Bridges for concatenation hierarchies

Jean-Éric Pin

LIAFA, CNRS and Université Paris VII
2 Place Jussieu 75251 Paris Cedex O5, FRANCE
e-mail: Jean-Eric.Pin@liafa.jussieu.fr

Abstract. In the seventies, several classification schemes for the rational languages were proposed, based on the alternate use of certain operators (union, complementation, product and star). Some thirty years later, although much progress has been done, several of the original problems are still open. Furthermore, their significance has grown considerably over the years, on account of the successive discoveries of surprising links with other fields, like non commutative algebra, finite model theory, structural complexity and topology. In this article, we solve positively a question raised in 1985 about concatenation hierarchies of rational languages, which are constructed by alternating boolean operations and concatenation products. We establish a simple algebraic connection between the Straubing-Thérien hierarchy, whose basis is the trivial variety, and the group hierarchy, whose basis is the variety of group languages. Thanks to a recent result of Almeida and Steinberg, this reduces the decidability problem for the group hierarchy to a property stronger than decidability for the Straubing-Thérien hierarchy.

The reader is referred to [20] for undefined terms and a general overview of the motivations of this paper.

1 Introduction

In the seventies, several classification schemes for the rational languages were proposed, based on the alternate use of certain operators (union, complementation, product and star). Some thirty years later, although much progress has been done, several of the original problems are still open. Furthermore, their significance has grown considerably over the years, on account of the successive discoveries of surprising links with other fields, like non commutative algebra [7], finite model theory [34], structural complexity [4] and topology [11, 16, 19]. In this article, we solve positively a question left open in [11].

We are interested in hierarchies constructed by alternating union, complementation and concatenation products. All these hierarchies are indexed by half integers (i.e. numbers of the form n or $n + \frac{1}{2}$, where n is a non-negative integer) and follow the same construction scheme. The languages of level $n + \frac{1}{2}$ are the finite union of products of the form

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

where $L_0, L_1, \ldots, L_k$ are languages of level n and $a_1, \ldots, a_k$ are letters. The languages of level $n+1$ are the boolean combinations[1] of languages of level $n+\frac{1}{2}$.

Thus a concatenation hierarchy $\mathcal{H}$ is fully determined by its level zero $\mathcal{H}_0$. For the sake of simplicity, levels of the form $\mathcal{H}_n$ will be called *full levels*, and levels of the form $\mathcal{H}_{n+\frac{1}{2}}$, *half levels*.

Three concatenation hierarchies have been intensively studied in the literature. The *dot-depth hierarchy*, introduced by Brzozowski [5], takes the finite or cofinite languages of A^+ as a basis. The *Straubing-Thérien hierarchy* [33, 28, 29] is based on the empty and full languages of A^*. The *group hierarchy*, considered in [11], is built on the group-languages, the languages recognized by a finite permutation automaton. It is the main topic of this paper.

These three hierarchies are infinite [6] and share another common feature : their basis is a variety of languages in the sense of Eilenberg [7]. It can be shown in general that, if the basis of a concatenation hierarchy is a variety of languages, then every level is a positive variety of languages, and in particular, is closed under intersection [2, 3, 22].

The main problems concerning these hierarchies are decidability problems : given a concatenation hierarchy $\mathcal{H}$, a half integer n and a rational language L, decide

(1) whether L belongs to $\mathcal{H}$,
(2) whether L belongs to $\mathcal{H}_n$.

The first problem has been solved positively for the three hierarchies [24, 11], but the second one is solved positively only for $n \leq \frac{3}{2}$ for the Straubing-Thérien hierarchy [25, 2, 3, 22] and for $n \leq 1$ for the two other hierarchies [9–11, 8]. It is still open for the other values of n although some partial results for the level 2 of the Straubing-Thérien hierarchy are known [21, 30, 32, 22, 36]. These problems are, together with the generalized star-height problem, the most important open problems on rational languages. Their logical counterpart is also quite natural : it amounts to decide whether a first order formula of Büchi's sequential calculus is equivalent to a Σ_n-formula on finite words models. See [14, 18] for more details.

Depending on the reader's favorite domain, a combinatorial, algebraic or logical approach of these problems is possible. The algebraic approach will be used in this paper. Since every level is a positive variety of languages, the variety theorem [7, 17] tells us there is a corresponding variety of finite ordered monoids (semigroups in the case of the dot-depth hierarchy) for each level. Let us denote these varieties by $\mathbf{V}_n$ for the Straubing-Thérien hierarchy, $\mathbf{B}_n$ for the dot-depth, and $\mathbf{G}_n$ for the group hierarchy (for any half integer n). Problem (2) now reduces to know whether the variety $\mathbf{V}_n$ (resp. $\mathbf{B}_n$, $\mathbf{G}_n$) is decidable. That is, given a finite ordered monoid (or semigroup) M decide whether it belongs to $\mathbf{V}_n$ (resp. $\mathbf{B}_n$, $\mathbf{G}_n$).

