Journal of Pure and Applied Algebra 41 (1986) 233-253
North-Holland

ON VARIETIES OF RATIONAL LANGUAGES AND VARIABLE LENGTH CODES II

Stuart W. MARGOLIS
Department of Computer Science, University of Nebraska-Lincoln, Lincoln, Nebraska 68588-0115, USA

Jean-Eric PIN
C.N.R.S. and University of Paris VI, Paris, France

Communicated by J. Rhodes
Received 1 October 1984

This article is a continuation of the work of the second author on the connections between the theory of varieties of languages and the theory of codes. We show that every variety of languages closed under concatenation product is described by its finite prefix codes. We also consider the operation which associates to any variety of monoids V the variety $V * W$ generated by all semidirect products of a monoid of V by a monoid of W, for various varieties W, and we describe the corresponding operation on varieties of languages.

Cet article poursuit l'étude des rapports entre la théorie des variétés de langages et la théorie des codes entreprise par le second auteur. On montre que toute variété de langages fermée par produit de concaténation est décrite par ses codes préfixes finis. On considère également l'opération qui associe à chaque variété de monoides V la variété $V * W$ engendrée par les produits semidirects d'un monoide de V par un monoide de W, pour diverses valeurs de W, et on décrit l'opération correspondante sur les variétés de langages.

Introduction

This article is a continuation of the work of the second author [13] on the connections between the theory of varieties of languages and the theory of codes.

Varieties of languages were introduced by S. Eilenberg [4] to provide a common framework to a certain number of isolated results characterizing recognizable languages in terms of their syntactic semigroups. Kleene's theorem on rational languages, Schützenberger's theorem on star-free languages and Simon's theorem on piecewise testable languages are the basic examples of this theory [4].

The theory of codes originated in the seminal work of Schützenberger in the fifties. It is well known that, as opposed to the case of a free group, submonoids of a free monoid need not be free. If this is the case, the basis of such a submonoid is called a code. The theory of codes has grown considerably in recent years and is

one of the leading parts of the combinatorial theory of semigroups. The (future) book of Berstel, Perrin and Schutzenberger [2] gives a complete survey of this theory.

The connection between codes and varieties first originated in the fact that certain combinatorial properties of codes are reflected by algebraic properties of their syntactic monoids. Thus there is a hope to classify certain codes by means of syntactic properties. However, only a few number of results of this type are known at present [18, 19, 6, 7]. Another connection between codes and varieties was discovered by Eilenberg and Schützenberger [4, Chapter 10] in their study of decomposition algorithms for rational sets [4]. This algorithm produces for every recognizable language a rational expression in which, in particular, the star operation is applied only to prefix codes. Eilenberg and Schützenberger used this fact to give new descriptions of certain varieties of languages. Further characterizations of varieties making use of codes were given in [13].

The purpose of this paper is to extend these results by characterizing a great number of varieties of languages. The description of a variety of languages is usually achieved as follows. One gives a 'basic' class of languages and a certain number of operations to construct the languages from this basic class. For example the rational languages are obtained from the letters by means of finite union, concatenation and star. The star-free languages are obtained from the letters and the empty word by means of boolean operations and concatenation. The interest of such descriptions depends on how 'natural' are the basic class and the operations. Our paper rests on the subjective claim that finite prefix codes are a natural basic class and that the operation of coding is a natural operation. Indeed our first main result can be informally stated as follows: Every variety of languages $\mathscr{V}$ closed under concatenation product is described by its finite prefix codes. More precisely the languages of $\mathscr{V}$ are obtained from the languages P^* where P is a finite prefix code such that P^* is in $\mathscr{V}$ by means of the variety operations: namely boolean operations, inverse morphisms and left and right quotients. Notice that concatenation is *not* needed in this description of $\mathscr{V}$. This theorem solves a conjecture of [13] and implies, in particular, that, for a given $n \geq 0$, the variety of languages whose syntactic monoids have complexity $\leq n$ [26] is described by its finite prefix codes.

