A COUNTEREXAMPLE TO A CONJECTURE CONCERNING CONCATENATION HIERARCHIES

JORGE ALMEIDA AND ONDŘEJ KLÍMA

Abstract. We give a counterexample to the conjecture which was originally formulated by Straubing in 1986 concerning a certain algebraic characterization of regular languages of level 2 in the Straubing-Thérien concatenation hierarchy of star-free languages.

1. Introduction

This paper contributes to one of the most interesting open problems in the theory of regular languages, namely the effective characterization of languages of the second level in the Straubing-Thérien concatenation hierarchy of star-free languages. For a detailed overview, missing definitions, and complete references we refer to the basic survey paper [5], where Section 8.1 is devoted to the Straubing-Thérien hierarchy and Section 8.5 contains a conjecture concerning the second level, which is the main topic of this paper.

The individual levels of the Straubing-Thérien hierarchy are defined inductively by alternately taking polynomial and Boolean closures, starting from the trivial variety of languages. It is decidable whether a given regular language belongs to each of the levels 0, 1/2, 1 and 3/2, but no algorithm is known to decide the same problem for level 2 or higher. Since the class $\mathcal{V}_2$ of all languages from the second level forms a variety of languages, one can consider the corresponding pseudovariety of monoids V_2 according to Eilenberg's correspondence. Thus the main longstanding open problem in the topic is the membership problem for the pseudovariety of monoids V_2, i.e. the problem of deciding whether a given finite monoid belongs to V_2.

It was proved in [6] that the languages from the second level $\mathcal{V}_2$ are the finite Boolean combinations of the languages of the form $A_0^*a_1A_1^*a_2\dots a_kA_k^*$, where the a_i's are letters and the A_j's are subsets of A. Also some algebraic characterizations of the class V_2 were established in [6]. Here we only recall that $\mathsf{V}_2 = \mathsf{PJ}$, where PJ is the pseudovariety of finite monoids generated by all power monoids $\mathcal{P}(M)$, where M is an arbitrary finite $\mathcal{J}$-trivial monoid. Unfortunately, there is no general algorithm to compute the power operator [2], even though many computations on specific pseudovarieties have been carried out [1, Chapter 11]. Perhaps surprisingly, none of the mentioned nontrivial results has so far led to any solution of the membership problem for V_2.

Straubing conjectured that V_2 is equal to a certain pseudovariety CJ which is given by an effective description. Straubing's conjecture was originally formulated in [12], while it was not repeated in the full version of that paper [13]. Later Straubing and Weil corrected a certain technical error contained in [13] and stated the equality $\mathsf{V}_2 = \mathsf{CJ}$ as Conjecture 2.5 in [14]. Straubing proved the inclusion $\mathsf{V}_2 \subseteq \mathsf{CJ}$ and that the classes V_2 and CJ do not differ on monoids generated by two elements. It has also been shown by Cowan [3, 4] that the two classes contain precisely the same inverse monoids. The notation CJ is not used in [5], but it is used for example in [1, page 400], where defining pseudoidentities for the variety CJ can be found.

We will use the alternative formulation of the Straubing conjecture based on the equality $\mathsf{CJ} = \mathsf{B}_1 ⓜ \mathsf{Sl}$, which follows from non-trivial general results given by Pin and Weil in [7] (see also [5, Theorem 6.5]). Here ⓜ is the Mal'cev product [5, Section 6], B_1 is the pseudovariety of finite semigroups corresponding to the variety of languages of dot-depth one, and Sl is the pseudovariety of finite semilattices. A general conjecture [5, Conjecture 8.1] concerning the Boolean-polynomial closure was formulated by Pin and Weil originally in [9]. Although the general conjecture was corrected in [10], all these adaptations did not change the original Straubing conjecture for the class V_2, so the present-day conjecture is the following.

Conjecture (Pin, Straubing, Weil [9, 10, 12, 13, 14])**.** $\mathsf{V}_2 = \mathsf{B}_1 ⓜ \mathsf{Sl}$.

In this paper we provide a counterexample to this conjecture, and consequently also to the generalization from [10].

Theorem. $\mathsf{V}_2 \neq \mathsf{B}_1 ⓜ \mathsf{Sl}$.

This of course does not solve the original problem of effectively characterizing $\mathcal{V}_2$ and in a sense it shows that the situation is rather more complicated than expected. Yet, our method does provide a tighter upper bound for V_2 which opens new paths for attacking the problem.

In Section 2, we recall the few preliminaries that are required for the proof of the theorem, which is presented in Section 3.

