (although non effective) characterization of $\mathbf{V}_2$ was also given in [48]. Given a variety of monoids $\mathbf{V}$, denote by $\mathbf{PV}$ the variety generated by all monoids of the form $\mathcal{P}(M)$, where $M \in \mathbf{V}$. Then $\mathbf{V}_2 = \mathbf{PJ}$. Unfortunately, no algorithm is known to decide whether a finite monoid divides the power monoid of a $\mathcal{J}$-trivial monoid. In other words, the decidability problem for level 2 is still open, although much progress has been made in recent years [11, 14, 20, 48, 68, 70, 74, 77]. This problem is actually a particular case of a more general question discussed in Section 9.

In the case of languages whose syntactic monoid is an inverse monoid, a complete characterization was given by Cowan [20], completing partial results of Straubing and the second author [70, 73, 74, 77]. We give here a much shorter proof of Cowan's result. It is shown in [74] (Section 3) and [77] (Proposition 5.2) that the membership problem in $\mathbf{V}_2$ for inverse monoids reduces to deciding whether the transition monoid of a so-called inverse automaton lies in $\mathbf{V}_2$. An *inverse automaton* is an automaton $\mathcal{A} = (Q, A \cup \bar{A}, i, F)$ over a symmetrized alphabet $A \cup \bar{A}$, which is deterministic and co-deterministic and which satisfies, for all $a \in A$, $q, q' \in Q$

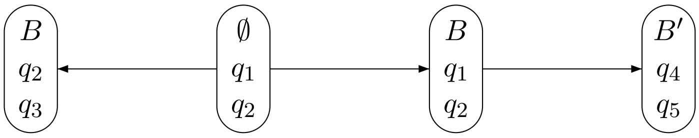
$$q \cdot a = q' \text{ if and only if } q' \cdot \bar{a} = q$$

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

Theorem 8.11 (Cowan) *The language recognized by an inverse automaton* $\mathcal{A} = (Q, A \cup \bar{A}, i, F)$ *is of level* 2 *in the Straubing hierarchy if and only if, for all* $q, q' \in Q$, $u, v \in (A \cup \bar{A})^*$, *such that* $q \cdot u$ *and* $q' \cdot u$ *are defined,* $q \cdot v = q'$ *and* $c(v) \subseteq c(u)$ *imply* $q = q'$.

Proof The necessary condition satisfied by the inverse automata recognizing a language of level 2 is proved in [74]. We now prove that this condition is sufficient. Let L be the language recognized by $\mathcal{A}$. First assume that $\mathcal{A}$ is complete. Then in view of the hypothesis, $\mathcal{A}$ has only one state, and L is either the empty set or equal to $(A \cup \bar{A})^*$, which are both languages of level 0. We now assume that $\mathcal{A}$ is not complete. The completion $\mathcal{A}'$ of $\mathcal{A}$ is the automaton $\mathcal{A}' = (Q \cup \{0\}, A \cup \bar{A}, i, F)$ obtained from $\mathcal{A}$ by adding a new state 0 ($0 \notin Q$) and by completing the transitions by setting $q \cdot a = 0$ if $q \cdot a$ was not defined in $\mathcal{A}$. The automaton $\mathcal{A}'$ recognizes L.

Let $\mathcal{A}_1$ (resp. $\mathcal{A}_2$) be the automaton obtained from $\mathcal{A}'$ by choosing $F_1 = F \cup \{0\}$ (resp. $F_2 = \{0\}$) as set of final states. Let L_1 and L_2 be the languages recognized by $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively. Then $L = L_1 \setminus L_2$ by construction. We claim that L_1 and L_2 are of level 3/2. Using the notation of Theorem 8.9, we consider a subgraph of $G'(\mathcal{A}')$ of the form



with $q_4 \in F_1$ (resp. $q_4 \in F_2$). There exist words $x, y, t \in (A \cup \bar{A})^*$ such that $c(x) = c(y) = B$, $q_1 \cdot x = q_1$, $q_2 \cdot x = q_2$, $q_1 \cdot y = q_2$, $q_1 \cdot t = q_4$ and $q_2 \cdot t = q_5$. First suppose that $q_2 \neq 0$. Then the paths $q_1 \xrightarrow{x} q_1$, $q_2 \xrightarrow{x} q_2$ and $q_1 \xrightarrow{y} q_2$ never visit state 0 in $\mathcal{A}'$. It follows from the hypothesis on $\mathcal{A}$ that $q_1 = q_2$ and hence $q_5 = q_4 \in F_1$ (resp. F_2).

