A negative answer to a question of Wilke on varieties of ω-languages

Jean-Eric Pin [(*)]

Abstract. In a recent paper, Wilke asked whether the boolean combinations of ω-languages of the form $\overrightarrow{L}$, for L in a given $+$-variety of languages, form the ω-part of an ∞-variety. We provide a negative answer to this question.

Varieties were introduced by Eilenberg in 1976 to give a unified framework to the algebraic theory of recognizable languages. Recall that a $+$-variety associates to every alphabet A a set $A^+\mathcal{V}$ of recognizable languages of A^+ satisfying the following properties:

(1) for each alphabet A, $A^+\mathcal{V}$ is closed under finite union and complement,

(2) for each morphism $\varphi : A^+ \to B^+$, $L \in B^+\mathcal{V}$ implies $L\varphi^{-1} \in A^+\mathcal{V}$,

(3) if $X \in A^+\mathcal{V}$, then, for all $u \in A^*$, $u^{-1}X, Xu^{-1} \in A^+\mathcal{V}$.

Eilenberg's well-known variety theorem states that $+$-varieties are in one-to-one correspondence with varieties of finite semigroups, that is, classes of finite semigroups closed under taking subsemigroups, quotients and finite direct products.

This result is such a powerful tool for classyfying recognizable languages that it was natural to try to extend it to ω-languages. After some pioneering work by Perrin [3] and Pécuchet [1,2], the right definition was given by Wilke [4,6]. It turns out that it does not suffice to work only with infinite words, but that finite words have to be considered at the same time. More precisely, when A is a finite alphabet, denote by A^ω the set of infinite words on A, and set $A^\infty = A^+ \cup A^\omega$. A subset X of A^∞ is identified with the pair $(X_+, X_\omega) = (X \cap A^+, X \cap A^\omega)$. In particular, X is said to be *recognizable* if both X_+ and X_ω are. An extension of the usual notion of quotients is in

(*) LITP/IBP, Université Paris VI et CNRS, Tour 55-56, 2 place Jussieu, 75251 Paris Cedex 05, FRANCE. **e-mail:** `pin@litp.ibp.fr`

order for the main definition. For $u \in A^*$, put

$$u^{-1}X = \{x \in A^\infty \mid ux \in X\}$$
$$Xu^{-\omega} = \{x \in A^+ \mid (xu)^\omega \in X\}$$

Similarly, for $u \in A^\omega$, put

$$Xu^{-1} = \{x \in A^+ \mid xu \in X\}$$

Now, an ∞-variety associates to every alphabet A a set $A^\infty\mathcal{V}$ of recognizable sets of A^∞ satisfying the following properties:

(1) for each alphabet A, $A^\infty\mathcal{V}$ contains $\emptyset$, A^+ and A^ω and is closed under finite union and complement,

(2) for each map $\varphi : A \to B^+$, which defines a morphism from $\varphi : A^\infty \to B^\infty$, $X \in B^\infty\mathcal{V}$ implies $X\varphi^{-1} \in A^\infty\mathcal{V}$,

(3) if $X \in A^\infty\mathcal{V}$, then, for all $u \in A^*$, $u^{-1}X, Xu^{-\omega} \in A^\infty\mathcal{V}$ and, for all $u \in A^\omega$, $Xu^{-1} \in A^\infty\mathcal{V}$.

Note that by property (1), if $X \in A^\infty\mathcal{V}$, then $X_+ = X \cap A^+$ and $X_\omega = X \cap A^\omega$ are also in $A^\infty\mathcal{V}$. Furthermore, the class of languages which associates to each alphabet A, the set of languages of the form X_+, where $X \in A^\infty\mathcal{V}$, is a $+$-variety, called the $+$-part of $\mathcal{V}$. Similarly, the ω-part of $\mathcal{V}$ associates to each alphabet A, the set of languages of the form X_ω, where $X \in A^\infty\mathcal{V}$.

There is also a variety theorem for ∞-varieties, but the algebraic counterpart is more involved. Since it will not be used in this article, we just refer the interested reader to [4,6] for more information.

There are several natural connections between languages and ω-languages. One of the simplest one is to consider a finite deterministic automaton as a Büchi automaton. In terms of languages, this defines, for a given language L of A^+, the set $\overrightarrow{L}$ of ω-words of A^ω that have an infinite number of prefixes in L. It is well-known that an ω-language is of the form $\overrightarrow{L}$ for some language L if and only if it is accepted by a deterministic Büchi automaton. For this reason, we will use the term *deterministic* to designate the ω-languages of this form.

