Duality and equational theory of regular languages

Mai Gehrke[*] Serge Grigorieff[†] Jean-Éric Pin[‡]

To appear in ICALP 2008

This paper presents a new result in the equational theory of regular languages, which emerged from lively discussions between the authors about Stone and Priestley duality. Let us call *lattice of languages* a class of regular languages closed under finite intersection and finite union. The main results of this paper (Theorems 5.2 and 6.1) can be summarized in a nutshell as follows:

> *A set of regular languages is a lattice of languages if and only if it can be defined by a set of profinite equations.*
>
> *The product on profinite words is the dual of the residuation operations on regular languages.*

In their more general form, our equations are of the form $u \to v$, where u and v are profinite words. The first result not only subsumes Eilenberg-Reiterman's theory of varieties and their subsequent extensions, but it shows for instance that any class of regular languages defined by a fragment of logic closed under conjunctions and disjunctions (first order, monadic second order, temporal, etc.) admits an equational description. In particular, the celebrated McNaughton-Schützenberger characterisation of first order definable languages by the aperiodicity condition $x^\omega = x^{\omega+1}$, far from being an isolated statement, now appears as an elegant instance of a very general result.

How is this equational theory related to duality? The connection between profinite words and Stone spaces was already discovered by Almeida [2], [3, Theorem 3.6.1], but Pippenger [14] was the first to formulate it in terms of Stone duality. Almeida (implicitely) and Pippenger (explicitely) both observed that the Boolean algebra of regular languages over A^* is dual to the Stone space $\widehat{A^*}$, the set of profinite words. Pippenger actually came very close to our first result, since he mentioned that this duality extends to a one-to-one correspondence between Boolean algebras of regular languages and quotients of $\widehat{A^*}$. Our first result is the full-fledged consequence of the similar one-to-one correspondence for all lattices of languages provided by Priestley duality.

However, this link to duality theory is in fact much stronger and encompasses not only the underlying lattices and spaces involved but also the algebraic operations including the product of profinite words. That is the content of our

[*]Radboud University Nijmegen, The Netherlands

[†]LIAFA, University Paris-Diderot and CNRS, France.

[‡]The authors acknowledge support from the AutoMathA programme of the European Science Foundation.

second result. It means that the profinite semigroup structure, in its entirety, is a dual structure and thus the entire theory is a special case of duality theory. In particular, the deep and highly evolved theory of duality and relational semantics from modal logic applies, and, in the other direction, the wealth of knowledge and examples from semigroup theory enriches our understanding of general duality theory. In this sense, the results described here are just the tip of an iceberg yet to be explored.

Due to the lack of space, most of the proofs are omitted.

1 Historical background

Our starting point was Eilenberg's variety theorem [7]. Recall that a *variety of languages* is a class of regular languages closed under Boolean operations, inverses of morphisms and left and right quotients by words. Eilenberg's theorem states that varieties of languages are in one-to-one correspondence with *varieties of finite monoids*, that is, classes of finite monoids closed under taking submonoids, quotient monoids and finite direct products.

The notion of a variety of finite monoids is similar to that of variety of monoids introduced by Birkhoff: a *variety of monoids* is a class of monoids closed under taking submonoids, quotient monoids and direct products. Birkhoff proved in [6] that his varieties can be characterized by sets of identities: for instance the identity $xy = yx$ characterizes the variety of commutative monoids. Almost fifty years later, Reiterman [18] extended Birkhoff's theorem to varieties of finite monoids: any variety of finite monoids can be characterized by a set of profinite identities. A *profinite identity* is an identity between two profinite words. The precise definition of profinite words will be given in Section 2, but they can be viewed as limits of sequences of words for a certain metric, the profinite metric. For instance, one can show that the sequence $x^{n!}$ converges to a profinite word denoted by x^ω and the variety of finite aperiodic monoids can be defined by the identity $x^\omega = x^{\omega+1}$.

Eilenberg's and Reiterman's theorems have been extended several times over the last twenty years by relaxing the definition of a variety of languages. In [11], the third author considered *positive varieties*, for which the closure under complement is not required and showed they correspond to varieties of finite ordered monoids. The counterpart of Reiterman's theorem, obtained by Pin-Weil [13], makes use of identities of the form $u \leqslant v$, where u and v are profinite words.

