

Formations of finite monoids and formal languages: Eilenberg’s variety theorem revisited

Adolfo Ballester-Bolinches, Jean-Éric Pin and Xaro Soler-Escrivà

Communicated by Karl Strambach

Abstract. We present an extension of Eilenberg’s variety theorem, a well-known result connecting algebra to formal languages. We prove that there is a bijective correspondence between formations of finite monoids and certain classes of languages, the formations of languages. Our result permits to treat classes of finite monoids which are not necessarily closed under taking submonoids, contrary to the original theory. We also prove a similar result for ordered monoids.

Keywords. Group formations, regular languages, semigroups, automata theory.

2010 Mathematics Subject Classification. 20D10, 20M35.

This paper is the first step of a programme aiming at exploring the connections between the formations of finite groups and regular languages. The starting point is Eilenberg’s variety theorem [10], a celebrated result of the 1970’s which underscores the importance of *varieties of finite monoids* (also called *pseudovarieties*) in the study of formal languages. Since varieties of finite groups are special cases of varieties of finite monoids, varieties seems to be a natural structure to study languages recognized by finite groups. However, in finite group theory, varieties are challenged by another notion. Although varieties are incontestably a central notion, many results are better formulated in the setting of *formations*. This raised the question whether Eilenberg’s variety theorem could be extended to a “formation theorem”.

The aim of this paper is to give a positive answer to this question. To our surprise, the resulting theorem holds not only for group formations but also for formations of finite monoids. We also prove a similar result for formations of ordered finite monoids, extending in this way a theorem of [17].

Before stating these results more precisely, let us say a word on the aforementioned research programme and give a brief overview of the existing literature. One

The authors are supported by Proyecto MTM2010-19938-C03-01 from MICINN (Spain). The second author is supported by the project ANR 2010 BLAN 0202 02 FREC. The third author was supported by the Grant PAID-02-09 from Universitat Politècnica de València.

of our ultimate goals would be to give a complete classification of the Hall varieties as defined by Steinberg in [25] and studied by Auinger and Steinberg [1–3]. Such a classification was given in [4] for varieties of finite supersolvable groups. A key tool of this paper is the operation $\mathbf{H} \rightarrow \mathbf{G}_p * \mathbf{H}$, where \mathbf{G}_p is the variety of finite p -groups and \mathbf{H} is a variety of groups. The second step of our programme [6] has been precisely to study this operation, and the corresponding operation on languages, when \mathbf{H} is a formation. The definition of a Hall variety can be readily extended to formations and as we said earlier, formations are a more flexible tool than varieties in finite group theory. Our hope is that it might be easier to describe the Hall formations than the Hall varieties.

Let us come back to the present paper. A *variety of finite monoids* is a class of finite monoids closed under taking submonoids, quotients and finite direct products. Eilenberg's theorem states that varieties of finite monoids are in bijection with certain classes of recognizable languages, the *varieties of languages*. The most famous instances of this correspondence are two early results of automata theory: star-free languages are associated with aperiodic monoids [20] and piecewise testable languages correspond to \mathcal{J} -trivial monoids [23]. But many more results are known and there is a rich literature on the subject. We refer the reader to [7, 18, 19] for an account of recent progress and a comprehensive bibliography. In the case of groups, only a few varieties of languages have been investigated. They correspond to the following varieties of finite groups: abelian groups [10], p -groups [10, 29, 30], nilpotent groups [10, 28], soluble groups [26, 30] and supersoluble groups [8].

Several attempts were made to extend Eilenberg's variety theory to a larger scope. For instance, ordered syntactic semigroups were introduced in [17]. The resulting extension of Eilenberg's variety theory permits to treat classes of languages that are not necessarily closed under complement, contrary to the original theory. Other extensions were developed independently by Straubing [27] and Ésik and Ito [11] and more recently by Gehrke, Grigorieff and Pin [13].

A formation of groups is a class of finite groups closed under taking quotients and subdirect products. The significance of formations in group theory is apparent since they are the first remarkable step in the development of a generalised Sylow theory. Thus it was Gaschütz who began his pioneering work on the subject in 1963 [12] with a paper which has become a classic. Since that time the subject has proliferated and has played a fundamental role in studying groups [5, 9]. To our knowledge, the notion of formations of finite algebras was considered for the first time in [21, 22, 24] and has never been used in finite semigroup theory.

Just as formations of finite monoids extend the notion of a variety of finite monoids, formations of languages are more general than varieties of languages. Like varieties, formations are classes of regular languages closed under Boolean

operations and quotients. But while varieties are closed under inverse of morphisms, formations of languages enjoy only a weak version of this property – see Property (F₂) – and thus comprise more general classes of languages than varieties. Nevertheless, our main result shows that an appropriate extension of Eilenberg’s variety theorem still holds for formations.

Our paper is organised as follows. Formations of monoids are introduced in Section 1. Section 2 gives the definitions and basic results on formal languages needed in this paper. Formations of languages are defined in Section 3. Our main result, the Formation Theorem, is presented in Section 4. Its counterpart for ordered monoids is the topic of Section 5. Instances of the Formation Theorem are given in Section 6: in particular, we give two descriptions of the formation of languages corresponding to the formation generated by the group A_5 , the alternating group of degree 5.

In this paper, all groups are finite. All monoids are either finite or free.

1 Formations of monoids

Recall that a monoid M is a *subdirect product* of the product of a family of monoids $(M_i)_{i \in I}$ if M is a submonoid of the direct product $\prod_{i \in I} M_i$ and if each induced projection π_i from M onto M_i is surjective. In this case, the projections *separate* the elements of M , in the sense that, if $\pi_i(x) = \pi_i(y)$ for all $i \in I$, then $x = y$. It is a well-known fact that this property characterizes subdirect products (see for instance [14, p. 78]).

Proposition 1.1. *A monoid M is a subdirect product of a family of monoids $(M_i)_{i \in I}$ if and only if there is a family of surjective morphisms $(M \rightarrow M_i)_{i \in I}$ which separate the elements of M .*

Being a subdirect product is a transitive relation, in the following sense:

Proposition 1.2. *Let M be a subdirect product of a family of monoids $(M_i)_{i \in I}$. Suppose that, for each $i \in I$, M_i is a subdirect product of a family $(M_{i,j})_{j \in I_i}$. Then M is a subdirect product of the family $(M_{i,j})_{i \in I, j \in I_i}$.*

Proof. The projections $\pi_{i,j} : M_i \rightarrow M_{i,j}$ separate the elements of M_i and the projections $\pi_i : M \rightarrow M_i$ separate the elements of M . It follows that the projections $\pi_{i,j} \circ \pi_i : M \rightarrow M_{i,j}$ separate the elements of M . \square

The next proposition states, in essence, that every subdirect product of quotients is a quotient of a subdirect product (see [22, the proof of Lemma 3.2]).

Proposition 1.3. *Let N be a subdirect product of a family of monoids $(M_i)_{i \in I}$. Suppose that, for each $i \in I$, M_i is the quotient of a monoid \widehat{M}_i . Then N is a quotient of a subdirect product of the family $(\widehat{M}_i)_{i \in I}$.*

Proof. We use the following notation. We denote by M the product of the family $(M_i)_{i \in I}$ and by $\pi_i : M \rightarrow M_i$ the projections. Similarly, \widehat{M} denotes the product of the family $(\widehat{M}_i)_{i \in I}$ and $\widehat{\pi}_i$ is the projection from \widehat{M} to \widehat{M}_i . For each $i \in I$, let $\gamma_i : \widehat{M}_i \rightarrow M_i$ be the quotient morphism and let $\gamma : \widehat{M} \rightarrow M$ be the product of these morphisms. Finally, let $\widehat{N} = \gamma^{-1}(N)$.

$$\begin{array}{ccc}
 \widehat{N} \subseteq \widehat{M} & \xrightarrow{\widehat{\pi}_i} & \widehat{M}_i \\
 \downarrow \gamma & & \downarrow \gamma_i \\
 N \subseteq M & \xrightarrow{\pi_i} & M_i
 \end{array}$$

