Some operations and transductions that preserve rationality.

Jean-Éric Pin* and Jacques Sakarovitch†

November 1982

1 Introduction

When a language theorist encounters a new operation on languages, his first impulse is to know whether this operation preserves rational languages. If the answer appears to be positive, he proceeds immediately to the construction of a more or less complicated automaton to solve the problem. However there are many operations on languages, many language theorists (see the references) and many different constructions to study these many operations. The aim of this paper is to show that almost all these constructions are a particular case of a general and simple approach. It is fair to say immediately that a few operations are overlooked, such as the star operation, complementation and reversal. However, the scope of our method is quite broad, broader indeed than one would expect, and goes from "classical" operations such as union, intersection, concatenation, quotient, shuffle, inverse and direct morphisms, etc., to less classical ones such as infiltration, Dyck reduction, longest common prefix, Straubing's counting, etc. It includes also questions that are not expressed directly as operations on languages, as, for example, Reutenauer's theorem on TOL-systems. The interest of the method in not only to give a unified framework for all these results. Statements of the form "such an operation preserves rational languages" can be readily be refined into "such an an operation preserves star-free languages" or even more generally "such an operation preserves such a variety of rational languages".

The key idea of our construction is to consider an operation $\varphi : A_1^* \times \cdots \times A_n^* \to A^*$ as the inverse of a transduction $\tau : A^* \to A_1^* \times \cdots \times A_n^*$ (whenever it is possible). Then, given monoids $M_1, \ldots, M_n$, recognizing the languages $L_1, \ldots, L_n$, of $A_1^*, \ldots, A_n^*$, respectively, we are able to construct a monoid M recognizing $(L_1, \ldots, L_n)\varphi$ as soon as the transduction τ admits

*Université Paris VI et CNRS, LITP, 4 Place Jussieu, 75230 Paris Cedex 05, France.

a matrix representation. At this point the term "matrix representation" is intentionally imprecise. The following definition gives a first approach to the notion. Let μ be a morphism from A^* into the monoid of $k \times k$ matrices (for some k), whose entries are subsets of $A_1^* \times \cdots \times A_n^*$. Roughly speaking, the transduction τ is then asked to be such that, for every u in A^*, $u\tau$ is a fixed linear expression of the entries of $u\mu$. Thus, for instance, the Kleene-Schützenberger theorem says that every rational transduction admits a matrix representation with *rational* subsets as entries. However, one can replace "linear expression" by "polynomial expression" and even by "series" in the previous definition. In any case, the construction of a monoid M that recognizes $(L_1, \ldots, L_n)\varphi$ *only depends on the morphism* μ and on $M_1, \ldots, M_n$.

Therefore, if φ is an operation, we proceed as follows: we first check whether φ can be expressed as the inverse of a transduction τ. This works in most cases (except for star and complementation). Now the construction of M reduces to finding a matrix representation for τ. This is again possible in most cases (except for reversal) and in general τ even admits a "linear" matrix representation. However, the following example, which is an extension of a classical exercise in language theory, shows that non-linear matrix representations might be required: given a language L of A^*, divide the words of L into $2n+1$ equal segments (if possible) for any *prime* number $2n+1$: $\zeta(L)$ is the set of all medial segments one can obtain this way. Then if L is rational, $\zeta(L)$ is rational.

2 Matrix representations of transductions

We refer the reader to [3] for undefined terms of this article.

Let M be a monoid (with unit 1). We denote by $\mathcal{P}(M)$ the *power set* of M: $\mathcal{P}(M)$ is a semiring with union as addition and the usual product of subsets as multiplication. The set of *rational subsets* of M, denoted $\mathrm{Rat}(M)$, is the smallest subsemiring of $\mathcal{P}(M)$ containing the finite sets and closed under the star operation. As usual, we denote by $\mathcal{P}(M)^{n\times n}$ the set of matrices of size n with entries in $\mathcal{P}(M)$.

A subset P of M is *recognized by a morphism* $\eta : M \to N$ if $P = P\eta\eta^{-1}$, that is, if there exists a subset Q of N such that $P = Q\eta^{-1}$. In this case we also say that N *recognizes* P. Note that if N is a submonoid of N', then N' also recognizes P. A set P is *recognizable* if it is recognized by a *finite* monoid. Kleene's theorem states that a language is recognizable if and only if it is rational.

