

SOME OPERATIONS AND TRANSDUCTIONS THAT PRESERVE RATIONALITY

by J.-E. PIN and J. SAKAROVITCH

Université Paris VI et CNRS

Laboratoire d'Informatique Théorique

4, PLACE Jussieu 75230 PARIS Cedex 05

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 of 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 is not only to give a unified framework for all these results. Statements of the form "such an operation preserves rational languages" can readily be refined into "such 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

$$\varphi : A_1^* \times \dots \times A_n^* \rightarrow A^* \text{ as the inverse of a transduction}$$

$\tau : A^* \rightarrow A_1^* \times \dots \times A_n^*$ (when it is possible). Then, given monoids M_1, \dots, M_n recognizing the languages L_1, \dots, L_n of A_1^*, \dots, A_n^* , respectively, we are able to construct a monoid M recognizing $(L_1, \dots, L_n)\varphi$ as soon as the transduction τ admits 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 $A_1^* \times \dots \times A_n^*$. 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 $(L_1, \dots, L_n)\varphi$ only depends on the morphism μ and on M_1, \dots, 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 applies in most cases (except for star and complementation). Now the construction of M reduces to finding a matrix representation for τ . This is 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 $2n+1$: $\xi(L)$ is the set of all medial segments one can obtain this way. Then if L is rational, $\xi(L)$ 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 $P(M)$ the *power set* of M : $P(M)$ is a semiring with union as addition and the usual product of subsets as multiplication. The set of *rational subsets* of M , denoted $\text{Rat } M$ is the smallest sub-semiring of $P(M)$ containing the finite sets and closed under the star operation. As usual we denote by $P(M)^{n \times n}$ the set of matrices of size n with entries in $P(M)$.

A subset P of M is *recognized by a morphism* $\eta : M \rightarrow N$ 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 .

P is a *recognizable* if it is recognized by a *finite* monoid.

Kleene's theorem states that a language is recognizable iff it is rational.

Let M and N be two monoids. A *transduction* $\tau : M \rightarrow N$ is a mapping from M into $P(N)$. One extends τ to a mapping $P(M) \rightarrow P(N)$ by setting $P\tau = \bigcup_{m \in P} m\tau$. The *inverse transduction* $\tau^{-1} : N \rightarrow M$ is defined by $Q\tau^{-1} = \{m \in M \mid m\tau \cap Q \neq \emptyset\}$. 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 $\tau : A^* \rightarrow M$ admits a *linear matrix representation* (λ, μ, ν) if there exist $n > 0$, a morphism $\mu : A^* \rightarrow P(M)^{n \times n}$, a row vector $\lambda \in P(M)^{1 \times n}$, a column vector $\nu \in P(M)^{n \times 1}$ such that for all $f \in A^*$, $f\tau = \lambda \cdot f\mu \cdot \nu$.

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

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

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

Theorem 2.1 Let $\tau : A^* \rightarrow M$ be a transduction that admits a linear matrix representation (λ, μ, ν) and let P be a subset of M recognized by a morphism $\eta : M \rightarrow N$. 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 $R = \{m \in P(N)^{n \times n} \mid \lambda\eta \cdot m \cdot \nu\eta \cap Q \neq \emptyset\}$.

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 $\tau : A^* \rightarrow M$ 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 Σ be an alphabet. We denote by $M * \Sigma^*$ the free product (or coproduct) of the monoids M and Σ^* . The monoid $M * \Sigma^*$ can be identified with the set of words of the form $m_0 \xi_1 m_1 \dots \xi_n m_n$ (where $m_0, \dots, m_n \in M$ and $\xi_1, \dots, \xi_n \in \Sigma$) equipped with the product

$$(m_0 \xi_1 \dots \xi_n m_n) (m'_0 \xi'_1 m'_1 \dots \xi'_n m'_n) = m_0 \xi_1 m_1 \dots \xi_n (m_n m'_0) \xi'_1 m'_1 \dots \xi'_n m'_n.$$

Let $\theta : \Sigma^* \rightarrow P(M)$ be a morphism, that is, a substitution from Σ^* to M . We denote by $[\theta] : P(M * \Sigma^*) \rightarrow P(M)$ the semiring morphism defined by