By construction, \widehat{N} is a submonoid of \widehat{M} . To prove that \widehat{N} is a subdirect product, it suffices to verify that, for each $i \in I$ and for each $\widehat{m}_i \in \widehat{M}_i$, there is an element $\widehat{m} \in \widehat{N}$ such that $\widehat{\pi}_i(\widehat{m}) = \widehat{m}_i$. Let $m_i = \gamma_i(\widehat{m}_i)$. Since N is a subdirect product of the family $(M_i)_{i \in I}$, there is an element n of N such that $\pi_i(n) = m_i$. For $j \neq i$, let \widehat{m}_j be an element of \widehat{M}_j such that $\gamma_j(\widehat{m}_j) = \pi_j(n)$. Finally let \widehat{m} be the element of \widehat{M} defined by $\widehat{\pi}_i(\widehat{m}) = \widehat{m}_i$ and $\widehat{\pi}_j(\widehat{m}) = \widehat{m}_j$ for $j \neq i$. Then the formulas

$$\pi_i(\gamma(\widehat{m})) = \gamma_i(\widehat{\pi}_i(\widehat{m})) = \gamma_i(\widehat{m}_i) = m_i = \pi_i(n)$$

and for each $j \neq i$,

$$\pi_j(\gamma(\widehat{m})) = \gamma_j(\widehat{\pi}_j(\widehat{m})) = \gamma_j(\widehat{m}_j) = \pi_j(n)$$

show $\gamma(\widehat{m}) = n$. Consequently, \widehat{m} belongs to \widehat{N} and satisfies $\widehat{\pi}_i(\widehat{m}) = \widehat{m}_i$. □

The notion of a formation is a standard tool of group theory that has been extended to general algebraic systems by Shemetkov and Skiba [22]. However, its use in finite semigroup theory seems to be new.

A *formation of monoids* is a class of finite monoids \mathbf{F} satisfying the two conditions:

- (i) any quotient of a monoid of \mathbf{F} also belongs to \mathbf{F} ,
- (ii) the subdirect product of any finite family of monoids of \mathbf{F} is also in \mathbf{F} .

If S is a set of finite monoids, the formation *generated* by S is the smallest formation containing S . The following result is well-known for formations of groups

[9, II.2.2, p. 272] and was extended by Shemetkov and Skiba to general algebraic systems [22, Chapter I, Lemma 3.2] (see also [15]). For the convenience of the reader, we give here a self-contained proof in the case of monoids.

Proposition 1.4. *The formation generated by a class S of monoids consists of all quotients of subdirect products of members of S .*

Proof. Let \mathbf{F} be the class of all quotients of subdirect products of members of S . It suffices to prove that \mathbf{F} is a formation. It is clearly closed under quotient. Let us prove that it is closed under subdirect products. Let M be a subdirect product of a finite family $(M_i)_{i \in I}$ of monoids M_i of \mathbf{F} . Each M_i is a quotient of a subdirect product N_i of members of S . By Proposition 1.3, M is also a quotient of a subdirect product N of the monoids N_i . Now, Proposition 1.2 shows that N is a subdirect product of members of S . Therefore M belongs to \mathbf{F} . \square

A *variety* of finite monoids is a class of finite monoids closed under taking submonoids, quotients and finite direct products. It follows that a formation is a variety if and only if it is closed under taking submonoids. Therefore a formation is not necessarily a variety. For instance, the formation generated by A_5 is known to be the class of all direct products of copies of A_5 ([9, II.2.13]). Other very natural examples are given in the next proposition and its corollary.

Recall that a monoid M has a zero if there is an element 0 in M such that, for all $x \in M$, $x0 = 0 = 0x$.

Proposition 1.5. *Let \mathbf{F} be a formation of groups. The finite monoids whose minimal ideal is a group of \mathbf{F} constitute a formation, which is not a variety, even if \mathbf{F} is a variety of groups.*

Proof. Let \mathbf{E} be the class of finite monoids described in the proposition and let $M \in \mathbf{E}$. If N is a quotient of M , the minimal ideal of N is a quotient of the minimal ideal of M . Since M belongs to \mathbf{E} , its minimal ideal is a group of \mathbf{F} . It follows that the minimal ideal of N is also a group of \mathbf{F} . Consequently, \mathbf{E} is closed under quotients.

Let M be a subdirect product of a finite family $(M_i)_{i \in I}$ of monoids of \mathbf{E} . Then there is a family of surjective morphisms $(\pi_i : M \rightarrow M_i)_{i \in I}$ which separates the elements of M . Let I be the minimal ideal of M and let G_i be the minimal ideal of M_i . By definition of \mathbf{E} , G_i is a group of \mathbf{F} . Let e_i be the identity of G_i . If e is an idempotent of I , then $\pi_i(e)$ is an idempotent of G_i and thus is necessarily equal to e_i . Since the family $(\pi_i)_{i \in I}$ separates the elements of M , I contains a unique idempotent and hence is a group G . Each π_i induces a surjective group morphism

from G onto G_i and this family of morphisms separates the elements of G . Therefore G is a subdirect product of the G_i and thus belongs to \mathbf{F} . Consequently, M is in \mathbf{E} and hence \mathbf{E} is a formation of monoids.

We claim that the variety of finite monoids \mathbf{V} generated by \mathbf{E} is the variety of all monoids. Indeed, let M be a finite monoid. Then the monoid M^0 obtained by adjoining a zero to M belongs to \mathbf{E} and thus M , which is a submonoid of M^0 , belongs to \mathbf{V} . In particular, \mathbf{E} is not a variety of finite monoids. \square

Corollary 1.6. *Finite monoids with zero constitute a formation, which is not a variety of finite monoids.*

Proof. Apply Proposition 1.5 to the trivial formation of groups. \square

Corollary 1.6 gives rise to a large collection of examples of formations.

Corollary 1.7. *The monoids with zero of a given formation of monoids constitute a formation.*

2 Languages

A *language* is a subset of a free monoid A^* . Let us say that a monoid morphism $\varphi : A^* \rightarrow M$ *recognizes* a language L of A^* if there is a subset P of M such that $L = \varphi^{-1}(P)$. It is equivalent to say that L is saturated by φ , that is $L = \varphi^{-1}(\varphi(L))$. If φ is surjective, we say that φ *fully recognizes* L . By extension, one says that a language is [fully] *recognized by a monoid* M if there exists a morphism from A^* into M which [fully] recognizes L .

The results presented in the remainder of this section are more or less folklore (see [10, 16, 18] for references). However, we include their proofs for two reasons. First, to keep the article self-contained. Secondly, requiring all morphisms to be surjective induces some subtle differences with the standard statements, making references to the existing literature more difficult.

Let us start with an elementary but useful result.

Proposition 2.1. *Let L be a language of A^* and let $\varphi : A^* \rightarrow M$ be a morphism recognizing L . Then for each language R of A^* , one has $\varphi(L \cap R) = \varphi(L) \cap \varphi(R)$.*

Proof. The inclusion $\varphi(L \cap R) \subseteq \varphi(L) \cap \varphi(R)$ is clear. To prove the opposite inclusion, consider an element s of $\varphi(L) \cap \varphi(R)$. Then one has $s = \varphi(r)$ for some word $r \in R$. It follows that $r \in \varphi^{-1}(s)$, wherefore $r \in \varphi^{-1}(\varphi(L))$ and finally $r \in L$ since $\varphi^{-1}(\varphi(L)) = L$. Thus $r \in L \cap R$ and $s \in \varphi(L \cap R)$, which concludes the proof. \square

We shall also need an important consequence of the universal property of the free monoid (see [16, p. 10]).

Proposition 2.2. *Let $\eta : A^* \rightarrow M$ be a morphism and $\beta : N \rightarrow M$ be a surjective morphism. Then there exists a morphism $\varphi : A^* \rightarrow N$ such that $\eta = \beta \circ \varphi$.*

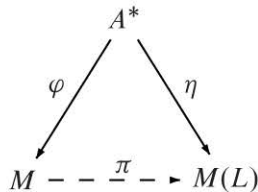
2.1 Syntactic morphism

Recall that the syntactic monoid of a language L of A^* is the quotient of A^* by the *syntactic congruence* of L , defined on A^* as follows: $u \sim_L v$ if and only if, for every $x, y \in A^*$,

$$xvy \in L \Leftrightarrow xuy \in L.$$

The natural morphism $\eta : A^* \rightarrow A^*/\sim_L$ is the *syntactic morphism* of L . Note that η fully recognizes L . Further, η has the following property.

Proposition 2.3. *Let L be a language of A^* and let $\eta : A^* \rightarrow M(L)$ be its syntactic morphism. A surjective morphism φ from A^* onto a monoid M fully recognizes L if and only if there is a surjective morphism $\pi : M \rightarrow M(L)$ such that $\eta = \pi \circ \varphi$.*



In other words, the syntactic monoid is a quotient of any monoid fully recognizing L and thus is the smallest monoid fully recognizing L .