Now if $q_2 = 0$, $q_5 = q_2 \cdot t = 0 \in F_1 \cap F_2$. In both cases, the condition of Theorem 8.9 are fulfilled, proving the claim. It follows that L is of level 2. □

Example 8.1 Let $A = \{a, b\}$ and $L = (ab)^*$. Its minimal automaton is represented below:

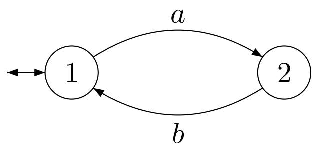


Figure 8.2: The minimal automaton of $(ab)^*$.

It satisfies the conditions of Theorem 8.11. In fact, by observing that $\emptyset^* = \{1\}$, L can be written in the form

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

It is interesting to remark that we have actually proved a little more than Cowan's theorem: each language recognized by an inverse automaton $\mathcal{A}$ is the difference of two languages of level 3/2 recognized by the completion of $\mathcal{A}$. It is proved in [74, 77] that Theorem 8.11 yields the following important corollary.

Corollary 8.12 *It is decidable whether an inverse monoid belongs to* $\mathbf{V}_2$.

Little is known beyond level 2: a semigroup theoretic description of each level of the hierarchy is known [36], but it is not an effective one. In other words, each level admits a description by identities, but these identities are not known for $n \geqslant 2$. Furthermore, even if these identities were known, this would not necessarily lead to a decision process for the corresponding variety. See also the conjecture discussed in Section 9.

8.2 Dot-depth hierarchy

The level 0 is the trivial +-variety. Therefore a language of A^+ is of level 0 if and only if it is empty or equal to A^+. As in the case of the Straubing hierarchy, one has the following easy characterization.

Proposition 8.13 *A language is of level* 0 *if and only if its syntactic semigroup is trivial.*

Therefore, one can decide in polynomial time whether the language of A^+ accepted by a deterministic n-state automaton is of level 0.

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 $k \geqslant 0$ and $u_0, \dots, u_k \in A^*$. But since $A^* = A^+ \cup \{1\}$, these languages can also be expressed as finite unions of languages of the form

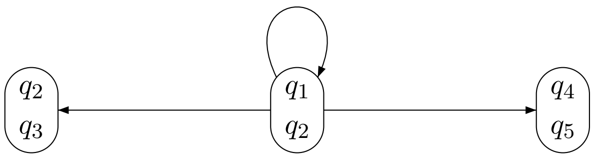
$$u_0A^*u_1A^* \cdots u_{k-1}A^*u_k$$

The syntactic characterization is a simple application of our main result.

Proposition 8.14 *A language of A^+ is of dot-depth $1/2$ if and only if its ordered syntactic semigroup satisfies the identity $x^\omega y x^\omega \leqslant x^\omega$.*

We can now mimic the algorithm given in the case of the Straubing hierarchy to decide whether the language accepted by a deterministic n-state automaton is of level $1/2$. The only difference is that empty paths are not allowed. In other words, instead of considering the reflexive and transitive closure of the graph of $\mathcal{C}$, one considers its transitive closure $G'_2(\mathcal{A})$. Nevertheless, the conclusion is the same.

