Theoretical Computer Science 76 (1990) 251-260
North-Holland

PRODUCTS OF LANGUAGES WITH COUNTER*

Pascal WEIL
L.I.T.P., tour 55-56, 2, place Jussieu, 75221 Paris Cedex 05, France

Communicated by D. Perrin
Received February 1988
Revised September 1988

Abstract. It is well known that varieties of rational languages are in one-to-one correspondence with varieties of finite monoids. This correspondence often extends to operations on languages and on monoids. We investigate the special case of the product of languages with counter, and describe the associated operations on monoids and varieties.

Résumé. On sait que les variétés de langages rationnels sont en correspondance bijective avec les variétés de monoïdes finis. Cette correspondance s'étend à de nombreuses opérations sur les langages d'une part, sur les monoïdes d'autre part. Nous étudions le cas particulier du produit de langages avec compteur, et nous décrivons l'opération associée, sur les monoïdes et sur les variétés.

Introduction

The theory of formal languages is one of the bases of theoretical computer science. A central problem of this theory has been, since the origins in the 1950s, the classification of rational languages. A very important tool for this task is the use of the syntactic monoid $M(L)$ of a rational language L on an alphabet A. Indeed, many combinatorial properties of L correspond to algebraic properties of $M(L)$. Eilenberg systematized this correspondence, and showed in 1975 that there is a one-to-one relation between certain families of finite monoids, called M-varieties, and certain families of rational languages, called $*$-varieties.

Numerous instances of this correspondence have been studied in detail since, such as the algebraic characterization of piecewise testable languages [14], locally testable languages [3, 7] or, conversely, the combinatorial characterization of languages whose syntactic monoid is $\mathcal{R}$-trivial [4, 2], whose syntactic monoid contains only solvable groups [15].

Following these ideas, the correspondence between certain operations on languages and certain operations on monoids or M-varieties has been studied, and in particular the relationship between concatenation products and Schützenberger products [18], unambiguous products and locally trivial morphisms [12] or inverse literal morphisms and power monoids [16].

* This work was supported by the P.R.C. Mathématique et Informatique.

This paper is a contribution to this study. We investigate the following operation of product of languages with counter:

$$L_1, \ldots, L_n \to (L_1 a_1 L_2 \ldots a_{n-1} L_n)_{r,p,k}$$

where $L_1, \ldots, L_n$ are languages, $a_1, \ldots, a_{n-1}$ are letters, r, p, k are integers and $(L_1 a_1 L_2 \ldots a_{n-1} L_n)_{r,p,k}$ is the set of words w such that the number of factorizations of w as $u_1 a_1 u_2 \ldots a_{n-1} u_n$ with u_i in L_i $(1 \leqslant i \leqslant n)$ is congruent to r mod k threshold p. We describe the associated monoid operation, and prove that these two operations are in correspondence (in Eilenberg's sense) at the variety level.

1. Preliminaries

All monoids and semigroups considered here are either free or finite. In this section we recall briefly basic concepts and classical results of rational language theory. For general surveys of the theory, including proofs for those stated here, see [4, 6, 8].

1.1. Recognizable and rational languages

Let A be a finite alphabet. A^* and A^+ denote respectively the free monoid and the free semigroup over A. A *language* is a subset L of A^*. A language L in A^* (resp. A^+) is said to be *recognizable* iff there exists a finite monoid (resp. semigroup) T and a morphism $\eta : A^* \to T$ (resp. $A^+ \to T$) such that $L = L\eta\eta^{-1}$. In that case, we say that T (or η) *recognizes* L. L is recognizable iff it is a union of classes of some finite-index congruence of A^* (resp. A^+). The *syntactic congruence* $\sim_L$ is the coarsest congruence that saturates L and the syntactic monoid $M(L)$ (resp. syntactic semigroup $S(L)$) is the quotient $A^*/\sim_L$ (resp. $A^+/\sim_L$). In particular, $M(L)$ (resp. $S(L)$) is the smallest monoid (resp. semigroup) recognizing L.

The class of *rational* languages of A^* (resp. A^+) is the smallest class containing the languages $\{a\}$ $(a \in A)$ and closed under union, concatenation product and star (resp. plus). Recall that if L is a language, then L^* (resp. L^+) is the submonoid of A^* (resp. subsemigroup of A^+) generated by L. Kleene [5] showed that a language is recognizable iff it is rational. A corollary of this fact is that the class of rational languages of A^* (resp. A^+) is closed under boolean operations.

1.2. Varieties

The fact that certain subclasses of rational languages correspond to certain classes of monoids or semigroups was first illustrated by Schützenberger [13]. He proved that star-free languages (those languages that can be obtained from finite languages using only boolean operations and concatenation products) are exactly the languages whose syntactic monoids (resp. semigroups) are aperiodic, i.e. group-free.

Eilenberg [4] showed that this fact is a particular case of a general phenomenon by introducing the concept of varieties.

