On Varieties of Automata Enriched
with an Algebraic Structure
(Extended Abstract)

Ondřej Klíma

Department of Mathematics and Statistics,
Masaryk University,
Brno, Czech Republic *
klima@math.muni.cz

Eilenberg correspondence, based on the concept of syntactic monoids, relates varieties of regular languages with pseudovarieties of finite monoids. Various modifications of this correspondence related more general classes of regular languages with classes of more complex algebraic objects. Such generalized varieties also have natural counterparts formed by classes of finite automata equipped with a certain additional algebraic structure. In this survey, we overview several variants of such varieties of enriched automata.

Algebraic theory of regular languages is a well established field in the theory of formal languages. A basic ambition of this theory is to obtain effective characterizations of various natural classes of regular languages. The fundamental concept is the notion of *syntactic monoid* of a given regular language L, which is the smallest possible monoid recognizing the language L, and which is isomorphic to the transition monoid of the minimal automaton of L. First examples of natural classes of languages, which were effectively characterized by properties of syntactic monoids, were the star-free languages [34] having aperiodic syntactic monoids and the piecewise testable languages [36] having $\mathscr{J}$-trivial syntactic monoids. A general framework for discovering relationships between properties of regular languages and properties of monoids was provided by Eilenberg [9], who established a one-to-one correspondence between so-called *varieties* of regular languages and *pseudovarieties* of finite monoids. Here varieties of languages are classes closed under taking quotients, preimages under morphisms and Boolean operations. On the other hand pseudovarieties of finite monoids are classes closed under taking finite direct products, submonoids and morphic images. Thus a membership problem for a given variety of regular languages can be translated to a membership problem for the corresponding pseudovariety of finite monoids. An advantage of this translation is that pseudovarieties of monoids are exactly classes of finite monoids which have equational description by pseudoidentities [32].

The goal of this contribution is not to overview all notions and applications of the algebraic theory of regular languages. For thorough introduction to that theory we refer to [25]. Other overviews are for example [26] and [44]. A more detailed information concerning the theory of pseudovarieties of finite monoids can be found in the survey [2] or in the books [1] and [33].

We should mention that many interesting classes of regular languages, which are studied by the algebraic methods, come from logic. It is well known that regular languages which are definable in the first order logic of finite linear orderings are exactly star-free languages [23]. Within the class of star-free languages, there were defined the so-called *dot-depth hierarchy* [7] and closely related *Straubing–Thérien hierarchy* [38, 41]. In [42] it was shown that a language belongs to the nth level of the latter

---

hierarchy if and only if it is definable by a formula with n alternations of quantifiers. Moreover, the class of star-free languages is exactly the class of all languages definable by linear temporal logic [15]. For a recent survey on the classes of languages given by fragments of first-order logic we refer to [8] and [40]. Some recent results can be found for example in [19] and [29].

Since not every natural class of languages is closed under all mentioned operations, various generalizations of the notion of varieties of languages were studied. One possible generalization is the notion of *positive varieties* of languages introduced in [24] for which an equational characterization was given in [27]; the positive varieties need not be closed under complementation. In the same direction one can consider varieties which need not be closed under taking unions (see [30]). We shall return to these concepts later. Another possibility is to weaken the closure property for preimages under morphisms. In this way one can consider $\mathscr{C}$-varieties of regular languages which were introduced in [39] and whose equational description was given in [20]. Here we require the presence of preimages under morphisms only for morphisms from a certain special class $\mathscr{C}$. An important example is the class formed by morphisms which map letters to letters; such varieties of languages (so-called *literal varieties)* and the corresponding pseudovarieties of monoids with marked generators (so-called *monoid-generator pairs*) were studied in [12]. Classes of languages with a complete absence of the preimages requirement were studied in [13].

In our contribution we would like to consider varieties of automata as another natural counterpart to varieties of regular languages. We should emphasize that the considered automata are deterministic finite automata. Characterizing of varieties of languages by properties of minimal automata is quite natural, since usually we assume that an input of a membership problem for a fixed variety of languages is given exactly by a minimal deterministic automaton. For example, if we want to effectively test whether an input language is piecewise testable, we do not want to compute its syntactic monoid which could be quite large[1]. Instead of that we consider a condition which must be satisfied by its minimal automaton and which was given in the original Simon's paper [36]. This characterization was used in [37] and [43] to obtain a polynomial and quadratic algorithm, respectively, for testing piecewise testability. In [18] Simon's condition was reformulated and the so-called *locally confluent acyclic automata*[2] were defined. Therefore we are looking for a general definition of a term *variety of automata*, to obtain a setting in which we could talk, for example, about the variety of locally confluent acyclic automata.