2. Preliminaries

There are two parallel theories of varieties of languages, namely $*$-varieties of languages and $+$-varieties of languages. In the first case, languages are subsets of A^* and $*$-varieties of languages correspond to pseudovarieties of monoids according to Eilenberg's correspondence, while, in the second case, languages are subsets of A^+ and $+$-varieties of languages correspond to pseudovarieties of semigroups. See [5, Section 1 and 4] for details. Here we note only that some operations have a different interpretation in these two theories, such as language complementation. Positive varieties of languages have been considered in both cases [5, Section 4].

By Reiterman's Theorem [11], pseudovarieties of algebras are defined by pseudoidentities. We will need this concept only for pseudovarieties of monoids. In this case, pseudoidentities are formal equalities of implicit operations, which are operations whose interpretation in finite monoids commutes with homomorphisms. Reiterman's Theorem has been extended by Pin and Weil [8] in particular to pseudovarieties of ordered monoids, the equality (of implicit operations) being replaced by the order relation. A pseudovariety of monoids can be viewed as a pseudovariety of ordered monoids by endowing its elements with all possible compatible partial orders.

Besides so-called explicit operations given by words, the most familiar example of implicit operation is the ω-power, which associates to each element m of a finite monoid its unique idempotent power $m^\omega = m^n$ $(n > 0)$. All implicit operations over a fixed finite set A form a compact monoid under a natural topology, namely the free profinite monoid F_A on the set A, in which the discrete submonoid generated by A, whose elements are the explicit operations, is a free monoid and thus is identified with A^*. In particular, if we let P_A denote the monoid of all subsets of A under the union operation, endowed with the discrete topology, then the natural mapping $A \to P_A$ induces a continuous homomorphism $\alpha : F_A \to P_A$. The restriction of α to A^* is the familiar content function and, more generally, a letter $a \in A$ belongs to $\alpha(u)$ for a given implicit operation $u \in F_A$ if and only if there is a factorization of the form $u = xay$ with $x, y \in F_A$. Thus, α is still called the *content function*. See [1, Section 8.1] for further details.

Trivially, P_A is a semilattice, i.e. a commutative and idempotent monoid. Moreover, since P_A is a relatively free profinite monoid on the set A in the class Sl, for $u, v \in F_A$, we have $\mathsf{Sl} \models u = v$ if and only if $\alpha(u) = \alpha(v)$. Note that the $*$-variety of languages corresponding to Sl is formed by finite Boolean combinations of languages of the form A^*aA^*, where a is a letter from an alphabet A. The pseudovariety Sl is also often denoted J_1, for example in [5].

A subset of A^+ is a *language of dot-depth one* if it is a finite Boolean combination of languages of the form $w_0A^*w_1A^* \dots A^*w_{k-1}A^*w_k$, where $k \geq 0$, $w_0, \dots, w_k \in A^*$, and at least one $w_i \in A^+$. The languages of dot-depth one constitute a $+$-variety of languages and the corresponding pseudovariety of semigroups is the B_1 of Straubing's conjecture. For the definition of the dot-depth hierarchy, connections to the Straubing-Thérien hierarchy, and details on the pseudovariety B_1, we refer to [5, Section 8.2].

Finally, we recall the characterization of $\mathcal{V}_{3/2}$, the level 3/2 of the Straubing-Thérien hierarchy. It consists of finite unions of languages of the form $A_0^*a_1A_1^*a_2 \dots a_kA_k^*$, where each $A_i \subseteq A$ and each $a_j \in A$, and it forms a positive $*$-variety of languages.

Proposition 1 ([9, Theorem 8.7], [5, Theorem 8.9]). *A language is of level* $3/2$ *if and only if its ordered syntactic monoid satisfies the pseudoidentity* $u^\omega v u^\omega \leq u^\omega$ *for all* u, v *such that* $\alpha(u) = \alpha(v)$.

Note that the aperiodicity pseudoidentity $x^{\omega+1} = x^\omega$ is a consequence of pseudoidentities from Proposition 1. The pseudovariety of ordered monoids corresponding to $\mathcal{V}_{3/2}$ is denoted $\mathsf{V}_{3/2}$.

3. Proof of the theorem

For the proof of the theorem, we give a certain pseudoidentity which is satisfied in V_2 and a monoid $M \in \mathsf{B}_1 \circledm \mathsf{SI}$ such that M does not satisfy this pseudoidentity. In other words we have $M \in \mathsf{B}_1 \circledm \mathsf{SI}$ and $M \notin \mathsf{V}_2$.

The following proposition gives new pseudoidentities for the pseudovariety V_2.

Proposition 2. *Let* u *and* v *be implicit operations such that* $\mathsf{V}_{3/2} \models u \leq v$. *Then* $\mathsf{V}_2 \models u^\omega = u^\omega v u^\omega$.