A nice connection between $\mathbf{V}_n$ and $\mathbf{B}_n$ was found by Straubing [29]. It is expressed by the formula

$$\mathbf{B}_n = \mathbf{V}_n * \mathbf{LI} \quad (n > 0) \tag{$*$}$$

[1] Boolean operations comprise union, intersection and complement.

which tells that the variety $\mathbf{B}_n$ is generated by semidirect products of the form $M * S$, where M is in $\mathbf{V}_n$ and S is a so-called "locally trivial" semigroup. Formula $(*)$ was established by Straubing for the full levels, but it still holds for the half levels.

In some sense, this formula reduces the study of the hierarchy $\mathbf{B}_n$ (the dot-depth) to that of $\mathbf{V}_n$ (the Straubing-Thérien's). Actually, things are not that easy, and it still requires a lot of machinery to show that $\mathbf{B}_n$ is decidable if and only if $\mathbf{V}_n$ is decidable [29]. Furthermore, this latter result is not yet formally proved for half levels.

A similar formula, setting a bridge between the varieties $\mathbf{G}_n$ and $\mathbf{V}_n$, was conjectured in [11] :

$$\mathbf{G}_n = \mathbf{V}_n * \mathbf{G} \quad (n \geq 0) \tag{$**$}$$

It tells that the variety $\mathbf{G}_n$ is generated by semidirect products of the form $M * G$, where M is in $\mathbf{V}_n$ and G is a group. The proof of this conjecture is the main result of this paper.

Actually, we show that a similar result holds for any hierarchy based on a group variety (such as commutative groups, nilpotent groups, solvable groups, etc.).

Does this result reduce the study of the group hierarchy to that of the Straubing-Thérien's? Yes and no. Formally, our result doesn't suffice to reduce the decidability problem of $\mathbf{G}_n$ to that of $\mathbf{V}_n$. However, a recent result of Almeida and Steinberg [1] gives a reduction of the decidability problem of $\mathbf{G}_n$ to a strong property of $\mathbf{V}_n$. More precisely, Almeida and Steinberg showed that if the variety of finite categories $\mathbf{gV}_n$ generated by $\mathbf{V}_n$ has a recursively enumerable basis of (pseudo)identities, then the decidability of $\mathbf{V}_n$ implies that of $\mathbf{G}_n$. Of course, even more algebra is required to use (and even state !) this result, but it is rather satisfactory for the following reason: although the decidability of $\mathbf{V}_n$ is still an open problem for $n \geq 2$, recent conjectures tend to indicate that a good knowledge of the identities of $\mathbf{gV}_n$ will be required to prove the decidability of $\mathbf{V}_n$. In other words, it is expected that the proof of the decidability of $\mathbf{V}_n$ will require the knowledge of the identities of $\mathbf{gV}_n$, giving in turn the decidability of $\mathbf{G}_n$.

2 Preliminaries and notations

2.1 Monoids

In this paper, all monoids are finite or free.

A relation $\leq$ on a monoid M is *stable* if, for all $x, y, z \in M$, $x \leq y$ implies $xz \leq yz$ and $zx \leq zy$. An *ordered monoid* is a monoid equipped with a stable order relation. An *order ideal* of M is a subset I of M such that, if $x \leq y$ and $y \in I$, then $x \in I$.

Note that every monoid, equipped with the equality relation, is an ordered monoid. This remark allows to consider any monoid as an ordered monoid.

Given two elements m and n of a monoid M, we put

$$m^{-1}n = \{x \in M \mid mx = n\}$$

Note that if M is a group, the set $m^{-1}n$ is equal to the singleton $\{m^{-1}n\}$, where this time m^{-1} denotes the inverse of m. This observation plays an important role in the proof of the main result.

Let M and N be monoids. A *monoid morphism* $\varphi : M \to N$ is a map from M into N such that $\varphi(xy) = \varphi(x)\varphi(y)$ for all $x, y \in M$. If M and N are ordered, φ is a morphism of ordered monoids if, furthermore, $x \leq y$ implies $\varphi(x) \leq \varphi(y)$ for all $x, y \in M$.

Let M and N be two ordered monoids. Then M is a *quotient* of N if there exists a surjective morphism of ordered monoids from N onto M. And M *divides* N if it is a quotient of a submonoid of N. Division is an order on finite ordered monoids (up to isomorphism).

A *variety of ordered monoids* is a class of finite ordered monoids closed under taking ordered submonoids, quotients and finite direct products. A *variety of monoids* is defined analogously.

Let M and N be ordered monoids. We write the operation of M additively and its identity by 0 to provide a more transparent notation, but it is not meant to suggest that M is commutative. A *left action* of N on M is a map $(t, s) \mapsto t{\cdot}s$ from $N \times M$ into M such that, for all $s, s_1, s_2 \in M$ and $t, t_1, t_2 \in N$,

(1) $(t_1t_2){\cdot}s = t_1(t_2{\cdot}s)$

(2) $1{\cdot}s = s$

(3) if $s \leq s'$ then $t{\cdot}s \leq t{\cdot}s'$

(4) $t{\cdot}(s_1 + s_2) = t{\cdot}s_1 + t{\cdot}s_2$

(5) $t{\cdot}0 = 0$

(6) if $t \leq t'$ then $t{\cdot}s \leq t'{\cdot}s$

The *semidirect product* of M and N (with respect to the given action) is the ordered monoid $M * N$ defined on $M \times N$ by the multiplication

$$(s, t)(s', t') = (s + t{\cdot}s', tt')$$

and the product order:

$$(s, t) \leq (s', t') \quad \text{if and only if } s \leq s' \text{ and } t \leq t'$$

Given two varieties of ordered monoids $\mathbf{V}$ and $\mathbf{W}$, denote by $\mathbf{V} * \mathbf{W}$ the variety of finite monoids generated by the semidirect products $M * N$ with $M \in \mathbf{V}$ and $N \in \mathbf{W}$.