Let us consider a minimal automaton $\mathscr{A}_L$ of a regular language L. A first easy observation is the following: if we change the initial state in $\mathscr{A}_L$ then the resulting automaton recognizes a left quotient of the original language L. Similarly but not trivially, if we change the final states, the resulting automaton recognizes a Boolean combination of right quotients of the original language L. Since we are interested in characterizations of varieties of languages, the choice of an initial state and final states can be left free and we can consider only underlying labeled graphs[3] which will form our varieties of automata. Furthermore, since varieties of languages are closed under taking unions and intersections, we need to include direct products of automata in our varieties of automata. Considering a preimage of a given regular language L under some morphism f, one can construct an automaton from the minimal automaton $\mathscr{A}_L$ of L, so-called f-*subautomaton*, where states form a subset and a new action by each letter a is the same as the action by the word $f(a)$ in the original automaton. Since these constructions generate new automata, namely products of automata and f-subautomata, and since we are mainly interested in minimal automata, we also include into our variety of automata all morphic images of existing automata. Finally, from technical reasons we add disjoint unions of automata. Thus a variety of automata will be a class of automata closed

---

[1]More than $(n-1)!$ where n is the number of states of the minimal automaton, see [5] for precise bounds.

[2]These automata recognize exactly piecewise testable languages and paper [18] contains a new (purely automata based) proof which does not use Simon's original result.

[3]Such automata without initial and final states are sometimes called semiautomata in the literature.

under taking products, disjoint unions, morphic images and f-subautomata. And of course, when we are limited to morphisms from a certain class $\mathscr{C}$, we can even talk about $\mathscr{C}$-varieties of automata. Then one can prove an Eilenberg type correspondence: varieties of languages correspond to varieties of automata. This concept occurred in [10] in the case of literal morphisms and in [6] under the name varieties of $\mathscr{C}$-actions. In particular, one can consider the variety of all counter-free automata [23] characterizing star-free languages or the variety of all locally confluent acyclic automata.

Now we enrich automata by an algebraic structure. If we start with a deterministic automaton where all states are reachable from the initial one then we can assign to each state q the set L_q consisting of all words which are acceptable if the computation starts from this state. Sometimes L_q is called the *future* of the state q. It is known [4] that identifying the states with the same future produces a minimal automaton. Thus a state q in the minimal automaton $\mathscr{A}_L$ can be identified with its future L_q and therefore it is a subset of A^*. Then such states are ordered by inclusion, which means that each minimal automaton is implicitly equipped with a partial order. Moreover, final states[4] form an upward closed subset. This leads to a notion of partially ordered automata where actions by letters are isotone mappings and languages are recognized by final states which form an upward closed subset.

Furthermore, varieties, or more generally $\mathscr{C}$-varieties, of partially ordered automata[5] can be defined once again as classes which are closed under taking products, disjoint unions, morphic images and f-subautomata. Now one can prove that these varieties of partially ordered automata correspond to positive varieties of languages. A well known example is the level $1/2$ in the Straubing-Thérien hierarchy of star-free languages. The effective characterization of the level $1/2$ can be found in [3]. This characterization can be equivalently stated as validity of the identity $1 \leq x$ in the syntactic ordered monoid of a language [28]. Therefore, the corresponding variety of partially ordered automata is formed by automata where actions are increasing mappings (for a state q and a letter a we have $q \cdot a \geq q$).

Now we return to the representation of the minimal automaton $\mathscr{A}_L$ of a regular language L where a state $q = L_q$ is a subset of A^* and we consider all possible intersections of states. Since we have only finitely many states in $\mathscr{A}_L$, we obtain finitely many intersections. The resulting meet-semilattice $\mathscr{S}_L$ can be naturally equipped with actions by letters: applying a letter a to an intersection $\bigcap_{i \in I} q_i$, i.e. a state in $\mathscr{S}_L$, is the intersection of all states $q_i \cdot a$. If we use as final states those which contain the empty word, then final states form a principal filter in the semilattice $\mathscr{S}_L$. This idea leads to a notion of a *meet automaton* which was introduced in [16]. Here the corresponding varieties of languages are not closed under taking unions, since in the product of automata the corresponding set of final states is not a principal filter. Therefore the corresponding classes of regular languages are conjunctive varieties which were defined in [30]. We have already mentioned that the syntactic (ordered) monoid of a language is isomorphic to the transition (ordered) monoid of the (ordered) minimal automaton of the language. Analogous statement is valid in the case of meet automata. In particular, the *canonical meet automaton* $\mathscr{S}_L$ of a language L is a minimal meet automaton of a given language. Moreover, its transition structure is a *syntactic semiring* which is a minimal semiring recognizing the language and which can be defined analogously to a syntactic monoid (see [30]). In the paper [16] there are mentioned some examples of $\mathscr{C}$-varieties of languages which can be characterized via varieties of meet automata. There is also a close connection between the notion of a canonical meet automaton and a notion of a *universal automaton* which contains all minimal non-deterministic automata of a given regular language (see [31] and [22]).