Proof. Since $\mathcal{V}_2$ is the Boolean closure of $\mathcal{V}_{3/2}$, it is clear that $\mathsf{V}_2 \models s = t$ if and only if $\mathsf{V}_{3/2} \models s = t$, i.e. if and only if $\mathsf{V}_{3/2} \models s \leq t$ and $\mathsf{V}_{3/2} \models t \leq s$.

From the assumption $\mathsf{V}_{3/2} \models u \leq v$, we deduce that $\alpha(u) = \alpha(v)$ because $\mathsf{SI} \subseteq \mathsf{V}_{3/2}$. From Proposition 1, we obtain immediately $\mathsf{V}_{3/2} \models u^\omega v u^\omega \leq u^\omega$.

When we multiply $u \leq v$ by u^ω from both sides, we obtain $u^\omega u u^\omega \leq u^\omega v u^\omega$. Since $\mathsf{V}_{3/2} \models x^{\omega+1} = x^\omega$, we conclude that $\mathsf{V}_{3/2} \models u^\omega \leq u^\omega v u^\omega$. □

We consider the following implicit operations over the set of variables $X = \{x, y, z\}$:

$$\pi = (xy)^\omega x \ , \quad \rho = z\,\pi\,\pi\,z \ , \quad \sigma = z\,\pi\,z \ .$$

Proposition 3. *The pseudovariety of finite monoids* V_2 *satisfies the following pseudoidentity*

$$\rho^\omega = \rho^\omega \sigma\, \rho^\omega. \tag{1}$$

Proof. Applying Proposition 1 to the pair of explicit operations xy and xxy, we obtain that $\mathsf{V}_{3/2}$ satisfies the pseudoidentity $(xy)^\omega xxy(xy)^\omega \leq (xy)^\omega$. If we multiply it by x on the right, then we deduce that $\mathsf{V}_{3/2} \models \pi\pi \leq \pi$. Hence $\mathsf{V}_{3/2} \models \rho \leq \sigma$ and the statement follows from Proposition 2. □

In the sequel, we consider a monoid M which is the transformation monoid of the automaton over the alphabet $A = \{a, b, c\}$ given in Figure 1.

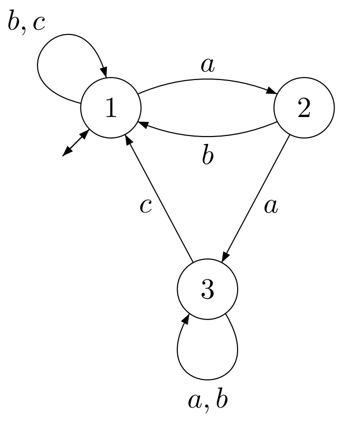

Figure 1. An automaton representation of the monoid M.

Note that the automaton is incomplete since there is no action of the letter c on state 2. Hence the elements of the monoid M are partial transformations. In Figure 2 we can see the structure of the monoid M using the usual eggbox representation of $\mathcal{J}$-classes, where a * marks

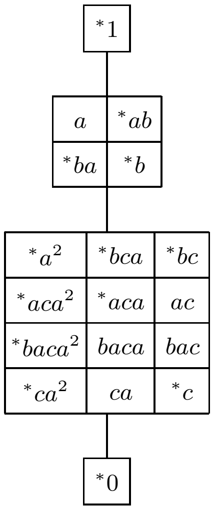

FIGURE 2. The eggbox picture of the monoid M.

a subgroup $\mathcal{H}$-class. A crucial observation is that the partial transformation c has incomplete domain and one-element range. Hence each transformation given by a word containing the letter c has one-element range or it is the empty transformation, i.e. the element 0. An elementary calculation shows that all partial transformations from M which have one-element ranges are $\mathcal{J}$-related. Further, the ideal generated by the element c, denoted by McM, consists of the two bottom $\mathcal{J}$-classes of M and hence it is a completely 0-simple semigroup.

Proposition 4. $M \in \mathsf{B}_1 \textcircled{m} \mathsf{SI}$.

Proof. To prove the statement, we describe a relational morphism [5, Section 6] τ from M to the semilattice P_A. Let $\varphi : A^* \to M$ be the morphism identifying letters from the fixed alphabet $A = \{a, b, c\}$ with the corresponding elements of M, i.e. with partial transformations on the three element set $\{1, 2, 3\}$.