Let M and N be two monoids. A *transduction* $\tau : M \to N$ is a mapping from M into the $\mathcal{P}(N)$. One extends τ to a mapping $\mathcal{P}(M) \to \mathcal{P}(N)$ by setting $P\tau = \bigcup_{m\in P} m\tau$. The inverse transduction $\tau^{-1} : N \to M$ is defined by $Q\tau^{-1} = \{m \in M \mid \tau \cap Q \neq \emptyset\}$. The transduction is rational if the set

$\{(m,n) \in M \times N \mid n \in m\tau\}$ is a rational subset of $M \times N$.

Let A be an alphabet and let M be a monoid.

Definition 2.1 A transduction $\tau : A^* \to M$ admits a *linear matrix representation* (λ, μ, ν) if there exist $n > 0$, a morphism $\mu : A^* \to \mathcal{P}(M)^{n\times n}$, a row vector $\lambda \in \mathcal{P}(M)^{1\times n}$, a column vector $\nu \in \mathcal{P}(M)^{n\times 1}$ such that for all $f \in A^*$, $f\tau = \lambda\cdot f\mu\cdot v$.

The theorem of Kleene-Schützenberger (cf. [3]) states that a transduction $\tau : A^* \to M$ is rational if and only if it admits a linear matrix representation with entries in $\mathrm{Rat}(M)$.

Every monoid morphism $M \to N$ can be extended to a morphism $\mathcal{P}(M) \to \mathcal{P}(N)$ and, for each $n > 0$, to a morphism $\mathcal{P}(M)^{n\times n} \to \mathcal{P}(N)^{n\times n}$.

The following elementary result is efficient for most of the applications we have in view.

Theorem 2.1 *Let $\tau : A^* \to M$ be a transduction that admits a linear matrix representation (λ, μ, ν) and let P be a subset of M recognized by a morphism $\eta : M \to N$. Then the language $P\tau^{-1}$ is recognized by the monoid of matrices $A^*\mu\eta$.*

Proof. Let $Q = P\eta$ and let R be the subset of $A^*\mu\eta$ defined by

$$R = \{m \in \mathcal{P}(N)^{n\times n} \mid \lambda\eta\cdot m\cdot \nu\eta \cap Q \neq \emptyset\}$$

Then by a routine calculation:

$$\begin{aligned}
R(\mu\eta)^{-1} &= \{f \in A^* \mid f\mu\eta \in R\} = \{f \in A^* \mid \lambda\eta\cdot f\mu\eta\cdot \nu\eta \cap Q \neq \emptyset\} \\
&= \{f \in A^* \mid f\tau\eta \cap Q \neq \emptyset\} = \{f \in A^* \mid f\tau \cap Q\eta^{-1} \neq \emptyset\} \\
&= \{f \in A^* \mid f\tau \cap P \neq \emptyset\} = P\tau^{-1}
\end{aligned}$$

Corollary 2.2 *Let $\tau : A^* \to M$ be a transduction that admits a linear matrix representation. If P is a recognizable subset of M, $P\tau^{-1}$ is a recognizable — hence rational — language of A^*.*

The extension of Theorem 2.1 requires some preliminaries.

Let M be a monoid and let Ξ be an alphabet. We denote by $M * \Xi^*$ the free product (or coproduct) of the monoids M and Ξ^*. The monoid $M * \Xi^*$ can be identified with the set of words of the form $m_0\xi_1 m_1 \cdots \xi_n m_n$ (where $m_0, \ldots, m_n \in M$ and $\xi_1, \ldots, \xi_n \in \Xi$) equipped with the product

$$(m_0\xi_1 m_1 \cdots \xi_n m_n)(m'_0\xi'_1 m'_1 \cdots \xi'_n m'_{n'}) = \\ m_0\xi_1 m_1 \cdots \xi_n (m_n m'_0)\xi'_1 m'_1 \cdots \xi'_n m'_{n'}$$

Let $\theta : \Xi^* \to \mathcal{P}(M)$ be a morphism, that is, a substitution from Ξ^* to M. We denote by $[\theta] : \mathcal{P}(M * \Xi^*) \to \mathcal{P}(M)$ the semiring morphism defined by

(a) $u[\theta] = m_0(\xi_1\theta)m_1 \cdots (\xi_n\theta)m_n$ for $u = m_0\xi_1 m_1 \cdots \xi_n m_n \in M * \Xi^*$,

(b) $X[\theta] = \bigcup_{u \in X} u[\theta]$ for $X \in \mathcal{P}(M * \Xi^*)$.