Eilenberg's variety theorem gives a one-to-one correspondence between varieties of languages and varieties of finite monoids. Thus it is not surprising that operations on varieties of languages correspond to operations on varieties of monoids. For example, an important theorem of Straubing [22] states that a variety of languages is closed under product if and only if the corresponding variety of monoids is closed under inverse of aperiodic morphisms. In this paper we consider the operation which associates to any variety of monoids $\mathbf{V}$ the variety $\mathbf{V} * \mathbf{A}$ generated by all semidirect products of a monoid of $\mathbf{V}$ by an aperiodic monoid on the right, and we describe the corresponding operation on varieties of languages. If $\mathscr{V}$ is the variety of languages corresponding to $\mathbf{V}$, then the variety of languages corresponding to $\mathbf{V} * \mathbf{A}$ is the smallest variety containing $\mathscr{V}$ and closed under the operations of prefix

pure coding and left concatenation with letters. More generally our second main result gives analogous descriptions for the operation $\mathbf{V} * \mathbf{W}$ where $\mathbf{W}$ is a fixed variety of monoids that is closed under inverse of aperiodic morphisms. As a by-product we obtain a description of the languages whose syntactic monoids have complexity (resp. abelian complexity) less than or equal to 1. Part of these results were announced in [14].

The proofs are based on an improvement of the simulation technique introduced in [13]. The basic idea is to 'approximate' a finite monoid by a syntactic monoid $M(P^*)$ where P is a finite prefix code. The most significant result in this direction is Corollary 2.8. Let $\mathbf{V}$ be a variety closed under inverse aperiodic morphisms. Then for every monoid M of $\mathbf{V}$ one can effectively construct a finite prefix code P such that M divides $M(P^*)$ and such that $M(P^*)$ is in $\mathbf{V}$. This result implies for example that the (open) problem of computing the complexity of a finite monoid M can be reduced to the case where M has the form $M(P^*)$ for some finite prefix code P.

The paper breaks up into four sections. Section 1 is a preliminary section and Section 2 presents the simulation technique. In Section 3 we show that a number of varieties are described by their finite codes. We also give some important counter-examples: the variety of languages of dot-depth one, for instance, cannot be described by its finite prefix codes. In Section 4 we discuss the operations $\mathbf{V} * \mathbf{W}$ for various values of $\mathbf{W}$ and we describe the corresponding operations on languages.

1. Preliminaries

In this section we recall some basic definitions and results. For all terms not defined in the text, see [4] or [8] or [17].

1.1. Semigroups

A semigroup is a set equipped with an associative law. A *monoid* is a semigroup with identity. An element $e \in S$ is *idempotent* if $e = e^2$. The set of all idempotents of S is denoted by $E(S)$.

In this paper all semigroups will be finite, except for free semigroups and free monoids.

A *variety of finite semigroups* (resp. monoids) is a class of semigroups (resp. monoids) closed under taking subsemigroups quotients and finite direct products.

Given a semigroup S, we denote by S^1 the monoid constructed as follows. If S is a monoid, then $S^1 = S$ and if S is not a monoid $S^1 = S \cup \{1\}$ where 1 is a new identity. U_1 denotes the two element idempotent semigroup. Thus $U_1 = \{1, 0\}$ where 1 is an identity and 0 is a zero.

A variety of semigroups $\mathbf{V}$ is called *monoidal* if $S \in \mathbf{V}$ implies $S^1 \in \mathbf{V}$. For example the variety $\mathbf{A}$ of all aperiodic (or group-free) semigroups is a monoidal variety of semigroups. Also the variety $\mathbf{J}_1$ of all idempotent and commutative semigroups

(or semilattices) and the variety $\boldsymbol{R}$ of $\mathscr{R}$-trivial semigroups are monoidal varieties of semigroups.

Given a variety of semigroups $\boldsymbol{V}$, $\boldsymbol{LV}$ denotes the variety of all semigroups S which are locally in $\boldsymbol{V}$, that is, such that for all $e \in E(S)$ the subsemigroup eSe is in $\boldsymbol{V}$. In particular, $\boldsymbol{LI}$ denotes the variety of all locally trivial semigroups and $\boldsymbol{LJ}_1$ denotes the variety of all locally idempotent and commutative semigroups.

Similarly $\boldsymbol{UV}$ denotes the variety of all semigroups S such that the subsemigroup $SE(S)S$ is in $\boldsymbol{V}$.

Let S and T be two semigroups. A relational morphism $\tau : S \to T$ is a relation from S to T such that

(1) For all $s \in S$, $s\tau \neq \emptyset$.

(2) For all $s_1, s_2 \in S$: $(s_1\tau)(s_2\tau) \subset (s_1 s_2)\tau$.