Pippenger [14] proposed to relax another condition by introducing *strains of languages*, which share the same properties as varieties of languages except for the closure under quotients by words, which is not required. Finally, Straubing [21] and independently, Esik [8], relaxed the closure under inverses of morphisms. Esik just required the closure under inverses of length-preserving morphisms. Straubing considered a class $\mathcal{C}$ of morphisms between free monoids containing the length-preserving morphisms and closed under composition and called $\mathcal{C}$-variety a class of regular languages closed under Boolean operations, quotients and inverses of morphisms from the class $\mathcal{C}$. The counterpart of Reiterman's theorem for this case was given by Kunc [10] (see also [12]).

2 Profinite topology

In this paper, A denotes a *finite* alphabet. A morphism $\varphi : A^* \to M$ *separates* two words u and v of A^* if $\varphi(u) \neq \varphi(v)$. By extension, we say that a monoid M *separates* two words if there is a morphism from A^* onto M that separates them. One can show that two distinct words can always be separated by a finite monoid. Given two words $u, v \in A^*$, we set

$$\begin{aligned} r(u,v) &= \min\{|M| \mid M \text{ is a monoid that separates } u \text{ and } v\} \\ d(u,v) &= 2^{-r(u,v)} \end{aligned}$$

with the usual conventions $\min \emptyset = +\infty$ and $2^{-\infty} = 0$. One can show that d is an *ultrametric*, that is, satisfies the following properties, for all $u, v, w \in A^*$,

(1) $d(u,v) = 0$ if and only if $u = v$,

(2) $d(u,v) = d(v,u)$,

(3) $d(u,w) \leqslant \max\{d(u,v), d(v,w)\}$.

Moreover, the relations $d(uv, u'v') \leqslant \max\{d(u,u'), d(v,v')\}$ hold for all $u, u', v, v' \in A^*$, so that the concatenation product on A^* is uniformly continuous.

Thus (A^*, d) is a metric space. Its completion, denoted by $\widehat{A^*}$, is called the *free profinite monoid* on A and its elements are called *profinite words*.

We now briefly review the main properties of $\widehat{A^*}$. The reader is referred to [22, 4] for more details. First, $\widehat{A^*}$ is compact. Second, the topology defined by d is the *profinite topology*, that is, the least topology which makes continuous every morphism from A^* onto a finite monoid (considered as a discrete metric space). It follows that every morphism φ from A^* onto a finite monoid F extends uniquely to a (uniformly) continuous morphism $\hat\varphi : \widehat{A^*} \to F$. Thirdly, since the product on A^* is uniformly continuous, it can be extended in a unique way to a uniformly continuous product on $\widehat{A^*}$. This product makes $\widehat{A^*}$ a monoid.

Recall that a set is *clopen* if it is both open and closed. There is a strong connection between clopen sets of $\widehat{A^*}$ and regular languages of A^*. Indeed, a language L is regular if and only if $\overline{L}$ is clopen in $\widehat{A^*}$ and $L = \overline{L} \cap A^*$ [4]. The languages of the form $\overline{L}$, where L is a regular language, actually form a basis for the topology and hence $\widehat{A^*}$ is *zero-dimensional*. It is also *totally disconnected* since its connected components are singletons.

What about sequences? First, every profinite word is the limit of some converging sequence of words. Next, a sequence of profinite words $(u_n)_{n \geqslant 0}$ is converging to a profinite word u if and only if, for every morphism φ from A^* onto a finite monoid, $\hat\varphi(u_n)$ is ultimately equal to $\hat\varphi(u)$.

For instance, if u is a word (or even a profinite word), one can prove that the sequence $u^{n!}$ is converging. Its limit is denoted by u^ω for the following reason: if φ is a morphism from A^* onto a finite monoid M, the sequence $\hat\varphi(u)^{n!}$ is ultimately equal to the unique idempotent power of $\hat\varphi(u)$, which is traditionally denoted by $\hat\varphi(u)^\omega$ in semigroup theory. Thus the notation u^ω is justified by the formula $\hat\varphi(u^\omega) = \hat\varphi(u)^\omega$.

The closure in $\widehat{A^*}$ of a regular language of A^* can be characterized as follows.

Proposition 2.1 *Let L be a regular language of A^* and let $u \in \widehat{A^*}$. The following conditions are equivalent:*

(1) $u \in \overline{L}$,

(2) $\hat{\varphi}(u) \in \varphi(L)$, *for all morphisms* φ *from* A^* *onto a finite monoid,*

(3) $\hat{\varphi}(u) \in \varphi(L)$, *for some morphism* φ *from* A^* *onto a finite monoid that recognizes* L,

(4) $\hat{\eta}(u) \in \eta(L)$, *where* η *is the syntactic morphism of* L.