Of course we identify in this definition the element m and the set $\{m\}$.

Let $\eta : M \to N$. We also denote by η the induced morphisms $M * \Xi^* \to N * \Xi^*$ and $\mathcal{P}(M) \to \mathcal{P}(N)$. A formal verification suffices to prove the following lemma:

Lemma 2.3 *Let $\eta : M \to N$ and $\theta : \Xi^* \to \mathcal{P}(M)$ be two monoid morphisms. The following diagram is commutative:*

$$\begin{tikzcd}
\mathcal{P}(M * \Xi^*) \arrow[r, "{[\theta]}"] \arrow[d, "\eta"'] & \mathcal{P}(M) \arrow[d, "\eta"] \\
\mathcal{P}(N * \Xi^*) \arrow[r, "{[\theta\eta]}"] & \mathcal{P}(N)
\end{tikzcd}$$

For any positive integer n, let $\Xi_n = \{\xi_{1,1}, \xi_{1,2}, \ldots, \xi_{n,n}\}$ be an alphabet with n^2 letters. Any matrix $m \in \mathcal{P}(M)^{n \times n}$ defines a morphism $m : \Xi_n^* \to \mathcal{P}(M)^{n \times n}$ by $\xi_{i,j} m = m_{i,j}$. We can now generalize Definition 2.1.

Definition 2.2 A transduction $\tau : A^* \to M$ admits a *matrix representation* (s, μ) if there exist $n > 0$, a morphism $\mu : A^* \to \mathcal{P}(M)^{n \times n}$, and an element $s \in \mathcal{P}(M * \Xi_n^*)$ such that for all $f \in A^*$, $f\tau = s[f\mu]$.

Then Theorem 2.1 can be generalized as follows.

Theorem 2.4 *Let $\tau : A^* \to M$ be a transduction that admits a matrix representation (s, μ) and let P be a subset of M recognized by a morphism $\eta : M \to N$. Then the language $P\tau^{-1}$ is recognized by the monoid of matrices $A^*\mu\eta$.*

Proof. The proof mimics the proof of Theorem 2.1. Let $Q = P\eta$ and let R be the subset of $A^*\mu\eta$ defined by

$$R = \{m \in \mathcal{P}(N)^{n \times n} \mid s\eta[m] \cap Q \neq \emptyset\}$$

Thus $R(\mu\eta)^{-1} = \{f \in A^* \mid s\eta[f\mu\eta] \cap Q \neq \emptyset\}$, hence by Lemma 2.3:

$$\begin{aligned}
R(\mu\eta)^{-1} &= \{f \in A^* \mid s\eta[f\mu\eta] \cap Q \neq \emptyset\} \\
&= \{f \in A^* \mid s[f\mu] \cap Q\eta^{-1} \neq \emptyset\} \\
&= \{f \in A^* \mid f\tau \cap P \neq \emptyset\} = P\tau^{-1}
\end{aligned}$$

3 Operations on languages

Let $L_1, \ldots, L_n$ be languages of $A_1^*, \ldots, A_n^*$, respectively and let $\eta_i : A_i^* \to M_i$ be a morphism recognizing L_i, for $1 \leq i \leq n$. Then clearly

$$\eta = \prod_{1 \leq i \leq n} \eta_i : \prod_{1 \leq i \leq n} A_i^* \to \prod_{1 \leq i \leq n} M_i$$

recognizes the subset $L_1 \times \cdots \times L_n$ of $A_1^* \times \cdots \times A_n^*$. Then Theorem 2.1 (or Theorem 2.4) gives the construction of a monoid recognizing $(L_1, \ldots, L_n)\varphi$ provided that φ satisfies the following conditions:

(1) φ is the inverse of a transduction $\tau : A^* \to A_1^* \times \cdots \times A_n^*$.

(2) τ admits a linear matrix representation (resp. a matrix representation)

As we claimed in the introduction, these two conditions hold for a lot of classical operations.

