JOURNAL OF COMPUTER AND SYSTEM SCIENCES **29**, 153–159 (1984)

On Cancellation Properties of Languages Which Are Supports of Rational Power Series

over a field

ANTONIO RESTIVO

Universitá di Palermo, Palermo, Italy

AND

CHRISTOPHE REUTENAUER

LITP, Institut de Programmation, Université Paris VI, Paris 75230, France

Received February 24, 1983; revised November 2, 1983

Two properties of languages which are supports of rational power series are proved: (i) if two supports are complementary, then they are regular languages; (ii) the Ehrenfeucht conjecture is true for these languages.

1. INTRODUCTION

In this paper, we study properties of *supports*, that is, formal languages that are supports of rational power series. We answer affirmatively a conjecture quoted in [14]: if two supports are complementary, then they are regular languages (Theorem 3.1). Secondly we solve, for this special class of languages, the Ehrenfeucht conjecture (cf. [9]): given a language it contains some finite test set (Theorem 4.1).

Recall that supports were introduced in [16], as a natural generalization of regular languages. They possess some properties similar to the properties of regular languages, such as pumping and closure by usual operations (but not complementation). For a survey of these questions, see [14]. The techniques of proof here rely on cancellation properties of supports. For Theorem 3.1, we use a characterization of regularity, through a cancellation property, as proved by Ehrenfeucht, *et al.* [5]. For Theorem 4.1 we establish a more delicate cancellation property, which allows us to prove the Ehrenfeucht conjecture in a similar way as for regular languages.

We study in this paper rational power series with coefficients in a field and not in a semiring as is customary. Actually, it is not reasonable to expect any interesting property of supports, when no assumption is made on the semiring of coefficients; indeed, a general semiring is a very loose structure. Recall that, in order to obtain a basic property such as the pumping lemma for supports, it is necessary to suppose that it is a field, as did Jacob [8] (see also [12]). As another example, let us mention

0022-0000/84 $3.00

a result of E. Sontag [17], who showed that if a rational power series with coefficients in a commutative ring has only a finite number of coefficients, then for each scalar a, the language of those words having a as coefficient is a regular language; this is no longer true for a general semiring (even commutative), as shown by C. Choffrut, see [3, p. 207]. Moreover, Sontag showed that for each language, it is possible to a find a (noncommutative) *ring* such that the characteristic series of this language is rational, see [17, p. 380]. A similar construction shows that there exists a noncommutative ring such that for any language, its characteristic series is a rational power series over this ring. (uniform)

2. Rational Power Series

Let A be a finite alphabet and k a field. A formal power series is a mapping $S: A^* \to k$. The image of a word w through S will be denoted (S, w). The series S is denoted by the infinite sum

$$S = \sum_{w \in A^*} (S, w)\, w.$$

The sum of two series S and T is defined by

$$(S + T, w) = (S, w) + (T, w).$$

The product of a series S by a scalar $\alpha \in k$ is defined by

$$(\alpha S, w) = \alpha(S, w).$$

The product of S by T is defined by

$$(ST, w) = \sum_{uv = w} (S, u)(T, v).$$

With these operations, the set of all formal power series gets a structure of algebra over k, denoted by $k\langle\langle A \rangle\rangle$. It contains A^* and k.

The *support* of a series S is the language

$$\operatorname{supp}(S) = \{w \in A^*, (S, w) \neq 0\}.$$

A *polynomial* is a series with finite support. The set of all polynomials, denoted by $k\langle A \rangle$, is a subalgebra of $k\langle\langle A \rangle\rangle$.

The *star* of a series S such that $(S, 1) = 0$, where 1 stands for the empty word, is defined by

$$S = \sum_{k \geqslant 0} S^k.$$

This infinite sum is well defined because $(S, 1) = 0$.

The set of *rational* power series is the least subalgebra of $k\langle\langle A\rangle\rangle$ containing $k\langle A\rangle$ and closed for the star operation.

A formal power series S is *recognizable* if there exists an integer n, a monoid homomorphism μ from the free monoid A^* into the multiplicative monoid $k^{n\times n}$ of n-by-n matrices over k, a row matrix $\lambda \in k^{1\times n}$ and a column matrix $\gamma \in k^{n\times 1}$ such that for any word w

$$(S, w) = \lambda\mu w\gamma. \tag{2.1}$$

By the Kleene–Schützenberger theorem, *a series is recognizable if and only if it is rational.*

In the sequel, we study languages that are supports of some rational power series; such a language will simply be called a *support*, for brevity. For a proof of the above-cited theorem and properties of supports, see [7] or [15].

3. Complementary Supports

We solve a conjecture quoted in [14].