3 Duality for distributive lattices

In Stone duality, the dual space of a bounded distributive lattice D is based on the set S_D of prime filters of D. As identified already by Birkhoff, there is a lattice embedding e of D into $\mathcal{P}(S_D)$, defined by:

$$e(d) \text{ is the set of prime filters containing } d.$$

A description of the range of e, both for Boolean algebras and then for distributive lattices was first provided by Stone [19, 20]. He showed that if one generates a topology on the space of prime filters with the sets in the image of the embedding e, then the resulting space is, in the Boolean case, a compact 0-dimensional space, and in the distributive lattice case a *spectral space*, i.e. a compact (not necessarily Hausdorff), sober space with a ring of compact-open sets as a basis. An answer in complete lattice theoretic terms is the result by Jónsson and Tarski on canonical extensions. This is the most advantageous point of view when considering additional structure on lattices and spaces such as the semigroup operation.

For distributive lattices, Priestley [16] gave a slightly different topological characterization of the range of e than Stone. If one generates a topology τ, not just with the sets in the range of e, but also with their complements, one obtains the dual space of the free Boolean extension of the lattice and, crucially, one may reconstruct the original lattice if one remembers, in addition to the dual space of the free Boolean extension of the lattice, also the inclusion order on the space of prime filters. Thus in Priestley duality the dual of a distributive lattice is the ordered topological space $(S_D, \subseteq, \tau)$. It is characterized by the property that it is compact and totally order disconnected. An ordered topological space is *totally order disconnected* provided the points of the space are separated by the upwards saturated clopen subsets. This is the duality we will use here.

One of the most powerful facts about dualities is that we get a complete correspondence between subobjects on one side and quotients on the other. Here we are interested in sublattices of regular languages, and these will of course correspond, under Priestley duality, to Priestley space quotients or equivalently, certain compatible preorders on the dual space of the lattice of all regular languages. Working out this correspondence dates back to work by M. E. Adams [1]. If D is a subalgebra of B, we obtain a dual quotient $S_B \twoheadrightarrow S_D$ by mapping a prime filter p of B to $p \cap D$. The topological condition that is needed is that the quotient is a continuous (and, in the DL case, order preserving) map. An equivalence relation (preorder for DL subalgebras) on the space S_B corresponds to a subalgebra provided the clopen subsets that are saturated with respect to the equivalence relation (preorder for DL subalgebras) separate the equivalence classes (of the equivalence relation corresponding to the preorder in the DL case).

4 Duality applied to $\mathrm{Reg}(A^*)$

The proof that the dual space of $\mathrm{Reg}(A^*)$ is none other than the space $\widehat{A^*}$ of profinite words can be found in Pippenger's paper [14]. It relies on two facts. First, given a prime filter p of $\mathrm{Reg}(A^*)$, there is a unique profinite word u such that, for every morphism from A^* onto a finite monoid, $\varphi(u)$ is the unique element m of M such that $\varphi^{-1}(m) \in p$. In the opposite direction, if u is a profinite word, the set

$$\begin{aligned} p_u = \{L \in \mathrm{Reg}(A^*) \mid \varphi^{-1}(\hat{\varphi}(u)) \subseteq L \\ \text{for some morphism } \varphi \text{ from } A^* \text{ onto a finite monoid } \} \end{aligned} \tag{1}$$

is a prime filter of $\mathrm{Reg}(A^*)$.

Theorem 4.1 (See [14]) *The topological space underlying the profinite completion $\widehat{A^*}$ is equal to the dual space of the Boolean algebra $\mathrm{Reg}(A^*)$. Furthermore, the canonical embedding is given by the topologial closure: $e(L) = \overline{L}$.*

5 Equational characterization of lattices

Formally, a *profinite equation* is a pair (u, v) of profinite words of $\widehat{A^*}$. We also use the term *explicit equation* when both u and v are words of A^*. We say that a regular language L of A^* *satisfies the profinite equation* $u \to v$ (or $v \leftarrow u$) if the condition $u \in \overline{L}$ implies $v \in \overline{L}$. Proposition 2.1 immediately gives some equivalent definitions:

Corollary 5.1 *Let L be a regular language of A^*, let η be its syntactic morphism and let φ be any morphism onto a finite monoid recognizing L. The following conditions are equivalent:*

(1) *L satisfies the equation $u \to v$,*
(2) *$\hat{\eta}(u) \in \eta(L)$ implies $\hat{\eta}(v) \in \eta(L)$,*
(3) *$\hat{\varphi}(u) \in \varphi(L)$ implies $\hat{\varphi}(v) \in \varphi(L)$.*

Given a set E of equations of the form $u \to v$, the set of all regular languages of A^* satisfying all the equations of E is called the set of languages *defined by* E. It is not hard to see that the set of languages defined by a set E of equations is a lattice. Our first result states that the converse is true as well.

