ARCHIVUM MATHEMATICUM (BRNO)
Tomus 40 (2004), 395 – 406

A CLASSIFICATION OF RATIONAL LANGUAGES BY SEMILATTICE-ORDERED MONOIDS

LIBOR POLÁK

ABSTRACT. We prove here an Eilenberg type theorem: the so-called conjunctive varieties of rational languages correspond to the pseudovarieties of finite semilattice-ordered monoids. Taking complements of members of a conjunctive variety of languages we get a so-called disjunctive variety. We present here a non-trivial example of such a variety together with an equational characterization of the corresponding pseudovariety.

0. Introduction

Syntactic characterizations of certain significant classes of rational languages were obtained by Schützenberger, Simon, Brzozowski-Simon and McNaughton. It was Eilenberg [2] who discovered the appropriate framework to formulate this type of results. Books by Pin [4] and by Almeida [1] collect the basic examples and are designed to be starting points for an extensive literature.

Our main result states that the so-called conjunctive varieties of rational languages correspond to the pseudovarieties of finite semilattice-ordered monoids. This is a modification of the classical Eilenberg Theorem – he uses syntactic monoids, boolean varieties of languages and pseudovarieties of finite monoids. In Pin's generalization [5] of the Eilenberg result, ordered syntactic monoids, positive varieties of languages and pseudovarieties of finite ordered monoids are used (see Section 3).

Historically the present paper is the first one dealing with syntactic semirings; in fact the computer science community led the author to prefer the term "idempotent semiring" for "semilattice-ordered monoid". Although our correspondence is not to be neglected from the algebraic point of view we had waited for significant examples. In between the concept of the syntactic semiring proved its viability: when dealing with the ordered version of the power operator [7] or when considering

---

2000 *Mathematics Subject Classification*: Primary: 68Q70; Secondary: 20M07, 06F05.

Key words and phrases: syntactic semilattice-ordered monoid, conjunctive varieties of rational languages.

Supported by the Ministry of Education of the Czech Republic under the project MSM 143100009.

Received November 26, 2002.

language equations [8]. Certain examples of conjunctive varieties of languages are presented in [6]. We still do not have examples the computer scientists are waiting for. The recent work by Straubing [9] has reopened the significance of considerations of variants of Eilenberg-type correspondences. Note that in [9] the syntactic information about a given language is not only the syntactic monoid but the whole syntactic homomorphism.

We added to our previous version Section 6 explaining modifications of our theory to +-languages and we also pass to complements of conjunctive varieties of languages to get so-called disjunctive varieties. The last section deals with a disjunctive variety of +-languages $\mathcal{L}_d$ formed by all $XA^* \cup A^*Y \cup Z$ where $X, Y, Z \subseteq A^+$ are finite.

1. Main result

A structure $(S, \cdot, \vee)$ is called a *semilattice-ordered semigroup* if

(i) $(S, \cdot)$ is a semigroup,
(ii) $(S, \vee)$ is a semilattice,
(iii) $a, b, c \in S$ implies $a(b \vee c) = ab \vee ac$ and $(a \vee b)c = ac \vee bc$

and a subset I of S is its *ideal* if

(i) $a \in I,\ b \in S,\ b \leq a$ implies $b \in I$,
(ii) $a, b \in I$ implies $a \vee b \in I$.

An ideal I of a semilattice-ordered semigroup defines a relation $\sim_I$ on the set S by

$$a \ \sim_I \ b \quad \Longleftrightarrow \quad (\ \forall\, p, q \in S^1\ )\ (\ paq \in I \Leftrightarrow pbq \in I\ )\ .$$

This relation is a congruence of $(S, \cdot, \vee)$ (see Lemma 2 (i)) and the corresponding factor-structure is called *the semilattice-ordered syntactic semigroup* of I in $(S, \cdot, \vee)$.

Let B^f stand for the set of all non-empty finite subsets of a set B. We often identify $b \in B$ with $\{b\} \in B^f$, so $B \subseteq B^f$. Now let L be a language over an alphabet A. Then L^f is an ideal of the semilattice-ordered monoid $((A^*)^f, \cdot, \cup)$ and we can apply the above construction to get the so-called *semilattice-ordered syntactic monoid of a language* L and we denote it by $\mathsf{S}(L)$.

An operator $\mathcal{L}$ is called *a conjunctive variety of languages* if for every finite set A a set $\mathcal{L}(A)$ of rational languages over the alphabet A is given in such a way that

(i) for every A, the set $\mathcal{L}(A)$ is closed with respect to finite meets (in particular, $A^* \in \mathcal{L}(A)$) and $\emptyset \in \mathcal{L}(A)$,
(ii) for every $A,\ a \in A$ and $L \in \mathcal{L}(A)$ we have $a^{-1}L,\ La^{-1} \in \mathcal{L}(A)$,
(iii) for all sets A and B, semilattice-ordered semigroup homomorphism $\phi : (A^*)^f \to (B^*)^f$ and $L \in \mathcal{L}(B)$ we have $\phi^{-1}(L^f) \cap A^* \in \mathcal{L}(A)$.

$\mathcal{L}$ is called *a positive variety of languages* if it satisfies (i), (ii) and

(iv) for every A, the set $\mathcal{L}(A)$ is closed with respect to finite joins,
(v) for all sets A and B, semigroup homomorphism $\phi : A^* \to B^*$ and $L \in \mathcal{L}(B)$ we have $\phi^{-1}(L) \in \mathcal{L}(A)$

and such a variety is called *a boolean variety of languages* if it satisfies in addition

(vi) for every A, the set $\mathcal{L}(A)$ is closed with respect to complements.

A class of finite semilattice-ordered monoids is called *a pseudovariety* if it is closed with respect to forming finite products, substructures and homomorphic images.

