Literally idempotent languages and their varieties - two letter case

Ondřej Klíma and Libor Polák *
Department of Mathematics, Masaryk University,
Janáčkovo nám 2a, 662 95 Brno, Czech Republic

Abstract

A language $L \subseteq A^*$ is literally idempotent in case that $ua^2v \in L$ if and only if $uav \in L$, for each $u, v \in A^*$, $a \in A$. We already studied classes of such languages closely related to the (positive) varieties of the famous Straubing-Thérien hierarchy. In the present paper we start a systematic study of literal varieties of literally idempotent languages, namely we deal with the case of two letter alphabet. First, we consider natural canonical expressions for such languages. Secondly, we describe all possible classes of the form $\mathscr{V}(\{a, b\})$ where $\mathscr{V}$ is a literal variety of literally idempotent languages.
MSC 2000 Classification: 68Q45 Formal languages and automata

1 Introduction

Papers by Straubing [12] on $\mathbb{C}$-varieties and Ésik and el. [3], [4] on literal varieties of languages enable us to consider new significant classes of languages. Due to the result by Kunc [7] we have also an equational logic for those classes.
(Positive) varieties of languages corresponding to pseudovarieties of (ordered) idempotent semigroups/monoids are not very important from the point of the language theory. This is far from being the case for languages corresponding to pseudovarieties of literally idempotent homomorphisms.
We started their study in [6]. Most of our classes resulted by considering intersections of well-known classical (positive) varieties with the class of all literally idempotent languages. Our new classes nicely fitted into the table in Section 8 by Pin [11]. For instance, the intersection of the piecewise testable languages with literally idempotent languages are exactly the boolean combinations of the languages

$$A^*a_1A^*a_2 \dots a_kA^*,\ k \in \mathbb{N}_0,\ a_1, \dots, a_k \in A,\ a_1 \neq a_2 \neq \dots \neq a_k\ .$$

Other results concern the classes of literally idempotent languages corresponding to literal pseudovarieties of monoid homomorphisms onto groups – see [9] and [5].
The main results of the present paper concern literally idempotent languages in the case of two letter alphabet. First, we consider natural canonical expressions for such

*Both authors were supported by the Ministry of Education of the Czech Republic under the project MSM 0021622409 and by the Grant Agency of the Czech Republic grant no. 201/06/0936.

languages. Secondly, we describe all possible classes of the form $\mathscr{V}(\{a, b\})$ where $\mathscr{V}$ is a literal variety of literally idempotent languages.
The paper is organized as follows. After introductory Section 1, we present in Section 2 the basics concerning literal varieties of languages and the literal pseudovarieties of homomorphisms from finitely generated free monoids onto finite monoids. Section 3 recalls how to obtain literal varieties of languages from literal varieties of homomorphisms from free monoids onto monoids. The next section is an introduction to literally idempotent languages and it explains how to get varieties of homomorphisms from certain monoid congruences. Section 5 deals with natural canonical regular expressions for literally idempotent languages over $\{a, b\}$. The next section describes all literally invariant congruences on certain monoid, denoted by M_2, which leads to the second main result. At the end of this section we briefly discuss non-equational literal varieties of languages. The last section describes the congruence on M_2 corresponding to $\mathscr{V}(\{a, b\})$ for the variety $\mathscr{V}$ of languages generated by a literally idempotent regular language L in terms of the minimal complete deterministic finite automata for L.

2 Preliminaries

Valuable treatments on syntactic methods in language theory are books by Almeida [1], Pin [10] and his chapter [11].
Let $\mathcal{M}$ be the class of all surjective homomorphisms from free monoids over nonempty finite sets onto finite monoids. A class $\mathcal{V} \subseteq \mathcal{M}$ is a *literal pseudovariety* if it is closed with respect to the homomorphic images, literal substructures and products of finite families – see Ésik and el. [3], [4] or Straubing [12] for a more general notion of a $\mathbb{C}$-pseudovariety.
More precisely, such class $\mathcal{V}$ satisfies the following

(H) for each $(\varphi : A^* \to M) \in \mathcal{V}$ and a surjective monoid homomorphism $\sigma : M \to N$, we have $\sigma\varphi \in \mathcal{V}$,

(S) for each $(\varphi : A^* \to M) \in \mathcal{V}$ and for each $f : B^* \to A^*$ with $f(B) \subseteq A$, we have $(\varphi f : B^* \to (\varphi f)(B^*)) \in \mathcal{V}$,

(P) each mapping of A^* onto one element monoid is in $\mathcal{V}$, and for each $(\varphi : A^* \to M)$, $(\psi : A^* \to N) \in \mathcal{V}$, the natural homomorphism of A^* onto $A^*/(\ker \varphi \cap \ker \psi)$ is in $\mathcal{V}$.