Theorem 5.2 *A set of regular languages of A^* is a lattice of languages if and only if it can be defined by a set of equations of the form $u \to v$, where $u, v \in \widehat{A^*}$.*

Proof. The proof is an instantiation of the duality between sublattices of $\mathrm{Reg}(A^*)$ and preorders on its dual space $\widehat{A^*}$. Given a lattice D of regular languages, we get dually a quotient map $q_D : \widehat{A^*} \twoheadrightarrow S_D$ given by $p_u \mapsto p_u \cap D$, where p_u is defined by Formula (1). Equivalently, we may describe this quotient map by the preorder Q_D on $\widehat{A^*}$ given by $u \; Q_D \; v$ if and only if $q_D(p_u) \subseteq q_D(p_v)$. But this latter condition is equivalent to requiring that, for all $L \in D$, $u \in \overline{L}$ implies

$v \in \overline{L}$. That is, in our terminology, the preorder on $\widehat{A^*}$ determining the quotient dual to D is exactly the equational theory of D:

$$Q_D = \{(u, v) \mid \text{for all } L \in D \ (L \text{ satisfies } u \to v)\}.$$

On the other hand, in the duality, given a preorder Q on $\widehat{A^*}$ giving rise to a Priestley quotient $\widehat{A^*}/Q$, the corresponding lattice is the set of all $L \in \mathrm{Reg}(A^*)$ so that their representation $\overline{L}$ is saturated with respect to the preorder. That is, $u \in \overline{L}$ implies $v \in \overline{L}$ for all $(u, v) \in Q$. But, by our earlier definition, this is exactly what we call the set of languages defined by Q if we identify each pair (u, v) in Q with the corresponding equation $u \to v$.

Since, coming from D, going to the preorder Q_D, and then going back to the set of languages defined by Q_D under duality gives us back D, we see that D is the set of languages defined by Q_D. $\square$

Writing $u \leftrightarrow v$ for ($u \to v$ and $v \to u$), we get an equational description of the Boolean algebras of languages.

Corollary 5.3 *A set of regular languages of A^* is a Boolean algebra of languages if and only if it can be defined by a set of equations of the form $u \leftrightarrow v$, where $u, v \in \widehat{A^*}$.*

6 Duality for quotienting operations

As announced in the introduction, our second main result is that the product on $\widehat{A^*}$ itself is dual to operations on $\mathrm{Reg}(A^*)$. The pertinent operations are the *residuals* of the product of languages, $\backslash$ and $/$, defined, for all $L, M, N \in \mathrm{Reg}(A^*)$, by the conditions

$$LM \subseteq N \iff M \subseteq L\backslash N \iff L \subseteq N/M.$$

More explicitely, the *right* and *left residuals* of N by M are given by:

$$M\backslash N = \{u \in A^* \mid Mu \subseteq N\} = \{u \in A^* \mid \text{ for all } v \in M,\, vu \in N\}$$
$$N/M = \{u \in A^* \mid uM \subseteq N\} = \{u \in A^* \mid \text{ for all } v \in M,\, uv \in N\}.$$

In extended Priestley duality [9], the additional operations are captured by additional relational structure on the dual space. A well-known case of this is the capture of a modality on the dual frame by its binary Kripke relation. More generally, n-ary relations on lattices are captured by $(n + 1)$-ary relations on their dual spaces. Remarkably, in the case of the algebra $(\mathrm{Reg}(A^*), \backslash, /)$, the dual relation common to the two additional operations is functional and turns out to be the product on profinite words.

Theorem 6.1 *The dual space of the algebra* $(\mathrm{Reg}(A^*), \backslash, /)$ *under extended duality is the topological monoid of profinite words $(\widehat{A^*}, \tau, \cdot)$. The relational dual of the operations $\backslash$ and $/$ is the product of profinite words. The closure of* $\mathrm{Reg}(A^*)$ *under $\backslash$ and $/$ accounts for the right and left continuity of the product, respectively, and the equational property $(H\backslash K)/L = H\backslash(K/L)$ of* $(\mathrm{Reg}(A^*), \backslash, /)$ *corresponds to the associativity of the product.*

The proof of Theorem 6.1 requires advanced machinery from duality theory and space does not allow us to give even a sketch of the proof here.