Then, if T' is a subsemigroup of T, the set $T'\tau^{-1} = \{s \in S \mid s\tau \cap T' \neq \emptyset\}$ is a subsemigroup of S.

Let $\boldsymbol{V}$ be a variety of semigroups. A relational morphism (resp. morphism) $\tau : S \to T$ is a relational $\boldsymbol{V}$-morphism if for all subsemigroups T' of T, $T' \in \boldsymbol{V}$ implies $T'\tau^{-1} \in \boldsymbol{V}$. In particular, (relational) $\boldsymbol{A}$-morphisms, also called (relational) aperiodic morphisms play a central role in the theory of finite semigroups. If $\boldsymbol{V}$ and $\boldsymbol{W}$ are two varieties of semigroups, $\boldsymbol{V}^{-1}\boldsymbol{W}$ denotes the variety of all semigroups S such that there exists a semigroup $T \in \boldsymbol{W}$ and a relational $\boldsymbol{V}$-morphism $\tau : S \to T$.

Given two varieties $\boldsymbol{V}$ and $\boldsymbol{W}$, $\boldsymbol{V} * \boldsymbol{W}$ denotes the variety generated by all semidirect products of the form $S * T$ where $S \in \boldsymbol{V}$ and $T \in \boldsymbol{W}$.

1.2. Languages

Let A be a finite alphabet and let A^* (resp. A^+) be the free monoid (resp. free semigroup) over A. The empty word is denoted by 1. Subsets of A^+ (resp. A^*) are called *languages*. Given two languages K and L of A^*, we set

$$K^{-1}L = \{v \in A^* \mid Kv \cap L \neq \emptyset\} \quad \text{and} \quad KL^{-1} = \{v \in A^* \mid vL \cap K \neq \emptyset\}.$$

In particular, if u is a word of A^*, we set

$$u^{-1}L = \{v \in A^* \mid uv \in L\} \quad \text{and} \quad Lu^{-1} = \{v \in A^* \mid vu \in L\}.$$

Observe that if u and v are words, then $(uv)^{-1}L = v^{-1}(u^{-1}L)$. Let L be a language of A^+ (resp. A^*). The *syntactic semigroup* (monoid) $S(L)$ (resp. $M(L)$) of L is the quotient of A^+ (resp. A^*) by the congruence $\sim_L$ defined by $u \sim_L v$ iff for all $x, y \in A^*$, $uxv \Leftrightarrow uyv \in L$. Recall that a language L of A^+ (resp. A^*) is recognizable iff $S(L)$ (resp. $M(L)$) is finite. In the sequel every language is assumed to be recognizable and thus every syntactic semigroup is finite.

A +-variety of languages $\mathscr{V}$ associates to every alphabet A a set $A^+\mathscr{V}$ of recognizable languages of A^+ such that

(1) For every alphabet A, if $a \in A$ and $L \in A^+\mathscr{V}$, then $a^{-1}L, La^{-1} \in A^+\mathscr{V}$.

(2) For every alphabet A, $A^+\mathscr{V}$ is closed under finite boolean operations.

(3) For any semigroup morphism $\psi: A^+ \to B^+$, $L \in B^+\mathscr{V}$ implies $L\psi^{-1} \in A^+\mathscr{V}$.

$*$-varieties of languages are defined in the same way by replacing every occurrence of '+' by '$*$' and 'semigroup' by 'monoid' in the previous definition.

Eilenberg's variety theorem states that there exists a one-to-one correspondence between +-varieties of languages and varieties of semigroups (resp. between $*$-varieties of language and varieties of monoids). In the sequel we will always refer implicitly to this correspondence by saying that such a +-variety of languages $\mathscr{V}$ corresponds to such a variety of semigroups $\mathbf{V}$ or vice-versa. For example the variety of all semigroups corresponds to the variety of rational languages. Also the variety $\mathbf{A}$ of aperiodic semigroups corresponds to the variety $\mathscr{A}$ of star-free languages (Schützenberger) [4,8,17]. Recall that for any alphabet A, $A^+\mathscr{A}$ is the smallest set of languages containing the letters and closed under finite boolean operations and concatenation.

A +-variety (resp. $*$-variety) $\mathscr{V}$ is closed under concatenation product iff for every alphabet A and for every $n \geq 0$, $L_1, \ldots, L_n \in A^+\mathscr{V}$ (resp. $A^*\mathscr{V}$) implies $L_1 \cdots L_n \in A^+\mathscr{V}$ (resp. $A^*\mathscr{V}$). The following theorem characterizes varieties under closed product.