For a variety $\mathcal{L}$ of languages we put

$$\mathscr{S}(\mathcal{L}) = \langle \; \{ \; \mathsf{S}(L) \mid A \text{ a finite set}, \, L \in \mathcal{L}(A) \; \} \; \rangle$$

– the pseudovariety of finite semilattice-ordered monoids generated by syntactic semilattice-ordered monoids of members of $\mathcal{L}$.

Let $\mathcal{V}$ be a pseudovariety of semilattice-ordered monoids. For a finite set A we put

$$\mathscr{L}(\mathcal{V})(A) = \{ \, L \subseteq A^* \mid \mathsf{S}(L) \in \mathcal{V} \, \} \, .$$

Our main result states

Theorem 1. *The assignments*

$$\mathcal{L} \mapsto \mathscr{S}(\mathcal{L}) \quad \textit{and} \quad \mathcal{V} \mapsto \mathscr{L}(\mathcal{V})$$

are mutually inverse bijections between conjunctive varieties of languages and pseudovarieties of finite semilattice-ordered monoids.

The main body of this paper is devoted to the proof.

2. Semilattice-ordered semigroups

A structure $(S, \cdot, \leq)$ is called *an ordered semigroup* if

(i) $(S, \cdot)$ is a semigroup,

(ii) $(S, \leq)$ is an ordered set,

(iii) $a, b, c \in S, \; a \leq b$ implies $ac \leq bc$ and $ca \leq cb$.

A semilattice-ordered semigroup becomes an ordered semigroup with respect to the relation $\leq$ defined by $a \leq b \;\Leftrightarrow\; a \vee b = b, \; a, b \in S$.

We have defined an ideal I of a semilattice-ordered semigroup $(S, \cdot, \vee)$ and the relation $\sim_I$ in Section 1. We denote by $a \sim_I$ the class of $\sim_I$ containing $a \in S$ and by ρ_I the assignment $a \mapsto a \sim_I, \; a \in S$.

The ideal $(a] = \{b \in S \mid b \leq a\}$ is called *the principal* ideal generated by an element $a \in S$.

We put $S^1 = S$ for a monoid S and $T^1 = T \cup \{\lambda\}$ for a semigroup T without a neutral element; we make λ neutral by setting $\lambda a = a\lambda = a, \; a \in T^1$.

For a subset I and an element $a \in S$ we put $a^{-1}I = \{b \in S \mid ab \in I\}$ and $Ia^{-1} = \{c \in S \mid ca \in I\}$. We speak about *left* and *right quotients* of I and the sets of the form $a^{-1}Ib^{-1}$ $(a, b \in S^1)$ are called *quotients* of I.

Let Δ_B denote the diagonal relation $\{(a, a) \mid a \in B\}$ on a set B.

We denote by $\mathcal{OS}$, $\mathcal{OM}$, $\mathcal{SOS}$, $\mathcal{SOM}$ the classes of all ordered semigroups, ordered monoids, semilattice-ordered semigroups and semilattice-ordered monoids, respectively, and by $\mathcal{FOS}, \ldots$ the classes of all finite ordered semigroups, ...

We say that $T \in \mathcal{SOS}$ *divides* $S \in \mathcal{SOS}$ if T is a homomorphic image of a substructure of S.

Lemma 2. *Let I be an ideal of a semilattice-ordered semigroup S. Then*

(i) *The relation $\sim_I$ is a congruence relation on S and $\rho = \rho_I$ is a surjective homomorphism of S onto the factor-structure $T = S/{\sim_I}$.*

(ii) *The induced order $\leq_T$ on T is given by*

$$a \sim_I \quad \leq_T \;\; b \sim_I \iff \;\; (\; \forall\, p, q \in S^1 \;)\; (\; pbq \in I \;\Rightarrow\; paq \in I \;)\;.$$

(iii) *$J = \rho(I)$ is an ideal in T.*

(iv) *$\rho^{-1}(J) = I$.*

(v) *$\sim_J = \Delta_T$.*

Proof. (i) Let $a, b, c \in S$, $a \sim_I b$. Then $\forall\, p, q \in S^1$ ($pacq \in I \Leftrightarrow pbcq \in I$) and therefore $ac \sim_I bc$; similarly $ca \sim_I cb$.

Now $p, q \in S^1$, $p(a \vee c)q \in I$ implies gradually $paq \vee pcq \in I$, $paq, pcq \in I$, $pbq, pcq \in I$, $p(b \vee c)q \in I$ and therefore $a \vee c \sim_I b \vee c$.

Basics of universal algebra give the rest.

(ii) Let $a, b \in S$, $c = a \sim_I$, $d = b \sim_I$. A sequence of equivalent formulas follows: $c \leq_T d$, $c \vee d = d$, $a \vee b \sim_I b$, $\forall\, p, q \in S^1$ ($p(a \vee b)q \in I \Leftrightarrow pbq \in I$), the formula in statement (ii).

(iii) Let $c \in J$, $d \in T$, $d \leq c$, i.e., $d \vee c = c$. There are $a \in I$, $b \in S$ such that $\rho(a) = c$, $\rho(b) = d$. Then by (ii), $\forall\, p, q \in S^1$ ($paq \in I \Rightarrow pbq \in I$). Now $p = q = \lambda$ gives $b \in I$ and $d \in J$. Clearly $c, d \in J$ yields $c \vee d \in J$.

(iv) The inclusion "$\supseteq$" is clear. Let $a \in S$, $\rho(a) \in J$. Then $\rho(a) = \rho(b)$ for some $b \in I$ and $\forall\, p, q \in S^1$ ($paq \in I \Leftrightarrow pbq \in I$). Now $p = q = \lambda$ gives $a \in I$.