More generally, given a subset P of a monoid M , the *syntactic congruence* of P is the congruence defined on M as follows: $u \sim_P v$ if and only if, for every $x, y \in M$,

$$xvy \in P \Leftrightarrow xuy \in P$$

The next result explains the behaviour of syntactic congruences under surjective morphisms.

Proposition 2.4. *Let $\pi : M \rightarrow N$ be a surjective morphism of monoids. Let P be a subset of N and let $Q = \pi^{-1}(P)$. Then for all $u, v \in M$, $u \sim_Q v$ if and only if $\pi(u) \sim_P \pi(v)$.*

Proof. Suppose that $u \sim_Q v$ and let us prove that $\pi(u) \sim_P \pi(v)$. By symmetry, it suffices to prove that, for all $(s, t) \in N \times N$, the condition $s\pi(u)t \in P$ implies $s\pi(v)t \in P$. Since π is surjective, one has $s = \pi(x)$ and $t = \pi(y)$ for some $x, y \in M$. Therefore if $s\pi(u)t \in P$, one gets $\pi(xuy) = \pi(x)\pi(u)\pi(y) \in P$ and thus $xuy \in Q$. Since $u \sim_Q v$, it follows that $xvy \in Q$ and thus $\pi(xvy) = s\pi(v)t \in P$. Thus $\pi(u) \sim_P \pi(v)$.

Suppose now that $\pi(u) \sim_P \pi(v)$ and let us prove that $u \sim_Q v$. By symmetry, it suffices to prove, for all $(x, y) \in M \times M$, that $xuy \in Q$ implies $xvy \in Q$. Suppose that $xuy \in Q$. Then

$$\pi(xuy) = \pi(x)\pi(u)\pi(y) \in \pi(Q) = P.$$

Since $\pi(u) \sim_P \pi(v)$, it follows that

$$\pi(x)\pi(v)\pi(y) \in P,$$

which gives $xvy \in \pi^{-1}(P) = Q$. Thus $u \sim_Q v$. \square

A subset P of M is called *disjunctive* if the congruence \sim_P is the equality relation. It follows readily from the definition that a subset P of M is disjunctive if and only if its complement in M is disjunctive. We shall need the following elementary proposition.

Proposition 2.5. *Let P be a disjunctive subset of a monoid M and $\gamma : A^* \rightarrow M$ be a surjective morphism. Then γ is the syntactic morphism of the language $\gamma^{-1}(P)$.*

Proof. Let $L = \gamma^{-1}(P)$. By construction, γ fully recognizes L . Therefore, the condition $\gamma(u) = \gamma(v)$ implies $u \sim_L v$. Conversely, if $u \sim_L v$, then $\gamma(u) \sim_P \gamma(v)$ by Proposition 2.4 and thus $\gamma(u) = \gamma(v)$. This proves that γ is the syntactic morphism of L . \square

A frequently asked question is whether any monoid is the syntactic monoid of some language. The answer is negative in the general case (see Example 2.8 below), but it is positive for groups.

Proposition 2.6. *Let π be a surjective morphism from A^* onto a group G . Then π is the syntactic morphism of the language $\pi^{-1}(1)$.*

Proof. Let $L = \pi^{-1}(1)$. We claim that G is the syntactic group of L . Indeed, let $u, v \in G^*$ and suppose that $u \sim_L v$. Then, for every $x, y \in G^*$, $\pi(xuy) = 1$ if and only if $\pi(xvy) = 1$. Take for y the empty word and for x a word such that $\pi(x) = \pi(u)^{-1}$. We get $\pi(xuy) = 1$ and hence $\pi(xvy) = 1$. But $\pi(xvy) = \pi(xv) = \pi(x)\pi(v) = \pi(u)^{-1}\pi(v)$ and thus $\pi(u) = \pi(v)$. This proves the claim and the proposition. \square

Corollary 2.7. *Every group is the syntactic monoid of some language.*

Example 2.8. Consider the 4-element monoid $M = \{1, a, b, c\}$, where 1 is the identity and the multiplication is defined by $xy = y$ for all $x, y \in M \setminus \{1\}$. Then M contains no disjunctive subset and hence cannot be the syntactic monoid of any language.

2.2 Operations on languages

Simple operations on languages have a natural algebraic counterpart. We study in this order complement, intersection, union, inverse of surjective morphisms and left and right quotients. We denote by L^c the complement of a language L of A^* .

Proposition 2.9. *Let L be a language of A^* . If L is fully recognized by a monoid M , then L^c is also fully recognized by M .*

Proof. Let $\varphi : A^* \rightarrow M$ be a morphism [fully] recognizing L and let $P = \varphi(L)$. Then $L = \varphi^{-1}(P)$ and hence $A^* \setminus L = \varphi^{-1}(M \setminus P)$. Thus M recognizes L^c . \square

Let $(M_i)_{1 \leq i \leq n}$ be a family of monoids and, for $1 \leq i \leq n$, let $\varphi_i : A^* \rightarrow M_i$ be a surjective monoid morphism. The product of these morphisms is the surjective morphism

$$\varphi : A^* \rightarrow \text{Im}(\varphi) \subseteq M_1 \times \cdots \times M_n$$

defined by $\varphi(x) = (\varphi_1(x), \dots, \varphi_n(x))$.

Proposition 2.10. *The monoid $\text{Im}(\varphi)$ is a subdirect product of the family of monoids $(M_i)_{1 \leq i \leq n}$.*

Proof. Let $M = \text{Im}(\varphi)$. By construction, M is a submonoid of the direct product $M_1 \times \cdots \times M_n$. Let $\pi_i : M \rightarrow M_i$ be the natural projection. One has by construction $\varphi_i = \pi_i \circ \varphi$ and thus π_i is surjective. It follows that M is a subdirect product. \square

Proposition 2.11. *Let L_1, \dots, L_n be languages of A^* and let, for $1 \leq i \leq n$, M_i be a monoid fully recognizing L_i . Then the sets $\bigcap_{1 \leq i \leq n} L_i$ and $\bigcup_{1 \leq i \leq n} L_i$ are fully recognized by a subdirect product of the monoids M_i .*

Proof. By hypothesis, each language L_i is fully recognized by a morphism φ_i from A^* onto a monoid M_i . Setting $P_i = \varphi_i(L_i)$, one gets $L_i = \varphi_i^{-1}(P_i)$. Let $\varphi : A^* \rightarrow M$ be the product of these morphisms. Proposition 2.10 shows that M

is a subdirect product of the monoids M_i . Then the formula

$$\bigcap_{1 \leq i \leq n} L_i = \varphi^{-1}((P_1 \times \dots \times P_n) \cap M)$$

shows that the intersection $\bigcap_{1 \leq i \leq n} L_i$ is fully recognized by M . Since union and intersection interchange under complementation, Proposition 2.9 shows that M also fully recognizes the set $\bigcup_{1 \leq i \leq n} L_i$. □

Corollary 2.12. *Let L_1, \dots, L_n be languages of A^* and let, for $1 \leq i \leq n$, M_i be the syntactic monoid of L_i . Then the syntactic monoid of $\bigcap_{1 \leq i \leq n} L_i$ and $\bigcup_{1 \leq i \leq n} L_i$ is a quotient of a subdirect product of the monoids M_i .*

Proof. By Proposition 2.11, the languages $\bigcap_{1 \leq i \leq n} L_i$ and $\bigcup_{1 \leq i \leq n} L_i$ are fully recognized by a subdirect product M of the monoids M_i , and by Proposition 2.3, their syntactic monoid is a quotient of M . □

Proposition 2.13. *Let $\alpha : A^* \rightarrow B^*$ be a monoid morphism and let L be a language of B^* recognized by a morphism φ from B^* onto a monoid M . Then $\varphi \circ \alpha$ recognizes the language $\alpha^{-1}(L)$. In particular, if φ is the syntactic morphism of L , then $\varphi \circ \alpha$ is the syntactic morphism of $\alpha^{-1}(L)$.*

Proof. Since φ recognizes L , one has $L = \varphi^{-1}(\varphi(L))$ and hence

$$\alpha^{-1}(L) = \alpha^{-1}(\varphi^{-1}(\varphi(L))) = (\varphi \circ \alpha)^{-1}(\varphi(L)).$$

If $\varphi \circ \alpha$ is surjective, it fully recognizes $\alpha^{-1}(L)$.