- (a) $u[\theta] = m_0 (\xi_1 \theta) m_1 \dots (\xi_n \theta) m_n$ for $u = m_0 \xi_1 \dots \xi_n m_n \in M * \Sigma^*$
- (b) $x[\theta] = \bigcup_{u \in X} u[\theta]$ for $X \in P(M * \Sigma^*)$

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

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

Lemma 2.3 Let $\eta : M \rightarrow N$ and $\Theta : \Sigma^* \rightarrow P(M)$ be two monoid morphisms. The following diagram is commutative :

$$\begin{array}{ccc} P(M * \Sigma^*) & \xrightarrow{[\Theta]} & P(M) \\ \eta \downarrow & & \downarrow \eta \\ P(N * \Sigma^*) & \xrightarrow{[\Theta\eta]} & P(N) \end{array}$$

For any positive integer n let $\Sigma_n = \{\xi_{11}, \xi_{12}, \dots, \xi_{nn}\}$ be an alphabet with n^2 letters. Any matrix $m \in P(M)^{n \times n}$ defines a morphism $m : \Sigma_n^* \rightarrow P(M)^{n \times n}$ by $\xi_{ij} m = m_{ij}$. We can now generalize definition 2.1.

Definition 2.2 A transduction $\tau : A^* \rightarrow M$ admits a *matrix representation* (s, μ) if there exist $n > 0$, a morphism $\mu : A^* \rightarrow P(M)^{n \times n}$ and an element $s \in P(M * \Sigma_n^*)$ 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 $\tau : A^* \rightarrow M$ be a transduction that admits a matrix representation (s, μ) and let P be a subset of M recognized by a morphism $\eta : M \rightarrow N$. Then the language $P\tau^{-1}$ is recognized by the monoid of matrices $A^*\mu\eta$.

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

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

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

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

3.- Operations on languages

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

$\eta = \prod_{1 \leq i \leq n} \eta_i : \prod_{1 \leq i \leq n} A_i^* \rightarrow \prod_{1 \leq i \leq n} M_i$ recognizes the subset $L_1 \times \dots \times L_n$ of $A_1^* \times \dots \times A_n^*$.

Then theorem 2.1 (or theorem 2.4) gives the construction of a monoid recognizing $(L_1, \dots, L_n)\varphi$ provided that φ satisfies the following conditions :

- (1) φ is the inverse of a transduction $\tau : A^* \rightarrow A_1^* \times \dots \times A_n^*$
- (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 a classical operations.

3.1 Inverse morphisms and inverse substitutions

In this example conditions (1) and (2) are trivially satisfied. For example let $\sigma : A^* \rightarrow B^*$ be a substitution. Then by definition σ induces a morphism $A^* \rightarrow P(B^*)$ and $u\sigma = 1.u\sigma.1$ 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 $P(M)$. Therefore we have obtained the following result proved in [13] for rational languages.

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

In the same way, we get easily

Proposition 3.2 Let $\varphi : A^* \rightarrow B^*$ 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 $\tau : A^* \rightarrow A^* \times A^*$ 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, $A^*\tau\eta$ 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 $u\mu = \begin{pmatrix} u \times A^* & o \\ o & A^* \times u \end{pmatrix}$ 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 $P^{-1}L = \{f \in A^* \mid Pf \cap L \neq \emptyset\}$. 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 $(P, \zeta, 1)$ - where ζ 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 [7] 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 $L_1 \dots L_n = (L_1 x \dots x L_n) \tau^{-1}$ where $\tau : A^* \rightarrow A^* x \dots x A^*$ is the transduction defined by

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

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

$$(\mu u)_{ij} = 0 \text{ if } i > j$$

$$(\mu u)_{ij} = \{u_1, \dots, u_n\} \in A^* x \dots x A^* \mid u_1 = \dots = u_{i-1} = u_{j+1} = \dots = u_n = 1$$

$$\text{and } u_i u_{i+1} \dots u_j = u \text{ if } i \leq j$$

In particular τ is a rational transduction. By theorem 2.1, $L_1 \dots L_n$ 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 $M_1 x \dots x M_n$ such that :

- (a) P is upper triangular, that is $P_{ij} = 0$ for $i > j$
- (b) The i^{th} entry of the diagonal is an element of M_i . More precisely $P_{ii} = \{(1, \dots, 1, m_i, 1, \dots, 1)\}$ for some $m_i \in M_i$
- (c) If $(m_1, \dots, m_n) \in P_{ij}$, then $m_k = 1$ for $k < i$ and $k > j$.