McNaughton has shown that any recognizable ω-language is a boolean combination of deterministic recognizable ω-languages. In view of a possible extension of McNaugthon's theorem to varieties, Pécuchet considered, for a given $+$-variety, the class $\overrightarrow{\mathcal{V}}$ defined as follows: for each alphabet A, $A^\omega\overrightarrow{\mathcal{V}}$ is the set of all boolean combinations of ω-languages of the form $\overrightarrow{L}$, where $L \in A^+\mathcal{V}$. Wilke [6] asked whether such classes $\overrightarrow{\mathcal{V}}$ form the ω-part of an ∞-variety. As it was shown by Perrin [3], the answer is positive if $\mathcal{V}$ is closed under product, that is, if, for each alphabet A, $L, L' \in A^+\mathcal{V}$ implies $LL' \in \mathcal{V}$. The answer is also positive for all the varieties studied by Pecuchet [1,2]. The aim of this paper is to provide a negative answer to the question raised by Wilke.

Let BA_2 be the five element Brandt aperiodic semigroup, that is, the semigroup with zero presented on $\{a, b\}$ by the relations $aba = a$, $bab = b$ and $a^2 = b^2 = 0$. This semigroup can also be defined as the syntactic semigroup of the language $(ab)^+$ over the two-letter alphabet $\{a, b\}$, or as the semigroup of partial functions given in the following table

	a	b	ab	ba	a^2
1	2	–	1	–	–
2	–	1	–	2	–

Alternatively, BA_2 is the semigroup of two-by-two matrices

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}$$

under the usual matrix multiplication.

Let $\mathbf{V}$ be the variety of finite semigroups generated by BA_2. It was shown in [5] that $\mathbf{V}$ is defined by the identities

$$xyx = xyxyx \qquad x^2y^2 = y^2x^2 \qquad x^2 = x^3 \tag{0.1}$$

Let $\mathcal{V}$ be the $+$-variety corresponding to $\mathbf{V}$. The aim of this paper is to prove the following negative result.

Theorem The class $\overrightarrow{\mathcal{V}}$ is not closed under inverse morphisms.

In particular, this gives a negative answer to the question proposed by Wilke.

Corollary The class $\overrightarrow{\mathcal{V}}$ is not the ω-part of an ∞-variety.

Proof. Let $A = \{a, b\}$, $B = \{a, b, c\}$ and $\varphi : A^+ \to B^+$ be the morphism defined by $a\varphi = ac$ and $b\varphi = b$. Let $L = \{b, ac\}^*a$. A simple computation shows that the syntactic semigroup of L is BA_2 and thus $L \in B^+\mathcal{V}$.

Let $X = \overrightarrow{L}\varphi^{-1}$. It is easily verified that X is the set of ω-words over A containing an infinite number of a's, that is $X = \overrightarrow{A^*a}$. We claim that X is not a boolean combination of ω-languages of the form $\overrightarrow{K}$, with $K \in A^+\mathcal{V}$.

Let us first describe the languages of $A^+\mathcal{V}$. Every semigroup of $\mathbf{V}$ generated by A is a quotient of the relatively free semigroup $F_A\mathbf{V}$, that is, the semigroup presented on A by the relations

$$(1)\ xyx = xyxyx \qquad (2)\ x^2y^2 = y^2x^2 \qquad (3)\ x^2 = x^3 \qquad \text{for all } x \in A^+$$

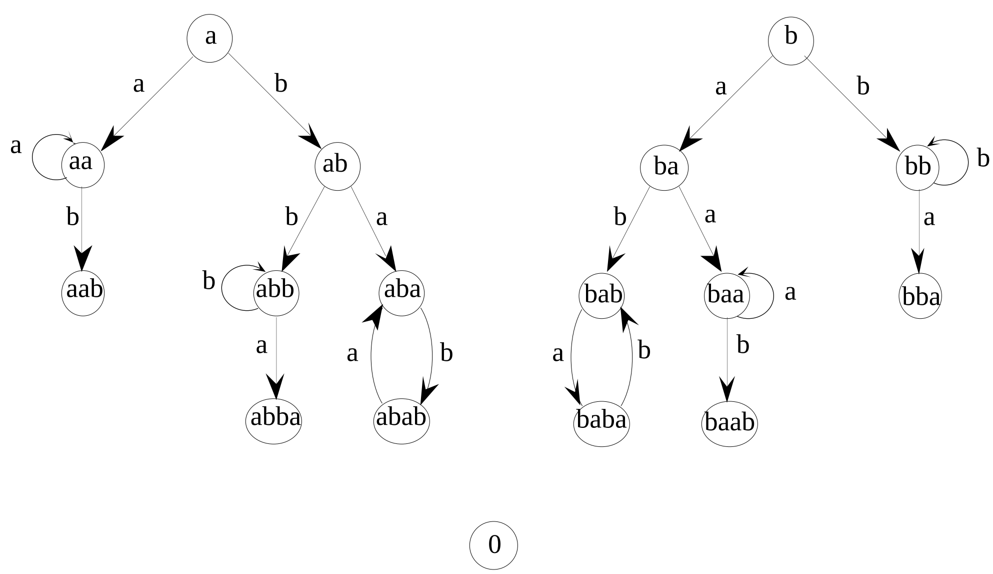
It is easy to derive from (1), (2), (3) the identities

$$(4)\ xyxzx = xzxyx \qquad (5)\ x^2y^2 = x^2y^2x = x^2y^3 \qquad (6)\ x^2yx = x^2y^2$$