Let $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Let I_n, for $n \in \mathbb{N}$, be the set of all n-ary implicit operations for the class of finite monoids – see e.g. [1]. We write $\pi^M : M^n \to M$ for the realization of $\pi \in I_n$ on a finite monoid M. A pseudoidentity $\pi = \rho$, where $\pi, \rho \in I_n$, is *literally* satisfied in

$$(\phi : A^* \to M) \in \mathcal{M}$$

$$\text{if} \ \ (\ \forall\ a_1, \dots, a_n \in A\ )\ \pi^M(\phi(a_1), \dots, \phi(a_n)) = \rho^M(\phi(a_1), \dots, \phi(a_n))\ .$$

We write $\phi \models_{\mathscr{L}} \pi = \rho$ in this case.

Usually we fix an alphabet $X_n = \{x_1, \dots, x_n\}$ of variables and we identify a word $u = x_{i_1} \dots x_{i_k} \in X_n^*$ with the implicit operation given by $u^M(b_1, \dots, b_n) = b_{i_1} \dots b_{i_k}$, where M is a finite monoid and $b_1, \dots, b_n \in M$. Examples of implicit operations which are not of this form are u^ω, for $u \in X_n^+$. We define

$$((x_{i_1} \dots x_{i_k})^\omega)^M(b_1, \dots, b_n) = b^\omega \ ,$$

where $b = b_{i_1} \dots b_{i_k}$ and b^ω is the unique idempotent in the set $\{b, b^2, b^3, \dots\}$.
A special case of the main result of Kunc [7] follows.

Result 1. (Kunc) *The literal pseudovarieties of homomorphisms onto finite monoids are exactly the subclasses of $\mathcal{M}$ defined by the literal satisfaction of sets of pseudoidentities.*

By a *quotient* of $L \subseteq A^*$ we mean any set $u^{-1}Lv^{-1} = \{ w \in A^* \mid uwv \in L \}$ where $u, v \in A^*$.
A *class* of (regular) languages is an operator $\mathscr{V}$ assigning to each non-empty finite set A a set $\mathscr{V}(A)$ of regular languages over the alphabet A.
Such a class is a *variety* if

(0) for each A, we have $\emptyset$, $A^* \in \mathscr{V}(A)$,

(i) each $\mathscr{V}(A)$ is closed with respect to finite unions, finite intersections, complements and quotients, and

(ii) for each non-empty finite sets A and B and a homomorphism $f : B^* \to A^*$, $K \in \mathscr{V}(A)$ implies $f^{-1}(K) \in \mathscr{V}(B)$.

A modification of (ii) to

(ii') for each non-empty finite sets A and B and a homomorphism $f : B^* \to A^*$ with $f(B) \subseteq A$, $K \in \mathscr{V}(A)$ implies $f^{-1}(K) \in \mathscr{V}(B)$

leads to the notions of *literal* variety of languages.
Let $L \subseteq A^*$ be a regular language. Recall that the *syntactic congruence* $\sim_L$ on A^* is defined by

$$u \sim_L v \text{ if and only if } (\ \forall\, p, q \in A^*\ )\ (\ puq \in L \iff pvq \in L\ )\ .$$

Further, the structure $\mathsf{O}\,(L) = A^*/\sim_L$ is called the *syntactic monoid* of L and the mapping $\phi_L : A^* \to \mathsf{O}\,(L),\ u \mapsto u \sim_L$ is the *syntactic homomorphism.*
For a class $\mathscr{V}$ of languages, let

$$\mathsf{M}\,(\mathscr{V}) = \ \langle\ \{\ \phi_L : A^* \to \mathsf{O}\,(L) \mid A \text{ non-empty finite},\ L \in \mathscr{V}(A)\ \}\ \rangle$$

be the literal pseudovariety generated by the syntactic homomorphisms of members of $\mathscr{V}$, and conversely, for $\mathcal{V} \subseteq \mathcal{M}$,

$$\mathcal{V} \mapsto \mathsf{L}\,(\mathcal{V}),\ \text{ where } (\mathsf{L}\,(\mathcal{V}))(A) = \{\ L \subseteq A^* \mid \phi_L \in \mathcal{V}\ \} \text{ for each } A\ .$$

Result 2. (Ésik and Larsen [4], Straubing [12]) *The operators* M *and* L *are mutually inverse bijections between the classes of literal varieties of languages and literal pseudovarieties of homomorphisms onto finite monoids.*