Theorem 8.15 *Let $\mathcal{A} = (Q, A, E, \{i\}, F)$ be a complete automaton recognizing a language L. Then L is of dot-depth $1/2$ if for every subgraph of $G'_2(\mathcal{A})$ of the form*

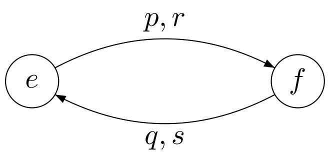


where the q_i's are states of $\mathcal{A}$, the condition $q_4 \in F$ implies $q_5 \in F$. The condition is also necessary if $\mathcal{A}$ is minimal.

Corollary 8.16 *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 finite 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 8.17 (Knast [27, 28]) *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 $(pq)^\omega ps(rs)^\omega = (pq)^\omega(rs)^\omega$.*



The variety of finite semigroups satisfying Knast's condition is usually denoted $\mathbf{B_1}$ ($\mathbf{B}$ refers to Brzozowski and 1 to level 1). Thus $\mathbf{B_1}$ is defined by the identities

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

The corresponding algorithm was analyzed by Stern [63]. One can decide in polynomial time whether the language accepted by a deterministic n-state automaton is of dot-depth 1.

It is not known yet whether level 3/2 of the dot-depth hierarchy is decidable.

8.3 The group hierarchy

We consider in this section the concatenation hierarchy based on the group languages, or group hierarchy. A part of the results of this section was presented in [41] in a slightly different form. By definition, a language of A^* is of level 0 in this hierarchy if and only if its syntactic monoid is a finite group. This can be easily checked on any deterministic automaton recognizing the language.

Proposition 8.18 *One can decide in polynomial time whether the language accepted by a deterministic n-state automaton is a group language.*

Proof It suffices to check whether the minimal automaton of the given language is a permutation automaton. $\square$

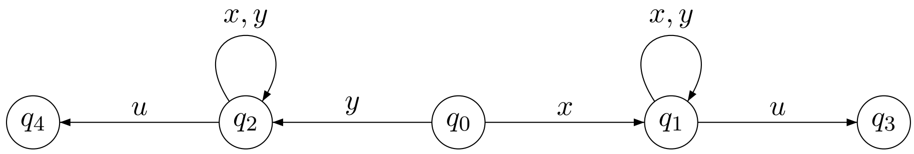
Level 1/2 is studied in detail in [41, 46]. By definition, the languages of level 1/2 are finite union of languages of the form $L_0a_1L_1 \cdots a_kL_k$ where the a_i's are letters and the L_i's are group languages. By Theorem 5.9, a language is of level 1/2 if and only if its ordered syntactic monoid belongs to the variety $[\![x^\omega yx^\omega \leqslant x^\omega]\!] \circledM \mathbf{G}$, which can be defined by the identity $x^\omega \leqslant 1$. This yields a polynomial time algorithm to check whether the language accepted by a deterministic n-state automaton is of level 1/2 in the group hierarchy.

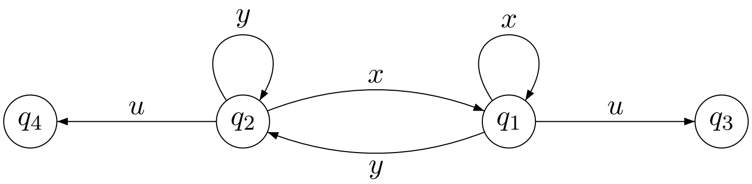
The study of the languages of level 1 in the group hierarchy started in 1985 [31] and was completed in [25] (see also [24]). A few more definitions are in order to state the algebraic characterization of this class of languages. A *block group* is a monoid such that every $\mathcal{R}$-class (resp. $\mathcal{L}$-class) contains at most one idempotent. Block groups form a variety of monoids, denoted $\mathbf{BG}$, and defined by the identity $(x^\omega y^\omega)^\omega = (y^\omega x^\omega)^\omega$. Thus $\mathbf{BG}$ is a decidable variety.

Theorem 8.19 *A language is of level* 1 *in the group hierarchy if and only if its syntactic monoid belongs to* **BG**.

In terms of automata, one gets the following result.