An **M**-*variety* (resp. **S**-*variety*) is a class of finite monoids (resp. semigroups) that is closed under sub, homomorphic image and finite direct product.

A $*$-*variety* (resp. $+$-*variety*) $\mathcal{V}$ is a family $A^*\mathcal{V}$ (resp. $A^+\mathcal{V}$) of classes of languages of A^* (resp. A^+) defined for all finite alphabets A, and such that

(1) $A^*\mathcal{V}$ (resp. $A^+\mathcal{V}$) is a boolean algebra;

(2) if $\varphi: A^* \to B^*$ (resp. $A^+ \to B^+$) is a morphism and if $L \in B^*\mathcal{V}$ (resp. $B^+\mathcal{V}$), then $L\varphi^{-1} \in A^*\mathcal{V}$ (resp. $A^+\mathcal{V}$);

(3) if $L \in A^*\mathcal{V}$ (resp. $A^+\mathcal{V}$) and $a \in A$, then both $a^{-1}L$ and La^{-1} are in $A^*\mathcal{V}$ (resp. $A^+\mathcal{V}$).

Theorem 1.1. *Let* **V** *be an* **M**-*variety* (*resp.* **S**-*variety*) *and* $\mathcal{W}$ *be a* $*$-*variety* (*resp.* $+$-*variety*).

(1) *If* $A^*\mathcal{V}$ (*resp.* $A^+\mathcal{V}$) *is the class of all languages that are recognized by some element of* **V**, *then* $\mathcal{V}$ *is a* $*$-*variety* (*resp.* $+$-*variety*).

(2) *If* **W** *is the class of all syntactic monoids of languages of* $A^*\mathcal{W}$ (*for all* A), *then* **W** *is an* **M**-*variety* (*resp. an* **S**-*variety*).

(3) *The correspondence* $\mathbf{V} \to \mathcal{V}$ *is one-to-one and onto between the class of all* **M**-*varieties* (*resp.* **S**-*varieties*) *and the class of all* $*$-*varieties* (*resp.* $+$-*varieties*).

1.3. Operations on languages and varieties

A large field of investigation of Eilenberg's variety correspondence is provided by the study of operations on languages (see in particular [11]). For a given operation

$$(L_1, \ldots, L_n) \to \mathrm{Op}(L_1, \ldots, L_n)$$

one tries to describe a monoid operation Op_M such that $\mathrm{Op}(L_1, \ldots, L_n)$ be recognized by $\mathrm{Op}_M(M(L_1), \ldots, M(L_n))$, and such that all languages recognized by $\mathrm{Op}_M(M_1, \ldots, M_n)$ can be described using only Op and languages recognized separately by $M_1, \ldots, M_n$.

Among the main results of this kind, we may cite the following. To the operation [16] $L \to L\varphi$ where φ is a literal morphism, is associated the monoid operation $M \to \mathcal{P}(M)$. To the operation [18]

$$(L_1, \ldots, L_n) \to L_1 a_1 L_2 \ldots a_{n-1} L_n$$

where $a_1, \ldots, a_n$ are letters, is associated the Schützenberger product

$$(M_1, \ldots, M_n) \to \Diamond_n(M_1, \ldots, M_n).$$

We shall not explicitly define this product in this section, as it is a special case of the operation that we investigate.

2. Products with counters

2.1. An operation on languages

The operation on languages presented here generalizes Straubing's product [18] and extends slightly an operation introduced in [9, 10]. It was also considered in [19].

Recall that, if $p \geq 0$ and $k \geq 1$ are integers, $\mathbf{Z}_{p,k}$ denotes the semiring of integers mod k threshold p. If r, s are non-negative integers, we write $r \equiv s \pmod{p, k}$ if r is congruent to s mod k threshold p.

Let $L_1, \ldots, L_n$ be languages over A ($n \geq 2$), and $a_1, \ldots, a_{n-1}$ be letters in A. Let also $r, p \geq 0$, $k \geq 1$. We define $(L_1 a_1 L_2 \ldots a_{n-1} L_n)_{r,p,k}$ to be the set of all words u in A^* such that the number of factorizations of u in the form

$$u = u_1 a_1 u_2 \ldots a_{n-1} u_n$$

with $u_i \in L_i$ for all $1 \leq i \leq n$, is congruent to r mod k threshold p.

If the L_is are all in A^+ and we are considering recognition by semigroups, we shall implicitly consider, in the definition of $(L_1 a_1 L_2 \ldots a_{n-1} L_n)_{r,p,k}$, the word u to be in A^+. A special case of this operation is the (counter-free) product $L_1 a_1 L_2 \ldots a_{n-1} L_n$, which is equal to $(L_1 a_1 L_2 \ldots a_{n-1} L_n)_{1,1,1}$.