The second part of the statement follows from Proposition 2.5. □

Recall that, for each subset X of a monoid M and for each element s of M , the left [right] quotient $s^{-1}X$ [Xs^{-1}] of X by s is defined as follows:

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

More generally, for any subset K of M , the left [right] quotient $K^{-1}X$ [XK^{-1}] of X by K is

$$K^{-1}X = \bigcup_{s \in K} s^{-1}X = \{t \in M \mid \text{there exists } s \in K \text{ such that } st \in X\},$$

$$XK^{-1} = \bigcup_{s \in K} Xs^{-1} = \{t \in M \mid \text{there exists } s \in K \text{ such that } ts \in X\}.$$

Proposition 2.14. *If a morphism fully recognizes a language L of A^* , it also fully recognizes the languages $K^{-1}L$ and LK^{-1} for every language K of A^* .*

Proof. Let φ be a morphism from A^* into a monoid M [fully] recognizing L and let $P = \varphi(L)$ and $R = \varphi(K)$. We claim that $\varphi^{-1}(R^{-1}P) = K^{-1}L$. Indeed, one has the following sequence of equivalent statements:

$$\begin{aligned}
 u \in \varphi^{-1}(R^{-1}P) &\iff \varphi(u) \in R^{-1}P \\
 &\iff \text{there exists } r \in R \text{ such that } r\varphi(u) \in P \\
 &\iff \text{there exists } k \in K \text{ such that } \varphi(k)\varphi(u) \in P \\
 &\iff \text{there exists } k \in K \text{ such that } ku \in \varphi^{-1}(P) \\
 &\iff \text{there exists } k \in K \text{ such that } ku \in L \\
 &\iff u \in K^{-1}L.
 \end{aligned}$$

Thus φ [fully] recognizes $K^{-1}L$. A similar proof works for LK^{-1} . □

2.3 Regular languages

A language is *recognizable* (or *regular*) if it is recognized by a finite deterministic automaton. This is equivalent to saying that the language is recognized by some finite monoid or that its syntactic monoid is finite. It is a well-known fact that a language is regular if and only if it has finitely many left (or right) quotients.

As a preparation to our main result, we now prove two results interesting on their own right. The first one follows essentially from [10, formula (4.2), p. 199].

Proposition 2.15. *If L is a recognizable language, every language recognized by the syntactic morphism of L belongs to the Boolean algebra generated by the quotients of L .*

Proof. Let $\eta : A^* \rightarrow M$ be the syntactic morphism of L . Since L is recognizable, M is finite. Let $P = \eta(L)$ and let u be an element of M . We claim that

$$\{u\} = \bigcap_{\{(x,y) \in M^2 \mid xuy \in P\}} x^{-1}Py^{-1} \setminus \bigcup_{\{(x,y) \in M^2 \mid xuy \notin P\}} x^{-1}Py^{-1}. \tag{2.1}$$

Let R be the right hand side of (2.1). It is clear that u belongs to R . Let now r be an element of R . Then, by construction, the conditions $xuy \in P$ and $xry \in P$ are equivalent. It follows that $u \sim_P r$. Since M is the syntactic monoid of L , the

syntactic congruence \sim_P is the equality. Thus $r = u$, which proves (2.1). Since η^{-1} commutes with Boolean operations and quotients, $\eta^{-1}(u)$ is a Boolean combination of quotients of L .

If K is a language recognized by η , then $K = \eta^{-1}(Q)$ with $Q = \eta(K)$. Therefore, since

$$K = \eta^{-1}(Q) = \bigcup_{u \in Q} \eta^{-1}(u),$$

the language K belongs to the Boolean algebra generated by the quotients of L . \square

Proposition 2.16. *Let $\varphi : A^* \rightarrow M$ be a surjective morphism and let, for each $s \in M$, M_s be the syntactic monoid of the language $\varphi^{-1}(s)$. Then each M_s is a quotient of M and M is subdirect product of the monoids M_s .*

Proof. Let $s \in M$ and let $\eta_s : A^* \rightarrow M_s$ be the syntactic morphism of $\varphi^{-1}(s)$. Since φ fully recognizes $\varphi^{-1}(s)$, there exists by Proposition 2.3 a surjective morphism $\pi_s : M \rightarrow M_s$ such that $\eta_s = \pi_s \circ \varphi$.

We claim that the projections π_s separate the elements of M . Let $x, y \in M$ and let u, v be words of A^* such that $\varphi(u) = x$ and $\varphi(v) = y$. If $\pi_s(x) = \pi_s(y)$, then $\eta_s(u) = \eta_s(v)$ and thus $u \sim_{\varphi^{-1}(s)} v$. It follows by Proposition 2.4 that

$$\varphi(u) \sim_{\{s\}} \varphi(v),$$

that is $x \sim_{\{s\}} y$. This property holds for all $s \in M$ and for $s = x$, gives in particular $x \sim_{\{x\}} y$. Since $1x1 \in \{x\}$, one gets $1y1 \in \{x\}$, that is $x = y$, which proves the claim. Therefore M is subdirect product of the monoids M_s . \square

3 Formations of languages

A class of regular languages \mathcal{C} associates with each finite alphabet A a set $\mathcal{C}(A^*)$ of regular languages of A^* . A *formation of languages* is a class of regular languages \mathcal{F} satisfying the following conditions:

- (F₁) for each alphabet A , $\mathcal{F}(A^*)$ is closed under Boolean operations and quotients,
- (F₂) if L is a language of $\mathcal{F}(B^*)$ and $\eta : B^* \rightarrow M$ denotes its syntactic morphism, then for each monoid morphism $\alpha : A^* \rightarrow B^*$ such that $\eta \circ \alpha$ is surjective, the language $\alpha^{-1}(L)$ belongs to $\mathcal{F}(A^*)$.

Observe that a formation of languages is closed under inverse of surjective morphisms, but this condition is not equivalent to (F₂). However, one could also use

another condition:

(F'₂) if L is a language of $\mathcal{F}(B^*)$ and $\varphi : B^* \rightarrow M$ is a morphism fully recognizing L , then for each monoid morphism $\alpha : A^* \rightarrow B^*$ such that $\varphi \circ \alpha$ is surjective, the language $\alpha^{-1}(L)$ belongs to $\mathcal{F}(A^*)$.

Proposition 3.1. *Conditions (F₂) and (F'₂) are equivalent.*

Proof. Since the syntactic morphism of a language fully recognizes this language, it is clear that (F'₂) implies (F₂).

Suppose that (F₂) holds and let $\varphi : B^* \rightarrow M$ be a morphism fully recognizing a language L . Let $\eta : B^* \rightarrow N$ be the syntactic morphism of L . By Proposition 2.3, there is a surjective morphism $\pi : M \rightarrow N$ such that $\eta = \pi \circ \varphi$. Let $\alpha : A^* \rightarrow B^*$ be a monoid morphism such that $\varphi \circ \alpha$ is surjective. Then $\eta \circ \alpha = \pi \circ (\varphi \circ \alpha)$ and thus $\eta \circ \alpha$ is surjective. It follows by (F₂) that $\alpha^{-1}(L)$ belongs to $\mathcal{F}(A^*)$, which proves (F'₂). \square

Let us now give an alternative definition of a formation of languages.

Proposition 3.2. *A class of regular languages \mathcal{F} is a formation of languages if and only if it satisfies conditions (F₁) and (F₃):*

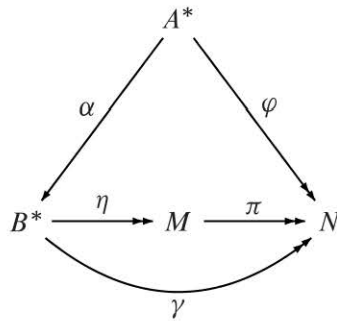
(F₃) if L is a language of $\mathcal{F}(B^*)$ and K is a language of A^* whose syntactic monoid is a quotient of the syntactic monoid of L , then K belongs to $\mathcal{F}(A^*)$.

Proof. Let \mathcal{F} be a class of regular languages. We first show that if \mathcal{F} satisfies (F₁) and (F₃), then it also satisfies (F₂). Let L be a language of $\mathcal{F}(B^*)$ and let $\eta : B^* \rightarrow M$ be its syntactic morphism. Let $\alpha : A^* \rightarrow B^*$ be a morphism such that $\eta \circ \alpha$ is surjective. By Proposition 2.13, the morphism $\eta \circ \alpha$ is the syntactic morphism of $\alpha^{-1}(L)$ and it follows from (F₃) that $\alpha^{-1}(L)$ belongs to $\mathcal{F}(A^*)$.