3 Literal Universal Algebra

In this section and in Lemmas 2 and 3, we consider letters rather as variables of an infinite set X, so let $X = \{x_1, x_2, \dots\}$ and let, for $n \in \mathbb{N}$, $X_n = \{x_1, \dots, x_n\}$.
Let $\boldsymbol{\mathcal{M}}$ be the class of all surjective homomorphisms from free monoids over non-empty sets onto monoids. A class $\boldsymbol{\mathcal{V}} \subseteq \boldsymbol{\mathcal{M}}$ is a *literal variety* if the following conditions are satisfied

(**H**) for each $(\varphi : A^* \to M) \in \boldsymbol{\mathcal{V}}$ and a surjective monoid homomorphism $\sigma : M \to N$, we have $\sigma\varphi \in \boldsymbol{\mathcal{V}}$,

(**S**) for each $(\varphi : A^* \to M) \in \boldsymbol{\mathcal{V}}$ and for each $f : B^* \to A^*$ with $f(B) \subseteq A$, we have $(\varphi f : B^* \to (\varphi f)(B^*)) \in \boldsymbol{\mathcal{V}}$,

(**P**) For each system $(\varphi_i : A^* \to M) \in \boldsymbol{\mathcal{V}},\ i \in I$, the natural homomorphism of A^* onto $A^*/\bigcap_{i\in I} \mathsf{ker}\,\varphi_i$ is in $\boldsymbol{\mathcal{V}}$.

A homomorphism $(\varphi : A^* \to M) \in \boldsymbol{\mathcal{M}}$ is *finite* if both A and M are finite. Let $\mathsf{Fin}\,\boldsymbol{\mathcal{V}}$ denote the class of all finite members from $\boldsymbol{\mathcal{V}}$.

Result 3. (Polák [9], Theorem 3) *Literal pseudovarieties of (finite) homomorphisms onto monoids are exactly classes of the form* $\mathsf{Fin}\,\boldsymbol{\mathcal{V}}$ *where* $\boldsymbol{\mathcal{V}}$ *is a union of a chain of literal varieties of homomorphisms onto monoids.*

A congruence ξ on Y^* is *literally invariant* if for each $u, v \in Y^*$ and each $g : Y^* \to Y^*$ with $g(Y) \subseteq Y$, the fact $u\ \xi\ v$ implies $g(u)\ \xi\ g(v)$. Similarly, for each monoid with an explicitly given set of generators.
Let $u, v \in X_n^*$. A homomorphism $(\ \phi : Y^* \to M\ ) \in \boldsymbol{\mathcal{M}}$ *literally satisfies* the identity $u = v$ if

$$(\ \forall\ f : X_n^* \to Y^* \text{ with } f(X_n) \subseteq Y\ )\ (\phi f)(u) = (\phi f)(v)\ .$$

We write $\phi \models_{\mathscr{L}} u = v$. For a class $\boldsymbol{\mathcal{V}} \subseteq \boldsymbol{\mathcal{M}}$, we put

$$\mathsf{Id}_{\mathscr{L}}\,\boldsymbol{\mathcal{V}} = \{\ (u, v) \in X^* \times X^* \mid (\ \forall\ \varphi \in \boldsymbol{\mathcal{V}}\ )\ \varphi \models_{\mathscr{L}} u = v\ \}\ .$$

Let $\Pi \subseteq X^* \times X^*$ be a set of identities. We set

$$\varphi \models_{\mathscr{L}} \Pi \ \text{ if } \ (\forall\ \pi \in \Pi\ )\ \varphi \models_{\mathscr{L}}\ \pi\ , \text{ and}$$

$$\mathsf{Mod}_{\mathscr{L}}\,\Pi = \{\ \varphi \in \boldsymbol{\mathcal{M}} \mid \varphi \models_{\mathscr{L}} \Pi\ \}\ .$$

Result 4. ([9], Theorem 2.1) *The mappings*

$$\boldsymbol{\mathcal{V}} \mapsto \mathsf{Id}_{\mathscr{L}}\,\boldsymbol{\mathcal{V}} \ \textit{and}\ \ \Pi \mapsto \mathsf{Mod}_{\mathscr{L}}\,\Pi$$

are mutually inverse bijections between the class of all literal varieties of homomorphisms onto monoids and the class of all literally invariant congruences on X^*.

We write $\xi_{\boldsymbol{\mathcal{V}}}$ instead of $\mathsf{Id}_{\mathscr{L}}\,\boldsymbol{\mathcal{V}}$. Further, for $\xi \subseteq X^* \times X^*$, we write $\xi_n = \xi \cap (X_n^* \times X_n^*)$.