Theorem 1.1 (Straubing [22]). *A variety of languages is closed under product iff the corresponding variety of semigroups (monoids) $\mathbf{V}$ satisfies $\mathbf{A}^{-1}\mathbf{V} = \mathbf{V}$.*

A *+-class* ($*$-class) of languages associates to any alphabet A a set $A^+\mathscr{C}$ of languages of A^+ (resp. a set $A^*\mathscr{C}$ of languages of A^*). We will say that a variety $\mathscr{V}$ contains a class $\mathscr{C}$ of languages if for every alphabet A, $A^+\mathscr{C} \subset A^+\mathscr{V}$ (resp. $A^*\mathscr{C} \subseteq A^*\mathscr{V}$ if we deal with $*$-classes and $*$-varieties).

1.3. Codes

A language C of A^+ is a *code* iff the the subsemigroup C^+ of A^+ generated by C is free of base C, that is if every element of C^+ has a unique factorization in elements of C. A code C is said to be *prefix* if for every $u, v \in A^*$, $uv \in C^+$ and $u \in C^+$ implies $v \in C^+$. It is well known that a code C is prefix iff no word of C has a proper left factor in C, that is, if for every $u, v \in A^*$, $uv \in C$ and $u \in C$ implies $v = 1$.

A code C is *pure* if for every $u \in A^+$, $u^n \in C^+$ for some $n > 0$ implies $u \in C^+$. The following result was first stated and proved in [18] (see also [19] and [23] for various extensions).

Proposition 1.2. *Let C be a finite code. The following conditions are equivalent*

(1) *C is pure.*

(2) *C^+ is star-free.*

(3) *$S(C^+)$ is aperiodic.*

1.4. Automata

In this paper we will consider only deterministic automata. However, an auto

maton need not be complete. Thus an automaton $\mathscr{A}=(Q,A)$ consists of a set of states Q, an alphabet A and a partial function $Q\times A\to Q$. This function defines an action of each letter on Q. We simply denote by qa the result of the action of the letter a on the state q. Thus qa is either a state of Q or the empty set. The action can be extended to an action of A^* on Q by the following induction rules

$$q1=q \quad \text{for all } q\in Q,$$

$$q(ua)=(qu)a \quad \text{for all } q\in Q,\ u\in A^* \text{ and } a\in A.$$

Thus each word of A^* defines partial function from Q to Q. The *rank* of u in $\mathscr{A}$ is the cardinality of the image of the function defined by u. More formally

$$\text{rank}(u)=\text{Card}\{qu \mid q\in Q\}$$

Given a language $L\subset A^*$, the *minimal automaton* of L is $\mathscr{A}(L)=(Q,A)$ where $Q=\{u^{-1}L \mid u\in A^* \text{ and } u^{-1}L\neq\emptyset\}$ and the action of A^* on Q is given by

$$(u^{-1}L)=\begin{cases}(uv)^{-1}L=v^{-1}(u^{-1}L) & \text{if } (uv)^{-1}L\neq\emptyset,\\ \text{undefined} & \text{otherwise.}\end{cases}$$

Note that this definition is slightly different from the definition given in [13] where only complete automata were considered.

2. Simulation of automata

The following definition was introduced in [13] for complete automata.

Definition. Let $\mathscr{A}_1=(Q_1,A_1)$ and $\mathscr{A}_2=(Q_2,A_2)$ be two finite automata. Then $\mathscr{A}_1$ simulates $\mathscr{A}_2$ if there exists a subset Q of Q_1, a bijection $\psi: Q\to Q_2$ and an injection $\pi: A_2\to A_1^+$ such that for all $q\in Q$ and for all $a\in A_2$, $q\psi a=q(a\pi)\psi$.

Informally this definition says that every letter a of A_2 has the same action in $\mathscr{A}_2$ as a certain word all of A_1^+ *acting on a fixed subset* Q of Q_1. The following result was proved in [13] for complete automata.