THEOREM 3.1. *Let L_1, L_2 be two complementary languages which are supports of rational power series. Then they are regular languages.*

Note that the converse is also true, because each regular language L is a support; even the characteristic series $\mathbf{L}$ of L,

$$\mathbf{L} = \sum_{w\in L} w$$

is rational, see, e.g., [14, Theorem 2.5.1]. Moreover the complementary of a regular language is regular.

To prove the theorem, we use a result of [5]. In this paper a property of languages is introduced as follows: a language L has the *cancellation* property if there exists an integer $n \geqslant 1$ such that for any words $w, x, u_1, ..., u_n, y$ verifying

$$w = xu_1 \cdots u_n y$$

there exists i, j, $1 \leqslant i \leqslant j \leqslant n$, such that

$$w \in L \Leftrightarrow xu_1 \cdots u_{i-1}u_{j+1} \cdots u_n y \in L.$$

In other words, by cancelling $u_i \cdots u_j$ in w, one obtains a word w' such that w and w' are simultaneously in or out of L.

The following theorem is due to Ehrenfeucht *et al.*

THEOREM. *If a language has the cancellation property, then it is regular.*

In analogy with the cancellation property, we say that a language L has the *weak cancellation property* if there exists an integer n such that for each word w in L such that $w = xu_1 \cdots u_n y$ for some words $x, u_1, \ldots, u_n, y$, there exists i, j, $1 \leqslant i \leqslant j \leqslant n$, such that $xu_1 \cdots u_{i-1}u_{j+1} \cdots u_n y$ is in L (the weak property is obtained from the strong one by replacing $\Leftrightarrow$ with $\Rightarrow$).

Note that if this property holds for n, then it holds also for any $n' \geqslant n$.

COROLLARY. *Let L_1, L_2 be two complementary languages. If they both have the weak cancellation property, then they are regular.*

Proof. By the theorem of Ehrenfeucht *et al.*, it suffices to show that L_1 has the cancellation property. Let n be such that both L_1 and L_2 have the weak cancellation property for n (see the previous remark). Let $w = xu_1 \cdots u_n y$ be some word. Define i, j, $1 \leqslant i \leqslant j \leqslant n$, by:

If $w \in L_1$ let i, j be such that $xu_1 \cdots u_{i-1}u_{i+1} \cdots u_n \in L_1$ (weak property for L_1).

If $w \in L_2$ let i, j be such that $xu_1 \cdots u_{i-1}u_{j+1} \cdots u_n \in L_2$ (weak property for L_2).

Thus $w \in L_1$ implies $xu_1 \cdots u_{i-1}u_{j+1} \cdots u_n \in L_1$, and $w \notin L_1$ implies $w \in L_2$ hence $xu_1 \cdots u_{i-1}u_{j+1} \cdots u_n y \in L_2$, hence $w \notin L_1$. Thus $w \in L_1 \Leftrightarrow xu_1 \cdots u_{i-1}u_{j+1} \cdots u_n \in L_1$ and L_1 has the cancellation property. ∎

Proof of Theorem 3.1. By the corollary, it suffices to show that each support has the weak cancellation property. Let $L = \mathrm{supp}(S)$ where S is defined by (2.1). Let $w = xu_1 \cdots u_n y \in L$. The vectors

$$\lambda\mu x,\ \lambda\mu xu_1,\ \lambda\mu xu_1u_2, \ldots, \lambda\mu xu_1 \cdots u_n$$

belong to the n-dimensional space $k^{1\times n}$.

Moreover $\lambda\mu x \neq 0$, otherwise $(S, w) = \lambda\mu\, x\mu\, u_1 \cdots u_n y = 0$ and $w \notin \mathrm{supp}(S)$. Hence there exists j, $1 \leqslant j \leqslant n$, such that $\lambda\mu\, xu_1 \cdots u_j$ is a linear combination of $\lambda\mu\, x, \ldots, \lambda\mu\, xu_1 \cdots u_{j-1}$:

$$\lambda\mu\, xu_1 \cdots u_j = \sum_{1 \leqslant i \leqslant j} \alpha_i\, \lambda\mu\, xu_1 \cdots u_{i-1}$$

with α_i in k.

Multiplying on the right by $\mu u_{j+1} \cdots u_n y\gamma$ we obtain

$$(S, w) = \sum_{1 \leqslant i \leqslant j} \alpha_i (S, xu_1 \cdots u_{i-1}u_{j+1} \cdots u_n y).$$

Because $(S, w) \neq 0$, there exists some i, $1 \leqslant i \leqslant j$, such that $(S, xu_i \cdots u_{i-1}u_{j+1} \cdots u_n y) \neq 0$. Hence $xu_1 \cdots u_{i-1}u_{j+1} \cdots u_n y \in L$ and L has the weak cancellation property. ∎

Remark. A quite similar proof shows that if $A^* = L_1 \cup \cdots \cup L_k$ is a partition of A^* into a finite number of supports, then they are all regular.