Proposition 8.20 *Let $\mathcal{A} = (Q, A, E, i, F)$ be a complete deterministic automaton recognizing a language L. Then L is of level 1 in the group hierarchy if there exist no subgraph of $\mathcal{A}$ of one of the following forms*





with $x, y, u \in A^*$, $q_3 \notin F$ and $q_4 \in F$. This condition is also necessary if $\mathcal{A}$ is minimal.

Proof Suppose that $\mathcal{A}$ has no subgraphs of the form above. We show that the syntactic monoid M of L contains no pair of $\mathcal{R}$-related idempotents. Let ω be the exponent of M and let f, g be words such that $f^\omega \;\mathcal{R}\; g^\omega$ in M. Set $x = f^\omega$ and $y = g^\omega$. Then, by a standard argument (see, for instance, Proposition 1.4 of [38]) $xy = y$ and $yx = x$ in M. We claim that x and y are syntactically equivalent. Let t and u be words such that $txu \in L$. Let $q_0 = i \cdot t$, $q_1 = q_0 \cdot x$ and $q_2 = q_0 \cdot y$. The relations $x = x^2$, $y = y^2$, $xy = y$ and $yx = x$ show that $q_1 \cdot x = q_1$, $q_1 \cdot y = q_2$, $q_2 \cdot x = q_1$ and $q_2 \cdot y = q_2$. Set $q_1 \cdot u = q_3$ and $q_2 \cdot u = q_4$. Since $txu \in L$, $i \cdot txu = q_3 \in F$. Therefore $q_4 \in F$ otherwise $\mathcal{A}$ would contain a subgraph of the second type. It follows that $tyu \in L$ since $i \cdot tyu = q_4$. A dual argument would show that $tyu \in L$ implies $txu \in L$, proving the claim. Thus $x = y$ in M. We let the reader verify, by using the first subgraph, that M contains no pair of $\mathcal{L}$-related idempotents. Thus $M \in$ **BG**.

Assume now that $\mathcal{A}$ is minimal. Suppose that $M \in$ **BG**, and suppose that $\mathcal{A}$ contains the first subgraph. Since $\mathcal{A}$ is minimal, every state of $\mathcal{A}$ is accessible and in particular, there exists a word $t \in A^*$ such that $i \cdot t = q_0$.

On the one hand, $i \cdot t(y^\omega x^\omega)^\omega u = q_4 \in F$ and thus $t(y^\omega x^\omega)^\omega u \in L$. On the other hand, $i \cdot t(x^\omega y^\omega)^\omega u = q_3 \notin F$. It follows that $t(x^\omega y^\omega)^\omega u \notin L$ and thus the identity $(x^\omega y^\omega)^\omega = (y^\omega x^\omega)^\omega$ doesn't hold in M, a contradiction. A similar argument would work for the second subgraph. $\square$

The previous result yields a polynomial time algorithm to check whether the language accepted by a deterministic n-state automaton is of level 1 in the group hierarchy. The proof is similar to the proof of Corollary 8.10 and it is left to the reader.

Corollary 8.21 *There is a polynomial time algorithm for testing whether the language accepted by a deterministic n-state automaton is of level* 1 *in the group hierarchy.*

Several other descriptions of **BG** are known. One of them describes **BG** as the variety generated by all Schützenberger products of groups. Another relates **BG** to the variety generated by all power monoids of groups. A third one gives a decomposition of **BG** as a Mal'cev product. The reader is referred to the survey article [43] for a more detailed discussion.

9 Boolean-polynomial closure

Let $\mathbf{V}$ be a variety of finite semigroups and let $\mathcal{V}$ be the corresponding $+$-variety. We have shown that the algebraic counterpart of the operation $\mathcal{V} \to \mathrm{Pol}\, \mathcal{V}$ on varieties of languages is the operation $\mathbf{V} \to [\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$. Similarly, the algebraic counterpart of the operation $\mathcal{V} \to \text{Co-Pol}\, \mathcal{V}$ is the operation $\mathbf{V} \to [\![x^\omega \leqslant x^\omega y x^\omega]\!] \circledM \mathbf{V}$. It is tempting to guess that the algebraic counterpart of the operation $\mathcal{V} \to \mathrm{BPol}\, \mathcal{V}$ is also of the form $\mathbf{V} \to \mathbf{W} \circledM \mathbf{V}$ for some variety $\mathbf{W}$. In this section, we give a precise statement of this conjecture and we discuss its consequences. Theorem 5.8 and Proposition 3.9 lead to a first characterization of the variety of finite semigroups corresponding to BPol $\mathcal{V}$.