The *wreath product* is closely related to the semidirect product. The wreath product $M \circ N$ of two ordered monoids M and N is the semidirect product $M^N * N$ defined by the action of N on M^N given by

$$(t{\cdot}f)(t') = f(t't)$$

for $f : N \to M$ and $t, t' \in N$. In particular, the multiplication in $M \circ N$ is given by

$$(f_1, t_1)(f_2, t_2) = (f, t_1t_2) \text{ where } f(t) = f_1(t) + f_2(tt_1) \text{ for all } t \in N$$

and the order on $M \circ N$ is given by

$$(f_1, t_1) \leq (f_2, t_2) \text{ if and only if } t_1 \leq t_2 \text{ and } f_1(t) \leq f_2(t) \text{ for all } t \in N$$

One can show that $\mathbf{V} * \mathbf{W}$ is generated by all wreath products of the form $M \circ N$, where $M \in \mathbf{V}$ and $N \in \mathbf{W}$.

2.2 Varieties of languages

Let A be a finite alphabet. The free monoid on A is denoted by A^* and the free semigroup by A^+. A language L of A^* is said to be *recognized* by an ordered monoid M if there exists a monoid morphism from A^* onto M and an order ideal I of M such that $L = \varphi^{-1}(I)$. In this case, we also say that L is recognized by φ. It is easy to see that a language is recognized by a finite ordered monoid if and only if it is recognized by a finite automaton, and thus is a rational (or regular) language. However, ordered monoids provide access to a more powerful algebraic machinery, that will be required for proving our main result. We start with an elementary result, the proof of which is omitted.

Proposition 1. *If a language L of A^* is recognized by M and if M divides N, then L is recognized by N.*

A set of languages closed under finite intersection and finite union is called a *positive boolean algebra*. Thus a positive boolean algebra always contains the empty language and the full language A^* since $\emptyset = \bigcup_{i \in \emptyset} L_i$ and $A^* = \bigcap_{i \in \emptyset} L_i$. A positive boolean algebra closed under complementation is a *boolean algebra*.

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

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

(1) for every alphabet A, $\mathcal{V}(A^*)$ is a positive boolean algebra,
(2) if $\varphi : A^* \to B^*$ is a monoid morphism, $L \in \mathcal{V}(B^*)$ implies $\varphi^{-1}(L) \in \mathcal{V}(A^*)$,
(3) if $L \in \mathcal{V}(A^*)$ and if $a \in A$, then $a^{-1}L$ and La^{-1} are in $\mathcal{V}(A^*)$.

A *variety of languages* is a positive variety closed under complement.

To each variety of ordered monoids $\mathbf{V}$, is associated the corresponding positive variety of languages $\mathcal{V}$. For each alphabet A, $\mathcal{V}(A^*)$ is the set of all languages of A^* recognized by an ordered monoid of $\mathbf{V}$. Similarly, to each variety of monoids $\mathbf{V}$, is associated the corresponding variety of languages $\mathcal{V}$. For each alphabet A, $\mathcal{V}(A^*)$ is the set of all languages of A^* recognized by a monoid of $\mathbf{V}$, also called $\mathbf{V}$-*languages*.

The variety theorem [7, 17] states that the correspondence $\mathbf{V} \to \mathcal{V}$ between varieties of ordered monoids and positive varieties of languages (resp. between varieties of monoids and varieties of languages) is one-to-one.

We refer the reader to [7, 13, 15, 20] for more details on varieties.

3 Algebraic tools

The aim of this section is to introduce an ordered version of several standard algebraic tools. We start with power monoids.

3.1 Power 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. Inclusion and reverse inclusion define two stable orders on $\mathcal{P}(M)$. For reasons that will become apparent in the next sections, we denote by $\mathcal{P}^+(M)$ the ordered monoid $(\mathcal{P}(M), \supseteq)$ and by $\mathcal{P}^-(M)$ the ordered monoid $(\mathcal{P}(M), \subseteq)$. The following proposition shows that the operator $\mathcal{P}$ preserves submonoids and quotients.

Proposition 2. *Let M be a submonoid (resp. a quotient) of N. Then $\mathcal{P}^+(M)$ is an ordered submonoid (resp. a quotient) of $\mathcal{P}^+(N)$.*

3.2 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 [24], and extended by Straubing [28] for any number of monoids. We give an ordered version of this definition.