(v) Let $c, d \in T$, $c \sim_J d$. There are $a, b \in S$ such that $\rho(a) = c$, $\rho(b) = d$. Then for arbitrary $p, q \in S^1$, $paq \in I$ gives $\rho(p)c\rho(q) \in J$ which is equivalent to $\rho(p)d\rho(q) \in J$ and thus $pbq \in I$ and conversely. This means $c = d$. $\square$

Lemma 3. *Let I and J be ideals of a semilattice-ordered semigroup S and let $a \in S$. Then*

(i) *$\sim_{I \cap J} \supseteq \sim_I \cap \sim_J$ and $S/{\sim_{I\cap J}}$ divides the product $S/{\sim_I} \times S/{\sim_J}$.*

(ii) *$a^{-1}I$ is an ideal of S, $\sim_{a^{-1}I} \supseteq \sim_I$ and $S/{\sim_{a^{-1}I}}$ is a homomorphic image of $S/{\sim_I}$. The same holds for $\sim_{Ia^{-1}}$.*

Proof. (i) The inclusion follows from the definitions of $\sim_I$, $\sim_J$, $\sim_{I \cap J}$ and the rest is a well-known universal algebra fact.

(ii) Let $b \in a^{-1}I$, $c \in S$, $c \leq b$. Then $ab \in I$, $ac \leq ab$ and therefore $ac \in I$, $c \in a^{-1}I$.

Let $b, c \in a^{-1}I$. Then $a(b \vee c) = ab \vee ac \in I$ and $b \vee c \in a^{-1}I$.

The conclusion follows from the definitions of $\sim_{a^{-1}I}$, $\sim_I$. $\square$

Lemma 4. *Let T be a substructure of a semilattice-ordered semigroup S and let I be an ideal in S. Then $T \cap I$ is an ideal in T, $\sim_I|\, T \times T \subseteq \sim_{T\cap I}$ and $T/{\sim_{T\cap I}}$ divides $S/{\sim_I}$.*

Proof. It is immediate that $T \cap I$ is an ideal in T and that $\sim_I \mid T \times T \subseteq \sim_{T\cap I}$. Thus $T/\sim_{T\cap I}$ is a homomorphic image of a substructure $T/\sim_I$ of the structure $S/\sim_I$. □

Lemma 5. *Let $\phi : S \to T$ be a homomorphism of semilattice-ordered semigroups and let J be an ideal in T. Then*
- (i) *$I = \phi^{-1}(J)$ is an ideal of S,*
- (ii) *in case of surjective ϕ, the factor-structure $S/\sim_I$ is a homomorphic image of $T/\sim_J$,*
- (iii) *in case of surjective ϕ, there also exists a (unique) homomorphism $\psi : T \to S/\sim_I$ such that $\psi \circ \phi = \rho_I$.*

Proof. (i) is clear.

(ii) Define $\alpha : b \sim_J \mapsto a \sim_I$ where $\phi(a) = b$, $a \in S$, $b \in T$. This assignment is really a mapping since $\phi(a') = b' \sim_J b$, $p, q \in S^1$ gives

$paq \in I$ implies $\phi(p)b\phi(q) = \phi(paq) \in J$ which is equivalent to $\phi(p)b'\phi(q) = \phi(pa'q) \in J$. Thus $pa'q \in I$ and conversely. Therefore $a' \sim_I a$.

Now it is obvious that α is a surjective homomorphism of $T/\sim_J$ onto $S/\sim_I$.

(iii) Due to $\psi \circ \phi = \rho_I$, the mapping ψ should send an element $b \in T$ to $\rho_I(a)$ where $a \in S$ is such that $\phi(a) = b$.

Correctness: Let $a' \in S$, $\phi(a') = b$, $p, q \in S^1$. Then $pa'q \in I$ iff $paq \in I$ since $\phi(pa'q) = \phi(paq)$. Clearly, ψ is a homomorphism. □

Lemma 6. *Let $I \neq \emptyset$ be an ideal of S such that $\sim_I = \Delta_S$. Then an arbitrary principal ideal $(a]$ is an intersection of quotients of I.*

Proof. By Lemma 2 (ii), $b \le a$ iff $b \in \bigcap_{p,q \in S^1, paq \in I} p^{-1}Iq^{-1}$. □

3. Languages

Let L be a language over a finite set A. We can express the syntactic congruence of L^f in $(A^*)^f$ more explicitly

$$\{u_1, \dots, u_k\} \sim_{L^f} \{v_1, \dots, v_l\} \iff$$
$$(\ \forall\, p, q \in A^*\ )\ (\, pu_1q, \dots, pu_kq \in L \ \Leftrightarrow\ pv_1q, \dots, pv_lq \in L\,)\,.$$

An element a of a semilattice $(S, \vee)$ is *join-irreducible* if $a = b \vee c$, $b, c \in S$ implies $a = b$ or $a = c$.

Let $(S, \le)$ be an ordered set. A subset H of S is *hereditary* if $a \in H$, $b \in S$, $b \le a$ implies $b \in H$. A hereditary set of an ordered semigroup $(S, \cdot, \le)$ defines a relation $\approx_H$ on S by

$$a \ \approx_H\ b \iff (\forall\, p, q \in S^1\,)\ (\, paq \in H \Leftrightarrow pbq \in H\,)$$

This relation is a congruence on $(S, \cdot)$ and the factor-structure is called *the syntactic semigroup* of H in $(S, \cdot, \le)$. It is ordered by

$$a \approx_H\, \le\ b \approx_H \iff (\ \forall\, p, q \in S^1\,)\ (\, pbq \in H \ \Rightarrow\ paq \in H\,)\,.$$

Let σ_H denote the mapping $a \mapsto a \approx_H$, $a \in S$.

Any language L over A is a hereditary subset of the trivially ordered monoid $(A^*, \cdot, \leq)$. The above construction gives *the ordered syntactic monoid* of the language L; we denote it by $\mathsf{O}(L)$.