Theorem 3.1 leaves open the following conjecture.

regular separability of disjoint supports

Conjecture. Let L_1, L_2 be two disjoint supports. Then there exist disjoint regular languages K_1, K_2 such that $L_1 \subset K_1$, $L_2 \subset K_2$. Note that a positive answer would imply Theorem 3.1. This conjecture is true for languages over a one-letter alphabet: if char $(k) = 0$, it is a trivial consequence of the theorem of Skolem–Mahler–Lech, see [10], and if $\text{char}(k) \neq 0$, it is proved in [13].

4. On the Ehrenfeucht Conjecture

The following conjecture is due to Ehrenfeucht, see [9]:

Let $L \subset A^*$ be a language. Then there exists a finite subset K of L such that for any alphabet B and any homomoprhisms f, $g: A^* \to B^*$, the condition $f|K = g|K$ implies $f|L = g|L$.

In other words, to test whether two homomorphisms coincide on L it is enough to do the test on some finite subset of L (depending only on L). This conjecture was proved in the case where L is context-free [1], or when A has only two letters [5] or [6].

THEOREM 4.1. *The Ehrenfeucht conjecture is true for supports.*

As the proof will show, the finite test set may effectively be constructed. We need a lemma, which proves a kind of cancellation property.

LEMMA 4.2. *Let L be a support. Then there exists an integer N such that each word w in L, of length at least N, admits a factorization $w = xuyvz$, such that u, $v \neq 1$ and $xyvz$, $xuyz$, $xyz \in L$.*

Proof. Let $L = \text{supp}(S)$ where S is defined by (2.1). Let $N = n^4$. Let $w \in L$, of length at least $2N$. Then w may be written

$$w = a_1 \cdots a_N s b_N \cdots b_1$$

for some letters $a_1, \ldots, a_N$, $b_1, \ldots, b_N$ and some word s. Consider in the n^4-dimensional vector space $k^{1\times n} \otimes k^{n\times 1} \otimes k^{1\times n} \otimes k^{n\times 1}$ the $n^4 + 1$ vectors

$$\begin{aligned}
&\lambda \otimes \gamma \otimes \lambda \otimes \gamma \\
&\lambda\mu a_1 \otimes \mu b_1 \gamma \otimes \lambda\mu a_1 \otimes \mu b_1 \gamma \\
&\lambda\mu a_1 a_2 \otimes \mu b_2 b_1 \gamma \otimes \lambda\mu a_1 a_2 \otimes \mu b_2 b_1 \gamma \\
&\qquad\vdots \\
&\lambda\mu a_1 \cdots a_N \otimes \mu b_N \cdots b_1 \gamma \otimes \lambda\mu a_1 \cdots a_N \otimes \mu b_N \cdots b_1 \gamma.
\end{aligned}$$

Because $w \in L$, $\lambda\mu w\gamma \neq 0$, hence the first vector is nonzero. Thus, there exists some j, $1 \leqslant j \leqslant N$, such that one has the linear dependence relation

$$(\lambda\mu a_1 \cdots a_j \otimes \mu b_j \cdots b_1\gamma)^2 = \sum_{1 \leqslant i \leqslant j} \alpha_i(\lambda\mu a_1 \cdots a_{i-1} \otimes \mu b_{i-1} \cdots b_1\gamma)^2$$

where $\alpha_i \in k$ and where the square means the tensor square. Let $\gamma' \in k^{n\times 1}$, $\lambda' \in k^{1\times n}$ and $M \in k^{n\times n}$. Then the mapping

$$k^{1\times n} \otimes k^{n\times 1} \otimes k^{1\times n} \otimes k^{n\times 1} \to k$$

$$v_1 \otimes v_2 \otimes v_3 \otimes v_4 \mapsto v_1\gamma' \cdot \lambda' v_2 \cdot \lambda v_3 M v_4 \gamma$$

is linear. Put $\gamma' = \mu a_{j+1} \cdots a_N s b_N \cdots b_1\gamma$, $\lambda' = \lambda\mu a_1 \cdots a_N s b_N \cdots b_{j+1}$, $M = \mu a_{j+1} \cdots a_N s b_N \cdots b_{j+1}$.