This theorem has far-reaching consequences. To mention just two, the syntactic ordered monoid of a regular language is none other than the dual space of the subalgebra of $(\mathrm{Reg}(A^*), \backslash, /)$ generated by the singleton set $\{L\}$ under the lattice operations and the residuation operations with arbitrary denominators, and closure of $\mathrm{Reg}(A^*)$ under product of languages corresponds to the fact that product for profinite words is an open mapping. In the next section we use Theorem 6.1 to give an important specialisation of Theorem 5.2.

The following observations will come in handy in the next section: for each $a \in A$ the residuals with denominator $\{a\}$ are central in language theory. We denote them by $a^{-1}(\ )$ and $(\ )a^{-1}$ instead of $\{a\}\backslash(\ )$ and $(\ )/\{a\}$, respectively, and call them *quotienting operations*.

We call a lattice of languages a *quotienting algebra of languages* provided it is closed under the quotienting operations. For instance, the lattice $\mathrm{Reg}(A^*)$ is a quotienting algebra. It is easy to prove that, for sets of regular languages closed under finite intersections, closure under the residuals with arbitrary denominators amounts to the same as closure under the quotienting operators.

7 Lattices of languages closed under quotienting

In this section we characterise those lattices of languages for which the dual quotient is not only a topological quotient but also an ordered monoid quotient. Recall that an *ordered monoid* is a partially ordered monoid in which the monoid operation is order preserving in each coordinate. Note that the map $\widehat{A^*} \twoheadrightarrow S_D$ defined in the proof of Theorem 5.2 is an ordered monoid quotient if and only if the relation Q_D is a congruence of ordered monoid.

Let u and v be two profinite words of $\widehat{A^*}$. We say that L *satisfies the semigroup equation* $u \leqslant v$ if, for all $x, y \in \widehat{A^*}$, it satisfies the equation $xvy \to xuy$. Since A^* is dense in $\widehat{A^*}$, it is equivalent to state that L satisfies these equations only for all $x, y \in A^*$. But there is a much more convenient characterization using the syntactic ordered monoid of L.

Proposition 7.1 *Let L be a regular language of A^*, let $(M, \leqslant_L)$ be its syntactic ordered monoid and let $\eta : A^* \to M$ be its syntactic morphism. Then L satisfies the equation $u \leqslant v$ if and only if $\hat{\eta}(u) \leqslant_L \hat{\eta}(v)$.*

Proof. Corollary 5.1 shows that L satisfies the equation $u \leqslant v$ if and only if, for every $x, y \in A^*$, $\hat{\eta}(xvy) \in \eta(L)$ implies $\hat{\eta}(xuy) \in \eta(L)$. Since $\hat{\eta}(xvy) = \hat{\eta}(x)\hat{\eta}(v)\hat{\eta}(y) = \eta(x)\hat{\eta}(v)\eta(y)$ and since η is surjective, this is equivalent to saying that, for all $s, t \in M$, $s\hat{\eta}(v)t \in \eta(L)$ implies $s\hat{\eta}(u)t \in \eta(L)$, which exactly means that $\hat{\eta}(u) \leqslant_L \hat{\eta}(v)$. $\square$

Using the fact that in the extended duality, preservation of operations on the algebraic side corresponds to bounded morphisms [9] on the other, one can now prove the following specialisation of Theorem 5.2.

Theorem 7.2 *Let D be a lattice of languages of A^*. The following conditions are equivalent:*

(1) *D is a quotienting algebra of languages,*

(2) *D can be defined by a set of semigroup equations $u \leqslant v$, where $u, v \in \widehat{A^*}$,*

(3) *the corresponding dual quotient $\widehat{A^*} \twoheadrightarrow S_D$ is an ordered quotient monoid.*

Theorem 7.2 can be readily extended to Boolean algebras. Let u and v be two profinite words. We say that a regular language L *satisfies the equation* $u = v$ if it satisfies the equations $u \leqslant v$ and $v \leqslant u$. Proposition 7.1 now gives immediately:

Proposition 7.3 *Let L be a regular language of A^* and let η be its syntactic morphism. Then L satisfies the equation $u = v$ if and only if $\hat{\eta}(u) = \hat{\eta}(v)$.*

This leads to the following equational description of the Boolean algebras of languages closed under quotients.