3.1 Inverse morphisms and inverse substitutions

In this example, conditions (1) and (2) are trivially satisfied. For example, let $\sigma : A^* \to B^*$ be a substitution. Then by definition σ induces a morphism $A^* \to \mathcal{P}(B^*)$ and $u\sigma = 1 \cdot u\sigma 1$ for all $u \in A^*$. Thus $(1, \sigma, 1)$ is a linear matrix representation of σ. With the notations of Theorem 2.1, $A^*\sigma\eta$ is a submonoid of $\mathcal{P}(M)$. Therefore we have obtained the following result proved in [14] for rational languages.

Proposition 3.1 *Let $\sigma : A^* \to B^*$ be a substitution. If $L \subset B^*$ is recognized by M, then $L\sigma^{-1}$ is recognized by $\mathcal{P}(M)$. In particular, if L is rational, $L\sigma^{-1}$ is rational.*

In the same way, we get easily

Proposition 3.2 *Let $\varphi : A^* \to B^*$ be a morphism. Each monoid that recognizes $L \subset B^*$ also recognizes $L\varphi^{-1}$.*

3.2 Intersection and union

We note that $L_1 \cap L_2 = (L_1 \times L_2)\tau^{-1}$ where $\tau : A^* \to A^* \times A^*$ is defined by $u\tau = (u \times u) = (1 \times 1)(u \times u)(1 \times 1)$. Thus $(1 \times 1, \tau, 1 \times 1)$ is a linear matrix representation of τ. With the notations of Theorem 2.1, $A^*\mu\eta$ is a submonoid of $M_1 \times M_2$. Thus

Proposition 3.3 (see [6]) *Let L_1, L_2 be languages of A^* recognized by M_1, M_2 respectively. Then $L_1 \cap L_2$ is recognized by $M_1 \times M_2$.*

The case of the union is a little more involved. Indeed $L_1 \cup L_2 = (L_1 \times L_2)\tau^{-1}$ where $u\tau = (u \times A^*) \cup (A^* \times u)$ for all $u \in A^*$. Thus $((1,1), \mu, (1,1)^t)$ where $u\mu = \begin{pmatrix} u\times A^* & 0 \\ 0 & A^*\times u \end{pmatrix}$ is a linear matrix representation of τ. Now an instant of reflection shows that $A^*\mu\eta$ is isomorphic to a submonoid of $M_1 \times M_2$. Therefore

Proposition 3.4 (see [6]) *Let L_1, L_2 be languages of A^* recognized by M_1, M_2 respectively. Then $L_1 \cup L_2$ is recognized by $M_1 \times M_2$.*

3.3 Left and right quotients (or derivatives)

Let P and L be languages of A^*. Then the left quotient of L by P is the set $P^{-1}L = \{f \in A^* \mid Pf \cap L \neq \emptyset\}$. The set LP^{-1} is defined dually. Now if P is fixed, $P^{-1}L = L\tau^{-1}$ where $u\tau = Pu$ for all $u \in A^*$. Clearly $(P, \iota, 1)$ — where ι denotes the identity of A^* — is a linear matrix representation of τ. Thus

Proposition 3.5 (see [3]) *Let L be a language of A^* recognized by a monoid M. Then M recognizes $P^{-1}L$ and LP^{-1} for all languages P of A^*.*

Corollary 3.6 *If L is rational, $P^{-1}L$ and LP^{-1} are rational for all languages P.*

3.4 Concatenation product

It is easy to see that $L_1 \cdots L_n = (L_1 \times \cdots \times L_n)\tau^{-1}$ where $\tau : A^* \times \cdots \times A^*$ is the transduction defined by

$$f\tau = \{(f_1, \cdots, f_n) \in A^* \times \cdots \times A^* \mid f_1 \cdots f_n = f\}$$

Moreover τ admits the linear matrix representation (λ, μ, ν) where $\lambda = ((1, \cdots, 1), 0, \cdots, 0)$, $\nu = (0, \cdots, 0, (1, \cdots, 1))^t$ and $\mu : A^* \to \mathcal{P}(A^* \times \cdots \times A^*)^{n\times n}$ is defined, for all $u \in A^*$, by

$$\begin{aligned}
&u_{i,j} = 0 \quad \text{if } i > j \\
&u_{i,j} = \{(u_1, \cdots, u_n) \in A^* \times \cdots \times A^* \mid u_1 = \ldots = u_{i-1} \\
&\qquad = u_{j+1} = \ldots = u_n = 1 \text{ and } u_i u_{i+1} \cdots u_j = u\} \text{ if } i \leq j
\end{aligned}$$

In particular, τ is a rational transduction. By Theorem 2.1, $L_1 \cdots L_n$ is recognized by the monoid $A^*\mu\eta$. Now $A^*\mu\eta$ is a submonoid of the monoid of all square matrices of size n with entries in the finite subsets of $M_1 \times \cdots \times M_n$ such that:

(a) P is upper triangular, that is $P_{i,j} = 0$ for $i > j$.