Corollary 9.1 *Let $\mathbf{V}$ be a variety of finite semigroups (resp. monoids) and let $\mathcal{V}$ be the corresponding $+$-variety(resp. $*$-variety). Then the variety of finite semigroups (resp. monoids) corresponding to BPol $\mathcal{V}$ is the join of the two varieties $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$ and $[\![x^\omega \leqslant x^\omega y x^\omega]\!] \circledM \mathbf{V}$.*

Now, if we assume that the variety of finite ordered monoids corresponding to BPol $\mathcal{V}$ can be written as $\mathbf{W} \circledM \mathbf{V}$ for some variety of finite ordered semigroups **W** independent of **V**, it is easy to calculate **W** by taking **V** equal to the variety **I** of trivial semigroups. One gets

$$\begin{aligned}\mathbf{W} &= \mathbf{W} \text{Ⓜ} \mathbf{I} \\ &= \Big([\![x^\omega y x^\omega \leqslant x^\omega]\!] \text{Ⓜ} \mathbf{I}\Big) \vee \Big([\![x^\omega \leqslant x^\omega y x^\omega]\!] \text{Ⓜ} \mathbf{I}\Big) \\ &= [\![x^\omega y x^\omega \leqslant x^\omega]\!] \vee [\![x^\omega \leqslant x^\omega y x^\omega]\!]\end{aligned}$$

It turns out that this variety of ordered semigroups is the variety of ordered semigroups generated by the variety of finite semigroups $\mathbf{B_1}$ defined in section 8.2.

Theorem 9.2 *The variety* $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \vee [\![x^\omega \leqslant x^\omega y x^\omega]\!]$ *is the variety of all ordered semigroups* $(S, \leqslant)$ *such that* $S \in \mathbf{B_1}$.

Proof By Proposition 8.14, the languages corresponding to the variety $[\![x^\omega y x^\omega \leqslant x^\omega]\!]$ are the languages of dot-depth 1/2. It follows from Corollary 9.1 that the positive variety of languages corresponding to $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \vee [\![x^\omega \leqslant x^\omega y x^\omega]\!]$ is the positive variety of languages of dot-depth 1. The result now follows from Knast's Theorem 8.17. □

We can thus reformulate our conjecture as follows:

Conjecture 9.1 *Let* $\mathcal{V}$ *be a variety of languages and let* **V** *be the corresponding variety of semigroups (resp. monoids). Then the variety of semigroups (resp. monoids) corresponding to BPol* $\mathcal{V}$ *is* $\mathbf{B_1} \text{Ⓜ} \mathbf{V}$.

One inclusion in the conjecture is certainly true.

Proposition 9.3 *The variety of semigroups (resp. monoids) corresponding to BPol* $\mathcal{V}$ *is contained in* $\mathbf{B_1} \text{Ⓜ} \mathbf{V}$.

Proof By Corollary 9.1, the variety of finite ordered monoids corresponding to BPol $\mathcal{V}$ is the join of the varieties $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \text{Ⓜ} \mathbf{V}$ and $[\![x^\omega \leqslant x^\omega y x^\omega]\!] \text{Ⓜ} \mathbf{V}$. Now each of these two varieties of ordered semigroups is contained in the variety of ordered semigroup generated by $\mathbf{B_1} \text{Ⓜ} \mathbf{V}$. The proposition follows. □

Now, by Theorem 2.3, the identities of $\mathbf{B_1} \text{Ⓜ} \mathbf{V}$ are

$$(x^\omega p y^\omega q x^\omega)^\omega x^\omega p y^\omega s x^\omega (x^\omega r y^\omega s x^\omega)^\omega = (x^\omega p y^\omega q x^\omega)^\omega (x^\omega r y^\omega s x^\omega)^\omega \qquad (5)$$

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

Proposition 9.4 *The conjecture is true if and only if every finite semigroup (resp. monoid) satisfying the identities* (5) *is a quotient of an ordered semigroup (resp. ordered monoid) of the variety* $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \circledM \mathbf{V}$.