The idea of considering products with counter is not new. In particular, Eilenberg considered such products with all the L_is equal to A^* and proved [4] that the boolean algebra generated by the languages of the form $(A^* a_1 A^* \ldots a_k A^*)_{r,0,p}$ (p prime, $r \geq 0$ and $k > 0$) is a $*$-variety that corresponds to the M-variety of all nilpotent groups. In [21], Thérien considered similar ideas; from the point of view of a congruential definition of varieties and in [22], he refined Eilenberg's result to describe the M-variety of nilpotent groups of class m. More recently, Pin [9, 10] considered products of languages with counters to study the closure of the family of open sets of A^* under various operations. It was also considered in the work of Straubing et al. [19].

2.2. An operation on semigroups

Let now Z be a semiring with unit and $S_1, \ldots, S_n$ be semigroups. For all $1 \leq i \leq n$, we denote by S_i^1 the monoid equal to S_i if S_i is a monoid, to $S_i \cup \{1\}$ where 1 is an identity, otherwise. Also we denote by K the semiring $Z[S_1^1 \times \cdots \times S_n^1]$ of all polynomials over $S_1^1 \times \cdots \times S_n^1$ with coefficients in Z. (For the recognizing power of K when Z is some $\mathbf{Z}_{p,k}$, see [1].)

Finally, we define $Z \Diamond_n (S_1, \ldots, S_n)$ to be the subset of the semiring of (n, n)-matrices over K consisting of all matrices $m = (m_{i,j})_{1 \leq i,j \leq n}$ satisfying

- if $i > j$, then $m_{i,j} = 0$;
- if $i = j$, then $m_{i,j} = (1, \ldots, 1, s_i, 1, \ldots, 1)$ from some s_i in S_i;
- if $i < j$, then $m_{i,j} \in Z[1 \times \cdots \times 1 \times S_i^1 \times S_{i+1}^1 \times \cdots S_j^1 \times 1 \times \cdots \times 1]$.

Note that these matrices are exactly the upper-triangular matrices whose ith diagonal entry is an element of S_i (not S_i^1) and whose (i, j)-entry (if $i < j$) is a polynomial with support in $S_i^1 \times \cdots \times S_j^1$. It is easy to check that $Z \Diamond_n (S_1, \ldots, S_n)$ is a semigroup. It is a monoid if $S_1, \ldots, S_n$ are monoids.

Note also that two particular cases of this operation have been studied previously. The first is relative to the case where $Z = \mathbf{Z}_{1,1} = \mathbb{B}$, the boolean semiring. Polynomials in $\mathbb{B}[S_1^1 \times \cdots \times S_n^1]$ correspond to subsets of $S_1^1 \times \cdots \times S_n^1$ and $\mathbb{B} \Diamond_n (S_1, \ldots, S_n)$ is

the classical Schützenberger product $\Diamond_n(S_1, \ldots, S_n)$. The second occurs when $Z = \mathbf{Z}_{0,k} = \mathbf{Z}_k$ for some k and was considered in [9, 10].

In [21], Thérien presented an operation on congruences rather similar to the monoid operation proposed here. The n arguments of the operation were equal and it was used to give generating congruences for certain hierarchies within the **M**-varieties of group-free monoids, solvable groups, $\mathcal{J}$-trivial monoids and nilpotent groups. Note also that Eilenberg used the semiring of polynomials $\mathbf{Z}_n[S]$ in relation to products of languages with counter in his study of p-groups [4].

If $\mathbf{V}_1, \ldots, \mathbf{V}_n$ are **S**-(resp. **M**-) varieties, $Z\Diamond_n(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ denotes the **S**-(resp. **M**-) variety generated by all products of the form $Z\Diamond_n(S_1, \ldots, S_n)$ with $S_i \in \mathbf{V}_i$ for all $1 \leq i \leq n$.

2.3. Main results

The main theorems of this paper are the following.

Theorem 2.1. *Let $n \geq 2$, $p \geq 0$, $k \geq 1$ be integers, and let $M_1, \ldots, M_n$ be monoids. Let $1 \leq i_1 < \cdots < i_{l+1} \leq n$, and let $a_1, \ldots, a_l$ be letters in A. Let finally $L_{i_1}, \ldots, L_{i_{l+1}}$ be languages in A^* recognized respectively by $M_{i_1}, \ldots, M_{i_{l+1}}$. Then, for all $r \geq 0$, $(L_{i_1} a_1 L_{i_2} \ldots a_l L_{i_{l+1}})_{r,p,k}$ is recognized by $\mathbf{Z}_{p,k}\Diamond_n(M_1, \ldots, M_n)$.*

Theorem 2.2. *Let $n \geq 2$, $p \geq 0$, $k \geq 1$ be integers, and let $M_1, \ldots, M_n$ be monoids. The languages in A^* that are recognized by $\mathbf{Z}_{p,k}\Diamond_n(M_1, \ldots, M_n)$ are in the boolean algebra generated by the languages of the form L_h or $(L_{i_1} a_1 L_{i_2} \ldots a_l L_{i_{l+1}})_{r,p,k}$ where $r \geq 0$, $1 \leq i_1 < \cdots < i_{l+1} \leq n$, $a_1, \ldots, a_l$ are letters in A and L_q is a language of A^* recognized by M_q for all $1 \leq q \leq n$.*

Section 3 is devoted to the proof of these theorems. Note that in the case $p = k = 1$, i.e. $\mathbf{Z}_{p,k} = \mathbf{Z}_{1,1} = \mathbf{B}$, both results are well known [13, 18]. A weaker version of Theorem 2.1 was proved in [10] in the case where $p = 0$, i.e. $\mathbf{Z}_{p,k} = \mathbf{Z}_{0,k} = \mathbf{Z}_k$.