Proposition 2.1. *If $\mathscr{A}_1$ simulates $\mathscr{A}_2$; then the transition semigroup of $\mathscr{A}_2$ divides the transition semigroup of $\mathscr{A}_1$.*

Proof. The result follows from an exercise of [4], but we give a complete proof for the convenience of the reader. Let S_1 (resp. S_2) be the transition semigroup of $\mathscr{A}_1$ (resp. $\mathscr{A}_2$) and let $\pi_1: A_1^+\to S_1$ and $\pi_2: A_2^+\to S_2$ be the natural projections. The injection $\pi: A_2\to A_1^+$ can be extended into a morphism $\pi: A_2^+\to A_1^+$. Let $T=A_2^+\pi\pi_1$. Then T is a subsemigroup of S_1. We claim that S_2 is a quotient of T. Indeed let $u,v\in A_2^+$ and assume that $u\pi\pi_1=v\pi\pi_1$. Then by definition $q(u\pi)=$

$q(u\pi)$ for all $q \in Q_1$. It follows that, for all $q \in Q$, $qu\pi\psi = qv\pi\psi$ and thus $q\psi u = q\psi v$. Since ψ is a bijection from Q to Q_2, $qu = qv$ for all $q \in Q_2$, that is $u\pi_2 = v\pi_2$. Thus $u\pi\pi_1 = v\pi\pi_1$ implies $u\pi_2 = v\pi_2$ and therefore there exists a surjective morphism $T \to S_2$. □

In view of applications to the theory of varieties we are especially interested in the following problem. Given an automaton $\mathscr{A}$ with transition semigroup S find a finite prefix code P such that

(a) The minimal automaton $\mathscr{B}$ of P^+ simulates $\mathscr{A}$.

(b) The syntactic semigroup T of P^+ does not differ 'too much' from S.

Of course, the meaning of the expression 'does not differ too much' depends on the context. A typical condition will be that if S belongs to a given variety of semigroups V, then T also belongs to V. The following theorem summarizes some results of [13].

Theorem 2.2. *There exists an algorithm which, given a finite complete automaton $\mathscr{A}$ with transition semigroup S, produces a finite prefix code P such that*

(1) *The minimal automaton of P^+ simulates $\mathscr{A}$. In particular, S divides the syntactic semigroup T of P^+.*

(2) *If $\mathscr{A}$ is the minimal automaton of the language A^*wA^* for some $w \in A^*$, then T is locally idempotent and commutative.*

(3) *Let V be a variety of semigroups such that $A^{-1}V = V$ and $LV = V$. Then if $S \in V$, then $T \in V$.*

(4) *In particular, if S is aperiodic, T is aperiodic.*

We will show now that an almost identical construction works for incomplete automata and that Theorem 2.2 can be improved. Let us first describe the modified construction.

Let $\mathscr{A} = (\{1, \ldots, n\}, \Sigma, \cdot)$ be a finite (partial) automaton with n states. Assume that $\mathscr{A}$ contains a non-empty transition, that is there exists a state $i \in \{1, \ldots, n\}$ and a letter $\sigma \in \Sigma$ such that $i\sigma \neq \emptyset$. Without loss of generality we may suppose that there exists a letter $\sigma \in \Sigma$ such that $n\sigma \neq \emptyset$. Let $\tau : N \to N$ be the function defined by $k\tau = 2^k - 2$. The key property of this function for our purpose is that $i_1\tau + i_2\tau = j_1\tau + j_2\tau$ implies that $\{i_1, i_2\} = \{j_1, j_2\}$.

Let $A = \{a\} \cup \{a_\sigma \mid \sigma \in \Sigma\}$ be an alphabet with $1 + \operatorname{Card} \Sigma$ letters and let $P = \{a^{i\tau} a_\sigma a^{n\tau - i\sigma\tau}$ such that $1 \leq i \leq n$, $\sigma \in \Sigma$ and $i\sigma$ is defined$\}$. Then P is a prefix code and the minimal automaton $\mathscr{B}$ of P^+ simulates $\mathscr{A}$. The states and transitions of $\mathscr{B}$ are given in [13] in the case that $\mathscr{A}$ is complete. The same technique works in the general case and we describe this here.