Let $M_1, \ldots, M_n$ be monoids. Denote by M the product $M_1 \times \cdots \times M_n$ and by $\mathcal{M}_n$ the semiring of square matrices of size n with entries in the ordered semiring $\mathcal{P}^+(M)$. The *Schützenberger product* of $M_1, \ldots, M_n$, denoted by $\diamondsuit_n(M_1, \ldots, M_n)$, is the submonoid of the multiplicative monoid composed of all the matrices P of $\mathcal{M}_n$ satisfying the three following conditions:

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

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

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

The Schützenberger product can be ordered by simply inheriting the order on $\mathcal{P}^+(M)$: $P \leq P'$ if and only if for $1 \leq i \leq j \leq n$, $P_{i,j} \leq P'_{i,j}$ in $\mathcal{P}^+(M)$. The corresponding ordered monoid is denoted $\diamondsuit_n^+(M_1, \ldots, M_n)$ and is called the *ordered Schützenberger product* of $M_1, \ldots, M_n$.

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 $P_{2,3} \subseteq M_2 \times M_3$.

We first state without proof some elementary properties of the Schützenberger product. Let $M_1, \ldots, M_n$ be monoids and let M be their ordered Schützenberger product.

Proposition 3. *Each M_i is a quotient of M. Furthermore, for each sequence $1 \leq i_1 < \ldots < i_k \leq n$, $\Diamond_k^+(M_{i_1}, \ldots, M_{i_k})$ is an ordered submonoid of M.*

Proposition 4. *If, for $1 \leq i \leq n$, M_i is a submonoid (resp. a quotient, a divisor) of the monoid N_i, then M is an ordered submonoid (resp. a quotient, a divisor) of the ordered Schützenberger product of $N_1, \ldots, N_n$.*

Our next result gives an algebraic characterization of the languages recognized by a Schützenberger product. It is the "ordered version" of a result first proved by Reutenauer [23] for $n = 2$ and by the author [12] in the general case (see also [35]).

Theorem 1. *Let $M_1, \ldots, M_n$ be monoids. A language is recognized by the ordered Schützenberger product of $M_1, \ldots, M_n$ if and only if it is a positive boolean combination of languages recognized by one of the M_i's or of the form*

$$L_0 a_1 L_1 \cdots a_k L_k \tag{1}$$

where $k > 0$, $a_1, \ldots, a_k \in A$ and L_j is recognized by M_{i_j} for some sequence $1 \leq i_0 < i_1 < \cdots < i_k \leq n$.

Due to the lack of place, the proof is omitted, but follows the main lines of the elegant proof given by Simon [26].

3.3 The wreath product principle

Straubing's *wreath product principle* [27, 31] provides a description of the languages recognized by the wreath product of two monoids. We extend here this result to the ordered case.

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

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

Observe that σ_φ is not a morphism, but a sequential function.

Theorem 2. *(Wreath product principle) Every language recognized by η is a finite union of languages of the form $U \cap \sigma_\varphi^{-1}(V)$, where U is a language of A^* recognized by φ and V is a language of B^* recognized by M.*

Conversely, every language of the form $\sigma_\varphi^{-1}(V)$ is recognized by a wreath product.

Proposition 5. *If V is a language of B^* recognized by M, then $\sigma_\varphi^{-1}(V)$ is recognized by $M \circ N$.*

Since we are working with concatenation hierarchies, we will encounter expressions of the form $\sigma_\varphi^{-1}\big(L_0(m_1, a_1)L_1 \cdots (m_k, a_k)L_k\big)$. The inversion formula given below converts these expressions into concatenation products. It is the key result in the proof of our main result.

Define, for each $m \in N$, a morphism $\lambda_m : B^* \to B^*$ by setting $\lambda_m(n, a) = (mn, a)$. Then for each $u, v \in A^*$ and $a \in A$:

$$\sigma_\varphi(uav) = \sigma_\varphi(u)(\varphi(u), a)\lambda_{\varphi(ua)}(\sigma_\varphi(v)) \tag{2}$$

Let $m_1, \ldots, m_{k+1}$ be elements of N, $a_1, \ldots, a_k$ be letters of A and $L_1, \ldots, L_k$ be languages of B^*. Setting $n_0 = 1$ and $n_j = m_j\varphi(a_j)$ for $1 \le j \le k$, the following formula holds

Lemma 1. *(Inversion formula)*

$$\sigma_\varphi^{-1}\big(L_0(m_1, a_1)L_1 \cdots (m_k, a_k)L_k\big) \cap \varphi^{-1}(m_{k+1}) = K_0a_1K_1 \cdots a_kK_k$$

where $K_j = \sigma_\varphi^{-1}(\lambda_{n_j}^{-1}(L_j)) \cap \varphi^{-1}(n_j^{-1}m_{j+1})$ for $1 \le j \le k$.