(b) The i-th entry of the diagonal is an element of M_i. More precisely, $P_{i,i} = \{(1, \ldots, 1, m_i, 1, \ldots, 1)\}$ for some $m_i \in M_i$.

(c) If $(m_1, \ldots, m_n) \in P_{i,j}$, then $m_k = 1$ for $k < i$ and $k > j$.

This latter monoid is called by Straubing [20] the Schützenberger product of $M_1, \ldots, M_n$ and is denoted $\Diamond_n(M_1, \ldots, M_n)$. For $n = 2$ it can be identified with the original definition of Schützenberger (see [6]). Note that in general the three monoids $\Diamond_2(M_1, \Diamond_2(M_2, M_3))$, $\Diamond_3(M_1, M_2, M_3)$ and $\Diamond_2(\Diamond_2(M_1, M_2), M_3)$ are all different.

Therefore, we have

Proposition 3.7 *Let $L_1, \ldots, L_n$ be languages recognized by $M_1, \ldots, M_n$ respectively. Then $L_1 \cdots L_n$ is recognized by $\Diamond_n(M_1, \ldots, M_n)$.*

A slight modification of the previous construction leads to the following result (see [12] for applications)

Proposition 3.8 *If $a_1, \ldots, a_{n-1}$ are letters of A, then $L_1 a_1 L_2 \cdots a_{n-1} L_n$ is recognized by $\Diamond_n(M_1, \ldots, M_n)$.*

3.5 Shuffle, infiltration product

Given a word $h = a_1 \cdots a_n \in A^*$ (the a_i's are letters) and a subset $I = \{i_1, \cdots, i_r\}$ of $\{1, \ldots, n\}$ (where $i_1 < \ldots < i_r$), we denote by h_I the word $a_{i_1} \cdots a_{i_r}$. With this notation, the *shuffle* of two words f and g is defined by

$$f \circ g = \{h \in A^* \mid h_I = f \text{ and } h_J = g \text{ for some } I, J \text{ such that } I \cap J = \emptyset \text{ and } I \cup J = \{1, \ldots, |h|\}\}$$

and the *infiltration product* [4] is defined by

$$f \uparrow g = \{h \in A^* \mid h_I = f \text{ and } h_J = g \text{ for some } I, J \text{ such that } I \cup J = \{1, \ldots, |h|\}\}$$

Shuffle and infiltation product are extended as usual to languages by setting

$$L_1 \circ L_2 = \bigcup_{f_1 \in L_1, f_2 \in L_2} f_1 \circ f_2 \qquad L_1 \uparrow L_2 = \bigcup_{f_1 \in L_1, f_2 \in L_2} f_1 \uparrow f_2$$

Now $L_1 \circ L_2 = (L_1 \times L_2)\tau^{-1}$ where $\tau : A^* \to A^* \times A^*$ is defined by $f\tau = \{(f_1, f_2) \in A^* \times A^* \mid f \in f_1 \circ f_2\}$. Since τ induces a morphism $A^* \to \mathcal{P}(A^* \times A^*)$, τ admits the linear matrix representation $(1, \tau, 1)$. Thus one gets easily

Proposition 3.9 [11] *Let L_1 and L_2 be languages of A^* recognized by M_1 and M_2 respectively. Then $L_1 \circ L_2$ is recognized by* $\mathrm{Fin}(M_1 \times M_2)$*, the monoid of finite subsets of $M_1 \times M_2$.*

In the same way $L_1 \uparrow L_2 = (L_1 \times L_2)\tau^{-1}$ where $f\tau = \{(f_1, f_2) \in A^* \times A^* \mid f \in f_1 \uparrow f_2\}$ and by the same argument as above we obtain

Proposition 3.10 *Let L_1 and L_2 be languages of A^* recognized by M_1 and M_2 respectively. Then $L_1 \uparrow L_2$ is recognized by* $\mathrm{Fin}(M_1 \times M_2)$.