Theorems 2.1 and 2.2 can be reformulated in terms of varieties as follows.

Corollary 2.3. *Let $n \geq 2$, $p \geq 0$, $k \geq 1$ be integers and let $\mathbf{V}_1, \ldots, \mathbf{V}_n$ be **M**-varieties. Let also $\mathbf{W} = \mathbf{Z}_{p,k}\Diamond_n(\mathbf{V}_1, \ldots, \mathbf{V}_n)$. Finally, let $\mathcal{V}_i$ $(1 \leq i \leq n)$ denote the $*$-variety associated to $\mathbf{V}_i$, and let $A^*\mathcal{W}$ be the boolean algebra generated by the languages of the form L_h or $(L_{i_1} a_1 L_{i_2} \ldots a_l L_{i_{l+1}})_{r,p,k}$ with $r \geq 0$, $1 \leq i_1 < \cdots < i_{l+1} \leq n$, $a_1, \ldots, a_l \in A$ and $L_q \in A^*\mathcal{V}_q$ for all q. Then $\mathcal{W}$ is a $*$-variety and the corresponding **M**-variety is $\mathbf{W}$.*

Note that these three results are stated in terms of monoids and languages in A^*. Analogous statements relative to languages in A^+ and semigroups also hold. Their proofs are identical to the ones we give below, up to the obvious changes.

3. Proofs of the theorems

3.1. Proof of Theorem 2.1

Let $M_1, \ldots, M_n$ be monoids and $\eta_i : A^* \to M_i$ $(1 \leq i \leq n)$ be morphisms recognizing languages L_i. For each $1 \leq i \leq n$, let $P_i = L_i\eta_i$. Let also $1 \leq i_1 < \cdots < i_l \leq n$ and $a_1, \ldots, a_{l-1}$ be letters of A.

For each letter a of A, let us define $a\mu \in \mathbf{Z}_{p,k}\Diamond_n(M_1, \ldots, M_n)$ by

$a\mu_{i,i} = (1, \ldots, 1, a\eta_i, 1, \ldots, 1)$, if $1 \leq i \leq n$;

$a\mu_{ij} = (1, \ldots, 1)$, if $i = i_u$, $j = i_{u+1}$ and $a = a_u$ for some $1 \leq u < l$;

$a\mu_{i,j} = 0$, otherwise.

We extend μ naturally to a morphism from A^* into $\mathbf{Z}_{p,k}\Diamond_n(M_1, \ldots, M_n)$.

Lemma 3.1. *Let $w \in A^+$. If $1 \leq j < i \leq n$, then $w\mu_{i,j} = 0$. If $1 \leq i \leq n$, then $w\mu_{i,i} = w\eta_i$. If $1 \leq i < j \leq n$ and either i or j is not in $\{i_1, \ldots, i_l\}$, then $w\mu_{i,j} = 0$. Finally, if $i = i_u$ and $j = i_v$ for some $1 \leq u < v \leq l$, then $w\mu_{i,j}$ is equal to $\sum \lambda_m m$, where the sum is extended over all elements of the form*

$$m = (1, \ldots, 1, m_{i_u}, 1, \ldots, 1, m_{i_{u+1}}, 1, \ldots, 1, m_{i_v}, 1, \ldots, 1)$$

($m_h \in M_h$ for all h), and λ_m is the number (calculated in $\mathbf{Z}_{p,k}$) of factorizations of the form

$$w = w_u a_u w_{u+1} \ldots a_{v-1} w_v$$

with $w_u\eta_{i_u} = m_{i_u}, \ldots, w_v\eta_{i_v} = m_{i_v}$.

Proof. Let $w = b_1 \ldots b_m$, $m \geq 1$, $b_1, \ldots, b_m \in A$. Then $w\mu = (b_1\mu) \ldots (b_m\mu)$. Since the matrices $a\mu$ $(a \in A)$ are upper-triangular, the statements relative to $w\mu_{i,j}$ with $j < i$ or $j = i$ are immediate.