Proposition 7.4 *A set of regular languages of A^* is a Boolean quotienting algebra if and only if it can be defined by a set of semigroup equations of the form $u = v$, where $u, v \in \widehat{A^*}$.*

8 Classes of languages closed under inverses of morphisms

The results of this section and the previous section permit in particular to recover the equational characterization of Eilenberg's varieties and Straubing's $\mathcal{C}$-varieties.

Denote by $\mathcal{C}$ a class of morphisms between free monoids containing the length-preserving morphisms and closed under composition. These morphisms will be called $\mathcal{C}$-morphisms. Examples include the classes of all *length-preserving* morphisms (morphisms for which the image of each letter is a letter), all *length-multiplying* morphisms (morphisms such that, for some integer k, the length of the image of a word is k times the length of the word), all *non-erasing* morphisms (morphisms for which the image of each letter is a nonempty word), all *length-decreasing* morphisms (morphisms for which the image of each letter is either a letter of the empty word) and all morphisms.

A *class of language lattices* $\mathcal{L}$ associates with every finite alphabet A a lattice of languages $\mathcal{L}(A^*)$. Theorem 5.2 gives an equational description for each of these lattices, but these equations depend on the alphabet A. We now show that if $\mathcal{L}$ is closed under inverses of $\mathcal{C}$-morphisms, a single set of equations suffices to characterize the whole class $\mathcal{L}$.

Indeed, if $u \rightarrow v$ is an equation of $\mathcal{L}(A^*)$ and $\varphi : A^* \rightarrow B^*$ is a $\mathcal{C}$-morphism, then $\hat{\varphi}(u) \rightarrow \hat{\varphi}(v)$ is an equation of $\mathcal{L}(B^*)$. This leads naturally to the following definition. Let Σ be a countable alphabet. A regular language L of A^* satisfies the *$\mathcal{C}$-identity* $u \leqslant v$, where $u, v \in \widehat{\Sigma^*}$ if, for each $\mathcal{C}$-morphism $\varphi : \Sigma^* \rightarrow A^*$, L satisfies the equation $\hat{\varphi}(v) \rightarrow \hat{\varphi}(u)$. Then one gets the following result:

Theorem 8.1 *A class of language lattices is closed under quotienting and under inverses of $\mathcal{C}$-morphisms if and only if it can be defined by a set of $\mathcal{C}$-identities of the form $u \leqslant v$, where $u, v \in \widehat{\Sigma^*}$.*

In practice, one may consider a $\mathcal{C}$-identity as an equation in which each letter represents a variable. If $\mathcal{C}$ is the class of length-preserving morphisms, these variables can be replaced by letters, if it is the class of length-multiplying morphisms, they can be replaced by words of the same fixed length, etc.

Of course, similar results hold for identities of the form $u \leftrightarrow v$, $u \leqslant v$ or $u = v$. Our main result thus offers multifarious aspects, which are summarized in the following table. Reiterman's theorem corresponds to the strongest assumptions.

Closed under	**Equations**	**Definition**
$\cup, \cap$	$u \to v$	$\hat{\eta}(u) \in \hat{\eta}(L) \Rightarrow \hat{\eta}(v) \in \hat{\eta}(L)$
quotienting	$u \leqslant v$	for all x, y, $xvy \to xuy$
complement	$u \leftrightarrow v$	$u \to v$ and $v \to u$
quotienting and complement	$u = v$	for all x, y, $xvy \leftrightarrow xuy$
Closed under inverses of morphisms		**Interpretation of variables**
all morphisms		words
nonerasing morphisms		nonempty words
length multiplying morphisms		words of equal length
length preserving morphisms		letters

9 Examples of equational definitions

In this section, we give a few examples of equational characterizations for classes of languages that are not closed under inverses of morphisms and hence do not form a variety of languages. The language A^* is called the *full language*.

9.1 Languages with zero and nondense languages

A *language with zero* is a language whose syntactic monoid has a zero. The class of regular languages with zero is closed under Boolean operations and residuals. According to Proposition 7.4, it has an equational definition, but finding one explicitely requires a little bit of work.