Let us show now that if \mathcal{F} satisfies (F₁) and (F₂), it also satisfies condition (F₃). Let L be a language of $\mathcal{F}(B^*)$ and let $\eta : B^* \rightarrow M$ be its syntactic morphism. Let $\pi : M \rightarrow N$ be a surjective morphism and let K be a language of A^* whose syntactic monoid is N . Let $\varphi : A^* \rightarrow N$ be the syntactic morphism of K and let $\gamma = \pi \circ \eta$. Finally, let

$$R = \eta^{-1}(\pi^{-1}(\varphi(K))) = \gamma^{-1}(\varphi(K)).$$

By construction, this language is recognized by η and by Proposition 2.15, it belongs to the Boolean algebra generated by the quotients of L . It follows by (F₁) that R belongs to $\mathcal{F}(B^*)$. Since A^* is a free monoid and since γ is surjective, there exists a morphism $\alpha : A^* \rightarrow B^*$ such that $\gamma \circ \alpha = \varphi$. Our notation is summarized in the diagram below. A double-head arrow indicates a surjective morphism.



Since φ is the syntactic morphism of K , the set $\varphi(K)$ is disjunctive in N and by Proposition 2.5, γ is the syntactic morphism of R . Since $\gamma \circ \alpha$ is equal to φ , it is surjective and Condition (F₂) shows that $\alpha^{-1}(R)$ belongs to $\mathcal{F}(A^*)$. But

$$\alpha^{-1}(R) = \alpha^{-1}(\gamma^{-1}(\varphi(K))) = \varphi^{-1}(\varphi(K)) = K.$$

Thus K belongs to $\mathcal{F}(A^*)$, which proves (F₃). □

Notice that a similar result holds for varieties of languages in Eilenberg’s sense. Recall that a variety of languages is a class \mathcal{V} of regular languages satisfying the following conditions:

- (V₁) for each alphabet A , $\mathcal{V}(A^*)$ is closed under Boolean operations and quotients,
- (V₂) if L is a language of $\mathcal{V}(B^*)$, then for each monoid morphism $\alpha : A^* \rightarrow B^*$, the language $\alpha^{-1}(L)$ belongs to $\mathcal{V}(A^*)$.

The counterpart of Proposition 3.2 for varieties is the following:

Proposition 3.3. *A class of regular languages \mathcal{V} is a variety of languages if and only if it satisfies conditions (V₁) and (V₃):*

- (V₃) if L is a language of $\mathcal{V}(B^*)$ and K is a language of A^* whose syntactic monoid divides the syntactic monoid of L , then K belongs to $\mathcal{V}(A^*)$.

Proof. Suppose that \mathcal{V} is a variety of languages and let \mathbf{V} be the associated variety of monoids. If L is a language of $\mathcal{V}(B^*)$, then its syntactic monoid M belongs to \mathbf{V} . Therefore, if the syntactic monoid of K divides M , it also belongs to \mathbf{V} and by the Variety Theorem, K belongs to $\mathcal{V}(A^*)$.

Suppose now that \mathcal{V} is a class of languages satisfying (V₁) and (V₃). We claim that \mathcal{V} satisfies (V₂). Let L be a language of $\mathcal{V}(B^*)$ and let $\alpha : A^* \rightarrow B^*$ be a monoid morphism. Then the syntactic monoid of $\alpha^{-1}(L)$ divides that of L and thus by (V₃), the language $\alpha^{-1}(L)$ belongs to $\mathcal{V}(A^*)$. □

4 The Formation Theorem

To each formation of monoids \mathbf{F} , let us associate the class of languages $\mathcal{F}(\mathbf{F})$ defined as follows: for each alphabet A , $\mathcal{F}(\mathbf{F})(A^*)$ is the set of languages of A^* fully recognized by some monoid of \mathbf{F} , or, equivalently, whose syntactic monoid belongs to \mathbf{F} .

Proposition 4.1. *If \mathbf{F} is a formation of monoids, then $\mathcal{F}(\mathbf{F})$ is a formation of languages.*

Proof. Propositions 2.9, 2.11 and 2.14 show that, for each alphabet A , the set $\mathcal{F}(\mathbf{F})(A^*)$ is closed under Boolean operations and quotients. Proposition 2.13 shows that the second condition defining a formation of languages is also satisfied. \square

We are now ready to state the main result of this section. Given a formation of languages \mathcal{F} , let us denote by $\mathbf{F}(\mathcal{F})$ the formation of monoids generated by the syntactic monoids of the languages of \mathcal{F} .

Theorem 4.2 (Formation Theorem). *The correspondences $\mathbf{F} \rightarrow \mathcal{F}(\mathbf{F})$ and $\mathcal{F} \rightarrow \mathbf{F}(\mathcal{F})$ are two mutually inverse, order preserving, bijections between formations of monoids and formations of languages.*

Proof. We first prove that $\mathbf{F}(\mathcal{F}(\mathbf{F})) = \mathbf{F}$. Let $\mathbf{F}' = \mathbf{F}(\mathcal{F}(\mathbf{F}))$. Let M be a monoid of \mathbf{F} and let $\varphi : A^* \rightarrow M$ be a surjective morphism. By Proposition 2.16, M is a subdirect product of the syntactic monoids M_s of the languages $\varphi^{-1}(s)$, for $s \in M$. Each M_s is a quotient of M and thus belongs to \mathbf{F} . It follows that $\varphi^{-1}(s)$ belongs to $\mathcal{F}(\mathbf{F})(A^*)$ and that M_s belongs to \mathbf{F}' . Consequently, M belongs to \mathbf{F}' . This proves the inclusion $\mathbf{F} \subseteq \mathbf{F}'$.

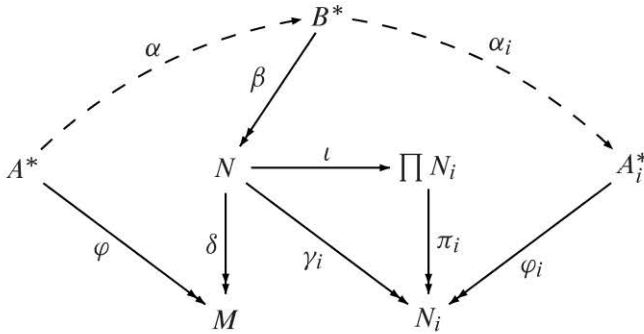
To prove the opposite inclusion, consider a monoid M of \mathbf{F}' . Then M is a quotient of a subdirect product of a finite family $(M_i)_{1 \leq i \leq n}$ of syntactic monoids of languages of $\mathcal{F}(\mathbf{F})$. Since a language belongs to $\mathcal{F}(\mathbf{F})$ if and only if its syntactic monoid belongs to \mathbf{F} , each M_i belongs to \mathbf{F} and thus M belongs to \mathbf{F} . Thus $\mathbf{F}' = \mathbf{F}$.

We now prove that $\mathcal{F}(\mathbf{F}(\mathcal{F})) = \mathcal{F}$. Let $\mathcal{F}' = \mathcal{F}(\mathbf{F}(\mathcal{F}))$. If L is a language of \mathcal{F} , then its syntactic monoid belongs to $\mathbf{F}(\mathcal{F})$ by definition and thus L belongs to \mathcal{F}' . This proves the inclusion $\mathcal{F} \subseteq \mathcal{F}'$.

Consider now a language L of $\mathcal{F}'(A^*)$ and let $\varphi : A^* \rightarrow M$ be its syntactic morphism. Since M belongs to $\mathbf{F}(\mathcal{F})$, it is a quotient of a subdirect product N of a finite family $(N_i)_{i \in I}$ of monoids, each monoid N_i being the syntactic monoid of a language L_i of $\mathcal{F}(A_i^*)$, for some alphabet A_i . Let $\varphi_i : A_i^* \rightarrow N_i$ be its syntactic morphism. Let δ be the surjective morphism from N onto M and ι the injective

morphism from N into $\prod_{i \in I} N_i$. We also denote by π_i the natural projection from $\prod_{i \in I} N_i$ onto N_i and we set $\gamma_i = \pi_i \circ \iota$. Since N is a subdirect product, each map γ_i is surjective. Finally, let $\beta : B^* \rightarrow N$ be a surjective monoid morphism.