Let $m = \max\{n\tau - i\sigma\tau \mid i\sigma$ is defined, $1 \leq i \leq n\}$. Then $\mathscr{B} = (Q, A)$ where $Q = \{q_j \mid -m \leq j \leq n\tau\}$ and the transitions are given by the following relations

$$(1)\qquad \begin{cases} q_j a = q_{j+1} & \text{if } j+1 \le n\tau, \\ q_{i\tau} a_\sigma = q_{-n\tau+i\sigma\tau} & \text{for } 1 \le i \le n \quad (\text{if } i\sigma \text{ is defined}). \end{cases}$$

Let us denote by S (resp. T) the transition semigroup of $\mathscr{A}$ (resp. $\mathscr{B}$) and let $W(S)$ (resp. $W(T)$) be the ideal of all elements of rank ≤ 1 in $\mathscr{A}$ (resp. $\mathscr{B}$). Note that $W(S)$ and $W(T)$ are both aperiodic ideals. The next theorem gives the relationship between S and T. Let N be the cyclic subsemigroup of T generated by a. It follows easily from (1) that $a^{n\tau+m+1}=0$ is the zero of N and that $a^{n\tau+m} \neq 0$. Thus $N = \{a, a^2, \ldots, a^{n\tau+m}, 0\}$ is nilpotent. Let

$$K = \{(n,u,s) \mid 0 \le r < n\tau + m, u \in S \setminus W(S), -m < s \le n\tau\}.$$

Now $R = N \cup K$ is a semigroup under the following multiplication:

$$(r,u,s)(r',u',s') = (r,uu',s') \quad \text{if } s = r' \text{ and } uu' \in S \setminus W(S),$$

$$a^i(r,u,s) = (i+r,u,s) \quad \text{if } i + r < n\tau + m,$$

$$(r,u,s)a^i = (r,u,s-i) \quad \text{if } -m < s - i.$$

For $x, y \in R$, $xy = 0$ in all other cases.

Now we can state:

Theorem 2.3. *S divides T and $T/W(T)$ is a subsemigroup of R.*

Proof. Since $\mathscr{B}$ simulates $\mathscr{A}$, s divides T by Proposition 2.1. The second part of the theorem is a consequence of the following lemma, where for $\sigma = \sigma_1 \cdots \sigma_p \in \Sigma^+$, u_σ denotes the word $a_{\sigma_1} a^{n\tau} n_{\sigma_2} a^{n\tau} \cdots a_{\sigma_p} a^{n\tau}$.

Lemma 2.4. *Let u be a word of rank ≥ 2 in $\mathscr{B}$. Then either $u = a^r$ for some $0 < r < n\tau + m$ or $u = a^r u_\sigma a^{-s}$ for some $\sigma \in \Sigma^+$ of rank ≥ 2 in $\mathscr{A}$ and for some $0 \le r < n\tau + m$ and $-m < s \le n\tau$ (s may be negative). Moreover, in this case*

$$(2)\qquad \begin{cases} q_{-r+i\tau} u = q_{-s+i\sigma\tau} & \text{if } -m \le -r + i\tau \le n\tau, -m \le -s + i\sigma\tau \le n\tau \text{ and if } i\sigma \neq \emptyset, \\ q_i u = \emptyset & \text{otherwise.} \end{cases}$$

Proof. Let u be a word of rank ≥ 2 in $\mathscr{B}$. Proposition 2.6 of [10] which was proved for complete automata, can be readily extended to our construction for incomplete automata. Therefore, if u contains a factor of the form $a_{\sigma_1} a^r a_{\sigma_2}$, then $r = n\tau$. Thus, either $u = a^r$ or $u = a^r u_\sigma a^{-s}$ for some $\sigma \in \Sigma^+$, $r > 0$ and $s \in \mathbb{Z}$. If $u = a^r$, then by (1), $q_j u = q_{j+r}$ and therefore $0 < r < n\tau + m$ since u has rank ≥ 2 in $\mathscr{B}$. If $u = a^r u_\sigma a^{-s}$, it follows easily from (1) that (2) holds. Finally the inequality $r < n\tau + m$ and $-m < s$ follow from the fact that u does not contain $a^{n\tau+m}$ as a factor, since $a^{n\tau+m}$ has rank 1. □

We now show that $T/W(T)$ is a subsemigroup of R. If $n = 1$, then $m = n\tau = 0$, $T/W(T)$ is trivial and the result is obvious. Thus we may assume $n \ge 2$. Let us denote

by $\bar{\sigma}$ the image in S of a word $\sigma \in \Sigma^+$ and by $u\pi$ the image in $T/W(T)$ of a word $u \in A^+$. Define a map $\phi : T/W(T) \to R$ by setting

$$z\phi = \begin{cases} (r, \bar{\sigma}, s) & \text{if } z = (a^r u_\sigma a^{-s})\pi \text{ for some } 0 \le r < n\tau + m \text{ and} \\ & -m < s \le n\tau \text{ and for some } \sigma \in \Sigma^+ \text{ of rank } \ge 2 \text{ in } \mathscr{A}, \\ a^p & \text{if } z = a^p \pi \text{ for some } 0 < p < n\tau + m, \\ 0 & \text{otherwise.} \end{cases}$$

