A COUNTEREXAMPLE TO SOME CONJECTURES 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.

1. Introduction

This paper contributes to one of the most interesting open problems in the theory of regular languages, namely the characterization of regular languages of the second level in the Straubing-Thérien concatenation hierarchy. For a detailed overview, missing definitions, and complete references we refer to the basic survey paper [5] (Section 8).

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 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 $A_0^* a_1 A_1^* a_2 \dots a_k A_k^*$ where the a_i's are letters and the A_j's are subsets of A. Since the class $\mathcal{V}_2$ forms a variety of languages, one can consider the corresponding pseudovariety of monoids $\mathbf{V}_2$ according to Eilenberg's correspondence. Some algebraic characterizations of the class $\mathbf{V}_2$ were established in [6]. Here we only recall that $\mathbf{V}_2 = \mathbf{PJ}$, where the pseudovariety of finite monoids $\mathbf{PJ}$ is the pseudovariety 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 specific computations have been carried out [1, Chapter 11].

It was conjectured by Straubing [10] (see also a discussion in the full version of that paper [11] and comments in [12]) that $\mathbf{V}_2$ is equal to a certain pseudovariety $\mathbf{CJ}$ which is given by pseudoidentities (see e.g. page 400 in [1]) and which can be effectively characterized. Straubing proved the inclusion $\mathbf{V}_2 \subseteq \mathbf{CJ}$ and that the classes $\mathbf{V}_2$ and $\mathbf{CJ}$ 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.

We will use the alternative formulation of the Straubing conjecture based on the equality $\mathbf{CJ} = \mathbf{B}_1 \circledm \mathbf{Sl}$ which follows from non-trivial general results given by Pin and Weil in [7] (see also [5, Theorem 6.5]). Here $\circledm$ is the Mal'cev product and $\mathbf{B}_1$ is the pseudovariety of finite semigroups corresponding to the variety of languages of dot-depth one and $\mathbf{Sl}$ is the pseudovariety of finite semilattices (commutative and idempotent monoids)[1]. Pin and Weil [8] have formulated a general conjecture[2] concerning the Boolean-polynomial closure, which was corrected in [9]. All these improvements did not change the original Straubing conjecture for the class $\mathbf{V}_2$, so the present-day conjecture is the following.

Conjecture (Pin, Straubing, Weil [8, 9, 10, 11, 12]). $\mathbf{V}_2 = \mathbf{B}_1 \circledm \mathbf{Sl}$.

In the rest of the paper we provide a counterexample to this conjecture, and consequently also to the generalization from [9].

Theorem. $\mathbf{V}_2 \neq \mathbf{B}_1 \circledm \mathbf{Sl}$.

[1] For the definition of the Mal'cev product see e.g. [5, Section 6] and for details on $\mathbf{B}_1$ see [5, Section 8.2]. The pseudovariety $\mathbf{Sl}$ is also often denoted $\mathbf{J}_1$.

[2] See also the table in [5, Section 10].

2. Proof of the theorem

We give a certain pseudoidentity which is satisfied in $\mathbf{V}_2$ and a monoid $M \in \mathbf{B}_1 \circledm \mathbf{Sl}$ such that M does not satisfy this pseudoidentity. In other words we have $M \in \mathbf{B}_1 \circledm \mathbf{Sl}$ and $M \notin \mathbf{V}_2$.

First of all, we recall the characterization of languages of level 3/2 of the Straubing-Thérien hierarchy. This level $\mathcal{V}_{3/2}$ consists of finite unions of languages of the form $A_0^* a_1 A_1^* a_2 \dots a_k A_k^*$, where each $A_i \subseteq A$ and each $a_j \in A$, and it forms a positive variety of languages. By $\alpha(u)$ we denote the set of all variables occurring in an implicit operation u.[3]

Proposition 1 ([8, 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 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 by $\mathbf{V}_{3/2}$ and if a pseudovariety $\mathbf{V}$ satisfies a pseudoidentity $\pi \leq \rho$ then we write $\mathbf{V} \models \pi \leq \rho$. The following proposition gives new pseudoidentities for the pseudovariety $\mathbf{V}_2$.