Result 5. ([9], Theorem 5.1) *Let* $\boldsymbol{\mathcal{V}}$ *be a literal variety of monoid homomorphisms and let* $n \in \mathbb{N}$. *Then* $L \in (\mathsf{L}\,(\mathsf{Fin}\,\boldsymbol{\mathcal{V}}))(X_n)$ *if and only if* L *is a regular language which is a union of classes of* $X_n^*/\xi_{\boldsymbol{\mathcal{V}},n}$.

Remark. If the structure $X_n^*/\xi_{\mathcal{V},n}$ is finite we do not need to write "regular". This will be our case for all non-trivial congruences.

Result 6. ([9], Theorem 5.3) *Let $\mathcal{V}$ be a union of a system of literal varieties of homomorphisms onto monoids, $\mathcal{V} = \bigcup_{i\in I} \mathcal{V}_i$. Then, for $n \in \mathbb{N}$,*

$$(\mathsf{L}\,(\mathsf{Fin}\,\mathcal{V}))(X_n) = \bigcup_{i\in I} (\mathsf{L}\,(\mathsf{Fin}\,\mathcal{V}_i))(X_n)\ .$$

Thus, being interested only in languages over n-letter alphabet it suffices to have a description of all possible $\xi_{\mathcal{V}_i,n}$'s.

4 Literally Idempotent Languages

A regular language L over a finite non-empty alphabet A is *literally idempotent* if its syntactic homomorphism $\varphi_L : A^* \to \mathsf{O}\,(L)$ satisfies the pseudoidentity $x^2 = x$ literally, which means

$$(\ \forall\ a \in A\ )\ a^2 \sim_L a\,,$$

or equivalently

$$(\ \forall\ u, v \in A^*,\ a \in A\ )\ (\ uav \in L \iff ua^2v \in L\ )\,.$$

Result 2 can be restricted to the literally idempotent case as follows.

Result 7. *The operators* M *and* L *are mutually inverse bijections between the classes of literal varieties of literally idempotent languages and literal pseudovarieties of literally idempotent homomorphisms onto finite monoids.*

Result 3 can be modified in an analogous way.

We can introduce a string rewriting system which is given by rules $pa^2q \to paq$ for each $a \in A$, $p, q \in A^*$. Let $\to^*$ be the reflexive-transitive closure of the relation $\to$. We say that a word $u \in A^*$ is a *normal form* of a word w if it satisfies the properties

$$w \to^* u \quad \text{and} \quad (\ u \to^* v \text{ implies } u = v\ )\ .$$

It is easy to see that this system is confluent and terminating. Consequently, for any word $w \in A^*$, there is the unique normal form $\overrightarrow{w} \in A^*$ of the word w. We will denote by $\sim$ the equivalence relation on A^* generated by the relation $\to$. In fact, this equivalence relation is a congruence of the monoid A^*.
For any language $L \subseteq A^*$, we define

$$\overline{L} = \{\, w \in A^* \mid (\ \exists\ u \in L\ )\ u \sim w \,\} \text{ which is } \{\, w \in A^* \mid (\ \exists\ u \in L\ )\ \overrightarrow{u} = \overrightarrow{w}\,\}\ .$$

A complete deterministic automaton $(Q, A, \cdot, i, T)$ is *literally idempotent* if, for each $q \in Q,\ a \in A$, we have $q \cdot a^2 = q \cdot a$. The following is obvious.

Lemma 1. *For a regular language $L \subseteq A^*$, the following statements are equivalent:*

(i) L *is literally idempotent,*

(ii) $\overline{L} = L$,

(iii) $\sim \subseteq \sim_L$,

(iv) the minimal complete deterministic finite automaton for L *is literally idempotent,*

(v) L *is a (disjoint) union (not necessarily finite !) of the languages of the form*

$$a_1^+ a_2^+ \dots a_k^+ ,\ k \in \mathbb{N}_0,\ a_1, \dots, a_k \in A,\ a_1 \neq a_2 \neq \dots \neq a_k\ .$$

□

Let M be the monoid with the presentation $< a_1, a_2, \dots \mid a_1^2 = a_1, a_2^2 = a_2, \dots >$ and similarly, for $n \in \mathbb{N}$, let $M_n = < a_1, \dots, a_n \mid a_1^2 = a_1, \dots, a_n^2 = a_n >$.
Result 4 can be adapted to literal varieties of literally idempotent homomorphisms using literally invariant congruences containing the relation $\sim$. Therefore we can use the following two statements whose proofs are straightforward.