Lemma 7. *Let L be a language over an alphabet A. The mapping*

$$\iota : u \approx_L \mapsto \{u\} \sim_{L^f},\ u \in A^*$$

is an injective semigroup homomorphism of $(\mathsf{O}(L), \cdot, \leq)$ into $(\mathsf{S}(L), \cdot, \vee)$ satisfying $a \leq b$ if and only if $\iota(a) \leq \iota(b)$ $(a, b \in \mathsf{O}(L))$. Moreover, $\iota(\sigma_L(L)) = \rho_{L^f}(L^f) \cap \iota(\mathsf{O}(L))$ and $\iota(\mathsf{O}(L))$ contains all join-irreducible elements of $(\mathsf{S}(L), \cdot, \vee)$.

Proof. The first part follows from the fact that

$$\{u\} \ \sim_{L^f} \ \{v\} \quad \Leftrightarrow \quad u \approx_L v, \quad u, v \in A^*$$

and from Lemma 2 (ii). Moreover, we have $u \approx_L \leq v \approx_L$ if and only if $\{u, v\} \sim_{L^f} \{v\}$, $(u, v \in A^*)$.

Now $\{u\} \sim_{L^f} \in \rho_{L^f}(L^f)$ iff $\{u\} \in L^f$, that is, $u \in L$.

Finally realize that $(U \cup V) \sim_{L^f} = U \sim_{L^f} \vee V \sim_{L^f}$ for any $U, V \in (A^*)^f$. □

Lemma 8. *The semilattice-ordered syntactic monoid of a language L over an alphabet A is isomorphic to the semilattice-ordered syntactic monoid of the ideal $I = (\sigma_L(L))^f$ of the semilattice-ordered monoid $((\mathsf{O}(L))^f, \cdot, \cup)$.*

Proof. Realize that, for any $u_1, \dots, u_k,\ v_1, \dots, v_l \in A^*$, we have

$$\{u_1 \approx_L, \dots, u_k \approx_L\} \sim_I \{v_1 \approx_L, \dots, v_l \approx_L\} \iff$$

$$\{u_1, \dots, u_k\} \sim_{L^f} \{v_1, \dots, v_l\}\ .$$

Thus $\{u_1, \dots, u_k\} \sim_{L^f} \mapsto \{u_1 \approx_L, \dots, u_k \approx_L\} \sim_I$ is the desired isomorphism. □

For languages $K, L \subseteq A^*$ we define $K \cdot L = \{\, uv \mid u \in K,\ v \in L \,\}$, $K^* = \{\, u_1 \dots u_k \mid k \geq 0,\ u_1, \dots, u_k \in K \,\}$.

Recall that the set of all *rational* languages over an alphabet A is the smallest family of subsets of A^* containing the empty set, all singletons $\{u\}$, $u \in A^*$, closed with respect to binary joins and the operations $\cdot$ and *.

Theorem 9. *Let L be a language over a finite alphabet A. The following are equivalent.*

(i) *L is rational,*
(ii) *the syntactic monoid $\mathsf{O}(L)$ of L is finite,*
(iii) *the semilattice-ordered syntactic monoid $\mathsf{S}(L)$ of L is finite.*

Proof. The equivalence of (i) and (ii) is Myhill's theorem [3].

Now: $\mathsf{S}(L)$ finite $\implies$ $\mathsf{O}(L)$ finite by Lemma 7,
$\mathsf{O}(L)$ finite $\implies$ $\mathsf{S}(L)$ finite by Lemma 8. □

Lemma 10. *The following are equivalent for a language K over a set A and a semilattice-ordered monoid U :*

(i) *There exists an ideal N in U and a semilattice-ordered semigroup homomorphism $\phi : (A^*)^f \to U$ such that $\phi^{-1}(N) = K^f$.*

(ii) *The syntactic semilattice-ordered monoid $\mathsf{S}(K)$ divides the semilattice-ordered monoid U.*

Proof. (ii) $\Rightarrow$ (i): Write ρ for ρ_{K^f}. Let T be a substructure of U and let ψ be a surjective homomorphism of T onto $\mathsf{S}(K)$.

For any $a \in A$ there is $t_a \in T$ such that $\rho(a) = \psi(t_a)$. Let ϕ be the extension of $a \mapsto t_a,\ a \in A$ to a homomorphism of the free semilattice-ordered monoid $(A^*)^f$ over the set A to U. Then $J = \psi^{-1}(\rho(K^f))$ is an ideal in T by Lemma 5 (i). Then $N = \{\, v \in U \mid \text{there exists } u \in J \text{ such that } v \le u \,\}$ is an ideal in U with the property $N \cap T = J$.

Finally,
$\phi^{-1}(N) = \phi^{-1}(J) = \phi^{-1}(\psi^{-1}(\rho(K^f))) = (\psi \circ \phi)^{-1}(\rho(K^f)) = \rho^{-1}(\rho(K^f)) = K^f$
by Lemma 2 (iv).

(i) $\Rightarrow$ (ii): Put $T = \phi((A^*)^f)$. Then $T \cap N$ is an ideal of T by Lemma 4. Combine Lemma 5 (ii) for $S = (A^*)^f,\ J = T \cap N$ with Lemma 4. $\square$

Lemma 11. *Let $K \neq \emptyset$ be a language over A and let L be a rational language over B. Let $\mathsf{S}(K)$ divide $\mathsf{S}(L)$. Then there exists a semilattice-ordered semigroup homomorphism $\psi : (A^*)^f \to (B^*)^f$ and a language M over B such that $\psi^{-1}(M^f) \cap A^* = K$ and M is a finite intersection of quotients of L.*

Proof. Applying (ii) $\Rightarrow$ (i) of Lemma 10 for $U = \mathsf{S}(L)$ we get the existence of a homomorphism $\phi : (A^*)^f \to \mathsf{S}(L)$ and an ideal $N \neq \emptyset$ in $\mathsf{S}(L)$ such that $\phi^{-1}(N) = K^f$.