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

Proof. It is clear that $\mathbf{V}_2 \models \pi = \rho$ if and only if $\mathbf{V}_{3/2} \models \pi = \rho$, i.e. if and only if $\mathbf{V}_{3/2} \models \pi \leq \rho$ and $\mathbf{V}_{3/2} \models \rho \leq \pi$.

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

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 $\mathbf{V}_{3/2} \models x^{\omega+1} = x^\omega$, we deduce that $\mathbf{V}_{3/2} \models u^\omega \leq u^\omega v u^\omega$. □

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

$$\pi = (xy)^\omega x\ , \quad \rho = z\,\pi\,\pi\,z\ , \quad \sigma = z\,\pi\,z\ . \tag{1}$$

Proposition 3. *The pseudovariety of finite monoids* $\mathbf{V}_2$ *satisfies the following pseudoidentity*

$$\rho^\omega = \rho^\omega \sigma\, \rho^\omega. \tag{2}$$

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

Note that the automaton is not deterministic since there is no action of the letter c on the 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 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. It is easy to see 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. $M \in \mathbf{B}_1 \circledm \mathbf{Sl}$.

[3]More precisely α is a morphism from the free profinite semigroup to the free profinite semilattice over the same set.

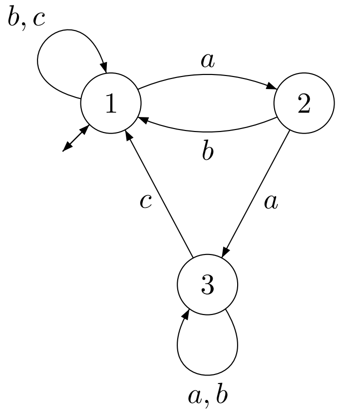

FIGURE 1. An automaton representation of the monoid M.

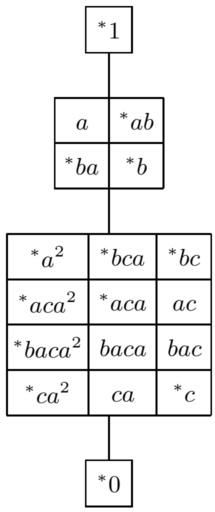

FIGURE 2. The eggbox picture of the monoid M.

Proof. To prove the statement, we describe a relational morphism τ from M to a certain semilattice[4]. 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 $\{1, 2, 3\}$. We denote by P_A the set of all subsets of A and consider the union operation on P_A. So, P_A is a monoid from the class $\mathbf{Sl}$, in particular each element of P_A is an idempotent. We consider a morphism α from A^* to P_A such that $\alpha(w)$ is the set of all letters occurring in $w \in A^*$.[5]

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

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

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

For $B = \emptyset$, the subsemigroup $\tau^{-1}(B)$ is a trivial monoid. Now assume that $B \neq \emptyset$, $c \notin B$. Then it is easy to see that the subsemigroup of M generated by the letters a and b is the syntactic semigroup of the language A^*aaA^* which is of dot-depth one[6]. Hence the subsemigroup $\tau^{-1}(B)$ belongs to $\mathbf{B}_1$.

[4]See [5, Section 6] for formal definitions.

[5]Note that this is a restriction of the content function α used in Proposition 1, which justifies using the same notation.

[6]This semigroup is usually denoted by A_2 in the literature.

If we assume that $c \in B$, then $\tau^{-1}(B)$ is a subsemigroup of the semigroup McM. Note that every aperiodic completely 0-simple semigroup is locally a semilattice. It is well known (cf. [5, Theorem 5.18]) that the pseudovariety of all local semilattices **LSl** corresponds to locally testable languages. Clearly, every locally testable language is of dot-depth one, so we conclude that $McM \in \mathbf{LSl} \subseteq \mathbf{B}_1$. The required property $\tau^{-1}(B) \in \mathbf{B}_1$ follows. □

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

Proposition 5. *The monoid M does not satisfy the pseudoidentity (2).*

Proof. We consider the following substitution $\psi : X \to A$ 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, Polynomial closure and unambiguous product, Theory Comput. Systems 30 (1997) 383–422.

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

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

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

[12] 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