Semigroup Forum Vol. 56 (1998) 225–227

RESEARCH ARTICLE

Modeling Literal Morphisms by Shuffle

Z. Ésik and I. Simon

Communicated by J.-E. Pin

Abstract

We show that literal morphisms on languages may be modeled by the shuffle operation and 2-testable languages. It follows that there exists no nontrivial non-commutative $*$-variety of regular languages closed under shuffle.

The shuffle $L_1 \otimes L_2 \subseteq (A \cup B)^*$ of two languages $L_1 \subseteq A^*$ and $L_2 \subseteq B^*$ is defined by[1]

$$L_1 \otimes L_2 \quad = \quad \{u_1 v_1 \dots u_n v_n : u = u_1 \dots u_n \in L_1,\ v = v_1 \dots v_n \in L_2\}.$$

Suppose that $\mathcal{V}$ is a $*$-variety. (For all notions not defined here we refer to [4].) We define:

- $\mathbf{L}\mathcal{V}$: the $*$-variety generated by the images of the languages in $\mathcal{V}$ under literal morphisms.
- $\mathbf{S}\mathcal{V}$: the $*$-variety generated by the languages $L_1 \otimes L_2$, for $L_1, L_2 \in A^*\mathcal{V}$.
- $\mathbf{S}_0\mathcal{V}$: the $*$-variety generated by the languages $L \otimes B^*$, for $L \in A^*\mathcal{V}$ and $B \subseteq A$.

Let A_0 denote the set $\{a, b\}$ and define

$$L_0 \quad = \quad (ab)^*.$$

Then L is a 2-testable language and the syntactic monoid of L_0 is isomorphic to the monoid BA_2^1 defined on page 109 in [4].

In this note we give a proof of the following result:

Theorem 1. *If $\mathcal{V}$ is a $*$-variety with $L_0 \in A_0^*\mathcal{V}$ then $\mathbf{S}_0\mathcal{V} = \mathbf{S}\mathcal{V} = \mathbf{L}\mathcal{V}$ is the $*$-variety of all regular languages.*

Corollary 2. *The only non-commutative $*$-variety closed under shuffle is the $*$-variety of all regular languages.*

Proof. If $\mathcal{V}$ is a non-commutative $*$-variety with $\mathbf{S}\mathcal{V} \subseteq \mathcal{V}$, then $\mathcal{V}$ contains the 2-testable languages. See [5]. ■

Corollary 2 answers a problem raised by J.-E. Pin [4].

---

[1] Different symbols are used by different authors to denote the shuffle product. Our notation is consistent with [2].

Remark 3. Using a different method, J. Almeida and J.-E. Pin [1] recently gave another proof of Corollary 2.

Theorem 1 follows immediately from the following facts. Theorem 4 may be derived by combining Theorems 1.13 and 1.18 in [4], Chapter 5.

Theorem 4. [4] *If $\mathcal{V}$ is a $*$-variety with $L_0 \in A_0^*\mathcal{V}$, then $\mathbf{L}\mathcal{V}$ is the $*$-variety of all regular languages.* ■

Lemma 5. *If $\mathcal{V}$ is a $*$-variety with $L_0 \in A_0^*\mathcal{V}$, then $\mathbf{L}\mathcal{V} \subseteq \mathbf{S}_0\mathcal{V}$.* ■

In the proof of Lemma 5, we make use of two simple facts. The first fact follows from the proof of Proposition 3.4 of Chapter II, in [3].

Lemma 6. [6] *Each $*$-variety $\mathcal{V}$ is closed under disjoint shuffle. In more detail, if $L_1 \in A^*\mathcal{V}$ and $L_2 \in B^*\mathcal{V}$ with $A \cap B = \emptyset$, then $L_1 \otimes L_2 \in (A \cup B)^*\mathcal{V}$.* ■

Lemma 7. *Suppose that $L_0 \in A_0^*\mathcal{V}$. Then for any finite set A and letter $c \notin A$, $(Ac)^*$ is in $(A \cup \{c\})^*\mathcal{V}$.*

Proof. Let h denote the homomorphism $A \cup \{c\} \to \{a, b\}$ mapping c to b and each letter of A to a. Then $(Ac)^* = h^{-1}((ab)^*)$, so that $(Ac)^* \in (A \cup \{c\})^*\mathcal{V}$. ■

Proof of Lemma 5. Suppose that $L \in A^*\mathcal{V}$ and that φ is a literal morphism $A^* \to B^*$. We need to show that the language $\varphi(L)$ belongs to $B^*\mathbf{S}_0\mathcal{V}$. Without loss of generality we may assume that A and B are disjoint and φ is surjective.

Let c be a letter not contained in $A \cup B$. Define $C = A \cup \{c\}$,

$$\begin{aligned} L_1 &= L \otimes c^* \\ L_2 &= L_1 \cap (Ac)^* \\ L_3 &= L_2 \otimes A^*. \end{aligned}$$

Then $L_1 \in C^*\mathcal{V}$, by Lemma 6, or since any $*$-variety is closed under inverse morphic images, and $L_2 \in C^*\mathcal{V}$, by Lemma 7. Moreover, $L_3 \in C^*\mathbf{S}_0\mathcal{V}$, by definition.

To end the proof, for each $b \in B$, let u_b be a word containing exactly one occurrence of each letter in $\varphi^{-1}(b)$, and no other letter. Consider the homomorphism $h : B^* \to C^*$ defined by $h(b) = u_b c$, $b \in B$. Then $\varphi(L) = h^{-1}(L_3)$, proving $\varphi(L) \in B^*\mathbf{S}_0\mathcal{V}$. ■

Remark 8. A direct proof of Corollary 2 can be shortened even further by simplifying the proof of Lemma 5 for the special case of a morphism $\varphi : (A \cup \{a, b\})^* \to (A \cup \{a\})^*$ which is the identity on A and maps both a and b to a.

Remark 9. It is known that shuffle may in turn be modeled by literal morphisms, see pages 19–21 [3].

Problem 10. Let $\mathcal{J}$ denote the $*$-variety of piecewise testable languages. Is $\mathbf{S}_0\mathcal{J} = \mathbf{L}\mathcal{J}$?

Acknowledgment. The results of these notes were obtained during the first author's visit to São Paulo in April 1995. The hospitality of the University of São Paulo is kindly acknowledged. Both authors would like to thank an anonymous referee for suggesting an improvement in the original argument. The first author was partially supported by grant no. T7383 of the National Foundation for Scinetific Research of Hungary, and the second author by FAPESP Proc. No. 93/0603-1 and by CNPq Proc. No. 523390/94-7(NV).

References

[1] Almeida, J., and J.-E. Pin, personal communication.

[2] Bloom, S. L., and Z. Ésik, *Free shuffle algebras for language varieties,* Theoret. Comput. Sci. **163** (1996), 55–98.

[3] Eilenberg, S., "Automata, Languages, and Machines," vol. A, Academic Press, New York, 1974.

[4] Pin, J.-E., "Varieties of Formal Languages," North Oxford Academic, 1986.

[5] Pin, J.-E., *Variétés de langages et monoïde des parties,* Semigroup Forum **20** (1980), 11–47.

Department of Computer Science
A. József University
Árpád tér 2
6720 Szeged, Hungary
email: esik@inf.u-szeged.hu

Instituto de Matemática e Estatística
Universidade de São Paulo
05508-900 São Paulo, SP, Brasil
email: is@ime.usp.br

Received August 1, 1995
and in final form June 1, 1997