If $1 \leq i < j \leq n$, we have

$$w\mu_{i,j} = \sum_{i = h_0 \leq \cdots \leq h_m = j} (b_1\mu_{h_0,h_1}) \ldots (b_m\mu_{h_{m-1},h_m}).$$

By definition of μ, all the terms of this sum are zero if either i or j is not in $\{i_1, \ldots, i_l\}$. Let us now assume that $i = i_u$ and $j = i_v$ for some $1 \leq u < v \leq l$. Then

$$w\mu_{i,j} = \sum (1, \ldots, 1, (b_1 \ldots b_{q_u - 1})\eta_{i_u}, 1, \ldots, 1, (b_{q_u + 1} \ldots b_{q_{u+1} - 1})\eta_{i_{u+1}},$$
$$1, \ldots, 1, (b_{q_v + 1} \ldots b_m)\eta_{i_v}, 1, \ldots, 1)$$

where the sum is extended over all sequences $1 \leq q_u < q_{u+1} < \cdots < q_v \leq m$ such that $b_{q_i} = a_i$ for all $u \leq i \leq v$. Thus the lemma is proved. □

We can now prove Theorem 2.1. Let $1 \leq h \leq n$ and

$$Q = \{m \in \mathbf{Z}_{p,k}\Diamond_n(M_1, \ldots, M_n) \mid m_{h,h} \in P_h\}.$$

Then $Q\mu^{-1} = L_h$. Indeed, for $w \in A^*$, $w\mu$ is in Q iff $w\mu_{h,h} = w\eta_h$ is in P_h, i.e. iff $w \in L_h$.

Let now $1 \leqslant i < j \leqslant n$ and $r \geqslant 0$. Let us denote by P the set

$$P = 1 \times \cdots \times 1 \times M_{i_1} \times 1 \times \cdots \times 1 \times M_{i_2} \times 1 \times \cdots \times 1 \times M_{i_l} \times 1 \times \cdots \times 1.$$

Finally, let Q be the set of all the elements m of $\mathbf{Z}_{p,k} \diamond_n (M_1, \ldots, M_n)$ such that

$$m_{1,n} = \sum_{s \in M_1 \times \cdots \times M_n} \nu_s s$$

with $\sum_{s \in P} \nu_s \equiv r \pmod{p, k}$. Then $Q\mu^{-1} = (L_{i_1} a_1 L_{i_2} \ldots a_{l-1} L_{i_l})_{r,p,k}$. Indeed, for $w \in A^*$, $w\mu$ is in Q iff the number of factorizations

$$w = w_1 a_1 w_2 \ldots a_{l-1} w_l$$

with $w_h \eta_{i_h} \in P_{i_h}$ for all $1 \leqslant h \leqslant l$ is congruent to r mod k threshold p. We conclude immediately by recalling that $P_{i_h} \eta_{i_h}^{-1} = L_{i_h}$.

3.2. Proof of Theorem 2.2

This proof is inspired by the techniques used to describe the finite free object over A of the variety $Z \diamond_n (\mathbf{V}_1, \ldots, \mathbf{V}_n)$, when $\mathbf{V}_1, \ldots, \mathbf{V}_n$ are locally finite varieties. This description will be the object of a later paper.

Let L be a language in A^* that is recognized by a morphism μ from A^* into $\mathbf{Z}_{p,k} \diamond_n (M_1, \ldots, M_n)$.

For each $1 \leqslant i \leqslant n$, let F_i be the subsemigroup of $M_i^{(M_i^A)}$ generated by the elements $a\eta_i$ $(a \in A)$, where $a\eta_i = (f(a))_{f \in M_i^A}$. We denote by $\eta_i : A^* \to F_i$ the induced morphism.

Lemma 3.2. *Let $\mu : A^* \to M_i$ be a morphism. There exists a morphism $\alpha : F_i \to M_i$ such that $\mu = \eta_i \alpha$.*

Proof. Let $f \in M_i^A$ be the restriction of μ to A. Then $\mu = \eta_i \pi_f$, where π_f is the projection of $M_i^{(M_i^A)}$ onto its f-component. □

Note that F_i is in fact the free object of the $\mathbf{M}$-variety generated by M_i.

We shall use Lemma 3.2 as follows. All elements of $\mathbf{Z}_{p,k} \diamond_n (M_1, \ldots, M_n)$ are upper-triangular matrices, and hence the mapping $\mu_{i,i}$ from A^* into M_i ($w\mu_{i,i}$ is the ith diagonal entry of $w\mu$) is a morphism. So $\mu_{i,i} = \eta_i \alpha_i$ for some morphism α_i from F_i into M_i.

For $1 \leqslant l \leqslant n-1$, let $R_l = A^l = \{(a_1, \ldots, a_l) \mid a_1, \ldots, a_l \in A\}$ and $S_l = \{(i_1, \ldots, i_{l+1}) \mid 1 = i_1 \leqslant i_2 \leqslant i_{l+1} = n\}$.