Write ρ for ρ_{L^f}. For $a \in A$ there is an element $t_a \in (B^*)^f$ such that $\phi(a) = \rho(t_a)$. Let ψ be the extension of $a \mapsto t_a,\ a \in A$ to a semilattice-ordered semigroup homomorphism $(A^*)^f \to (B^*)^f$. We have $\rho \circ \psi = \phi$.

The structure $\mathsf{S}(L)$ is finite by Theorem 9 and thus N is a principal ideal. Let $J = \rho(L^f)$. By Lemma 2 (v), $\sim_J = \Delta_{\mathsf{S}(L)}$ and therefore, by Lemma 6, $N = a_1^{-1} J b_1^{-1} \cap \cdots \cap a_r^{-1} J b_r^{-1}$ for some $a_1, \ldots, a_r, b_1, \ldots, b_r \in \mathsf{S}(L)$. Let $\rho(\{p_{i,1}, \ldots, p_{i,m_i}\}) = a_i,\ \rho(\{q_{i,1}, \ldots, q_{i,m_i}\}) = b_i,\ i = 1, \ldots, r$. Then

$$\rho^{-1}(N) = \{p_{1,1}, \ldots, p_{1,m_1}\}^{-1} \cdot L^f \cdot \{q_{1,1}, \ldots, p_{1,n_1}\}^{-1} \cap \ldots .$$

Now realize that

$$\{p_1, \ldots, p_m\}^{-1} \cdot L^f \cdot \{q_1, \ldots, q_n\}^{-1} = \Big(\bigcap_{i=1,\ldots,m,\ j=1,\ldots,n} p_i^{-1} L q_j^{-1} \Big)^f$$

and $L_1^f \cap L_2^f = (L_1 \cap L_2)^f$ for $L_1, L_2 \subseteq B^*$. Put $M = \rho^{-1}(N) \cap B^*$. $\square$

4. From varieties of languages to monoids

Let $\mathcal{L}$ be a conjunctive variety of languages. We put

$$\mathscr{D}(\mathcal{L}) = \{\, (S, \cdot, \vee) \in \mathcal{FSOM} \mid (A \text{ a finite set},\ L \subseteq A^*,\ (\mathsf{S}(L), \cdot, \vee) \text{ divides } (S, \cdot, \vee)\,) \;\Rightarrow\; L \in \mathcal{L}(A) \,\}.$$

Note that we can write here $L \neq \emptyset$.

Lemma 12. *For any conjunctive variety of languages $\mathcal{L}$, the class $\mathscr{D}(\mathcal{L})$ is a pseudovariety of finite semilattice-ordered monoids.*

Proof. Let T be a divisor of $S \in \mathscr{D}(\mathcal{L})$. Any monoid dividing T divides also S and thus $T \in \mathscr{D}(\mathcal{L})$.

It remains to show that $S, T \in \mathscr{D}(\mathcal{L})$ implies $S \times T \in \mathscr{D}(\mathcal{L})$. Let $L \neq \emptyset$ be a language over a finite alphabet A and let $\mathsf{S}(L)$ divide $S \times T$. By (ii) $\Rightarrow$ (i) of Lemma 10, there exist an ideal $N \neq \emptyset$ in $S \times T$ and a homomorphism $\phi : (A^*)^f \to S \times T$ such that $L^f = \phi^{-1}(N)$. Due to the finiteness of $S \times T$, $N = ((s,t)]$ for some $s \in S,\ t \in T$. We have $N = ((s] \times T) \cap (S \times (t])$. Now $\phi^{-1}((s] \times T),\ \phi^{-1}(S \times (t])$ are ideals in $(A^*)^f$, so they are of the forms $(L_1)^f$ and $(L_2)^f$, respectively. Let $\pi_1 : S \times T \to S,\ \pi_2 : S \times T \to T$ be the natural projections. We have $(L_1)^f = (\pi_1 \circ \phi)^{-1}((s]),\ (L_2)^f = (\pi_2 \circ \phi)^{-1}((t])$. By (i) $\Rightarrow$ (ii) of Lemma 10, $\mathsf{S}(L_1)$ divides S and $\mathsf{S}(L_2)$ divides T and therefore $L_1, L_2 \in \mathcal{L}(A)$. Also $L = L_1 \cap L_2 \in \mathcal{L}(A)$. Consequently $S \times T \in \mathscr{D}(\mathcal{L})$. $\square$

For any mapping $\phi : S \to T$, put $\ker \phi = \{(a, a') \in S \times S \mid \phi(a) = \phi(a')\}$.

Lemma 13. *For any conjunctive variety of languages $\mathcal{L}$, the classes $\mathscr{D}(\mathcal{L})$ and $\mathscr{S}(\mathcal{L})$ coincide.*

Proof. The monoids generating $\mathscr{S}(\mathcal{L})$ are in $\mathscr{D}(\mathcal{L})$ by Lemma 11 and thus $\mathscr{S}(\mathcal{L}) \subseteq \mathscr{D}(\mathcal{L})$.

Let $S \in \mathscr{D}(\mathcal{L})$. The structure $((A^*)^f, \cdot, \cup)$ is the free semilattice-ordered monoid over the set A and therefore there exist a finite set A and a surjective homomorphism $\phi : ((A^*)^f, \cdot, \cup) \to (S, \cdot, \vee)$. We have that $((A^*)^f, \cdot, \cup)/\ker \phi$ is isomorphic to $(S, \cdot, \vee)$.

For any $a \in S$, the set $L_a = \phi^{-1}((a])$ is an ideal in $(A^*)^f$ by Lemma 5 (i). We show that $\ker \phi = \bigcap_{a \in S} \sim_{L_a}$.