Let us fix a total order on the alphabet A. Let $u_0, u_1, \ldots$ be the ordered sequence of all words of A^+ in the induced shortlex order. For instance, if $A = \{a, b\}$ with $a < b$, the first elements of this sequence would be 1, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa, bab, bba, bbb, $aaaa$, ... It is proved in [17, 5] that the sequence of words $(v_n)_{n \geqslant 0}$ defined by $v_0 = u_0$, $v_{n+1} = (v_n u_{n+1} v_n)^{(n+1)!}$ converges to an idempotent ρ_A of the minimal ideal of $\widehat{A^*}$. We can now state:

Proposition 9.1 *A regular language has a zero if and only if it satisfies the equation $x\rho_A = \rho_A = \rho_A x$ for all $x \in A^*$.*

Proof. Let L be a regular language and let $\eta : A^* \to M$ be its syntactic monoid. Since ρ_A belongs to the minimal ideal of $\widehat{A^*}$, $\hat{\eta}(\rho_A)$ is an element of the minimal ideal of M. In particular, if M has a zero, $\hat{\eta}(\rho_A) = 0$ and L satisfies the equations $x\rho_A = \rho_A = \rho_A x$ for all $x \in A^*$.

Conversely, assume that L satisfies these equations. Let $m \in M$ and let $x \in A^*$ be such that $\eta(x) = m$. Then the equations $\hat\eta(x\rho_A) = \hat\eta(\rho_A) = \hat\eta(\rho_A x)$ give $m\hat\eta(\rho_A) = \hat\eta(\rho_A) = \hat\eta(\rho_A)m$, showing that $\hat\eta(\rho_A)$ is a zero of M. Thus L has a zero. □

In the sequel, we shall use freely the symbol 0 in equations to mean that a monoid has a zero. For instance the equation $x \leqslant 0$ of Theorem 9.2 below should be formally replaced by the three equations $x\rho_A = \rho_A = \rho_A x$ and $x \leqslant \rho_A$.

A language L of A^* is *dense* if, for every word $u \in A^*$, $L \cap A^*uA^* \neq \emptyset$. Note that dense languages are not closed under intersection: $(A^2)^*$ and $(A^2)^*A \cup \{1\}$ are dense, but their intersection is not dense. However, one can show that regular nondense or full languages form a lattice of languages closed under quotients.

We now give an equational description of the form foretold by Theorem 7.2.

Theorem 9.2 *A language of A^* is nondense or full if and only if it satisfies the equations $x \leqslant 0$ for all $x \in A^*$.*

9.2 Languages defined by density

The *density* of a language $L \subseteq A^*$ is the function which counts the number of words of length n in L. More formally, it is the function $d_L : \mathbb{N} \to \mathbb{N}$ defined by $d_L(n) = |L \cap A^n|$. See [23] for a general reference.

If $d_L(n) = O(1)$, then L is called a *slender language*. It is well known that a regular language is slender if and only if it is a finite union of languages of the form xu^*y, where $x, u, y \in A^*$. Regular slender languages form a lattice of languages closed under residuals and morphisms.

Note that if $|A| \leqslant 1$, all regular languages are slender. For $|A| \geqslant 2$, slender or full languages admit a simple equational characterization. Let us denote by $i(u)$ the first letter (or *initial*) of a word u.

Theorem 9.3 *Suppose that $|A| \geqslant 2$. A regular language of A^* is slender or full if and only if it satisfies the equations $x \leqslant 0$ for all $x \in A^*$ and the equation $x^\omega u y^\omega = 0$ for each $x, y \in A^+$, $u \in A^*$ such that $i(uy) \neq i(x)$.*

We now also consider the Boolean closure of slender languages. A language is called *coslender* if its complement is slender.

Theorem 9.4 *Suppose that $|A| \geqslant 2$. A regular language of A^* is slender or coslender if and only if its syntactic monoid has a zero and satisfies the equations $x^\omega u y^\omega = 0$ for each $x, y \in A^+$, $u \in A^*$ such that $i(uy) \neq i(x)$.*

Note that if $A = \{a\}$, the language $(a^2)^*$ is slender but its syntactic monoid, the cyclic group of order 2, has no zero. Therefore the condition $|A| \geqslant 2$ in Theorem 9.4 is mandatory.

A language is *sparse* if it has polynomial density, that is, if $d_L(n) = O(n^k)$ for some $k > 0$. It is well known that a regular language is sparse if and only if it is a finite union of languages of the form $u_0 v_1^* u_1 \cdots v_n^* u_n$, where u_0, v_1, $\dots$, v_n, u_n are words. Regular sparse languages from a lattice of languages and are closed under concatenation product, morphisms and residuals.

Theorem 9.5 *Suppose that $|A| \geqslant 2$. A regular language of A^* is sparse or full if and only if it satisfies the equations $x \leqslant 0$ for all $x \in A^*$ and the equations $(x^\omega y^\omega)^\omega = 0$ for each $x, y \in A^+$ such that $i(x) \neq i(y)$.*

Pursuing the analogy with slender languages, we consider now the Boolean closure of sparse languages. A language is *cosparse* if its complement is sparse.