One can make one step further. As we add intersections to the representation of minimal automaton,

---

[4] A state is final if and only if it contains the empty word.

[5] There exist several papers which use the term ordered (deterministic, non-deterministic or two-way) automaton in a different meaning, e.g. in [35] it is required that an action by a letter is increasing but need not be isotone.

we can try to add also unions. In other words, we consider the sublattice of the lattice 2^{A^*} generated by $\mathscr{A}_L$. Since this lattice is distributive, we define an abstract notion of a *distributive lattice automata* (*DL-automata*) which are automata enriched by a distributive lattice structure, where both operations are compatible with actions by letters. Note that this model differs from lattice automata defined in [21]. We want to define varieties of *DL*-automata as a natural counterpart of generalized varieties of languages which are not required to be closed under taking any of Boolean operations. Indeed, such classes naturally occur in the theory of formal languages: for example, many classes defined by models of quantum automata are of this kind. The goal is a characterization of such classes. Note that it is also possible to extend this principle, consider the Boolean subalgebra of 2^{A^*} generated by $\mathscr{A}_L$ and define a notion of a *BA-automaton*. Before developing this theory we prefer to clarify all aspects of the theory of *DL*-automata, since there are some difficulties. For example, in the case of meet automata, since actions by letters are morphism with respect to the meet operation, actions by sets of letters are also morphisms with respect to this operation. In the case of *DL*-automata, such an extension is not valid.

At the end we could mention that one can extend the construction in at least two natural directions. First, the theory of tree languages is a field where many fundamental ideas from the theory of deterministic automata were successfully generalized. Another recent notion of biautomata (see [17] and [14]) is based on considering both-sided quotients instead of left quotients only. In both cases one can try to apply the previous constructions and consider varieties of automata (enriched by an algebraic structure). Some papers in this direction already exist [11].

Acknowledgement

References

[1] J. Almeida (1994): *Finite semigroups and universal algebra*. World Scientific, Singapore, doi:`10.1142/2481`.

[2] J. Almeida (2005): *Profinite semigroups and applications*. In V.B. Kudryavtsev, I.G. Rosenberg & M. Goldstein, editors: *Structural theory of automata, semigroups, and universal algebra*, *NATO Science Series II: Mathematics, Physics and Chemistry* 207, Springer, pp. 1–45, doi:`10.1007/1-4020-3817-8_1`.

[3] M. Arfi (1987): *Polynomial Operations on Rational Languages*. In F.-J. Brandenburg, G. Vidal-Naquet & M. Wirsing, editors: *STACS*, *Lecture Notes in Computer Science* 247, Springer, pp. 198–206, doi:`10.1007/BFb0039607`.

[4] J. Brzozowski (1962): *Canonical regular expressions and minimal state graphs for definite events*. In: *Mathematical theory of Automata*, *Symposia series* 12, Research Institute, Brooklyn, pp. 529–561.

[5] J. Brzozowski & B. Li (2013): *Syntactic Complexity of R- and J-Trivial Regular Languages*. In H. Jürgensen & R. Reis, editors: *DCFS*, *Lecture Notes in Computer Science* 8031, Springer, pp. 160–171, doi:`10.1007/978-3-642-39310-5_16`.

[6] L. Chaubard, J.-É. Pin & H. Straubing (2006): *Actions, wreath products of C-varieties and concatenation product*. *Theor. Comput. Sci.* 356(1-2), pp. 73–89, doi:`10.1016/j.tcs.2006.01.039`.

[7] R. Cohen & J. Brzozowski (1971): *Dot-Depth of Star-Free Events*. *J. Comput. Syst. Sci.* 5(1), pp. 1–16, doi:`10.1016/S0022-0000(71)80003-X`.