Our notation is summarized in the diagram below.



Since φ_i is surjective, Proposition 2.2 shows that there is a monoid morphism $\alpha_i : B^* \rightarrow A_i^*$ such that $\varphi_i \circ \alpha_i = \gamma_i \circ \beta$. Then α_i is not necessary surjective, but $\varphi_i \circ \alpha_i$ is surjective. By the same type of argument, there is a monoid morphism $\alpha : A^* \rightarrow B^*$ such that $\varphi = \delta \circ \beta \circ \alpha$.

We need to show that L belongs to $\mathcal{F}(A^*)$. Let $P = \varphi(L)$. Since φ is the syntactic morphism of L , one has

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

Let $Q = \delta^{-1}(P)$ and let $m = (m_i)_{i \in I}$ be an element of N . Then the following formula holds:

$$\beta^{-1}(m) = \bigcap_{i \in I} \beta^{-1}(\gamma_i^{-1}(m_i)) = \bigcap_{i \in I} \alpha_i^{-1}(\varphi_i^{-1}(m_i)). \tag{4.1}$$

Now, since φ_i is the syntactic morphism of the language L_i , Proposition 2.15 shows that the language $\varphi_i^{-1}(m_i)$ belongs to the Boolean algebra generated by the quotients of L_i . Since L_i belongs to $\mathcal{F}(A_i^*)$, one also has $\varphi_i^{-1}(m_i) \in \mathcal{F}(A_i^*)$. Since the morphism $\varphi_i \circ \alpha_i$ is equal to $\gamma_i \circ \beta$, it is surjective and by (F₂) the language $\alpha_i^{-1}(\varphi_i^{-1}(m_i))$ belongs to $\mathcal{F}(B^*)$. It follows now from (4.1) and (F₁) that $\beta^{-1}(m)$ belongs to $\mathcal{F}(B^*)$. Further, since

$$\beta^{-1}(\delta^{-1}(P)) = \beta^{-1}(Q) = \bigcup_{m \in Q} \beta^{-1}(m),$$

one gets

$$\beta^{-1}(\delta^{-1}(P)) \in \mathcal{F}(B^*).$$

Finally,

$$L = \varphi^{-1}(P) = \alpha^{-1}(\beta^{-1}(\delta^{-1}(P)))$$

and since the morphism $\delta \circ \beta \circ \alpha$ is equal to φ , it is surjective, and thus by (F'_2) the language L belongs to $\mathcal{F}(A^*)$. Thus $\mathcal{F} = \mathcal{F}'$. \square

As a consequence of the previous theorem, we obtain Eilenberg’s variety theorem.

Corollary 4.3. *The restriction of the correspondences $\mathbf{F} \rightarrow \mathcal{F}(\mathbf{F})$ and $\mathcal{F} \rightarrow \mathbf{F}(\mathcal{F})$ to varieties of monoids and to varieties of languages are two mutually inverse bijections.*

Proof. Let \mathbf{V} be a variety of monoids. Then $\mathcal{V}(\mathbf{V})$ is a formation of languages. To show that $\mathcal{V}(\mathbf{V})$ is a variety of languages it is enough to see that $\mathcal{V}(\mathbf{V})$ satisfies condition (V_3) . Let L be a language of $\mathcal{V}(\mathbf{V})(B^*)$. Then the syntactic monoid $M(L)$ of L is in \mathbf{V} . Now, if K is a language of A^* whose syntactic monoid $M(K)$ divides $M(L)$, then $M(K) \in \mathbf{V}$. It follows that $K \in \mathcal{V}(\mathbf{V})(A^*)$.

Consider now a variety of languages \mathcal{V} . Let us prove that $\mathbf{V}(\mathcal{V})$ is a variety of monoids. Let M be a monoid of $\mathbf{V}(\mathcal{V})$ and let T be a submonoid of M . Let $\varphi : B^* \rightarrow M$ and $\tau : A^* \rightarrow T$ be two surjective monoid morphisms. By Proposition 2.2, there is a morphism $\alpha : A^* \rightarrow B^*$ such that $\iota \circ \tau = \varphi \circ \alpha$, where ι denotes the inclusion map from T into M . For each $t \in T$ the following equalities hold:

$$\tau^{-1}(t) = \tau^{-1}(\iota^{-1}(t)) = \alpha^{-1}(\varphi^{-1}(t)).$$

On the other hand, $\varphi^{-1}(t)$ is a language of B^* fully recognized by M and M belongs to $\mathbf{V}(\mathcal{V})$. Thus, by Theorem 4.2, $\varphi^{-1}(t) \in \mathcal{V}(\mathbf{V}(\mathcal{V}))(B^*) = \mathcal{V}(B^*)$. Now, since \mathcal{V} is a variety of languages, we obtain that $\alpha^{-1}(\varphi^{-1}(t)) \in \mathcal{V}(A^*)$. Therefore, the syntactic monoid M_t of $\tau^{-1}(t)$ belongs to $\mathbf{V}(\mathcal{V})$, for each $t \in T$. Since by Proposition 2.16, T is a subdirect product of the family of monoids $(M_t)_{t \in T}$, we conclude that $T \in \mathbf{V}(\mathcal{V})$. \square

5 Positive formations and ordered monoids

A generalization of Eilenberg’s variety theorem was proposed by the second author in [17]. This result provides a bijective correspondence between the varieties of finite ordered monoids and the so-called positive varieties of languages. We show in this section that the Formation Theorem can be extended in a similar way. To keep this paper to a reasonable size, we give only the definitions required to state the main result. The proof can be readily adapted from the proof of Theorem 4.2 by using the arguments of [17]. The reader is referred to [17–19] for more details on positive varieties.

5.1 Positive formations of languages

A *positive variety of languages* is defined by relaxing the definition of a variety of languages: only positive Boolean operations (union and intersection) are allowed – no complement. It is therefore natural to define a *positive formation of languages* as a class of regular languages \mathcal{F} satisfying (F_2) and

(F_1^+) for each alphabet A , $\mathcal{F}(A^*)$ is closed under finite union, finite intersection and quotients.

Examples of positive formations of languages will be given in Section 6.

5.2 Ordered monoids

An ordered monoid is a monoid equipped with a stable partial order relation, usually denoted by \leq . A *morphism* of ordered monoids is a morphism of monoids which preserves orders.

A subset I of an ordered monoid is an *order ideal* if $x \in I$ and $y \leq x$ imply $y \in I$.

We say that (S, \leq) is an *ordered submonoid* of an ordered monoid (T, \leq) if S is a submonoid of T and the order on S is the restriction to S of the order on T . Similarly, (T, \leq) is a *quotient* of (S, \leq) if there exists a surjective morphism of ordered monoids $\varphi : (S, \leq) \rightarrow (T, \leq)$.

Given a family $(M_i)_{i \in I}$ of ordered monoids, the *product* $\prod_{i \in I} M_i$ is the ordered monoid defined on the set $\prod_{i \in I} M_i$ by the law $(s_i)_{i \in I} (s'_i)_{i \in I} = (s_i s'_i)_{i \in I}$ and the order given by $(s_i)_{i \in I} \leq (s'_i)_{i \in I}$ if and only if, for all $i \in I$, $s_i \leq s'_i$.

An ordered monoid M is a *subdirect product* of a family of ordered monoids $(M_i)_{i \in I}$ if M is an ordered submonoid of the product $\prod_{i \in I} M_i$ and if each induced projection from M onto M_i is surjective.

A *formation of ordered monoids* is a class of finite ordered monoids closed under taking quotients and finite subdirect products.

5.3 Recognition by ordered monoids

Let L be a language of A^* and let M be an ordered monoid. Then L is *fully recognized* by M if and only if there exist an order ideal I of M and a surjective monoid morphism φ from A^* into M such that $L = \varphi^{-1}(I)$.

Let $\eta : A^* \rightarrow M$ be the syntactic morphism of L and let $P = \eta(L)$. The *syntactic order* \leq_P is defined on M as follows: $u \leq_P v$ if and only if for all $x, y \in M$,

$$xvy \in P \Rightarrow xuy \in P.$$

The partial order \leq_P is stable and the resulting ordered monoid (M, \leq_P) is called the *ordered syntactic monoid* of L .