For $a \in A$, $1 \leqslant l \leqslant n-1$, $r = (a_1, \ldots, a_l) \in R_l$ and $s = (i_1, \ldots, i_{l+1}) \in S_l$, let us define $a\theta^{r,s}$ in $\mathbf{Z}_{p,k} \diamond_n (F_1, \ldots, F_n)$ as follows. If $1 \leqslant i \leqslant n$, then $a\theta_{i,i}^{r,s} = a\eta_i$; if $1 \leqslant i < j \leqslant n$, then $a\theta_{i,j}^{r,s} = (1, \ldots, 1)$ if $i = i_u$, $j = i_{u+1}$, $a = a_u$ for some $1 \leqslant u \leqslant l$; $a\theta_{i,j}^{r,s} = 0$ otherwise. Then, we have the following.

Lemma 3.3. *Let $w \in A^+$. If $1 \leq i \leq n$, then $w\theta^{r,s}_{i,i} = w\eta_i$. If $1 \leq i < j \leq n$, then $w\theta^{r,s}_{i,j} = 0$ if i or j is not in s. Otherwise, $i = i_u$, $j = i_v$ for some $1 \leq u < v \leq l+1$ and*

$$w\theta^{r,s}_{i,j} = \sum_{(w_u,\ldots,w_v) \in F_{i_u} \times \cdots \times F_{i_v}} \lambda_{w_u,\ldots,w_v} \times (1, \ldots, 1, w_u, 1, \ldots, 1, w_{u+1}, 1, \ldots, 1, w_v, 1, \ldots, 1)$$

where $\lambda_{w_u,\ldots,w_v}$ is the number (calculated in $\mathbf{Z}_{p,k}$) of factorizations $w = x_u a_u x_{u+1} \ldots a_{v-1} x_v$ with $x_q \eta_{i_u} = w_q$ for all $u \leq q \leq v$.

Proof. The result concerning $w\theta^{r,s}_{i,i}$ is immediate. Let then $1 \leq i < j \leq n$ and let $w = b_1 \ldots b_m$ ($b_1, \ldots, b_m$ in A).

$$w\theta^{r,s}_{i,j} = \sum_{i = h_0 \leq h_1 \leq \cdots \leq h_m = j} (b_1 \theta^{r,s}_{h_0,h_1}) \ldots (b_m \theta^{r,s}_{h_{m-1},h_m}).$$

By definition of $\theta^{r,s}$, all the terms in this sum are zero if i or j is not in s. Let us then assume that $i = i_u$, $j = i_v$ for some $1 \leq u < v \leq l+1$. Then, by definition of $\theta^{r,s}$, we have

$$w\theta^{r,s}_{i,j} = \sum (1, \ldots, 1, (b_1 \ldots b_{q_1-1})\eta_{i_u}, 1, \ldots, 1, (b_{q_1+1} \ldots b_{q_2-1})\eta_{i_{u+1}}, 1, \ldots, 1, (b_{q_{v-u}+1} \ldots b_m)\eta_{i_v}, 1, \ldots, 1)$$

where the sum is extended over all sequences $1 \leq q_1 < \cdots < q_{v-u} \leq m$ such that $b_{q_1} = a_u$, $b_{q_2} = a_{u+1}, \ldots, b_{q_{v-u}} = a_v$. This proves the lemma. □

We shall now make precise the relationship between μ and the $\theta^{r,s}$ ($(r, s) \in R_l \times S_l$, $1 \leq l \leq n-1$).

Lemma 3.4. *Let w, $w' \in A^+$. Then, $w\mu = w'\mu$ if $w\theta^{r,s} = w'\theta^{r,s}$ for all $(r, s) \in R_l \times S_l$, $1 \leq l \leq n-1$.*

Proof. If $1 \leq i \leq n$, then $w\mu_{i,i} = w\theta^{r,s}_{i,i}\alpha_i$. So, $w\theta^{r,s} = w'\theta^{r,s}$ implies $w\mu_{i,i} = w'\mu_{i,i}$.

Let now $1 \leq i < j \leq n$. If $w = b_1 \ldots b_m$ ($m \geq 1$, $b_1, \ldots, b_m$ in A), we have

$$w\mu_{i,j} = \sum ((b_1 \ldots b_{q_1-1})\mu_{i_u,i_u})(b_{q_1}\mu_{i_u,i_{u+1}})((b_{q_1+1} \ldots b_{q_2-1})\mu_{i_{u+1},i_{u+2}}) \ldots (b_{q_{v-u}}\mu_{i_{v-1},i_v})((b_{q_{v-u}+1} \ldots b_m)\mu_{i_v,i_v})$$

where the sum is extended over all sequences $i = i_u < i_{u+1} < \cdots < i_v = j$ and $1 \leq q_1 < q_2 < \cdots < q_{v-u} \leq m$. This can be rewritten as

$$w\mu_{i,j} = \sum (w_u\alpha_{i_u})(a_u\mu_{i_u,i_{u+1}})(w_{u+1}\alpha_{i_{u+1}}) \ldots (a_{v-1}\mu_{i_{v-1},i_v})(w_v\alpha_{i_v})$$

where the sum is over all sequences $i = i_u < \cdots < i_v = j$, $1 \leq u < v \leq n$, and all factorizations $w = x_u a_u x_{u+1} \ldots a_{v-1} x_v$ such that $a_u, \ldots, a_{v-1} \in A$, $x_u, \ldots, x_v \in A^*$, and $x_u\eta_{i_u} = w_u, \ldots, x_v\eta_{i_v} = w_v$.