This last monoid is called by Straubing [19] the Schützenberger product of M_1, \dots, M_n and is denoted by $\Diamond_n(M_1, \dots, M_n)$. For $n = 2$ it can be identified with the original definition of Schützenberger (see [6]). Note that in general

$$\Diamond_2(M_1, \Diamond_2(M_2, M_3)) \neq \Diamond_3(M_1, M_2, M_3) \neq \Diamond_2(\Diamond_2(M_1, M_2), M_3).$$

Therefore, we have

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

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

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

3.5 Shuffle, infiltration product

Given a word $h = a_1 \dots a_n \in A^*$ (the a_i 's are letters) and a subset $I = \{i_1, \dots, i_r\}$ of $\{1, \dots, n\}$ (where $i_1 < \dots < i_r$) we denote by h_I the word $a_{i_1} \dots a_{i_r}$. With this notation the *shuffle* of two words f and g is defined by $f \circ g = \{h \in A^* \mid I, J \quad I \cup J = \{1, \dots, |h|\}, I \cap J = \emptyset, h_I = f \text{ and } h_J = g\}$ and the *infiltration product* [4] is defined by

$f \uparrow g = \{h \in A^* \mid I, J \quad I \cup J = \{1, \dots, |h|\}, h_I = f \text{ and } h_J = g\}$

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

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

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

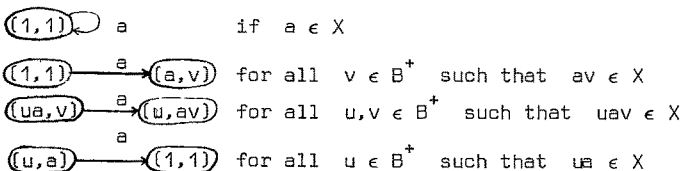
Proposition 3.9 [11] Let L_1 and L_2 be languages of A^* recognized by M_1 and M_2 respectively. Then $L_1 \circ L_2$ is recognized by $\text{Fin}(M_1 \times M_2)$, the monoid of finite subsets of $M_1 \times M_2$.

In the same way $L_1 \uparrow L_2 = (L_1 \times L_2)\tau^{-1}$ where $\tau = \{(f_1, f_2) \in A^* \times A^* \mid f \in f_1 \uparrow f_2\}$ 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 $\text{Fin}(M_1 \times M_2)$.

3.6 Morphisms

Let $\varphi : A^* \rightarrow B^*$ be a morphism. Then $\varphi^{-1} = \tau : B^* \rightarrow A^*$ 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 matrix 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) :



The relations defined by $a, a \in B$, generate a monoid of relations that recognizes X^* . This is precisely the petal monoid of X^* , denoted by $\text{Pet } X^*$. 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 $\varphi : A^* \rightarrow B^*$. 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 $\mu : B^* \rightarrow (\text{Rat } A^*)^{Q \times Q}$ is defined, for all $a \in B$, by

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

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

Now if L is recognized by $\eta : A^* \rightarrow M$, 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 $\text{Rat } M \circ \text{Pet } X^*$. Thus

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

In the case where X is a code the substitution product turns out to be a wreath product (in the sense of [15]). Therefore

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

If φ is injective one can be more precise

Proposition 3.13 Let $\varphi : A^* \rightarrow B^*$ 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 $M \circ \text{Pet } X^*$.

Finally if φ is literal (= length preserving, strictly alphabetic) we could prove the following result, obtained in [18] and [13] for rational languages.

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

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 $G = (A, \{\sigma_1, \dots, \sigma_n\})$ where A is an alphabet and $\sigma_1, \dots, \sigma_n$ are substitutions of A^* into A^* . Let $B = \{1, \dots, n\}$. To each word $u = i_1 \dots i_r$ of B^* one associates the substitution $\sigma_u = \sigma_{i_1} \dots \sigma_{i_r}$ (the substitution associated to 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 $\tau : B^* \rightarrow A^*$ 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.