Now we consider the relational morphism $\tau : M \to P_A$, given by the formula

$$\tau(m) = \{\alpha(w) \mid w \in A^*,\ \varphi(w) = m\}, \quad \text{for } m \in M.$$

It is clear that τ is indeed a relational morphism as $\tau = \alpha \circ \varphi^{-1}$. Since $P_A \in \mathsf{SI}$, it is enough to prove that for each $B \in P_A$ we have $\tau^{-1}(B) \in \mathsf{B}_1$.

For $B = \emptyset$, the subsemigroup $\tau^{-1}(B)$ is a singleton. Now assume that $B \neq \emptyset$ and $c \notin B$. The automaton obtained from that in Figure 1 by changing the action of the letter c to let it act like the letter b is the minimal automaton of the language A^*aaA^*, which is of dot-depth one. The transformation semigroup of this automaton, i.e. the syntactic semigroup of the language A^*aaA^*, is the subsemigroup of M generated by the letters a and b, which is usually denoted A_2 in the literature. This syntactic semigroup belongs to B_1, whence so does its subsemigroup $\tau^{-1}(B)$.

If we assume that $c \in B$, then $\tau^{-1}(B)$ is a subsemigroup of the completely 0-simple semigroup McM. Note that every aperiodic completely 0-simple semigroup is locally a semilattice. It is well known [5, Theorem 5.18] that the pseudovariety of semigroups consisting of all local semilattices LSI corresponds to the $+$-variety of all locally testable languages, which are Boolean combinations of languages of the form wA^*, A^*w or A^*wA^* with $w \in A^+$. Since every locally testable language is of dot-depth one, we conclude that $McM \in \mathsf{LSI} \subseteq \mathsf{B}_1$. The required property $\tau^{-1}(B) \in \mathsf{B}_1$ follows. □

We finish the proof of the theorem with the following observation.

Proposition 5. *The monoid M does not satisfy the pseudoidentity (1).*

Proof. We consider the substitution $\psi : F_X \to M$ given by the rules $\psi(x) = a$, $\psi(y) = b$, $\psi(z) = c$. Then it is easy to check that $\psi(\pi) = (ab)^\omega a = a$, $\psi(\rho) = caac = c$, $\psi(\sigma) = cac = 0$. Finally, we have $\psi(\rho^\omega) = c \neq 0 = \psi(\rho^\omega \sigma \rho^\omega)$. □

Acknowledgments

The work of the first author was supported, in part, by *Fundação para a Ciência e a Tecnologia* (FCT) through the *Centro de Matemática da Universidade do Porto* and by the FCT project PTDC/MAT/65481/2006, which is partly funded by the European Community Fund FEDER. The work of the second author was supported, in part, by the Ministry of Education of the Czech Republic under the project MSM 0021622409 and by the Grant no. 201/09/1313 of the Grant Agency of the Czech Republic. The work leading to this paper has also been carried out within the framework of the ESF programme "Automata: from Mathematics to Applications (AutoMathA)", benefiting in particular from an exchange grant for the visit of the second author to the University of Porto.

References

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

[2] K. Auinger and B. Steinberg, On the extension problem for partial permutations, Proc. Amer. Math. Soc. 131 (2003) 2693–2703.

[3] D. Cowan, Inverse monoids of dot-depth two, Int. J. Algebra Comput. 3 (1993) 411–424.

[4] D. Cowan, A result on the dot-depth hierarchy for inverse monoids, in *Semigroups, Automata and Languages*, J. Almeida et al (eds.), 41–57, World Scientific, 1996.

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

[6] J.-E. Pin and H. Straubing, Monoids of upper triangular matrices, in *Semigroups. Structure and Universal Algebraic Problems*, G. Pollák et al (eds.), Colloq. Math. Soc. János Bolyai 39, 259–272, North-Holland, 1985.

[7] J.-E. Pin and P. Weil, Profinite semigroups, Mal'cev products and identities, J. Algebra 182 (1996) 604–626.

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

[9] J.-E. Pin and P. Weil, Polynomial closure and unambiguous product, Theory Comput. Systems 30 (1997) 383–422.

[10] J.-E. Pin and P. Weil, A conjecture on the concatenation product, Theoret. Inform. Appl. 35 (2001) 597–618.

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

[12] H. Straubing, Semigroups and languages of dot-depth 2, Proc. ICALP'86, Lect. Notes in Comput. Sci. 226 (1986) 416–423.

[13] H. Straubing, Semigroups and languages of dot-depth two, Theoret. Comput. Sci. 58 (1988) 361–378.

[14] H. Straubing and P. Weil, On a conjecture concerning dot-depth two languages, Theoret. Comput. Sci. 104 (1992) 161–183.

(ja) Dep. Matemática Pura / CMUP, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal

(ok) Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Kotlářská 2, Brno 61137, Czech Republic