3.6 Morphisms

Let $\varphi : A^* \to B^*$ be a morphism. Then $\varphi^{-1} = \tau : B^* \to A^*$ is a rational transduction. It follows that φ satisfies the conditions (1) and (2) and Theorem 2.1 can be applied. However the explicit construction of a linear matric representation requires some machinery. We first define the *petal monoid* of X^* when X is a finite language of B^* [9]. Let $Q = \{(1,1)\} \cup \{(u,v) \in B^+ \times B^+ \mid uv \in X\}$. One associates to each letter $a \in B$ a relation as follows (for the sake of simplicity we use arrows instead of formal definitions):

$$\begin{tikzcd} (1,1) \arrow[loop right, "a"] \end{tikzcd} \qquad \text{if } a \in X$$

$$(1,1) \xrightarrow{a} (a,v) \qquad \text{for all } v \in B^+ \text{ such that } av \in X$$

$$(ua,v) \xrightarrow{a} (u,av) \qquad \text{for all } u,v \in B^+ \text{ such that } uav \in X$$

$$(u,a) \xrightarrow{a} (1,1) \qquad \text{for all } u,v \in B^+ \text{ such that } ua \in X$$

The relations defined by a, for $a \in B$, generate a monoid of relations that recognizes X^*. This is precisely the petal monoid of X^*, denoted by $\mathrm{Pet}\, X^*$. In the sequel we shall represent in the usual way relations on Q by boolean matrices of size $Q \times Q$.

Next we introduce the following

Definition 3.1 Let N be a monoid of relations on a set Q and let S be a semiring. The *substitution product* of S by N (denoted $S \circ N$) is the set of all matrices of size $Q \times Q$ obtained by substituting elements of S for the non-zero entries of matrices of N.

Let us come back to our morphism $\varphi : A^* \to B^*$. Set $X = A\varphi \cap B^*$, $I^* = 1\varphi^{-1}$ and for all $x \in X$, $C(x) = \{a \in A \mid a\varphi = x\}$. Then one can prove — we omit the details — that φ^{-1} admits the linear matrix representation

(λ, μ, ν) where $\mu : B^* \to (\operatorname{Rat} A^*)^{Q \times Q}$ is defined, for all $a \in B$, by

$$(a\mu)_{p,q} = \begin{cases} 1 & \text{if } q \neq (1,1) \text{ and } p \xrightarrow{a} q \\ I^*C(ua) & \text{if p = (u, a), q = (1, 1) and } p \xrightarrow{a} (1,1) \\ I^*C(a) & \text{if p = q = (1, 1) and } (1,1) \xrightarrow{a} (1,1) \\ 0 & \text{in all other cases} \end{cases}$$

$$\lambda_p = \begin{cases} 1 & \text{if } p = (1,1) \\ 0 & \text{if not} \end{cases} \qquad \nu_p = \begin{cases} I^* & \text{if } p = (1,1) \\ 0 & \text{if not} \end{cases}$$

Now if L is recognized by $\eta : A^* \to M$, Theorem 2.1 states that $L\varphi$ is recognized by $B^*\mu\eta$. But a straightforward verification shows that $A^*\mu\eta$ is a submonoid of $\operatorname{Rat} M \circ \operatorname{Pet} X^*$. Thus

Proposition 3.11 *Let $\varphi : A^* \to B^*$ be a morphism such that $X = A\varphi \cap B^*$. If $L \subset A^*$ is recognized by a monoid M, then $L\varphi$ is recognized by the substitution product* $\operatorname{Rat} M \circ \operatorname{Pet} X^*$.

In the case where X is a code the substitution product turns out to be a wreath product (in the sense of [16]). Therefore

Proposition 3.12 *Let $\varphi : A^* \to B^*$ be a morphism and let $X = A\varphi \cap B^*$ is a code. If $L \subset A^*$ is recognized by a monoid M, then $L\varphi$ is recognized by the wreath product* $\operatorname{Rat} M \circ \operatorname{Pet} X^*$.

If φ is injective one can be more precise

Proposition 3.13 *Let $\varphi : A^* \to B^*$ be an injective morphism and let $X = A\varphi$. If a language $L \subset A^*$ is recognized by a monoid M, then $L\varphi$ is recognized by the wreath product* $\operatorname{Rat} M \circ \operatorname{Pet} X^*$.