Proof. Denote respectively by L and R the left and the right hand sides of the formula. If $u \in L$, then

$$\sigma_\varphi(u) = v_0(m_1, a_1)v_1(m_2, a_2) \cdots (m_k, a_k)v_k$$

with $v_j \in L_j$. Let $u = u_0a_1u_1 \cdots a_ku_k$, with $|u_j| = |v_j|$ for $0 \le j \le k$. Then

$$\sigma_\varphi(u) = \underbrace{\sigma_\varphi(u_0)}_{v_0}\underbrace{(\varphi(u_0), a_1)}_{(m_1,a_1)}\underbrace{\lambda_{\varphi(u_0a_1)}(\sigma_\varphi(u_1))}_{v_1} \cdots \underbrace{(\varphi(u_0a_1 \cdots u_{k-1}), a_k)}_{(m_k,a_k)}\underbrace{\lambda_{\varphi(u_0a_1u_1\cdots u_{k-1}a_k)}(\sigma_\varphi(u_k))}_{v_k}$$

It follows $\sigma_\varphi(u_0) \in L_0$, $\lambda_{\varphi(u_0a_1)}(\sigma_\varphi(u_1)) \in L_1, \ldots, \lambda_{\varphi(u_0a_1u_1\cdots u_{k-1}a_k)}(\sigma_\varphi(u_k)) \in L_k$ and $(\varphi(u_0), a_1) = (m_1, a_1), \ldots, (\varphi(u_0a_1 \cdots u_{k-1}), a_k) = (m_k, a_k)$. These conditions, added to the condition $\varphi(u) = m_{k+1}$, can be rewritten as

$$n_j\varphi(u_j) = m_{j+1} \text{ and } \lambda_{n_j}(\sigma_\varphi(u_j)) \in L_j \text{ for } 0 \le j \le k$$

and thus, are equivalent to $u_j \in K_j$, for $0 \le j \le k$. Thus $u \in R$.

In the opposite direction, let $u \in R$. Then $u = u_0a_1u_1 \cdots a_ku_k$ with $u_0 \in K_0$, $\ldots, u_k \in K_k$. It follows $n_j\varphi(u_j) = m_{j+1}$, for $0 \le j \le k$. Let us show that $\varphi(u_0a_1 \cdots a_ju_j) = m_{j+1}$. Indeed, for $j = 0$, $\varphi(u_0) = n_0\varphi(u_0) = m_1$, and, by induction,

$$\varphi(u_0a_1 \cdots a_ju_j) = m_j\varphi(a_ju_j) = m_j\varphi(a_j)\varphi(u_j) = n_j\varphi(u_j) = m_{j+1}$$

Now, by formula (2):

$$\sigma_\varphi(u) = \sigma_\varphi(u_0)(m_1, a_1)\lambda_{n_1}(\sigma_\varphi(u_1))(m_2, a_2) \cdots (m_k, a_k)\lambda_{n_k}(\sigma_\varphi(u_k))$$

Furthermore, by the definition of K_j, $\sigma_\varphi(u_j) \in L_j$ and thus $u \in L$, concluding the proof.

4 Main result

Let $\mathbf{H}_0$ be a variety of groups and let $\mathcal{H}_0$ be the corresponding variety of languages. Let $\mathcal{H}$ be the concatenation hierarchy of basis $\mathcal{H}_0$. As was explained in the introduction, the full levels $\mathcal{H}_n$ of this hierarchy are varieties of languages, corresponding to varieties of monoids $\mathbf{H}_n$ and the half levels $\mathcal{H}_{n+\frac{1}{2}}$ are positive varieties of languages, corresponding to varieties of ordered monoids $\mathbf{H}_{n+\frac{1}{2}}$. Our main result can be stated as follows:

Theorem 3. *The equality* $\mathbf{H}_n = \mathbf{V}_n * \mathbf{H}_0$ *holds for any half integer* n.

The first step of the proof consists in expressing $\mathbf{H}_{n+\frac{1}{2}}$ in terms of $\mathbf{H}_n$. If $\mathbf{V}$ is a variety of monoids, and k is a positive integer, denote by $\diamondsuit_k(\mathbf{V})$ (resp. $\diamondsuit_k^+(\mathbf{V})$) the variety of (resp. ordered) monoids generated by the (resp. ordered) monoids of the form $\diamondsuit_k(M_1, \ldots, M_k)$ (resp. $\diamondsuit_k^+(M_1, \ldots, M_k)$), where $M_1, \ldots, M_k \in \mathbf{V}$. Finally, let $\diamondsuit(\mathbf{V})$ (resp. $\diamondsuit^+(\mathbf{V})$) be the union over k of all the varieties $\diamondsuit_k(\mathbf{V})$ (resp. $\diamondsuit_k^+(\mathbf{V})$). Theorem 1 and its non-ordered version give immediately

Theorem 4. *For every positive integer* n, $\mathbf{V}_{n+\frac{1}{2}} = \diamondsuit^+(\mathbf{V}_n)$ *and* $\mathbf{V}_{n+1} = \diamondsuit(\mathbf{V}_n)$. *Similarly,* $\mathbf{H}_{n+\frac{1}{2}} = \diamondsuit^+(\mathbf{H}_n)$ *and* $\mathbf{H}_{n+1} = \diamondsuit(\mathbf{H}_n)$.