5.4 The Positive Formation Theorem

To each formation of ordered monoids \mathbf{F} , let us associate the class of languages $\mathcal{F}(\mathbf{F})$ defined as follows: for each alphabet A , $\mathcal{F}(\mathbf{F})(A^*)$ is the set of languages of A^* fully recognized by some ordered monoid of \mathbf{F} , or, equivalently, whose ordered syntactic monoid belongs to \mathbf{F} .

Proposition 5.1. *If \mathbf{F} is a formation of ordered monoids, then $\mathcal{F}(\mathbf{F})$ is a positive formation of languages.*

Given a positive formation of languages \mathcal{F} , let us denote by $\mathbf{F}(\mathcal{F})$ the formation of ordered monoids generated by the ordered syntactic monoids of the languages of \mathcal{F} . We are now ready to state the Positive Formation Theorem:

Theorem 5.2 (Positive Formation Theorem). *The correspondences $\mathbf{F} \rightarrow \mathcal{F}(\mathbf{F})$ and $\mathcal{F} \rightarrow \mathbf{F}(\mathcal{F})$ are two mutually inverse, order preserving, bijections between formations of ordered monoids and positive formations of languages.*

6 Examples

This section presents three instances of the [Positive] Formation Theorem. The first two examples were first considered in [13]. The third example is related to group theory.

6.1 Languages with zero and nondense languages

A *language with zero* is a language whose syntactic monoid has a zero, or equivalently, a language recognized by a monoid with zero. By Corollary 1.6, finite monoids with zero constitute a formation. The Formation Theorem now gives immediately:

Proposition 6.1. *The class of recognizable languages with zero is a formation of languages.*

In particular languages with zero form a Boolean algebra.

A language L of A^* is *dense* if, for every word $u \in A^*$, $L \cap A^*uA^* \neq \emptyset$. The language A^* is called the *full language*. The class \mathcal{ND} of regular nondense or full languages was first considered in [13].

Proposition 6.2. *The class \mathcal{ND} is a positive formation of languages.*

Proof. Let L_1 and L_2 be two nondense languages of A^* . Then there exist two words $u_1, u_2 \in A^*$ such that $L_1 \cap A^*u_1A^* = \emptyset$ and $L_2 \cap A^*u_2A^* = \emptyset$. It follows

that

$$(L_1 \cap L_2) \cap A^*u_1A^* = \emptyset \quad \text{and} \quad (L_1 \cup L_2) \cap A^*u_1u_2A^* = \emptyset.$$

Thus $L_1 \cap L_2$ and $L_1 \cup L_2$ are nondense. If $L_1 = A^*$, then $L_1 \cap L_2 = L_2$ and $L_1 \cup L_2 = A^*$. Thus $\mathcal{ND}(A^*)$ is closed under finite union and finite intersection.

Let L be a nondense language. Then there exists a word $u \in A^*$ such that $L \cap A^*uA^* = \emptyset$. Let $x, y \in A^*$. We claim that $x^{-1}Ly^{-1} \cap A^*uA^* = \emptyset$. Otherwise, there exist two words s, t such that $sut \in x^{-1}Ly^{-1}$. Therefore $xsuty \in L$, a contradiction, since $L \cap A^*uA^* = \emptyset$. Thus $x^{-1}Ly^{-1}$ is nondense. If $L = A^*$, then $x^{-1}Ly^{-1} = A^*$ for all words $x, y \in A^*$. Therefore $\mathcal{ND}(A^*)$ is closed under quotients.

Let L be a language of $\mathcal{ND}(B^*)$ and let $\eta : B^* \rightarrow M$ denotes its syntactic morphism. Let $\alpha : A^* \rightarrow B^*$ be a monoid morphism such that $\eta \circ \alpha$ is surjective. If L is the full language B^* , then $\alpha^{-1}(B^*)$ is the full language A^* . If L is nondense, there exists a word $u \in B^*$ such that $B^*uB^* \cap L = \emptyset$. Let $x = \eta(u)$. Since η fully recognizes L , one has by Proposition 2.1

$$\emptyset = \eta(B^*uB^* \cap L) = \eta(B^*uB^*) \cap \eta(L) = MxM \cap \eta(L).$$

Since $\eta \circ \alpha$ is surjective, there is a word $v \in A^*$ such that $\eta(\alpha(v)) = x$. We claim that $A^*vA^* \cap \alpha^{-1}(L) = \emptyset$. Indeed suppose that $A^*vA^* \cap \alpha^{-1}(L)$ contains a word w . Then $\alpha(w) \in L$ and thus $\eta(\alpha(w)) \in \eta(L)$. Furthermore, one has

$$\eta(\alpha(w)) \in \eta(\alpha(A^*vA^*)) = MxM.$$

This leads to a contradiction since $MxM \cap \eta(L)$ is empty. Thus $\alpha^{-1}(L)$ is nondense and \mathcal{ND} satisfies (F_2) . □

Theorem 9.2 of [13] can now be rephrased as follows, via the Positive Formation Theorem:

Proposition 6.3. *The formation of ordered monoids corresponding to \mathcal{ND} consists of all finite ordered monoids with 0 in which 0 is the top element of the order.*

6.2 The formation generated by A_5

Let \mathbf{F} be the formation generated by A_5 , the alternating group of degree 5, and let \mathcal{F} be the associated formation of languages. By [9, II.2.13] \mathbf{F} is known to be the class of all direct products of copies of A_5 .

By definition, a language belongs to \mathcal{F} if and only if its syntactic monoid is a group of \mathbf{F} . Therefore, a language L of A^* is in $\mathcal{F}(\mathbf{F})(A^*)$ if and only if its syntactic monoid is a direct product of copies of A_5 .

The group A_5 can be generated for instance by one of the sets $A = \{a, b\}$ or $B = \{c, d, e\}$, where a, b, c, d and e are the permutations of the set $\{1, 2, 3, 4, 5\}$ defined as follows:

$$a = c = (1\ 2\ 3), \quad b = (2\ 4)(3\ 5), \quad d = (1\ 4\ 2), \quad e = (1\ 5\ 2).$$

These two sets of generators define the automata \mathcal{A} and \mathcal{B} represented in Figures 1 and 2. Taking 1 as initial and unique final state, a simple computation shows that \mathcal{A} recognizes the language of A^*

$$K = (b + a(ba^*b)^*a(ba^*b)^*a)^*$$

and that \mathcal{B} recognizes the language of B^*

$$L = (c(d + e)^*cB + d(c + e)^*dB + e(c + d)^*eB)^*.$$

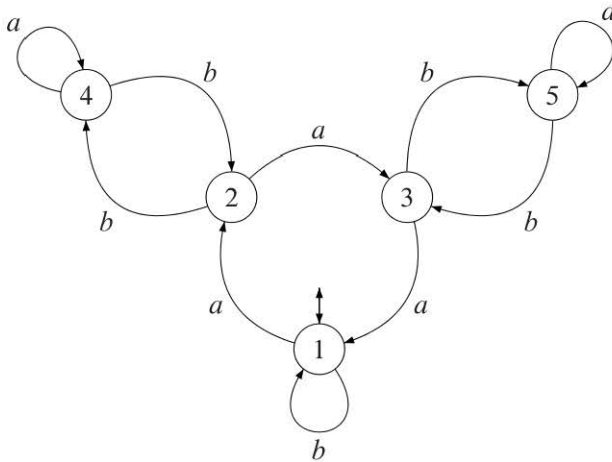


Figure 1. The minimal automaton \mathcal{A} of K .

By construction, \mathcal{F} is the formation of languages generated by K , or by L . Therefore, one should be able to express K from L (and L from K) by using (F_1) and (F_2) . This is actually quite simple.

Let $\varphi : A^* \rightarrow A_5$ be the syntactic morphism of K and let $\psi : B^* \rightarrow A_5$ be the syntactic morphism of L . Let also $\alpha : A^* \rightarrow B^*$ and $\beta : B^* \rightarrow A^*$ be the morphisms defined respectively by $\alpha(a) = c$ and $\alpha(b) = cdec^2$ and by $\beta(c) = a$, $\beta(d) = aba^2baba$ and $\beta(e) = a^2baba^2b$. A short computation shows that

$$\varphi = \psi \circ \alpha \quad \text{and} \quad \psi = \varphi \circ \beta. \tag{6.1}$$

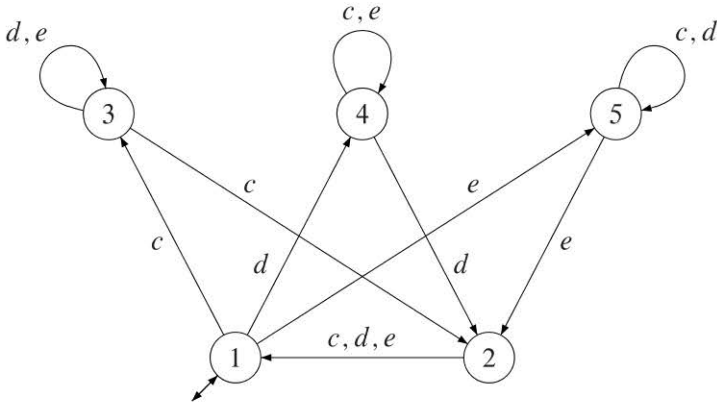


Figure 2. The minimal automaton \mathcal{B} of L .