But the condition $w\theta^{r,s}_{i,j} = w'\theta^{r,s}_{i,j}$ for all (r, s) implies that the number of factorizations $w = x_u a_u x_{u+1} \ldots a_{v-1} x_v$ with $x_u\eta_{i_u} = w_u, \ldots, x_v\eta_{i_v} = w_v$ on one hand, and the number of factorizations $w' = x'_u a_u x'_{u+1} \ldots a_{v-1} x'_v$ with $x'_u\eta_{i_u} = w_u, \ldots, x'_v\eta_{i_v} = w_v$ on the other hand are congruent to one another mod k threshold p. So $w\mu_{i,j} = w'\mu_{i,j}$. □

Let us now return to the language L recognized by μ. L is a finite union of $m\mu^{-1}$ where $m \in \mathbf{Z}_{p,k}\Diamond_n(M_1, \ldots, M_n)$. Each $m\mu^{-1}$ in its turn is a finite intersection of elements of the form $m(\theta_{i,j}^{r,s})^{-1}$, with m in $\mathbf{Z}_{p,k}\Diamond_n(F_1, \ldots, F_n)$.

If $1 \leqslant i \leqslant n$, then $m(\theta_{i,i}^{r,s})^{-1} = m_{i,i}\eta_{i,i}^{-1}$ and hence is a language recognized by F_i. Since F_i divides a finite direct product of copies of M_i, $m(\theta_{i,i}^{r,s})^{-1}$ is in the boolean algebra generated by the languages recognized by M_i.

Let now $1 \leqslant i < j \leqslant n$, $1 \leqslant l \leqslant n-1$, $r = (a_1, \ldots, a_l) \in R_l$ and $s = (i_1, \ldots, i_{l+1}) \in S_l$. Let us assume that

$$m_{i,j} = \sum \alpha_{w_u, \ldots, w_v}(1, \ldots, 1, w_u, 1, \ldots, 1, w_{u+1}, 1, \ldots, 1, w_v, 1, \ldots, 1)$$

where the sum is extended over all $i = i_u < i_{u+1} \cdots < i_v = j$ and all elements $(w_u, \ldots, w_v)$ of $F_{i_u} \times F_{i_{u+1}} \times \cdots \times F_{i_v}$. After Lemma 3.3, $w \in m(\theta_{i,j}^{r,s})^{-1}$ iff the number of factorizations of the form

$$w = x_u a_u x_{u+1} \ldots a_{v-1} x_v$$

with $x_u\eta_{i_u} = w_u, \ldots, x_v\eta_{i_v} = w_v$ is congruent to $\alpha_{w_u, \ldots, w_v}$ mod k threshold p, for all $w_u \in F_{i_u}$, $w_{u+1} \in F_{i_{u+1}}, \ldots, w_v \in F_{i_v}$. So $m(\theta_{i,j}^{r,s})^{-1}$ is a finite intersection of languages of the form

$$(L_u a_u L_{u+1} \ldots a_{v-1} L_v)_{r,p,k}$$

with $L_u, \ldots, L_v$ respectively recognized by $M_{i_u}, \ldots, M_{i_v}$.

4. Conclusion

The operation on monoids introduced in this paper, although somewhat complex to manipulate, corresponds to a simple and natural language operation. In a later paper, we will review the applications of this operation to the classification of rational languages.

However, we can mention briefly here that this uniform way of dealing with products of languages allows us to reformulate certain classical results on group languages (languages whose syntactic monoids are groups), and gives a new proof of some of their corollaries. Let $m \geqslant 1$ and let p be a prime. We let $\mathbf{G}_p$ (resp. $\mathbf{G}_{\text{nil},m}$, $\mathbf{G}_{\text{nil}}$, $\mathbf{G}_{\text{sol}}$) be the M-variety of p-groups (resp. nilpotent groups of class m, nilpotent groups, solvable groups) and we let $\mathbf{I}$ be the trivial M-variety, consisting only of $\{1\}$.

Straubing's characterization of the languages of solvable groups [17] can be restated as: $\mathbf{G}_{\text{sol}}$ is the least M-variety $\mathbf{V}$ such that $\mathbf{Z}_n\Diamond_2(\mathbf{V}, \mathbf{I}) = \mathbf{V}$ for all n. The semidirect product $\mathbf{G}_{\text{sol}} * \mathbf{V}$ is the least M-variety $\mathbf{W}$ containing $\mathbf{V}$ such that $\mathbf{Z}_n\Diamond_2(\mathbf{W}, \mathbf{I}) = \mathbf{W}$.