The second step is to prove the following formula

Theorem 5. *For every variety of monoids* $\mathbf{V}$, $\diamondsuit^+(\mathbf{V} * \mathbf{H}_0) = \diamondsuit^+(\mathbf{V}) * \mathbf{H}_0$ *and* $\diamondsuit(\mathbf{V} * \mathbf{H}_0) = \diamondsuit(\mathbf{V}) * \mathbf{H}_0$.

The proof of Theorem 5 is given in the next section. Let us first derive the proof of Theorem 3 by induction on n. The case $n = 0$ is trivial, since $\mathbf{V}_0$ is the trivial variety. By induction, $\mathbf{H}_n = \mathbf{V}_n * \mathbf{H}_0$ and thus $\diamondsuit^+(\mathbf{H}_n) = \diamondsuit^+(\mathbf{V}_n * \mathbf{H}_0)$. It follows, by Theorem 4 and by Theorem 5,

$$\mathbf{H}_{n+\frac{1}{2}} = \diamondsuit^+(\mathbf{H}_n) = \diamondsuit^+(\mathbf{V}_n * \mathbf{H}_0) = \diamondsuit^+(\mathbf{V}_n) * \mathbf{H}_0 = \mathbf{V}_{n+\frac{1}{2}} * \mathbf{H}_0$$

and similarly,

$$\mathbf{H}_{n+1} = \diamondsuit(\mathbf{H}_n) = \diamondsuit(\mathbf{V}_n * \mathbf{H}_0) = \diamondsuit(\mathbf{V}_n) * \mathbf{H}_0 = \mathbf{V}_{n+1} * \mathbf{H}_0$$

5 Proof of Theorem 5

The proof is given in the ordered case, since the proof of the non-ordered case is similar and easier. We will actually prove a slightly more precise result:

Theorem 6. *Let* $\mathbf{U}_1, \ldots, \mathbf{U}_n$ *be varieties of monoids and let* $\mathbf{H}$ *be a variety of groups. Then* $\diamondsuit_n^+(\mathbf{U}_1, \cdots, \mathbf{U}_n) * \mathbf{H} = \diamondsuit_n^+(\mathbf{U}_1 * \mathbf{H}, \cdots, \mathbf{U}_n * \mathbf{H})$.

We treat this equality as a double inclusion. The inclusion from left to right is easier to establish and follows from a more general result

Theorem 7. *Let* $\mathbf{U}_1, \ldots, \mathbf{U}_n$ *and* $\mathbf{V}$ *be varieties of monoids. Then*

$$\Diamond_n^+(\mathbf{U}_1, \cdots, \mathbf{U}_n) * \mathbf{V} \subseteq \Diamond_n^+(\mathbf{U}_1 * \mathbf{V}, \cdots, \mathbf{U}_n * \mathbf{V})$$

Proof. Let $\mathbf{X} = \Diamond_n^+(\mathbf{U}_1, \cdots, \mathbf{U}_n) * \mathbf{V}$ and let $\mathbf{Y} = \Diamond_n^+(\mathbf{U}_1 * \mathbf{V}, \cdots, \mathbf{U}_n * \mathbf{V})$. It suffices to prove that the $\mathbf{X}$-languages are $\mathbf{Y}$-languages. By Theorem 2, every $\mathbf{X}$-language of A^* is a positive boolean combination of $\mathbf{V}$-languages and of languages of the form $\sigma_\varphi^{-1}(L)$, where $\varphi : A^* \to N$ is a morphism from A^* into some monoid $N \in \mathbf{V}$, $\sigma_\varphi : A^* \to (N \times A)^*$ is the sequential function associated with φ and L is a language of $\Diamond_n^+(\mathbf{U}_1, \cdots, \mathbf{U}_n)$. Since $\mathbf{V} \subseteq \mathbf{Y}$, the $\mathbf{V}$-languages are $\mathbf{Y}$-languages. Now, by Theorem 1, L is a positive boolean combination of languages of the form

$$L_0(m_1, a_1)L_1(m_2, a_2) \cdots (m_k, a_k)L_k \tag{3}$$

where $L_j \in \mathcal{U}_{i_j}((N \times A)^*)$, $(m_i, a_i) \in N \times A$ and $1 \leq i_0 < \cdots < i_k \leq n$. Since boolean operations commute with σ_φ^{-1}, it suffices to check that $\sigma_\varphi^{-1}(L)$ is a $\mathbf{Y}$-language when L is of the form (3). Furthermore

$$\sigma_\varphi^{-1}(L) = \bigcup_{m_{k+1} \in N} \left(\sigma_\varphi^{-1}(L) \cap \varphi^{-1}(m_{k+1})\right)$$

and by Lemma 1, $\sigma_\varphi^{-1}(L) \cap \varphi^{-1}(m_{k+1})$ can be written as $K_0 a_1 K_1 \cdots a_k K_k$, where $K_j = \sigma_\varphi^{-1}(\lambda_{n_j}^{-1}(L_j)) \cap \varphi^{-1}(n_j^{-1} m_{j+1})$ for $1 \leq j \leq k$.