Lemma 2. *(i) The mapping* $\xi \mapsto \xi/\sim$*, where* $(u\sim)\ \xi/\sim\ (v\sim)$ *if and only if* $u\ \xi\ v$ $(u, v \in X^*)$ *is an isomorphism of the lattice of all literally invariant congruences on* X^* *corresponding to literal varieties of literally idempotent monoid homomorphisms (i.e. those congruences containing* $\sim$*) onto the lattice of all literally invariant congruences of the monoid* $X^*/\sim$.
(ii) The mapping $u(x_1, \dots, x_n)\sim\ \mapsto \overrightarrow{u(a_1, \dots, a_n)}$ *is an isomorphism of the lattice of all literal congruences of the monoid* $X^*/\sim$ *onto the lattice of all literally invariant congruences of* M. □

In fact, we are rather interested in the n-variable case.

Lemma 3. *Let* $n \in \mathbb{N}$*. Then*
(0) The relations on X_n^* *of the form* $\xi \cap X_n^* \times X_n^*$*, where* ξ *is a literally invariant congruence on* X^**, are exactly the literally invariant congruences on* X_n^*.
(i) The mapping $\xi \mapsto \xi/\sim$ *is an isomorphism of the lattice of all literally invariant congruences on* X_n^* *corresponding to literal varieties of literally idempotent monoid homomorphisms onto the lattice of all literally invariant congruences of the monoid* $X_n^*/\sim$.
(ii) The mapping $u(x_1, \dots, x_n)\sim\ \mapsto \overrightarrow{u(a_1, \dots, a_n)}$ *is an isomorphism of the lattice of all literally invariant congruences of the monoid* $X_n^*/\sim$ *onto the lattice of all literally invariant congruences of* M_n. □

5 Literally Idempotent Languages over Two Letter Alphabet

If we consider one letter alphabet $\{a\}$, then the literally idempotent languages are exactly $\emptyset, \{1\}, a^+, a^*$.

Now consider a regular language L over $A = \{a\}$ with the minimal deterministic automaton $\mathcal{D} = (Q, A, \cdot, i, T)$. Choose the minimal $d \in \mathbb{N}$ and then the minimal $k \in \mathbb{N}_0$ such that $i \cdot a^k = i \cdot a^{k+d}$. Let

$$R = L \cap \{1, a, \ldots, a^{k-1}\} \text{ and } S = L \cap a^k\{1, a, \ldots, a^{d-1}\} \ .$$

Then $R \cup S(a^d)^*$ is a "canonical" regular expression for L.
The situation for literally idempotent languages over $A = \{a, b\}$ is similar. Each regular language L over A is a disjoint union of the sets

$$L \cap (a\{a,b\}^* a \cup a),\ L \cap (a\{a,b\}^* b \cup \{1\}),\ L \cap b\{a,b\}^* a,\ L \cap (b\{a,b\}^* b \cup b) \ .$$

If L is literally idempotent each summand behaves similarly as a regular language over a single letter alphabet. We consider the first summand (the reasonings about the remaining ones are analogous). Let $\mathcal{A} = (Q, A, \cdot, i, T)$ be the minimal deterministic automaton for L. Choose the minimal $d \in \mathbb{N}$ and then the minimal $k \in \mathbb{N}_0$ such that $i \cdot a(ba)^k = i \cdot a(ba)^{k+d}$. Let

$$R = L \cap a\{1, ba, \ldots, (ba)^{k-1}\} \text{ and } S = L \cap a(ba)^k\{1, ba, \ldots, (ba)^{d-1}\} \ .$$

Then $\overline{R} \cup \overline{S}((b^+a^+)^d)^*$ is a "canonical" regular expression for the first summand of L.

6 Literal Varieties of Literally Idempotent Languages Over Two Letter Alphabet

Now we are going to describe all literally invariant congruences of the monoid M_2. For the sake of simplicity we put $a = a_1,\ b = a_2$. We can identify the elements of M_2 with the words

$$1,\ u_{2\ell+1} = a(ba)^\ell,\ u_{2\ell+2} = (ab)^{\ell+1},\ v_{2\ell+1} = b(ab)^\ell,\ v_{2\ell+2} = (ba)^{\ell+1},\ \ell \in \mathbb{N}_0 \ .$$

On M_2 we have the trivial congruence $\Delta = \{\, (w, w) \in M_2 \times M_2 \mid w \in M_2 \,\}$ and the universal congruence $\nabla = M_2 \times M_2$.
For $k \in \mathbb{N},\ d \in \mathbb{N}$, we put

$$U_{k,d} = \{u_k, u_{k+2d}, \ldots\} \quad \text{and} \quad V_{k,d} = \{v_k, v_{k+2d}, \ldots\} \ ;$$

and we write simply U_k instead of $U_{k,1}$ and V_k instead of $V_{k,1}$.
Notice that for each of the above sets we can choose regular expressions determining it within M_2 :

$$a(ba)^l((ba)^d)^* \text{ for } U_{2l+1,d},\ (ab)^{l+1}((ab)^d)^* \text{ for } U_{2l+2,d},\ l \in \mathbb{N}_0,$$

and similarly for $V_{k,d}$'s.
Such a regular expression r determines a (classical) regular expression $\overline{r}$ by substituting $a^+ = aa^*$ for a and $b^+ = bb^*$ for b.
For $k \in \mathbb{N}$ and $d \in \mathbb{N}$, consider the following equivalences on the set M_2 :