So let $\phi(u) = \phi(v),\ u, v \in (A^*)^f$. For arbitrary $a \in S,\ p, q \in (A^*)^f$ we have $puq \in L_a \ \Leftrightarrow\ pvq \in L_a$ due to $\phi(puq) = \phi(pvq)$. Conversely, let $(u, v) \in \bigcap_{a \in S} \sim_{L_a}$. In particular $(u, v) \in \sim_{L_{\phi(u)}}$ and $1 \cdot u \cdot 1 \in L_{\phi(u)}$ gives $v = 1 \cdot v \cdot 1 \in L_{\phi(u)}$, that is $\phi(v) \leq \phi(u)$. Similarly $\phi(u) \leq \phi(v)$.

Thus $(S, \cdot, \vee)$ is isomorphic to a substructure of the product $\prod_{a \in S} \mathsf{S}(L_a)$ and every $\mathsf{S}(L_a)$ is a homomorphic image of S.

Now $S \in \mathscr{D}(\mathcal{L})$ gives $L_a \in \mathcal{L}(A),\ \mathsf{S}(L_a) \in \mathscr{S}(\mathcal{L})$ and thus $S \in \mathscr{S}(\mathcal{L})$. $\square$

5. Proof of the Theorem

Let $\mathcal{V}$ be a pseudovariety of finite semilattice-ordered monoids and let $\mathcal{L}$ be a conjunctive variety of rational languages. Trivially $\mathscr{S}(\mathcal{L})$ is a pseudovariety and $\mathscr{L}(\mathcal{V})$ is a conjunctive variety of languages by Lemmas 3, 4 and 5 (ii). It remains to prove (i)–(iv).

(i) $\mathcal{V} \supseteq \mathscr{S}(\mathscr{L}(\mathcal{V}))$:

The right-hand side is generated by monoids $\mathsf{S}(L)$ where $L \in \mathscr{L}(\mathcal{V})(A)$ for some finite set A, that is, by monoids $\mathsf{S}(L) \in \mathcal{V}$.

(ii) $\mathcal{V} \subseteq \mathscr{S}(\mathscr{L}(\mathcal{V}))$:

Let $S \in \mathcal{V}$. By Lemma 13 it is enough to show that $S \in \mathscr{D}(\mathscr{L}(\mathcal{V}))$. So let A be a finite set and let $L \subseteq A^*$ be such that $\mathsf{S}(L)$ divides S. Then $\mathsf{S}(L) \in \mathcal{V},\ L \in \mathscr{L}(\mathcal{V})$ and consequently $S \in \mathscr{D}(\mathscr{L}(\mathcal{V}))$.

(iii) $\mathcal{L} \supseteq \mathscr{L}(\mathscr{S}(\mathcal{L}))$:

Let A be a finite set and let $L \in \mathscr{L}(\mathscr{S}(\mathcal{L}))(A)$, that is, $L \in \mathscr{L}(\mathscr{D}(\mathcal{L}))(A)$. Then $\mathsf{S}(L) \in \mathscr{D}(\mathcal{L})$ and $L \in \mathcal{L}(A)$ since $\mathsf{S}(L)$ divides $\mathsf{S}(L)$.

(iv) $\mathcal{L} \subseteq \mathscr{L}(\mathscr{S}(\mathcal{L}))$:

For a finite set A and $L \in \mathcal{L}(A)$ we have $\mathsf{S}(L) \in \mathscr{S}(\mathcal{L}),\ L \in \mathscr{L}(\mathscr{S}(\mathcal{L}))(A)$.

6. Modifications

6.1. **+-varieties of languages and pseudovarieties of semilattice-ordered semigroups.** We have considered languages as subsets of A^*. An alternative way is to exclude the empty word and consider languages as subsets of A^+. For $L \subseteq A^+$ the set L^f is an ideal of the semilattice-ordered semigroup $((A^+)^f, \cdot, \cup)$. The structure $\mathsf{S}'(L) = ((A^+)^f, \cdot, \cup)/ \sim_{L^f}$ is called the *syntactic semilattice-ordered semigroup* of L.

For a variety $\mathcal{L}$ of +-languages we put

$$\mathscr{S}'(\mathcal{L}) = \langle \{ \mathsf{S}'(L) \mid A \text{ a finite set, } L \in \mathcal{L}(A) \} \rangle$$

– the pseudovariety of finite semilattice-ordered semigroups generated by syntactic semilattice-ordered semigroups of members of $\mathcal{L}$.

Let $\mathcal{V}$ be a pseudovariety of semilattice-ordered semigroups. For a finite set A we put

$$\mathscr{L}'(\mathcal{V})(A) = \{ L \subseteq A^* \mid \mathsf{S}'(L) \in \mathcal{V} \} .$$

Straightforward modifications of our considerations lead to

Theorem 14. *The assignments*

$$\mathcal{L} \mapsto \mathscr{S}'(\mathcal{L}) \quad \text{and} \quad \mathcal{V} \mapsto \mathscr{L}'(\mathcal{V})$$

are mutually inverse bijections between conjunctive varieties of +-languages and pseudovarieties of finite semilattice-ordered semigroups.