4.2 Reduction of the free group

Let A be an alphabet, \bar{A} a copy of A and set $\hat{A} = A \cup \bar{A}$. As is well-known the free group $F(A)$ over A is the quotient of \hat{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 \hat{A}^* is congruent to a unique reduced word $u\delta$. This defines a function $\delta : \hat{A}^* \rightarrow \hat{A}^*$ 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 \hat{A}^* . Then $R\delta$ is rational.

Indeed let $\tau : \hat{A}^* \rightarrow \hat{A}^*$ be the transduction defined by $(a_1 \dots a_n)\tau = D^* a_1 D^* \dots a_n D^*$ (where a_1, \dots, 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 $aa = 1$ ($a \in A$) or other variants - see [16] -

4.3. Applications of the 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)

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

$$\mathfrak{L}(L) = \bigcup_{\{2n+1\} \text{ prime}} \frac{1}{2n+1} L. \text{ We shall prove the following result}$$

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

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

Proof Define for all $n > 0$, $\tau_n : A^* \rightarrow A^*$ by $u\tau_n = A^{n|u|} u A^{n|u|}$ and let $\tau : A^* \rightarrow A^*$ be defined by $u\tau = \bigcup_{\{2n+1\} \text{ prime}} u\tau_n$. Then $L\tau_n^{-1} = \frac{1}{2n+1} L$ and

$L\tau^{-1} = \mathfrak{L}(L)$. Moreover τ_n admits the (non-linear) matrix representation (s_n, μ) where $\mu = \begin{pmatrix} A^{|u|} & 0 \\ 0 & u \end{pmatrix}$ for all $u \in A^*$ and $s_n = \xi_{11}^n \xi_{22}^n \xi_{11}^n$. 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 $\eta : A^* \rightarrow M$, theorem 2.4 shows that $\frac{1}{2n+1} L$ and $\mathfrak{L}(L)$ 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 $P(M)$ 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 $\mu = u$ and $s = \xi^2$. 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 $\tau : A^* \rightarrow A^*$ is any rational transduction then

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

References

- [1] J. Beauquier On the structure of context-free languages. To appear in International Journal of Computer Mathematics.
- [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, IV Ann. of Math. 68, 1958, 81-95.
- [5] Ch. Choffrut A closure property of deterministic context-free languages. Information Processing Letters 12, 1981, 13-16.
- [6] S. Eilenberg Automata, Languages and Machines. Academic Press Vol. B 1976
- [7] S. Ginsburg and E.H. Spanier Quotients of context-free languages, J. Assoc. Computing Machinery 10, 1963, 487-492.
- [8] J.E. Hopcroft, J.D. Ullman. Introduction to automata theory, languages and computation, Addison Wesley, 1979.
- [9] E. Le Rest and M. Le Rest Sur le calcul du monoïde syntaxique d'un sous-monoïde finiment engendré. Semigroup Forum Vol 21, 1980 73-185.
- [10] M. Lothaire Combinatorics on words, Addison Wesley, to appear.
- [11] J.-F. Perrot Variétés de langages et opérations. Theoretical Computer Science 7, 1978, 197-210.
- [12] J.-E. Pin Hiérarchies de concaténation. Submitted.
- [13] Ch. Reutenauer Sur les variétés de langages et de monoïdes, 4th GI Conference - Lecture Notes in Computer Science 67, Springer, 1979, 260-265.
- [14] Ch. Reutenauer Sur les séries associées à certains systèmes de Lindenmayer. Theoretical Computer Science 9, 1979, 363-375
- [15] J. Sakarovitch Sur la définition du produit en couronne, in Colloque Codages et transductions (Pirillo ed.), Florence, 1981, 285-300.
- [16] J. Sakarovitch Syntaxe des langages de Chomsky, Thèse Sci. Paris, 1979
- [17] M.P. Schützenberger Sur les relations rationnelles entre monoïdes libres, Theoretical Computer Science 3, 1976, 243-259.

- [18] H..Straubing Recognizable sets and power sets of finite semigroups
Semigroup Forum 18, 1979, 331-340.
- [19] H. Straubing A generalization of the Schützenberger product of finite
monoids. Theoretical Computer Science 13, 1981, 137-150.
- [20] H. Straubing Families of recognizable sets corresponding to certain
varieties of finite monoids. J. Pure and Applied Algebra,
15, 1979, 305-318.