$\rho_{k,d}$ with non-trivial (= non-singleton) classes

$$U_{k,d},\ U_{k+1,d},\dots,U_{k+2d-1,d},V_{k,d},\ V_{k+1,d},\dots,V_{k+2d-1,d}\ ,$$

σ_k with the non-trivial classes $U_k \cup U_{k+1}$ and $V_k \cup V_{k+1}$,

τ_k with the non-trivial classes $U_k \cup V_{k+1}$ and $U_{k+1} \cup V_k$,

ν_k with the non-trivial class $U_k \cup U_{k+1} \cup V_k \cup V_{k+1}$.

For a presentation of an ordered set $(S, \leq)$, it is convenient to use the *covering* relation :

$$p \prec q \text{ if and only if } p < q \text{ and there is no } s \text{ with } p < s < q\ .$$

Proposition 1. *Literally invariant congruences of the monoid M_2 are exactly the relations listed above. Moreover, the covering relation consists exactly of the following pairs :*

$$\rho_{k,pd} \prec \rho_{k,d}\ ,\quad \rho_{k+1,d} \prec \rho_{k,d}\ ,$$

$$\rho_{k,1} \prec \sigma_k\ ,\quad \rho_{k,1} \prec \tau_k\ ,\quad \sigma_k \prec \nu_k\ ,\quad \tau_k \prec \nu_k\ ,$$

$$\sigma_{k+1} \prec \sigma_k\ ,\quad \tau_{k+1} \prec \tau_k\ ,\quad \nu_{k+1} \prec \nu_k\ ,$$

$$\nu_1 \prec \nabla\ ,$$

where $k, d \in \mathbb{N}$, p is a prime number.

Proof. Clearly, all the relations mentioned above are literally invariant congruences and the covering relation among them is as stated.

Let α be an arbitrary literally invariant congruence of the monoid M_2 different from Δ. Recall that the parameter k of words $u_k \in M_2$ and $v_k \in M_2$ is, in fact, the *length* of these words. Let $\ell \in \mathbb{N}$ be the smallest integer such that there is a word w of length ℓ which is not the shortest element in the class $w\alpha$. It is not hard to see that such ℓ exists. Indeed, because $\alpha \neq \Delta$ there is some non-singleton class. If $u_k\ \alpha\ v_k$ then $u_k = au_k\ \alpha\ av_k = u_{k+1}$. So there is a class which contains words of different length.

Observe that $\ell = 1$ implies $a\ \alpha\ 1\ \alpha\ b$ and $\alpha = \nabla$ in this case.

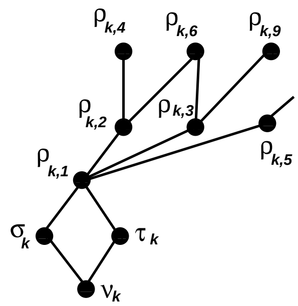
So, assume that $\ell > 1$ in the rest of the proof. If $w = u_\ell$ has the property $u_\ell\ \alpha\ 1$ then we have $a \cdot u_\ell\ \alpha\ a \cdot 1$. Hence $1\ \alpha\ u_\ell = a \cdot u_\ell\ \alpha\ a$ and this is a contradiction with the choice of $w = u_l$. So, $\{1\}$ is a singleton class in the congruence α.

Assume that for ℓ we have $u_\ell\ \alpha\ w'$ where $w' \in M_2$ has length smaller than ℓ. Note that each word which is shorter than ℓ is the shortest word in its class. Hence two words of different length shorter than ℓ are not α-related. We saw above what happens when the words of the same length are α-related. So only $u_{\ell-1}$ and $v_{\ell-1}$ could be α-related. Also each word which is longer than ℓ contains u_ℓ as a subword and consequently it is α-related to a shorter word.