Proof This follows immediately from Propositions 3.5 and 3.9, Theorem 5.8 and the fact that BPol $\mathcal{V}$ is the variety generated by Pol $\mathcal{V}$. $\square$

Conjecture 9.1 was proved to be true in a few particular cases. First, if $\mathbf{V}$ is the trivial variety of monoids, then $\mathbf{B_1} \circledM \mathbf{I} = \mathbf{J}$. In this case, the second form of the conjecture was also proved directly by Straubing and Thérien [69].

Theorem 9.5 *Every finite $\mathcal{J}$-trivial monoid is a quotient of an ordered monoid satisfying the identity* $x \leqslant 1$.

Second, if $\mathbf{V}$ is the trivial variety of semigroups, then $\mathbf{B_1} \circledM \mathbf{I} = \mathbf{B_1}$ is, by Knast's Theorem 8.17, the variety of finite semigroups corresponding to the languages of dot-depth 1. Therefore, the conjecture is true in this case, leading to the following corollary.

Corollary 9.6 *Every semigroup of* $\mathbf{B_1}$ *is a quotient of an ordered semigroup satisfying the identity* $x^\omega y x^\omega \leqslant x^\omega$.

Third, if $\mathbf{V} = \mathbf{G}$, the variety of monoids consisting of all finite groups, $\mathbf{B_1} \circledM \mathbf{G} = \mathbf{J} \circledM \mathbf{G} = \mathbf{PG} = \Diamond\mathbf{G} = \mathbf{BG}$ is the variety corresponding to the level 1 of the group hierarchy. Therefore, the conjecture is also true in this case.

Corollary 9.7 *Every semigroup of* $\mathbf{BG}$ *is a quotient of an ordered semigroup satisfying the identity* $x^\omega \leqslant 1$.

It is amusing to prove directly this result for powergroups. Given a group G, denote by $\mathcal{P}'(G)$ the monoid of all non-empty subsets of G under multiplication. Then $\mathcal{P}'(G)$ is naturally ordered by the relation $\leqslant$ defined by

$$X \leqslant Y \text{ if and only if } Y \subseteq X$$

Proposition 9.8 *Let G be a group. Then $(\mathcal{P}'(G), \leqslant)$ satisfies the identity* $x^\omega \leqslant 1$.

Proof The idempotents of $\mathcal{P}'(G)$ are the subgroups of G. They all contain the trivial subgroup, which is the identity of $\mathcal{P}'(G)$. Therefore the identity $x^\omega \leqslant 1$ is satisfied. $\square$

The level 2 of the Straubing hierarchy corresponds to the case $\mathbf{V} = \mathbf{J_1}$. Therefore, one can formulate the following conjecture for this level

Conjecture 9.2 *A recognizable language is of level* 2 *in the Straubing hierarchy if and only if its syntactic semigroup satisfies the identities*

$$\big((x^\omega py^\omega qx^\omega)^\omega x^\omega py^\omega sx^\omega (x^\omega ry^\omega sx^\omega)^\omega\big) = \big((x^\omega py^\omega qx^\omega)^\omega (x^\omega ry^\omega sx^\omega)^\omega\big) \quad (6)$$

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

If this conjecture was true, it would imply the decidability of the levels 2 of the Straubing hierarchy and of the dot-depth. It was shown [70, 74, 77] that Corollary 8.12 implies that Conjecture 9.2 is true for languages recognized by an inverse monoid.