□

6.2. **Disjunctive varieties of languages.** In our theory we can pass to complements of languages :

An operator $\mathcal{L}$ is called *a disjunctive variety* of *-languages if for every finite set A a set $\mathcal{L}(A)$ of rational *-languages over the alphabet A is given in such a way that

- (i$^{\mathsf{c}}$) for every A, the set $\mathcal{L}(A)$ is closed with respect to finite joins (in particular, $\emptyset \in \mathcal{L}(A)$) and $A^* \in \mathcal{L}(A)$,
- (ii) for every $A,\ a \in A$ and $L \in \mathcal{L}(A)$ we have $a^{-1}L,\ La^{-1} \in \mathcal{L}(A)$,
- (vii) for all sets A and B, a semilattice-ordered semigroup homomorphism $\phi : (A^*)^f \to (B^*)^f$ and $L \in \mathcal{L}(B)$ we have $\phi^{(-1)}(L) = \{ v \in A^* \mid \phi(v) \cap L^f \neq \emptyset \} \in \mathcal{L}(A)$,

In a definition of a disjunctive variety of +-languages each $\mathcal{L}(A)$ is a set of +-languages and we modify (vii) to (vii') :

(vii') for all sets A and B, a semilattice-ordered semigroup homomorphism $\phi : (A^+)^f \to (B^+)^f$ and $L \in \mathcal{L}(B)$ we have $\phi^{(-1)}(L) = \{ v \in A^+ \mid \phi(v) \cap L^f \neq \emptyset \} \in \mathcal{L}(A)$, respectively.

It is clear how to reformulate Theorems 1 and 14 for disjunctive *-varieties and +-varieties.

7. Examples

Consider the following classes of languages : for each finite set A put

$$\mathcal{L}_l(A) = \{ XA^* \cup Z \mid X, Z \subseteq A^+ \text{ finite} \},$$

$$\mathcal{L}_r(A) = \{ A^*Y \cup Z \mid Y, Z \subseteq A^+ \text{ finite} \},$$

$$\mathcal{L}_d(A) = \{ XA^* \cup A^*Y \cup Z \mid X, Y, Z \subseteq A^+ \text{ finite} \}, \text{ and}$$

$\mathcal{L}_b(A) = \{ r_1A^*s_1 \cup \cdots \cup r_kA^*s_k \cup Z \mid k \geq 0,\ r_1, \ldots, r_k, s_1, \ldots, s_k \in A^+,\ Z \subseteq A^+ \text{ finite}\}$.

Result 15 (Pin [4], Chapter 2, Theorems 3.4., 3.6. and Corollary 3.7.)**.** *The classes $\mathcal{L}_l$, $\mathcal{L}_r$ and $\mathcal{L}_b$ are boolean varieties of languages. The equational characterizations of the corresponding pseudovarieties of semigroups are*

$$x^\omega y = x^\omega,\ yx^\omega = x^\omega \textit{ and } x^\omega y x^\omega = x^\omega, \textit{respectively}.$$

The class $\mathcal{L}_b$ is the smallest boolean variety containing both $\mathcal{L}_l$ and $\mathcal{L}_r$.

In a similar spirit we obtain.

Theorem 16. *The classes $\mathcal{L}_l$, $\mathcal{L}_r$, $\mathcal{L}_d$ and $\mathcal{L}_b$ are disjunctive varieties of languages. The equational characterizations of the corresponding pseudovarieties of semilattice-ordered semigroups are*

$$x^\omega y = x^\omega,\ yx^\omega = x^\omega,\ (\ x^\omega y x^\omega = x^\omega,\ x^\omega y^\omega = x^\omega z \vee t y^\omega\ ) \textit{ and } x^\omega y x^\omega = x^\omega,$$

respectively. The class $\mathcal{L}_d$ is the smallest disjunctive variety containing both $\mathcal{L}_l$ and $\mathcal{L}_r$. Moreover, $\mathcal{L}_d \neq \mathcal{L}_b$.

Proof. We need to verify the condition (vii') :
If $\phi : (A^+)^f \to (B^+)^f$ is a semilattice-ordered semigroup homomorphism and $K, L \subseteq B^+$, then $\phi^{(-1)}(K \cup L) = \phi^{(-1)}(K) \cup \phi^{(-1)}(L)$. Therefore it is enough to consider only the summands r, rB^*, B^*s, rB^*s where $r, s \in B^+$. So let $A = \{a_1, \ldots, a_n\}$, $\phi(a_i) = \{v_{i1}, \ldots, v_{ik_i}\}$, $i = 1, \ldots, n$, $v_{ij} \in B^+$. Then

$\phi^{(-1)}(r) = \{ a_{i_1} \ldots a_{i_p} \in A^+ \mid$ there exist $j_1, \ldots, j_p$ such that $v_{i_1j_1} \ldots v_{i_pj_p} = r \}$,

$\phi^{(-1)}(rB^*) = \{ a_{i_1} \ldots a_{i_p} \in A^+ \mid$ there exist $j_1, \ldots, j_p$ such that r is a prefix of $v_{i_1j_1} \ldots v_{i_pj_p} \}$, and similarly for B^*s and rB^*s.

Clearly, for disjunctive varieties $\mathcal{K}$ and $\mathcal{L}$, the sets

$$(\mathcal{K} \vee \mathcal{L})(A) = \{ K \cup L \mid K \in \mathcal{K}(A),\ L \in \mathcal{L}(A) \},\ A \text{ finite},$$

form the smallest disjunctive variety $\mathcal{K}\vee\mathcal{L}$ containing both $\mathcal{K}$ and $\mathcal{L}$. Thus $\mathcal{L}_l\vee\mathcal{L}_r = \mathcal{L}_d$.

Next we show now that $\mathcal{L}_d$ is not closed with respect to intersections. Indeed, let $A = \{a,b,c\}$, $K = aA^*$, $L = A^*a$. Suppose that $K \cap L = XA^* \cup A^*Y \cup Z$, $X,Y,Z \subseteq A^+$ finite. Let n be greater than the lengths of all the words from $X \cup Y \cup Z$. Then $a^n \in K \cap L$ and $a^n \in XA^*$ would imply $a^{n-1}b \in XA^*$ and similarly $a^n \in A^*Y$ would imply $ba^{n-1} \in A^*Y$. We get a contradiction in both cases. Thus $\mathcal{L}_d \neq \mathcal{L}_b$.