Also, Eilenberg's result on p-groups and nilpotent groups [4] can be restated as: $\mathbf{G}_p$ is the least M-variety containing $\mathbf{Z}_p\Diamond_k(\mathbf{I}, \ldots, \mathbf{I})$ for all $k \geqslant 1$. $\mathbf{G}_{\text{nil}}$ is the least M-variety containing $\mathbf{Z}_q\Diamond_k(\mathbf{I}, \ldots, \mathbf{I})$ for all $k \geqslant 1$, q prime.

Thérien's refinement of this result [22] is equivalent to: $\mathbf{G}_{\mathrm{nil},m}$ is the least M-variety containing $\mathbf{Z}_n \diamond_{m+1}(\mathbf{I}, \ldots, \mathbf{I})$ for all $n \geqslant 2$.

This presentation of Thérien's result make its corollary (in [22]) both natural and immediate. Let $G_{n,k}$ be the set of (k, k)-upper triangular matrices with coefficients in $\mathbf{Z}_n$ and diagonal coefficients equal to 1. It is immediate that $G_{n,k} = \mathbf{Z}_n \diamond_k (1, \ldots, 1)$. So we have: $\mathbf{G}_p$ (resp. $\mathbf{G}_{\mathrm{nil}}$, $\mathbf{G}_{\mathrm{nil},m}$) is generated by the $G_{p,k}$, $k \geqslant 1$ (resp. by the $G_{q,k}$, $k \geqslant 1$ and q prime, by the $G_{n,m}$, $n \geqslant 2$).

References

[1] P. Blanchard, Morphismes et comptages sur les langages rationnels, *J. Inform. Process. Cybernet.* **23** (1987) 3-11.

[2] J. Brzozowski and F. Fich, Languages of $\mathscr{R}$-trivial monoids, *J. Comput. System Sci.* **20** (1980) 32-49.

[3] J. Brzozowski and I. Simon, Characterization of locally testable events, *Discrete Math.* **4** (1973) 243-271.

[4] S. Eilenberg, *Automata, Languages and Machines, Vol. B* (Academic Press, New York, 1976).

[5] S. Kleene, Representation of events in nerve nets and finite automata, in: Shannon and McCarthy, eds., *Automata Studies* (Princeton University Press, Princeton 1954) 3-51.

[6] G. Lallement, *Semigroups and Combinatorial Applications* (Wiley, New York, 1979).

[7] R. McNaughton, Algebraic decision procedures for local testability, *Math. Systems Theory* **8** (1974) 60-76.

[8] J.-E. Pin, *Variétés de Langages Formels* (Masson, Paris, 1984); *Varieties of Formal Languages* (North Oxford Academic, London, 1986) and (Plenum Press, New York, 1986).

[9] J.-E. Pin, Finite group topology and p-adic topology for free monoids, in: *12th ICALP*, Lecture Notes in Computer Science **194** (Springer, Berlin, 1985) 445-455.

[10] J.-E. Pin, Topologies for the free monoid, *J. Algebra*, to appear.

[11] J.-E. Pin and J. Sakarovitch, Une application de la représentation matricielle des transductions, *Theoret. Comput. Sci.* **35** (1985) 271-293.

[12] J.-E. Pin, H. Straubing and D. Thérien, Locally trivial categories and unambiguous concatenation, *J. Pure Appl. Algebra* **52** (1988) 297-311.

[13] M.-P. Schützenberger, On finite monoids having only trivial subgroups, *Inform. and Control* **8** (1965) 190-194.

[14] I. Simon, Piecewise testable events, in: *Proc. 2nd G.I. Conf.*, Lecture Notes in Computer Science **33** (Springer, Berlin, 1975) 214-222.

[15] H. Straubing, Varieties of recognizable sets whose syntactic monoids contain solvable groups, Ph.D. thesis, University of California-Berkeley, 1978.

[16] H. Straubing, Recognizable sets and power sets of finite semigroups, *Semigroup Forum* **18** (1979) 331-340.

[17] H. Straubing, Families of recognizable sets corresponding to certain varieties of finite monoids, *J. Pure Appl. Algebra* **15** (1979) 305-318.

[18] H. Straubing, A generalization of the Schützenberger product of finite monoids, *Theoret. Comput. Sci.* **13** (1981) 137-150.

[19] H. Straubing, D. Thérien and W. Thomas, Regular languages defined with generalized quantifiers, in: T. Lepistö and A. Salomaa, eds., *15th ICALP*, Lecture Notes in Computer Science **317** (Springer, Berlin, 1988).

[20] D. Thérien, Languages of nilpotent and solvable groups, in: *Proc. 6th ICALP*, Lecture Notes in Computer Science **71** (Springer, Berlin, 1979) 616-632.

[21] D. Thérien, Classification of finite monoids: the language approach, *Theoret. Comput. Sci.* **14** (1981) 195-208.

[22] D. Thérien, Subword counting and nilpotent groups, in: L. Cummings, ed., *Combinatorics on Words, Progress and Perspectives* (Academic Press, 1983) 297-305.