Finally, L_j, and hence $\lambda_{n_j}^{-1}(L_j)$, is a $\mathbf{U}_{i_j}$-language. Now, $\varphi^{-1}(n_j^{-1} m_{j+1})$ is by construction a $\mathbf{V}$-language, and by Proposition 5, $\sigma_\varphi^{-1}(\lambda_{n_j}^{-1}(L_j))$ is a $(\mathbf{U}_{i_j} * \mathbf{V})$-language. It follows that K_j is also a $(\mathbf{U}_{i_j} * \mathbf{V})$-language and by Theorem 1 and formula 5, $\sigma_\varphi^{-1}(L)$ is a $\mathbf{Y}$-language.

Let us now conclude the proof of Theorem 6. We keep the notations of the proof of Theorem 7, with $\mathbf{V} = \mathbf{H}$. This theorem already gives the inclusion $\mathbf{X} \subseteq \mathbf{Y}$. To obtain the opposite inclusion, it suffices now to show that each $\mathbf{Y}$-language is a $\mathbf{X}$-language.

Let K be a $\mathbf{Y}$-language. Then K is recognized by an ordered monoid of the form $\Diamond_n^+(M_1 \circ G_1, \ldots, M_n \circ G_n)$, where $M_1, \ldots, M_n \in \mathbf{U}_n$ and $G_1, \ldots, G_n \in \mathbf{H}$. Let $G = G_1 \times \cdots \times G_n$. Then $G \in \mathbf{H}$, each G_i is a quotient of G, each $M_i \circ G_i$ divides $M_i \circ G$ and, thus by Proposition 4, $\Diamond_n^+(M_1 \circ G_1, \ldots, M_n \circ G_n)$ divides $\Diamond_n^+(M_1 \circ G, \ldots, M_n \circ G)$. By Proposition 1, K is also recognized by the latter ordered monoid, and, by Theorem 1, K is a positive boolean combination of languages of the form

$$K_0 a_1 K_1 \cdots a_k K_k$$

where $a_1, \cdots a_k \in A$, and K_j is recognized by $M_{i_j} \circ G$ for some sequence $1 \leq i_0 < i_1 < \cdots < i_k \leq n$. Now, by Theorem 2, K_j is a finite union of languages of the form $\sigma_\varphi^{-1}(L_j) \cap \varphi^{-1}(g_j)$ where $\varphi : A^* \to G$ is a morphism, $g_j \in G$, $\sigma_\varphi : A^* \to (G \times A)^*$ is the sequential function associated with φ and L_j is recognized by M_{i_j}. Using distributivity of product over union, we may thus suppose that $K_j = \sigma_\varphi^{-1}(L_j) \cap \varphi^{-1}(g_j)$ for $0 \le j \le k$. Set $n_0 = 1$, $m_1 = g_0$ and, for $1 \le j \le k$, $n_j = m_j\varphi(a_j)$ and $m_{j+1} = n_j g_j$.

Two special features of groups will be used now. First, if $g, h \in G$, the set $g^{-1}h$, computed in the monoid sense, is equal to $\{g^{-1}h\}$, where this time g^{-1} denotes the inverse of g. Next, each function λ_g is a bijection, and $\lambda_g^{-1} = \lambda_{g^{-1}}$. With these observations in mind, one gets

$$K_j = \sigma_\varphi^{-1}\Big(\lambda_{n_j}^{-1}\big(\lambda_{n_j^{-1}}^{-1}(L_j)\big)\Big) \cap \varphi^{-1}(n_j^{-1}m_{j+1})$$

whence, by the inversion formula,

$$K = \sigma_\varphi^{-1}\big(L'_0(m_1, a_1)L'_1(m_2, a_2)\cdots(m_k, a_k)L'_k\big) \cap \varphi^{-1}(m_{k+1})$$

where $L'_j = \lambda_{n_j^{-1}}^{-1}(L_j)$. Now, L'_j is recognized by M_{i_j}, and by Theorem 1, the language $L'_0(m_1, a_1)L'_1(m_2, a_2)\cdots(m_k, a_k)L'_k$ is recognized by $\diamondsuit_n^+(M_1, \ldots, M_n)$. It follows, by Proposition 2, that K is a **X**-language. $\square$