[8] V. Diekert, P. Gastin & M. Kufleitner (2008): *A Survey on Small Fragments of First-Order Logic over Finite Words*. *Int. J. Found. Comput. Sci.* 19(3), pp. 513–548, doi:`10.1142/S0129054108005802`.

[9] S. Eilenberg (1976): *Automata, Languages and Machines, vol. B.* Academic Press.

[10] Z. Ésik & M. Ito (2003): *Temporal Logic with Cyclic Counting and the Degree of Aperiodicity of Finite Automata.* Acta Cybern. 16(1), pp. 1–28. Available at `http://www.inf.u-szeged.hu/actacybernetica/edb/vol16n1/Esik_2003_ActaCybernetica.xml`.

[11] Z. Ésik & S. Iván (2008): *Some Varieties of Finite Tree Automata Related to Restricted Temporal Logics.* Fundam. Inform. 82(1-2), pp. 79–103. Available at `http://iospress.metapress.com/content/4216mrh7r6477172/`.

[12] Z. Ésik & K.G. Larsen (2003): *Regular languages definable by Lindström quantifiers.* RAIRO - Theoretical Informatics and Applications 37(3), pp. 179–241, doi:`10.1051/ita:2003017`.

[13] M. Gehrke, S. Grigorieff & J.-É. Pin (2008): *Duality and Equational Theory of Regular Languages.* In L. Aceto, I. Damgård, L.A. Goldberg, M.M. Halldórsson, A. Ingólfsdóttir & I. Walukiewicz, editors: *ICALP (2),* Lecture Notes in Computer Science 5126, Springer, pp. 246–257, doi:`10.1007/978-3-540-70583-3_21`.

[14] M. Holzer & S. Jakobi (2013): *Minimization and characterizations for biautomata.* In S. Bensch, F. Drewes, R. Freund & F. Otto, editors: *NCMA,* 294, Österreichische Computer Gesellschaft, pp. 179–193.

[15] H. Kamp (1968): *Tense logic and theory of linear orders.* Ph.D. thesis, University of California.

[16] O. Klíma & L. Polák (2008): *On varieties of meet automata.* Theor. Comput. Sci. 407(1-3), pp. 278–289, doi:`10.1016/j.tcs.2008.06.005`.

[17] O. Klíma & L. Polák (2012): *On biautomata.* RAIRO - Theor. Inf. and Applic. 46(4), pp. 573–592, doi:`10.1051/ita/2012014`.

[18] O. Klíma & L. Polák (2013): *Alternative Automata Characterization of Piecewise Testable Languages.* In M.-P. Béal & O. Carton, editors: *Developments in Language Theory,* Lecture Notes in Computer Science 7907, Springer, pp. 289–300, doi:`10.1007/978-3-642-38771-5_26`.

[19] M. Kufleitner & P. Weil (2012): *On logical hierarchies within FO^2-definable languages.* Logical Methods in Computer Science 8(3), doi:`10.2168/LMCS-8(3:11)2012`.

[20] M. Kunc (2003): *Equational description of pseudovarieties of homomorphisms.* RAIRO - Theoretical Informatics and Applications 37(3), pp. 243–254, doi:`10.1051/ita:2003018`.

[21] O. Kupferman & Y. Lustig (2007): *Lattice Automata.* In B. Cook & A. Podelski, editors: *VMCAI,* Lecture Notes in Computer Science 4349, Springer, pp. 199–213, doi:`10.1007/978-3-540-69738-1_14`.

[22] S. Lombardy & J. Sakarovitch (2008): *The universal automaton.* In J. Flum, E. Grädel & T. Wilke, editors: *Logic and Automata,* Texts in Logic and Games 2, Amsterdam University Press, pp. 457–504.

[23] R. McNaughton & S. Papert (1971): *Counter-Free Automata.* M.I.T. Press.

[24] J.-É. Pin (1995): *A Variety Theorem Without Complementation.* Russian Mathematics 39, pp. 80–90. Available at `http://www.liafa.jussieu.fr/~jep/publications.html`.

[25] J.-É. Pin (1997): *Syntactic semigroups.* In G. Rozenberg & A. Salomaa, editors: *Handbook of Formal Languages,* 1, Springer, pp. 679–746, doi:`10.1007/978-3-642-59136-5_10`. Available at `www.liafa.jussieu.fr/~jep/publications.html`.