If the first letter of u_ℓ and the first letter of w' are different, i.e. w' starts with b, then $w'\ \alpha\ u_\ell = au_\ell\ \alpha\ aw'$. So the word aw' is not the shortest word in its class,

hence it has length at least ℓ and we see that $w' = v_{\ell-1}$. Analogically, if u_ℓ and w' end with different letters then $w' = u_{\ell-1}$. We will discuss these situations first.
i) If $u_\ell \ \alpha \ v_\ell$ then $u_\ell \ \alpha \ u_{\ell+1}$ as we showed at the beginning of the proof. Because α is a literally invariant congruence we have also $v_\ell \ \alpha \ v_{\ell+1}$ and we can prove inductively that $u_\ell \ \alpha \ u_m$ and $u_\ell \ \alpha \ v_m$ for each $m > \ell$. Hence $\alpha = \nu_{\ell-1}$.
ii) Assume that $u_\ell \ \alpha \ u_{\ell-1}$ and u_ℓ is not α-related to $v_\ell \ \alpha \ v_{\ell-1}$. Then for an arbitrary word w we have $wu_\ell \ \alpha \ wu_{\ell-1}$ and $wv_\ell \ \alpha \ wv_{\ell-1}$. In particular for $m > \ell$ we have $u_m \ \alpha \ u_{m-1}$ and $v_m \ \alpha \ v_{m-1}$. So, $U_{\ell-1} \cup U_\ell \subseteq u_\ell\alpha$ and $V_{\ell-1} \cup V_\ell \subseteq v_\ell\alpha$. This means $\sigma_{\ell-1} \subseteq \alpha$. Each class of $\sigma_{\ell-1}$ contains a word shorter than ℓ, hence these words belongs to different classes of α, i.e. $\sigma_{\ell-1} = \alpha$
iii) The case $u_\ell \ \alpha \ v_{\ell-1}$ is dual to the case ii) when we read words from right to left. In this case we obtain $\alpha = \tau_{\ell-1}$.
Finally, assume that u_ℓ and w' have the same first and last letters, i.e. $w' = u_{\ell-2d}$ for some integer d (recall $w' \neq 1$). We put $k = \ell - 2d$. It is clear that $\rho_{k,d} \subseteq \alpha$. Each class of $\rho_{k,d}$ contains a word shorter than ℓ, hence $\rho_{k,d} = \alpha$. □

A part of the dual of the k-th level of the lattice of all literally invariant congruences on M_2 is depicted below.



The whole lattice is the product of such a level with the chain $1 < 2 < \dots$ of all positive natural numbers with ∇ and Δ adjoined. We draw the dual since we are primarily interested in the corresponding varieties of languages.

According Lemma 3 we have found all literally invariant congruences on X_2^* corresponding to literal varieties of literally idempotent monoid homomorphisms. Therefore we know all the ingrediences $\xi_{\boldsymbol{\mathcal{V}},2}$ used in Result 5 which leads to a description of all "equational" varieties of languages restricted to the alphabet $\{a, b\}$.
Finally, Result 3 together with Result 6 allows us to deal also with arbitrary literal varieties of literally idempotent languages over $\{a, b\}$. A complete description is possible but rather bulky. Considerations of decreasing chains of ρ's is similar to those in [8] concerning mono-unary algebras. We present here all other chains which does not use ρ's and which lead to new varieties of languages given by (real) pseudoidentities. In this way we obtain just three new varieties of languages. At first, if we work with an arbitrary decreasing chains of ν's then the resulting variety $\mathscr{V}_\nu$ is given by literal pseudoidentities $(xy)^\omega x = (xy)^\omega$ and $y(xy)^\omega = (xy)^\omega$ (together with literal idempotency). If we take decreasing chains of σ's then the resulting

variety $\mathscr{V}_\sigma$ of languages is given by the literal pseudoidentity $(xy)^\omega x = (xy)^\omega$ and analogically we obtain the variety of languages $\mathscr{V}_\tau$ corresponding to the pseudoidentity $y(xy)^\omega = (xy)^\omega$ in the case of τ's. The description of the ordered set of all congruences in Proposition 1 implies that there are no other reasonable decreasing chains of congruences.

The considerations of the ordered case lead to slightly more complicated results.

7 Generating Varieties by Languages Using Their Automata

We know that $\mathscr{V}(\{a,b\})$'s for equational literal varieties $\mathscr{V}$ of literally idempotent languages correspond to literally invariant congruences on M_2 which were described in Proposition 1. Let L be a literally idempotent language over the alphabet $\{a,b\}$. Let $\mathcal{A} = (Q, \{a,b\}, \cdot, i, T)$ be the minimal complete deterministic automaton for L. We would like to find the smallest possible $\mathscr{V}(\{a,b\})$ containing L. So we are looking for the greatest literally invariant congruence on M_2 contained in $\sim_L$.
It is well-known that the syntactic homomorphism ϕ_L can be identified with the mapping which maps $u \in \{a,b\}^*$ onto the transformation of Q induced by the word u.
Recall that the automaton $\mathcal{A}$ is literally idempotent. We distinguish several cases :

(i) There is the only cycle in $\mathcal{A}$ and it is of length 1.