Theorem 9.6 *Suppose that $|A| \geqslant 2$. A regular language of A^* is sparse or cosparse if and only if its syntactic monoid has a zero and satisfies the equations $(x^\omega y^\omega)^\omega = 0$ for each $x, y \in A^+$ such that $i(x) \neq i(y)$.*

10 Conclusion

We proved that every lattice of regular languages is given by an equational theory, a result that subsumes Eilenberg's variety theorem and its extensions to positive varieties and $\mathcal{C}$-varieties. One could further extend this result to classes of regular languages only closed under finite intersection by using the syntactic semiring introduced by Polák [15]. Our result could also be adapted to languages of infinite words, words over ordinals or linear orders, and even perhaps to tree languages.

Our second main result does not in itself give a new result in the theory of automata and semigroups, but it reveals a very strong link between two theories pertaining to the foundations of computer science: the theory of relational semantics for non-classical (modal, intuitionistic, many-valued, etc.) logics on the one side and the algebraic theory of automata on the other. We have indicated how the fundamental tools of semigroup theory fit into the duality perspective, obtaining an extensive repertoire of equational theories as a modular family of results so typical of modal correspondence theory. Further duality results will be presented in the full version of this paper.

References

[1] M. E. Adams, The Frattini sublattice of a distributive lattice, *Alg. Univ.* **3** (1973), 216–228.

[2] J. Almeida, Residually finite congruences and quasi-regular subsets in uniform algebras, *Partugaliæ Mathematica* **46** (1989), 313–328.

[3] J. Almeida, *Finite semigroups and universal algebra*, World Scientific Publishing Co. Inc., River Edge, NJ, 1994.

[4] J. Almeida, Profinite semigroups and applications, in *Structural theory of automata, semigroups, and universal algebra*, pp. 1–45, *NATO Sci. Ser. II Math. Phys. Chem.* vol. 207, Springer, Dordrecht, 2005. Notes taken by Alfredo Costa.

[5] J. Almeida and M. V. Volkov, Profinite identities for finite semigroups whose subgroups belong to a given pseudovariety, *J. Algebra Appl.* **2**,2 (2003), 137–163.

[6] G. Birkhoff, On the structure of abstract algebras, *Proc. Cambridge Phil. Soc.* **31** (1935), 433–454.

[7] S. Eilenberg, *Automata, languages, and machines. Vol. B*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976.

[8] Z. Ésik, Extended temporal logic on finite words and wreath products of monoids with distinguished generators, in *DLT 2002, Kyoto, Japan*, Ito, Masami et al. (éd.), Berlin, 2002, pp. 43–58, *Lect. Notes Comp. Sci.* n° 2450, Springer.

[9] R. Goldblatt, Varieties of complex algebras, *Ann. Pure App. Logic* **44** (1989), 173–242.

[10] M. Kunc, Equational description of pseudovarieties of homomorphisms, *Theoretical Informatics and Applications* **37** (2003), 243–254.

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

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

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

[14] N. Pippenger, Regular languages and Stone duality, *Theory Comput. Syst.* **30**,2 (1997), 121–134.

[15] L. Polák, Syntactic semiring of a language, in *MFCS 2001*, pp. 611–620, *Lect. Notes Comp. Sci.* vol. 2136, Springer, Berlin, 2001.

[16] H. A. Priestley, Representation of distributive lattices by means of ordered Stone spaces, *Bull. London Math. Soc.* **2** (1970), 186–190.

[17] N. R. Reilly and S. Zhang, Decomposition of the lattice of pseudovarieties of finite semigroups induced by bands, *Algebra Universalis* **44**,3-4 (2000), 217–239.

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

[19] M. Stone, The theory of representations for Boolean algebras, *Trans. Amer. Math. Soc.* **40** (1936), 37–111.

[20] M. H. Stone, Applications of the theory of Boolean rings to general topology, *Trans. Amer. Math. Soc.* **41**,3 (1937), 375–481.

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

[22] P. Weil, Profinite methods in semigroup theory, *Int. J. Alg. Comput.* **12** (2002), 137–178.

[23] S. Yu, Regular languages, in *Handbook of language theory*, G. Rozenberg and A. Salomaa (éd.), vol. 1, ch. 2, pp. 679–746, Springer, 1997.