Indeed, consider the following derivations, where the identity used at each step is indicated above the equality symbol.

$$\begin{aligned}
xyxzx &\overset{(1)}{=} xyxyxzxzx \overset{(2)}{=} xzxzxyxyx \overset{(1)}{=} xzxyx\\
x^2y^2 &\overset{(2)}{=} y^2x^2 \overset{(3)}{=} y^2x^3 \overset{(2)}{=} x^2y^2x\\
x^2y^2 &\overset{(3)}{=} x^2y^3\\
x^2yx &\overset{(1)}{=} xxyxyx \overset{(4)}{=} xyxxyx \overset{(1)}{=} xyxxyxyx \overset{(2)}{=} xyyxyxxx\\
&\overset{(3)}{=} xyyxyxx \overset{(4)}{=} xyxyyxx \overset{(2)}{=} xyxxxyy \overset{(3)}{=} xyxxyy\\
&\overset{(2)}{=} xyyyxx \overset{(3)}{=} xyyxx \overset{(2)}{=} xxxyy \overset{(3)}{=} xxyy
\end{aligned}$$

Using these identities, one gets a 17 element semigroup whose right representation is shown in the graph below. The edges of this graph are of the form $s \xrightarrow{a} sa$ for $s \in F_A\mathbf{V}$ and $a \in A$, but the arrows ending in 0 are omitted. Thus $aaba = aabb = abbaa = abbab = abaa = ababb = babb = babaa = baaba = baabb = bbaa = bbab = 0$.



For instance, $babab = bab$ since there is an arrow of label b from $baba$ to bab.

It is not difficult to see from this diagram that the languages recognized by $F_A\mathbf{V}$ (that is, the languages of $A^+\mathcal{V}$) are unions of languages of one of the following categories:

(1) aa^+, bb^+, abb^+, baa^+, $ab(ab)^+$, $(ab)^+a$, $ba(ba)^+$, $(ba)^+b$

(2) a finite language

(3) $R = A^*(a^2ba + a^2b^2 + aba^2 + bab^2 + b^2a^2 + b^2ab + b^3)A^*$

Now, if F is finite, $\overrightarrow{F} = \emptyset$. Therefore, every ω-language of the form $\overrightarrow{K}$, where $K \in A^+\mathcal{V}$, can be written as a union of ω-languages of the form a^ω, b^ω, ab^ω, ba^ω, $(ab)^\omega$, $(ba)^\omega$ or $\overrightarrow{R}$. But now, if Z is one of these ω-languages, one has $(aab)^\omega \in Z$ if and only if $aab^\omega \in Z$. Therefore this property also holds if Z is a boolean combination of these ω-languages. Now, since $(aab)^\omega \in X$ but $aab^\omega \notin X$, X cannot be expressed as a boolean combination of ω-languages of the form $\overrightarrow{K}$, where $K \in A^+\mathcal{V}$. Thus $X \notin A^\omega\overrightarrow{\mathcal{V}}$. □

Acknowledgements

The author would like to thank the anonymous referee for pointing out an error in the computation of $F_A\mathbf{V}$ in the first version of this article.

References

[1] J. P. Pécuchet, Variétés de semigroupes et mots infinis, in B. Monien and G. Vidal-Naquet eds., STACS 86, Lecture Notes in Comput. Sci. **210**, Springer, (1986), 180–191.

[2] J. P. Pécuchet, Etude syntaxique des parties reconnaissables de mots infinis, in Proc. 13th ICALP, Kott ed.) Lecture Notes in Comput. Sci. **226**, (1986), 294–303.

[3] D. Perrin, Variétés de semigroupes et mots infinis, *C.R. Acad. Sci. Paris* **295**, (1982), 595–598.

[4] D. Perrin and J.-E. Pin, Semigroups and automata on infinite words, *Semigroups, Formal Languages and Groups, J. Fountain (ed.)*, Kluwer Academic Publishers (1995), 49–72.

[5] A. N. Trahtman, Identities of a five-element 0-simple semigroup, *Semigroup Forum* **48**, (1994), 385–387.

[6] T. Wilke, An algebraic theory for regular languages of finite and infinite words, *Algebra and Computation* **3**, (1993), 447–489.