We first show that ϕ is well defined, that is, if two words u and u' of A^+ define the same transformation of rank ≥ 2 in $\mathscr{B}$, then $u\phi = u'\phi$. The previous lemma gives a description of these words of rank ≥ 2. If $u = a^p$ and $u' = a^{p'}$ for some $0 < p \le p' < n\tau + m$, then $q_{n\tau - p} a^p = q_{n\tau}$ and since u and v have the same actions, $q_{n\tau - p} a^{p'} = q_{n\tau}$. Therefore $p = p'$ and $u = u'$. Assume now that $u = a^r u_\sigma a^s$, $u' = a^{r'} u_\sigma a^{s'}$ for some $0 \le r, r' < n\tau + m$, $-m < s, s' \le n\tau$ and $\sigma, \sigma' \in \Sigma$ of rank ≥ 2 in $\mathscr{A}$. Let q_x and q_y be two states of $\mathscr{B}$ such that $q_x u = q_x u' \ne \emptyset$, $q_y u' \ne \emptyset$ and $q_x u \ne q_y u$. Then by (2) there exist some indices i_1, i_2, j_1, j_2 such that

$$x = -r + i_1\tau = -r' + i_2\tau,$$

$$y = -r + j_1\tau = -r' + j_2\tau.$$

It follows that $i_1\tau + j_2\tau = i_2\tau + j_1\tau$ and thus by the property of τ, $\{i_1, j_2\} = \{i_2, j_1\}$. Since $x \ne y$, $i_1 \ne j_1$ and hence $i_1 = i_2$, $j_1 = j_2$ and $r = r'$. Now by (2), $q_{-r+i_1\tau} u = q_{-s+i_1\sigma\tau}$ and $q_{-r+i_1\tau} u' = q_{-s'+i_1\sigma'\tau}$ and thus $-s + i_1\sigma\tau = -s' + i_1\sigma'\tau$ and similarly $-s + j_1\sigma\tau = -s' + j_1\sigma'\tau$. Therefore $i_1\sigma\tau + j_1\sigma'\tau = i_1\sigma'\tau + j_1\sigma\tau$ and thus $\{i_1\sigma, j_1\sigma'\} = \{i_1\sigma', j_1\sigma\}$. Since $q_x u \ne q_y u$, $i_1\sigma \ne j_1\sigma$ and consequently $i_1\sigma = i_1\sigma'$, $j_1\sigma = j_1\sigma'$ and $s = s'$. Finally (2) shows that σ and σ' have the same action on $\mathscr{A}$, that is $\bar{\sigma} = \bar{\sigma}'$.

Now assume that $u = a^p$ for some $0 < p < n\tau + m$ and that $u' = a^r u_\sigma a^{-s}$ for some $0 < r < n\tau + m$, $-m < s \le n\tau$ and $\sigma \in \Sigma^+$. Then $q_{-m} a^p \ne \emptyset$ and $q_{-m+1} a^p \ne \emptyset$ imply $q_{-m} u' \ne \emptyset$ and $q_{-m+1} u' \ne \emptyset$. Therefore by (2) there exist two distinct indices i and j such that $-m = -r + i\tau$ and $-m + 1 = -r + j\tau$. It follows that $j\tau - i\tau = 1$, a contradiction. Thus ϕ is well defined.

We now show that ϕ is a morphism. Let σ, σ' be two words of Σ^+ of rank ≥ 2 in $\mathscr{A}$ and let $0 \le r, r' < n\tau + m$ and $-m < s, s' \le n\tau$. Then

$$(a^r u_\sigma a^{-s} a^{r'} u_{\sigma'} a^{-s'})\phi = \begin{cases} (r, \bar{\sigma}\bar{\sigma}', s') & \text{if } s = r' \text{ and if } \sigma\sigma' \text{ has rank } \ge 2 \text{ in } \mathscr{A}, \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand

$$(r, \bar{\sigma}, s)(r', \bar{\sigma}, s) = \begin{cases} (r, \bar{\sigma}\bar{\sigma}', s') & \text{if } s = r' \text{ and if } \bar{\sigma}\bar{\sigma}' \in S \setminus W(S), \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $(a^r u_\sigma a^{-s} a^{r'} u_{\sigma'} a^{-s'})\phi = (a^r u_\sigma a^{-s})\phi\,(a^{r'} u_{\sigma'} a^{-s'})\phi$ in any case.