Finally if φ is litteral (= length preserving, strictly alphabetic) we could prove the following result, obtained in [19] and [14] for rational languages.

Proposition 3.14 *Let $\varphi : A^* \to B^*$ be a litteral morphism. If a language $L \subset A^*$ is recognized by a monoid M, then $L\varphi$ is recognized by* $\operatorname{Fin} M$.

3.7 Miscellaneous

We just mention here a list of operations for which Theorem 2.1 applies:

- The longer common prefix of two words: see [5]
- The nabla operation, a cousin of the shuffle: see [1]
- Straubing's counting: see [20]
- Inverse of rational functions: see [17]

etc.

4 Other applications

4.1 A result on TOL-systems

We call TOL a set $G = (A, \{\sigma_1, \ldots, \sigma_n\})$ where A is an alphabet and $\sigma_1, \ldots, \sigma_n$ are substitutions of A^* into A^*. Let $B = \{1, \ldots, n\}$. With each word $u = i_1 \cdots i_r$ of B^* one associates the substitution $\sigma_u = \sigma_{i_1} \cdots \sigma_{i_r}$ (the substitution associated with the empty word is the identity). The following proposition is a slight extension of a result of [14].

Proposition 4.1 *Let G be a TOL-system and let K, L be two rational languages of A^*. Then $G(K, L) = \{u \in B^* \mid K\sigma_u \cap L \neq \emptyset\}$ is rational.*

Indeed $G(K, L) = K\tau^{-1}$ where $\tau : B^* \to A^*$ is the transduction defined by $u\tau = L\sigma_u^{-1}$ for all $u \in B^*$. The remainder of the proof consists of verifying that τ admits a linear matrix representation. Then one applies Corollary 2.2. $\square$

4.2 Reduction of the free group

Let A be an alphabet, $\bar{A}$ a copy of A and set $\tilde{A} = A \cup \bar{A}$. As is well-known the free group $F(A)$ over A is the quotient of $\tilde{A}^*$ by the congruence generated by the relations $a\bar{a} = 1$ and $\bar{a}a = 1$. We denote by D^* the set of all words congruent to 1. A word is *reduced* if it contains no occurrence of factors of the form $a\bar{a}$ or $\bar{a}a$. One can prove that every word u of A^* is congruent to a unique reduced word $u\delta$. This defines a function $\delta : \tilde{A}^* \to \tilde{A}^*$ called the *Dyck reduction*. The classical result of Benois (see [2]) can be restated as follows:

Proposition 4.2 *Let R be a rational subset of $\tilde{A}^*$. Then $R\delta$ is rational.*

Indeed let $\tau : \tilde{A}^* \to \tilde{A}^*$ be the transduction defined by $(a_1 \cdots a_n)\tau = D^* a_1 D^* \cdots a_n D^*$ (where $a_1, \ldots, a_n$ are letters). Then one can prove — this is the difficult part of the proof — that $\tau^{-1} = \delta$. Now it is not difficult to find a linear matrix representation for τ and the result follows from Corollary 2.2. $\square$

Of course a similar result holds for the congruence generated by $a\bar{a} = 1$ $(a \in A)$ or other variants — see [15] —

4.3 Applications of Theorem 2.4

We first return to the example of the introduction. Let L be a language and let n be a positive integer. Set (cf. [8] for instance)

$$\frac{1}{2n+1}L = \{u \in A^* \mid \text{there exist } x, y \in A^* \ |x| = |y| = n \text{ and } xuy \in L\}$$

and

$$\zeta(L) = \bigcup_{(2n+1) \text{ prime}} \frac{1}{2n+1}L.$$

We shall prove the following result

Proposition 4.3 *If L is recognized by a monoid M, then $\frac{1}{2n+1}L$ (for every $n > 0$) and $\zeta(L)$ are recognized by $M \times C$ where C is a one-generator submonoid of $\mathcal{P}(M)$.*

Corollary 4.4 *If L is rational, $\frac{1}{2n+1}L$ (for every $n > 0$) and $\zeta(L)$ are rational.*