More generally, the conjecture $\mathbf{V}_{n+1} = \mathbf{B_1} \text{Ⓜ} \mathbf{V}_n$ would reduce the decidability of the Straubing hierarchy to a problem on the Mal'cev products of the form $\mathbf{B_1} \text{Ⓜ} \mathbf{V}$. However, except for a few exceptions (including $\mathbf{G}$, $\mathbf{J}$ and the finitely generated varieties, like the trivial variety or $\mathbf{J_1}$), it is not known whether the decidability of $\mathbf{V}$ implies that of $\mathbf{B_1} \text{Ⓜ} \mathbf{V}$.

10 The sequential calculus

This section is devoted to the consequences of our results in finite model theory, and more precisely, to Büchi's sequential calculus. We assume that the reader is familiar with the standard notations of formal logic.

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

$$\mathfrak{M}_u = \big(\{1, \dots, |u|\}, (R_a)_{a \in A}, <\big)$$

where $R_a = \{i \in \{1, \dots, |u|\} \mid u(i) = a\}$ is the set of positions of the letter a in u and $<$ is the usual order on $\{1, \dots, |u|\}$. 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 that satisfy φ:

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

For instance, if $\varphi = \exists i\ R_a i$, then $L(\varphi) = A^*aA^*$. The reader is referred to the survey article [45] for more detail on this logic. The first order definable languages were first characterized by McNaughton and Papert [32].

Theorem 10.1 *A recognizable subset of A^* is first-order definable if and only if it is star-free.*

This, combined with Schützenberger theorem, gives a syntactic characterization of first-order definable languages.

Corollary 10.2 *Let X be a recognizable subset of A^*. Then the following conditions are equivalent:*
(1) *X is first-order definable,*
(2) *X is star-free,*
(3) *the syntactic monoid of X is aperiodic.*

The correspondence between star-free languages is even tighter than indicated in Corollary 10.2. Indeed 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 $n \geqslant 0$, $\Delta_n = \Sigma_n \cap \Pi_n$. 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 10.3 (Thomas [72], Perrin and Pin [34])
(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.

Theorem 10.4 *A language of A^* is in Δ_{n+1} if and only if it is in UPol $\mathcal{V}_n$.*

Proof This follows immediately from Theorems 7.3 and 10.3. $\square$

This result reminds us of a result of Arnold [7] in a different context. A set of infinite words is Σ_1^1 (analytic) if and only if it is accepted by a countable Büchi automaton and it is a Borel set if and only if it is accepted by a countable unambiguous Büchi automaton. Now, by Suslin's theorem, $\Sigma_1^1 \cap \Pi_1^1 = \Delta_1^1$ is the class of Borel sets. Thus a set of words is Δ_1^1 if and only if it is accepted by a countable unambiguous Büchi automaton.

One can summarize our results in the following diagrams

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

Figure 10.3: The logical hierarchy

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

Figure 10.4: The Straubing-Thérien hierarchy

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 10 of this article.

References

[1] J. Almeida, Equations for pseudovarieties, *Formal properties of finite automata and applications, J.-E. Pin (ed.)*, Lecture Notes in Computer Science, Vol. 386, Springer-Verlag, Berlin, (1989), 148–164.

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

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

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

[5] M. Arfi, Polynomial operations and rational languages, *Proc. 4th STACS*, Lecture Notes in Computer Science, Vol. 247, Springer-Verlag, Berlin, 1987, 198–206.

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

[7] A. Arnold, Topological characterizations of infinite behaviours of transition systems, in *Automata, Languages and Programming*, J. Diaz (ed.), Lecture Notes in Computer Science, Vol. 154, Springer-Verlag, Berlin, 1983, 28–38.

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

[9] C. J. Ash, Inevitable Graphs: A proof of the type II conjecture and some related decision procedures, *Internat. J. Algebra Comput.* **1**, (1991), 127–146.

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

[11] F. Blanchet-Sadri, On dot-depth two, *Inform. Théor. Appl.* **24**, (1990), 521–529.

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

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

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

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

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

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

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

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

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

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

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

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

[24] K. Henckell, S. W. Margolis, J.-E. Pin and J. Rhodes, Ash's Type II Theorem, Profinite Topology and Malcev Products, *Internat. J. Algebra Comput.* **1**, (1991) 411–436.