(ii) There are exactly two cycles in $\mathcal{A}$, both of the length 1.

(iii) There is the only cycle in $\mathcal{A}$ and it is of length 2.

(iv) There is the only cycle in $\mathcal{A}$ and it is of length $2d$ where $d \geq 2$

or there are exactly two cycles of lengths $2d_1$ and $2d_2$ where $d_1, d_2 \in \mathbb{N}$

or there are exactly two cycles of lengths $2d$ and 1.

Notice that exactly one case of (i) – (iv) happens. Let k be the smallest such that all words w of length $\geq k$ transform the initial state i into a cycle. In the second subcase of (iv), let d equals to the least common multiple of d_1 and d_2.

Proposition 2. *The literally invariant congruence on M_2 corresponding to the language L is*

ν_k *in the case (i),*

σ_k *in the case (ii),*

τ_k *in the case (iii),*

$\rho_{k,d}$ *in the case (iv).*

Proof. Case (i) : All the words with the length $\geq k$ transform all the states into the cycle. Hence $\sim_L$ contains ν_k. Now, for $k \geq 2$, u_{k-1} or v_{k-1} does not transform the initial state i into the cycle, so using Proposition 1, we get that $\sim_L$ does not contain ν_{k-1}.

Case (ii) : For each $m \in \mathbb{N}$, we have that u_m and v_m transform the initial state differently. Therefore $\sim_L$ contains no ν_m. For each $m \geq k$ we have that u_m and u_k induce the same transformation. Similarly for v_m and v_k. Hence $\sim_L$ contains σ_k. Now, for $k \geq 2$, u_{k-1} or v_{k-1} does not transform the initial state i into a cycle and we get that $\sim_L$ does not contain σ_{k-1}.

Case (iii) : Similarly to Case (ii) we have that $\sim_L$ contains no ν_m. For each $m \geq k$ we have that u_m and v_{m+1} induce the same transformation. Similarly for v_m and u_{m+1}. Hence $\sim_L$ contains τ_k. Now, for $k \geq 2$, u_{k-1} or v_{k-1} does not transform the initial state i into the cycle and we get that $\sim_L$ does not contain τ_{k-1}.

Case (iv) : Using similar arguments as above we can show that :

$$\sim_L \supseteq \rho_{k,d}, \quad \sim_L \not\supseteq \rho_{k-1,d}, \quad \sim_L \not\supseteq \rho_{k,d/p} \quad \text{for each prime divisor } p \text{ of } d \ ,$$

$$\text{and} \quad \sim_L \not\supseteq \sigma_k, \quad \sim_L \not\supseteq \tau_k \quad \text{for } d = 1 \ .$$

□

Remark. Notice that the automata corresponding to $\mathscr{V}_\nu$, $\mathscr{V}_\sigma$, $\mathscr{V}_\tau$ are exactly the automata satisfying (i), (ii), (iii) above, respectively.

References

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

[2] Z. Ésik, Extended temporal logic on finite words and wreath product of monoids with distinguished generators, *Proc. DLT 02*, LNCS 2450 (2003), 43–58

[3] Z. Ésik and M. Ito, Temporal logic with cyclic counting and the degree of aperiodicity of finite automata, *Acta Cybernetica* 16 (2003), 1–28, preprint BRICS 2001

[4] Z. Ésik and K.G. Larsen, Regular languages defined by Lindström quantifiers, *Theoretical Informatics and Applications* 37 (2003), 197–242, preprint BRICS 2002

[5] O. Klíma and L. Polák, On literal varieties of homomorphisms onto nilpotent groups. in *Proc. 2nd International Conference on Language and Automata Theory and Applications, Tarragona, Spain, 2008*

[6] O. Klíma and L. Polák, On varieties of literally idempotent languages, to appear in *Theoretical Informatics and Applications*

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

[8] L. Polák, On varieties, generalized varieties and pseudovarieties of homomorphisms. In *Contribution to General Algebra 16*, Verlag Johannes Heyn (2005), 173–188

[9] L. Polák, On varieties and pseudovarieties of homomorphisms onto abelian groups, *Proceedings Intern. Conference on Semigroups and Languages, Lisboa 2005*, World Scientific (2007), 255–264

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

[11] J.-E. Pin, Syntactic semigroups, Chapter 10 in *Handbook of Formal Languages*, G. Rozenberg and A. Salomaa eds, Springer, 1997

[12] H. Straubing, On logical descriptions of regular languages, *Proc. Latin 2002*, LNCS Vol. 2286 (2002), 528–538