[26] J.-É. Pin (2012): *Equational Descriptions of Languages.* Int. J. Found. Comput. Sci. 23(6), pp. 1227–1240, doi:`10.1142/S0129054112400497`.

[27] J.-É. Pin & P. Weil (1996): *A Reiterman theorem for pseudovarieties of finite first-order structures.* Algebra Universalis 35(4), pp. 577–595, doi:`10.1007/BF01243597`.

[28] J.-É. Pin & P. Weil (1997): *Polynominal Closure and Unambiguous Product.* Theory Comput. Syst. 30(4), pp. 383–422, doi:`10.1007/BF02679467`.

[29] T. Place & M. Zeitoun (2014): *Separating Regular Languages with First-Order Logic.* CoRR abs/1402.3277. Available at `http://arxiv.org/abs/1402.3277`.

[30] L. Polák (2004): *A classification of rational languages by semilattice-ordered monoids*. Archivum Mathematicum 40(4), pp. 395–406. Available at `http://emis.muni.cz/journals/AM/04-4/index.html`.

[31] L. Polák (2005): *Minimalizations of NFA using the universal automaton*. Int. J. Found. Comput. Sci. 16(5), pp. 999–1010, doi:`10.1142/S0129054105003431`.

[32] J. Reiterman (1982): *The Birkhoff theorem for finite algebras*. Algebra Universalis 14, pp. 1–10, doi:`10.1007/BF02483902`.

[33] J. Rhodes & B. Steinberg (2009): *The q-theory of Finite Semigroups*. Monographs in Mathematics, Springer, doi:`10.1007/b104443`.

[34] M. P. Schützenberger (1965): *On Finite Monoids Having Only Trivial Subgroups*. Information and Control 8(2), pp. 190–194, doi:`10.1016/S0019-9958(65)90108-7`.

[35] T. Schwentick, D. Thérien & H. Vollmer (2001): *Partially-Ordered Two-Way Automata: A New Characterization of DA*. In W. Kuich, G. Rozenberg & A. Salomaa, editors: Developments in Language Theory, Lecture Notes in Computer Science 2295, Springer, pp. 239–250, doi:`10.1007/3-540-46011-X_20`.

[36] I. Simon (1975): *Piecewise testable events*. In H. Barkhage, editor: Automata Theory and Formal Languages, Lecture Notes in Computer Science 33, Springer, pp. 214–222, doi:`10.1007/3-540-07407-4_23`.

[37] J. Stern (1985): *Complexity of Some Problems from the Theory of Automata*. Information and Control 66(3), pp. 163–176, doi:`10.1016/S0019-9958(85)80058-9`.

[38] H. Straubing (1981): *A Generalization of the Schützenberger Product of Finite Monoids*. Theor. Comput. Sci. 13, pp. 137–150, doi:`10.1016/0304-3975(81)90036-0`.

[39] H. Straubing (2002): *On Logical Descriptions of Regular Languages*. In S. Rajsbaum, editor: LATIN, Lecture Notes in Computer Science 2286, Springer, pp. 528–538, doi:`10.1007/3-540-45995-2_46`.

[40] H. Straubing & P. Weil (2012): *An introduction to finite automata and their connection to logic*. In D. D'Souza & P. Shankar, editors: Modern Applications of Automata Theory, IISc Research Monographs Series, World Scientific, doi:`10.1142/9789814271059_0001`. Available at `http://arxiv.org/abs/1011.6491`.

[41] D. Thérien (1981): *Classification of Finite Monoids: The Language Approach*. Theor. Comput. Sci. 14, pp. 195–208, doi:`10.1016/0304-3975(81)90057-8`.

[42] W. Thomas (1982): *Classifying Regular Events in Symbolic Logic*. J. Comput. Syst. Sci. 25(3), pp. 360–376, doi:`10.1016/0022-0000(82)90016-2`.

[43] A. Trahtman (2001): *Piecewise and Local Threshold Testability of DFA*. In R. Freivalds, editor: FCT, Lecture Notes in Computer Science 2138, Springer, pp. 347–358, doi:`10.1007/3-540-44669-9_33`.

[44] P. Weil (2004): *Algebraic Recognizability of Languages*. In J. Fiala, V. Koubek & J. Kratochvíl, editors: MFCS, Lecture Notes in Computer Science 3153, Springer, pp. 149–175, doi:`10.1007/978-3-540-28629-5_8`. Available at `http://arxiv.org/abs/cs/0609110`.