Apply this mapping to the above relation, obtaining

$$\begin{aligned}(S, w)^3 = \sum_{1 \leqslant i \leqslant j} &\alpha_i(S, a_1 \cdots a_{i-1}a_{j+1} \cdots a_N s b_N \cdots b_1)\\ &\cdot (S, a_1 \cdots a_N s b_N \cdots b_{j+1}b_{i-1} \cdots b_1)\\ &\cdot (S, a_1 \cdots a_{i-1}a_{j+1} \cdots a_N s b_N \cdots b_{j+1}b_{i-1} \cdots b_1).\end{aligned}$$

From this relation and $(S, w) \neq 0$ we deduce that there exists i, $1 \leqslant i \leqslant j$, such that the ith term of the above sum is nonzero. Let $x = a_1 \cdots a_{i-1}$, $u = a_i \cdots a_j$, $y = a_{j+1} \cdots a_N s b_N \cdots b_{j+1}$, $v = b_j \cdots b_i$, $z = b_{i-1} \cdots b_1$. Then we obtain: $xyvz$, $xuyz$, $xyz \in L$. ∎

Proof of the Theorem. Let $L \subset A^*$ be a support and N the integer of the lemma. Let

$$K = \{w \in L, |w| < N\}.$$

K is a finite subset of L. Let f, g be two homomorphisms $A^* \to B^*$ such that $f|K = g|K$. We show by induction on $|w|$, that for each $w \in L$, $f(w) = g(w)$.

This is surely true if $|w| < N$. Let $|w| \geqslant N$: then, by the lemma, $w = xuyvz$ for some words $u, v \neq 1$, x, y, z such that $xyvz$, $xuyz$, $xyz \in L$. By induction f and g coincide on these three words. Let $F(A)$ (resp. $F(B)$) be the free group generated by A (resp. B). Then f and g extend uniquely to homomorphisms $\bar{f}$, $\bar{g}$: $F(A) \to F(B)$. Because $w = xuyvz = xuyz\,(xyz)^{-1}\,xyvz$, w belongs to the subgroup generated by xyz, $xuyz$, and $xyvz$. Hence $\bar{f}(w) = \bar{g}(w)$, which implies $f(w) = g(w)$. ∎

Remark. If L is a regular language, it is easy to show that there exists an integer N such that each word w in L of length at least N admits a factorization $w = xuvz$ such that xz, xvz, $xuz \in L$. This raises the question of whether this property is also true for supports.

Other open questions concerning supports are the following: (i) If L is the support of a rational power series over $\mathbb{R}$, is L also a support of a rational power series over

$\mathbb{Q}$? (question raised in [14]). (ii) Is it possible to characterize bounded supports, in a way similar to the characterization of bounded regular or context-free languages, as proved in [2, 11]?

Acknowledgments

This work was done while the second author was a visiting professor at the University of Palermo, supported by the Consiglio Nazionale delle Ricerche.

References

1. J. Albert, K. Culik, and J. Karhumäki, Test sets for context-free languages and systems of equations over a free monoid, *Inform. and Control* **52** (1982), 172–186.
2. L. Boasson and A. Restivo, Une caractérisation des langages algébriques bornés, *RAIRO Inform. Théor.* **11** (1977), 203–205.
3. C. Choffrut, Sur les tranductions reconnaissables, *RAIRO Inform. Théo.* **12** (1978), 203–218.
4. K. Culik and A. Salomaa, Test sets and checking words for homomorphism equivalence, *J. Comput. System Sci.* **20** (1980), 379–395.
5. A. Ehrenfeucht, R. Parikh, and G. Rozenberg, Pumping lemmas for regular sets, *SIAM J. Comput.* **10** (1981), 536–541.
6. A. Ehrenfeucht, J. Karhumäki, and G. Rozenberg, On binary equality languages and a solution to the test set conjecture in the binary case, *J. Algebra* **85** (1983), 76–85.
7. S. Eilenberg, "Automatas, Languages and Machines," Vol. A, Academic Press, New York, 1974.
8. G. Jacob, Un théorème de factorisation des produits d'endomorphismes de K^n, *J. Algebra* **63** (1979), 389–412.
9. M. Karpinski (Ed.), New Scottish book of problems, in preparation.
10. C. Lech, A note on recurring series, *Ark. Math.* **2** (1952), 417–421.
11. A. Restivo, Mots sans répétitions et languages rationnels bornés, *RAIRO Inform. Théor.* **11** (1977), 197–202.
12. C. Reutenauer, An Ogden-like iteration lemma for rational power series, *Acta Informa.* **13** (1980), 189–197.
13. C. Reutenauer, Sur les éléments inversibles de l'algèbre de Hadamard des séries rationnelles, *Bull. Soc. Math. France* **110** (1982), 225–232.
14. A. Salomaa, Formal power series in noncommuting variables, *in* "18th Scand. Congr. Math., Proc., Aarhus 1980," *Prog. Math.* **11** (1981).
15. A. Salomaa and M. Soittola, "Automata-Theoretic Aspects of Formal Power Series," Springer-Verlag, New York/Berlin, 1978.
16. M. P. Schützenberger, On the definition of a family of automata, *Inform. and Control* **4** (1961), 245–270.
17. E. Sontag, On some questions of rationality and decidability, *J. Comput. System Sci.* **11** (1975), 375–381.