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

[26] S.C. Kleene, Representation of events in nerve nets and finite automata, in *Automata Studies*, C.E. Shannon and J. McCarthy (eds.), Princeton University Press, Princeton, NJ, 1956, 3–42.

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

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

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

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

[31] S. W. Margolis and J.E. Pin, Product of group languages, *Proc. FCT Conference*, Lecture Notes in Computer Science, Vol. 199, Springer-Verlag, Berlin, 1985, 285–299.

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

[33] D. Perrin, Automata, in *Handbook of Theoretical Computer Science* Vol B, J. Van Leeuwen, (ed.), Elsevier, Amsterdam, 1990, Chapter 1.

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

[35] J.-E. Pin, Propriétés syntactiques du produit non ambigu. *Proc. 7th ICALP*, Lecture Notes in Computer Science, Vol. 85, Springer-Verlag, Berlin, 1980, 483–499.

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

[37] J.-E. Pin, Finite group topology and p-adic topology for free monoids, *Proc. 12th ICALP*, Lecture Notes in Computer Science, Vol. 194, Springer-Verlag, Berlin, 1985, 445–455.

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

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

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

[41] J.-E. Pin, Polynomial closure of group languages and open sets of the Hall topology, *Proc. ICALP 1994*, Lecture Notes in Computer Science, Vol. 820, Springer-Verlag, Berlin, 1994, 424–435.

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

[43] J.-E. Pin, $\mathbf{BG} = \mathbf{PG}$, a success story, in *Semigroups, Formal Languages and Groups*, J. Fountain (ed.), NATO Advanced Study Institute Series, Kluwer, Dordrecht, 1995, 33–47.

[44] J.-E. Pin, A variety theorem without complementation, *Izv. Vyssh. Uchebn. Zaved. Mat.* **39** (1995), 80–90. English version, *Russian Mathem.* **39**, (1995), 74–83.

[45] J.-E. Pin, Logic, Semigroups and Automata on Words, *Ann. Math. Artificial Intel.* **16**, (1996), 343–384.

[46] J.-E. Pin, Polynomial closure of group languages and open sets of the Hall topology, *Theoret. Comput. Sci.* **169**, (1996), 185–200.

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

[48] J.-E. Pin and H. Straubing, Monoids of upper triangular matrices, in *Semigroups*, Colloquia Mathematica Societatis Janos Bolyai, Vol. 39, North Holland, Amsterdam, 1981, 259–272.

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

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

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

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

[53] Ch. Reutenauer, Sur les variétés de langages et de monoïdes, Lecture Notes in Computer Science, Vol. 67, Springer-Verlag, Berlin, 1979, 260–265.

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

[55] Ch. Reutenauer, Sur mon article "Une topologie du monoïde libre", *Semigroup Forum* **22**, (1981), 93–95.

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

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

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

[59] I. Simon, Piecewise testable events, *Proc. 2nd GI Conf.*, Lecture Notes in Computer Science, Vol. 33, Springer-Verlag, Berlin, 1975, 214–222.

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

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

[62] I. Simon, The product of rational languages, *Proc. ICALP 1993*, Lecture Notes in Computer Science, Vol. 700, Springer-Verlag, Berlin, 1993, 430–444.

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

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

[65] H. Straubing, A generalization of the Schützenberger product of finite monoids, *Theoret. Comp. Sci.* **13** (1981), 137–150.

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

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

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

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

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

[71] D. Thérien, Classification of finite monoids: the language approach, *Theoret. Comp. Sci.* **14**, (1981), 195–208.

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

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

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

[75] P. Weil, Implicit operations on pseudovarieties: an introduction, in *Monoids and semigroups with applications*, J. Rhodes (ed.), World Scientific, Singapore, 1991, 89–104.

[76] P. Weil, Closure of varieties of languages under products with counter, *J. Comput. System Sci.* **45**, (1992), 316–339.

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