References

1. J. Almeida and B. Steinberg, *On the decidability of iterated semidirect products with applications to complexity*, preprint.
2. M. Arfi, Polynomial operations and rational languages, 4th STACS, Lect. Notes in Comp. Sci. **247**, Springer, (1987), 198–206.
3. M. Arfi, Opérations polynomiales et hiérarchies de concaténation, *Theoret. Comput. Sci.* **91**, (1991), 71–84.
4. B. Borchert, D. Kuske, F. Stephan, On existentially first-order definable languages and their relation to NP. *Proceedings of ICALP 1998, Lect. Notes in Comp. Sci.* , Springer Verlag, Berlin, Heidelberg, New York, (1998), this volume.
5. J. A. Brzozowski, Hierarchies of aperiodic languages, *RAIRO Inform. Théor.* **10**, (1976), 33–49.
6. J.A. Brzozowski and R. Knast, The dot-depth hierarchy of star-free languages is infinite, *J. Comput. System Sci.* **16**, (1978), 37–55.
7. S. Eilenberg, *Automata, languages and machines*, Vol. B, Academic Press, New York, 1976.
8. K. Henckell and J. Rhodes, The theorem of Knast, the $PG = BG$ and Type II Conjectures, in J. Rhodes (ed.) *Monoids and Semigroups with Applications*, Word Scientific, (1991), 453–463.
9. R. Knast, A semigroup characterization of dot-depth one languages, *RAIRO Inform. Théor.* **17**, (1983), 321–330.
10. R. Knast, Some theorems on graph congruences, *RAIRO Inform. Théor.* **17**, (1983), 331–342.
11. S. W. Margolis and J.E. Pin, Product of group languages, FCT Conference, Lect. Notes in Comp. Sci. **199**, (1985), 285–299.
12. J.-E. Pin, Hiérarchies de concaténation, *RAIRO Informatique Théorique* **18**, (1984), 23–46.
13. J.-E. Pin, *Variétés de langages formels*, Masson, Paris, 1984; English translation: *Varieties of formal languages*, Plenum, New-York, 1986.

14. J.-E. Pin, Logic on words, *Bulletin of the European Association of Theoretical Computer Science* **54**, (1994), 145–165.
15. J.-E. Pin, Finite semigroups and recognizable languages: an introduction, in NATO Advanced Study Institute *Semigroups, Formal Languages and Groups, J. Fountain (ed.)*, Kluwer academic publishers, (1995), 1–32.
16. J.-E. Pin, $\mathbf{BG} = \mathbf{PG}$, a success story, in NATO Advanced Study Institute *Semigroups, Formal Languages and Groups, J. Fountain (ed.)*, Kluwer academic publishers, (1995), 33–47.
17. J.-E. Pin, A variety theorem without complementation, *Izvestiya VUZ Matematika* **39** (1995) 80–90. English version, *Russian Mathem. (Iz. VUZ)* **39** (1995) 74–83.
18. J.-E. Pin, Logic, semigroups and automata on words, *Annals of Mathematics and Artificial Intelligence*, (1996), **16**, 343–384.
19. J.-E. Pin, Polynomial closure of group languages and open sets of the Hall topology, *Theoretical Computer Science* **169**, (1996), 185–200.
20. J.-E. Pin, Syntactic semigroups, in *Handbook of formal languages*, G. Rozenberg et A. Salomaa eds., vol. 1, ch. 10, pp. 679–746, Springer (1997).
21. J.-E. Pin and H. Straubing, 1981, Monoids of upper triangular matrices, *Colloquia Mathematica Societatis Janos Bolyai 39, Semigroups, Szeged*, 259–272.
22. J.-E. Pin and P. Weil, Polynomial closure and unambiguous product, *Theory Comput. Systems* **30**, (1997), 1–39.
23. Ch. Reutenauer, Sur les variétés de langages et de monoïdes, *Lect. Notes in Comp. Sci.* **67**, (1979) 260–265.
24. M.P. Schützenberger, On finite monoids having only trivial subgroups, *Information and Control* **8**, (1965), 190–194.
25. I. Simon, Piecewise testable events, *Proc. 2nd GI Conf., Lect. Notes in Comp. Sci.* **33**, Springer Verlag, Berlin, Heidelberg, New York, (1975), 214–222.
26. I. Simon, The product of rational languages, *Proceedings of ICALP 1993, Lect. Notes in Comp. Sci.* **700**, Springer Verlag, Berlin, Heidelberg, New York, (1993), 430–444.
27. H. Straubing, Families of recognizable sets corresponding to certain varieties of finite monoids, *J. Pure Appl. Algebra*, **15**, (1979), 305–318.
28. H. Straubing, A generalization of the Schützenberger product of finite monoids, *Theoret. Comp. Sci.* **13**, (1981), 137–150.
29. H. Straubing, Finite semigroups varieties of the form $\mathbf{U} * \mathbf{D}$, *J. Pure Appl. Algebra* **36**, (1985), 53–94.
30. H. Straubing, Semigroups and languages of dot-depth two, *Theoret. Comp. Sci.* **58**, (1988), 361–378.
31. H. Straubing, The wreath product and its application, in *Formal properties of finite automata and applications, J.-E. Pin (ed.), Lect. Notes in Comp. Sci.* **386**, Springer Verlag, Berlin, Heidelberg, New York, (1989), 15–24.
32. H. Straubing and P. Weil, On a conjecture concerning dot-depth two languages, *Theoret. Comp. Sci.* **104**, (1992), 161–183.
33. D. Thérien, Classification of finite monoids: the language approach, *Theoret. Comp. Sci.* **14**, (1981), 195–208.
34. W. Thomas, Classifying regular events in symbolic logic, *J. Comput. Syst. Sci* **25**, (1982), 360–375.
35. P. Weil, Closure of varieties of languages under products with counter, *J. Comput. System Sci.* **45**, (1992), 316–339.
36. P. Weil, Some results on the dot-depth hierarchy, *Semigroup Forum* **46**, (1993), 352–370.