Proof. Define for all $n > 0$, $\tau_n : A^* \to A^*$ by $u\tau_n = A^{n|u|}uA^{n|u|}$ and let $\tau : A^* \to A^*$ be defined by $u\tau = \bigcup_{(2n+1) \text{ prime}} u\tau_n$. Then $L\tau_n^{-1} = \zeta(L)$. Moreover τ_n admits the (non-linear) matrix representation (s_n, μ) where $u\mu = \begin{pmatrix} A^{|u|} & 0 \\ 0 & u \end{pmatrix}$ for all $u \in A^*$ and $s_n = \zeta_{1,1}^n \zeta_{2,2} \zeta_{1,1}^n$. In the same way τ admits the matrix representation (s, μ) where $s = \sum_{(2n+1) \text{ prime}} s_n$. Now if L is recognized by $\eta : A^* \to M$, Theorem 2.4 shows that $\frac{1}{2n+1}L$ and $\zeta(L)$ are recognized by $A^*\mu\eta$. But $A^*\mu\eta$ is isomorphic to a submonoid of $M \times C$ where C is the submonoid of $\mathcal{P}(M)$ generated by $A\eta$. $\square$

Here is another example. Define $\sqrt{L} = \{u \in A^* \mid u^2 \in L\}$. Then $\sqrt{L} = L\tau^{-1}$ where $u\tau = u^2$ for all $u \in A^*$. Clearly τ admits the matrix representation (s, μ) where $u\mu = u$ and $s = \zeta^2$. Therefore

Proposition 4.5 *If a language L is recognized by a monoid M, then M also recognizes $\sqrt{L}$. In particular if L is rational (resp. star-free), so is $\sqrt{L}$.*

Finally the reader who likes more complicated examples may try to prove that if $L \subset \{a, b\}^*$ is rational and if $\tau : A^* \to A^*$ is any rational transduction, then

$$L' = \bigcup_{n \text{ square-free}} \{u \in A^* \mid D^* u^{\lfloor \sqrt{n} \rfloor} a^n (u\tau)^{n!} b \cap L \neq \emptyset\}$$

is rational.

References

[1] J. Beauquier, On the structure of context-free languages, *International Journal of Computer Mathematics* **11** (1982), 3–19.

[2] M. Benois, Parties rationnelles du groupe libre, *C. R. Acad. Sci. Paris Sér. A* **269** (1969), 1188–1190.

[3] J. Berstel, *Transductions and context-free languages*, Teubner, 1979.

[4] K. T. Chen, R. H. Fox and R. C. Lyndon, Free differential calculus, *Ann. of Math.* **68** (1958), 81–95.

[5] C. Choffrut, A Closure Property of Deterministic Context-Free Languages, *Ann. of Math.* **12** (1981), 13–16.

[6] S. Eilenberg, *Automata, Languages and Machines*, vol. B, Academic Press, New York, 1976.

[7] J. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison Wesley, 1979.

[8] E. Le Rest and M. Le Rest, Sur le calcul du monoïde syntaxique d'un sous-monoïde finiment engendré, *Semigroup Forum* **21** (1980), 73–185.

[9] J.-F. Perrot, Variétés de langages et operations, *Theoret. Comput. Sci.* **7** (1978), 197–210.

[10] J.-E. Pin, Hiérarchies de concaténation, *RAIRO Informatique Théorique* **18** (1984), 23–46.

[11] C. Reutenauer, Sur les variétés de langages et de monoïdes, in *Theoretical computer science (Fourth GI Conf., Aachen)*, vol. 67, pp. 260–265, Springer-Verlag, Berlin, 1979.

[12] J. Sakarovitch, *Syntaxe des langages de Chomsky*, PhD thesis, Université Paris 6, 1979.

[13] J. Sakarovitch, Sur la définition du produit en couronne, in *Codages et Transductions*, G. Pirillo (éd.), pp. 285–299, Institut U. Dini, Florence, C.N.R., 1981.

[14] M. P. Schützenberger, Sur les Relations Rationnelles Entre Monoïdes Libres, *Theoret. Comput. Sci.* **3** (1976), 243–259.

[15] H. Straubing, Recognizable Sets and Power Sets of Finite Semigroups, *Semigroup Forum* **18** (1979), 331–340.

[16] H. STRAUBING, Recognizable Sets and Power Sets of Finite Semigroups, *Theoret. Comput. Sci.* **13** (1981), 137–150.