Now, one has $\varphi(K) = \psi(L)$ and this subset P of A_5 consists of all permutations fixing the element 1. It follows that K and L are related by the formulas

$$\alpha^{-1}(L) = \alpha^{-1}(\psi^{-1}(P)) = (\psi \circ \alpha)^{-1}(P) = \varphi^{-1}(P) = K,$$

$$\beta^{-1}(K) = \beta^{-1}(\varphi^{-1}(P)) = (\varphi \circ \beta)^{-1}(P) = \psi^{-1}(P) = L.$$

Further (6.1) shows that $\alpha [\beta]$ satisfies (F_2) with respect to $L [K]$. This gives another proof that K and L generate the same formation of languages.

Let \mathbf{V} be the variety of groups generated by A_5 and let \mathcal{V} be the associated variety of languages. The cyclic group C_2 is a subgroup of A_5 and thus belongs to \mathbf{V} . But the description of \mathbf{F} given above shows that C_2 is not in \mathbf{F} . It follows that the language $(A^2)^*$ of all words of even length of A^* is in $\mathcal{V}(A^*)$, since its syntactic monoid is equal to C_2 , but is not in $\mathcal{F}(A^*)$. It would be a challenge to prove this result without the Formation Theorem.

7 Conclusion

We proved a Formation Theorem which extends Eilenberg’s variety theorem. This result allows one to study classes of regular languages which do not form varieties of languages, and in particular, languages recognized by groups belonging to a given formation. Indeed, many formations which are not varieties arise naturally in the structural study of the groups. For instance, given a class \mathbf{X} of simple non-abelian groups, the class of groups with all composition factors in \mathbf{X} is a formation which is not a variety. If attention is focused on soluble groups, the class of all soluble groups whose 2-chief factors are not central is a formation which is not

subgroup-closed either. On the other hand, there are important results which are well-known for varieties of groups and are still open for a general formation. One of the most remarkable examples concerns with the formation or variety generated by a group. It is known that a variety generated by a group contains only finitely many subvarieties [24]. The corresponding problem for formations is one of the most famous open questions in the theory. It would be interesting to explore the language theoretic counterpart of this problem.

Bibliography

- [1] K. Auinger and B. Steinberg, The geometry of profinite graphs with applications to free groups and finite monoids, *Trans. Amer. Math. Soc.* **356** (2004), 805–851 (electronic).
- [2] K. Auinger and B. Steinberg, Constructing divisions into power groups, *Theoret. Comput. Sci.* **341** (2005), 1–21.
- [3] K. Auinger and B. Steinberg, On power groups and embedding theorems for relatively free profinite monoids, *Math. Proc. Cambridge Philos. Soc.* **138** (2005), 211–232.
- [4] K. Auinger and B. Steinberg, Varieties of finite supersolvable groups with the M. Hall property, *Math. Ann.* **335** (2006), 853–877.
- [5] A. Ballester-Bolinches and L. M. Ezquerro, *Classes of Finite Groups*, Mathematics and Its Applications 584, Springer-Verlag, Dordrecht, 2006.
- [6] A. Ballester-Bolinches, J.-É. Pin and X. Soler-Escrivà, Languages associated with saturated formations of groups, preprint (2012).
- [7] M. J. J. Branco, Varieties of languages, in: *Semigroups, Algorithms, Automata and Languages* (Coimbra 2001), World Scientific, River Edge (2002), 91–132.
- [8] O. Carton, J.-É. Pin and X. Soler-Escrivà, Languages recognized by finite supersolvable groups, *J. Autom. Lang. Comb.* **14** (2009), 149–161.
- [9] K. Doerk and T. Hawkes, *Finite Soluble Groups*, de Gruyter Expositions in Mathematics 4, Walter de Gruyter, Berlin, 1992.
- [10] S. Eilenberg, *Automata, Languages, and Machines. Volume B*, Academic Press, New York, 1976.
- [11] Z. Ésik and M. Ito, Temporal logic with cyclic counting and the degree of aperiodicity of finite automata, *Acta Cybernet.* **16** (2003), 1–28.
- [12] W. Gaschütz and U. Lubeseder, Kennzeichnung gesättigter Formationen, *Math. Z.* **82** (1963), 198–199.

- [13] M. Gehrke, S. Grigorieff and J.-É. Pin, Duality and equational theory of regular languages, in: *Automata, Languages and Programming – ICALP 2008* (Reykjavik 2008), Part 2, Lecture Notes in Computer Science 5126, Springer-Verlag, Berlin (2008), 246–257.
- [14] P. A. Grillet, *Semigroups, An Introduction to the Structure Theory*, Marcel Dekker, New York, 1995.
- [15] W. Guo and K. P. Shum, Formation operators on classes of algebras, *Comm. Algebra* **30** (2002), 3457–3472.
- [16] J.-É. Pin, *Varieties of Formal Languages*, North Oxford, London, and Plenum, New York, 1986.
- [17] J.-E. Pin, A variety theorem without complementation, *Russian Math. (Iz. VUZ)* **39** (1995), 80–90.
- [18] J.-É. Pin, Syntactic semigroups, in: *Handbook of Formal Languages*, Volume 1, Springer-Verlag, Berlin (1997), 679–746.
- [19] J.-É. Pin, Profinite methods in automata theory, in: *26th International Symposium on Theoretical Aspects of Computer Science (STACS 2009)*, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl (2009), 31–50.
- [20] M.-P. Schützenberger, On finite monoids having only trivial subgroups, *Inf. Control* **8** (1965), 190–194.
- [21] L. A. Shemetkov, Product of formations of algebraic systems, *Algebra Logic* **23** (1984), 484–490.
- [22] L. A. Shemetkov and A. N. Skiba, *Formations of Algebraic Systems* (in Russian), *Sovremennaya Algebra*, Nauka, Moscow, 1989.
- [23] I. Simon, Piecewise testable events, in: *Automata Theory and Formal Languages* Lecture Notes in Computer Science 33, Springer-Verlag, Berlin (1975), 214–222.
- [24] A. N. Skiba, *Finite Subformations of Varieties of Algebraic Systems* (in Russian), *Problems in Algebra 2*, “Universitet-skoe”, Minsk, 1986, 7–20, 126.
- [25] B. Steinberg, Inverse automata and profinite topologies on a free group, *J. Pure Appl. Algebra* **167** (2002), 341–359.
- [26] H. Straubing, Families of recognizable sets corresponding to certain varieties of finite monoids, *J. Pure Appl. Algebra* **15** (1979), 305–318.
- [27] H. Straubing, On logical descriptions of regular languages, in: *LATIN 2002*, Lecture Notes in Computer Science 2286, Springer-Verlag, Berlin (2002), 528–538.
- [28] D. Thérien, Subword counting and nilpotent groups, in: *Combinatorics on Words* (Waterloo 1982), Academic Press, Toronto (1983), 297–305.

- [29] P. Weil, An extension of the Schützenberger product, in: *Lattices, Semigroups, and Universal Algebra* (Lisbon 1988), Plenum, New York (1990), 315–321.
- [30] P. Weil, Products of languages with counter, *Theoret. Comput. Sci.* **76** (1990), 251–260.

Received September 19, 2011; revised May 13, 2012.

Author information

Adolfo Ballester-Bolinches, Departament d'Àlgebra, Universitat de València,
C/ Dr. Moliner, 50 46100-Burjassot (València), Spain.

Jean-Éric Pin, LIAFA, Université Paris VII and CNRS, Case 7014,
F-75205 Paris Cedex 13, France.

Xaro Soler-Escrivà, Dpt. d'Estadística i Investigació Operativa, Universitat d'Alacant,
Sant Vicent del Raspeig, Ap. Correus 99, E-03080 Alacant, Spain.