We need to show that, for each $r, s \in A^+$,

$$\mathsf{S}'(\{r\}^{\mathsf{c}}) \models x^\omega y = x^\omega,\ \mathsf{S}'(\{rA^*\}^{\mathsf{c}}) \models x^\omega y = x^\omega,\ \mathsf{S}'(\{rA^*s\}^{\mathsf{c}}) \models x^\omega y x^\omega = x^\omega\,.$$

Let $n \geq |r|, |s|$ (the lengths of r, s) and let $v_1, \ldots, v_k \in A^+$ with $|v_1|, \ldots, |v_k| \geq n$. Then, for each $p, q, w \in A^*$,

$$\begin{aligned} p\{v_1,\ldots,v_k\}wq \subseteq \{r\}^{\mathsf{c}} &\iff p\{v_1,\ldots,v_k\}q \subseteq \{r\}^{\mathsf{c}}\,,\\ p\{v_1,\ldots,v_k\}wq \subseteq (rA^*)^{\mathsf{c}} &\iff p\{v_1,\ldots,v_k\}q \subseteq (rA^*)^{\mathsf{c}}\,,\\ p\{v_1,\ldots,v_k\}w\{v_1,\ldots,v_k\}q \subseteq (rA^*s)^{\mathsf{c}} &\iff p\{v_1,\ldots,v_k\}\{v_1,\ldots,v_k\}q \subseteq (rA^*s)^{\mathsf{c}}\,. \end{aligned}$$

Each idempotent of $\mathsf{S}'(L)$ can be represented as $\{v_1,\ldots,v_k\} \sim_L$ with $|v_1|, \ldots, |v_k| \geq n$ and thus we get the result.

The syntactic semigroup of L is isomorphic to a substructure of $\mathsf{S}'(L)$ by Lemma 7. Therefore Result 15 implies

$$\begin{aligned} L \in \mathcal{L}_l(A) &\iff \mathsf{S}'(L^{\mathsf{c}}) \models x^\omega y = x^\omega\,,\\ L \in \mathcal{L}_b(A) &\iff \mathsf{S}'(L^{\mathsf{c}}) \models x^\omega y x^\omega = x^\omega\,. \end{aligned}$$

The statement concerning $\mathcal{L}_r(A)$ is dual to that for $\mathcal{L}_l(A)$.

Notice that each of $x^\omega y = x^\omega$ and $yx^\omega = x^\omega$ implies $x^\omega y^\omega \leq x^\omega z \vee yx^\omega$. Observe that $\mathsf{S}'(L^{\mathsf{c}}) \models x^\omega y^\omega \leq x^\omega z \vee yx^\omega$ if and only if (for each $p, q \in A^*$, $u, v, r, s \in A^+$ such that $u\sim_{L^{\mathsf{c}}}, v\sim_{L^{\mathsf{c}}}$ are idempotents)

$$(\,puvq \in L \implies purq \in L \text{ or } psvq \in L\,)\,.$$

Put $p = q = \lambda$. Then for each idempotents $u\sim_{L^{\mathsf{c}}}$, $v\sim_{L^{\mathsf{c}}}$ we have

$$uv \in L \implies uA^+ \subseteq L \text{ or } A^+v \subseteq L\,.$$

For every semigroup S satisfying $x^\omega y x^\omega = x^\omega$ also $S \models x^\omega y z^\omega = x^\omega z^\omega$ and S^m consists entirely of idempotents for $m = |S|$ (see [4], Chapter 2, Prop. 3.5.).

Now let $\mathsf{S}'(L^{\mathsf{c}}) \models x^\omega y x^\omega = x^\omega$, $x^\omega y^\omega \leq x^\omega z \vee t y^\omega$ and let $w \in L$ with $|w| \geq 2m$. Then $w = utv$ where $u, t, v \in A^*$, $|u| = |v| = m$. The elements $u\sim_{L^{\mathsf{c}}}, v\sim_{L^{\mathsf{c}}}$ are idempotents of $\mathsf{S}'(L^{\mathsf{c}})$ and thus $uv \in L$ and ($uA^* \subseteq L$ or $A^*v \subseteq L$).

Finally observe that $uA^+ = ua_1A^* \cup \cdots \cup ua_nA^*$ for $A = \{a_1, \ldots, a_n\}$, and similarly for A^+v. ∎

References

[1] Almeida, J., *Finite Semigroups and Universal Algebra*, World Scientific, 1994.

[2] Eilenberg, S., *Automata, Languages and Machines*, Vol. B, Academic Press, 1976.

[3] Myhill, J., *Finite automata and the representation of events*, WADD Techn. Report 57–624, Wright Patterson Air Force Base, 1957.

[4] Pin, J.-E., *Varieties of Formal Languages*, Plenum, 1986.

[5] Pin, J.-E., *A variety theorem without complementation*, Izvestiya VUZ Matematika **39** (1995), 80–90. English version: Russian Mathem. (Iz. VUZ) **39** (1995), 74–83.

[6] Polák, L., *Syntactic semiring of a language*, in Proc. Mathematical Foundation of Computer Science 2001, Lecture Notes in Comput. Sci., Vol. **2136** (2001), 611–620.

[7] Polák, L., *Operators on Classes of Regular Languages*, in Algorithms, Automata and Languages, J.P.G. Gomes and P. Silva (ed.), World Scientific (2002), 407–422.

[8] Polák, L., *Syntactic Semiring and Language Equations*, in Proc. of the Seventh International Conference on Implementation and Application of Automata, Tours 2002, Lecture Notes in Comput. Sci., Vol. **2608** (2003), 182–193.

[9] Straubing, H., *On logical descriptions of regular languages*, in Proc. LATIN 2002, Lecture Notes in Comput. Sci., Vol. **2286** (2002), 528–538.

Department of Mathematics, Masaryk University
Janáčkovo nám. 2a, 662 95 Brno, Czech Republic
E-mail: polak@math.muni.cz