Similarly, if $0 \le q < m + n\tau$,

$$(a^q a^r u_\sigma a^{-s})\phi = (a^{q+r} u_\sigma a^{-s})\phi = \begin{cases} (q+r, \bar{\sigma}, s) & \text{if } q+r < m+n\tau, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$(a^q\phi)(a^r u_\sigma a^{-s})\phi = a^q(r, \bar{\sigma}, s) = \begin{cases} (q+r, \bar{\sigma}, s) & \text{if } q+r < m+n\tau, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $(a^q a^r u_\sigma a^{-s})\phi = a^q\phi(a^r u_\sigma a^{-s})\phi$ and dually $(a^r u_\sigma a^{-s} a^q)\phi = (a^r u_\sigma a^{-s})\phi a^q\phi$. Finally

$$(a^q a^{q'})\phi = a^{q+q'} = (a^q\phi)(a^{q'}\phi).$$

The last step of the proof consists in showing that ϕ is injective. First, if $z \neq 0$, then $z = u\pi$ for some u of rank ≥ 2 in $\mathscr{B}$ and thus $z\phi \neq 0$ by Lemma 2.4. Assume $z\phi = z'\phi = a^p\pi$. Then $z = z' = a^p\pi$. Similarly, if $z\phi = z'\phi = (r, \bar{\sigma}, s)$, then $z = (a^r u_{\sigma_1} a^{-s})\pi$ and $z' = (a^r u_{\sigma_2} a^{-s})\pi$ for some $\sigma_1, \sigma_2 \in \Sigma^+$ such that $\bar{\sigma}_1 = \bar{\sigma}_2 = \bar{\sigma}$. Now, by (2), $a^r u_{\sigma_1} a^{-s}$ and $a^r u_{\sigma_2} a^{-s}$ have the same action in $\mathscr{B}$ and thus $z = z'$. □

The next proposition relates the structure of the semigroups R and S.

Proposition 2.5. *There exists an aperiodic relational morphism* $\psi : R \to [S/W(S)]^1$.

Proof. Define a relation $\psi : R \to [S/W(S)]^1$ by

$$a^p\psi = \{1\} \quad \text{for } 0 < p < n\tau + m,$$

$$(r, u, s)\psi = \{u\} \quad \text{for } 0 \leq r < n\tau + m,\ -m < s \leq n\tau, u \in S \setminus W(S),$$

$$0\psi = (S/W(S))^1.$$

It is easy to check that ψ is a relational morphism. We claim that ψ is aperiodic. For let $e = e^2 \in (S/W(S))^1$. If $e = 0$, then $e\psi^{-1} = \{0\}$ is aperiodic. If $e = 1$, then $1\psi^{-1}$ is the subsemigroup of R generated by a and thus $1\psi^{-1}$ is aperiodic. Finally if $e \in S \setminus W(S)$, then

$$e\psi^{-1} = \{(r, e, s) \mid 0 \leq r < n\tau + m,\ -m < s \leq n\tau\} \cup \{0\}.$$

Clearly $e\psi^{-1}$ satisfies the equation $x^2 = x^3$ and so is aperiodic. □

Consequently we obtain:

Theorem 2.6. *For every finite semigroup S there exists a finite prefix code P such that the syntactic semigroup T of P^+ satisfies the following properties*

(1) *S divides T.*

(2) *There exists an aperiodic relational morphism $\psi : T \to S^1$.*

Proof. Let $(S^1, S, \cdot)$ be the automaton induced by the right action of S on S^1, that is, for all $q \in S^1$ and $s \in S$, $q \cdot s = qs$. By Theorem 2.3 and Proposition 2.5 there ex-

ists a finite prefix code P such that, with the previous notation

(a) S divides T.

(b) $T/W(T)$ divides R.

(c) There exists an aperiodic relational morphism $R \to (S/W(S))^1$.

Since $W(T)$ is an aperiodic ideal, the projection $T \to T/W(T)$ is aperiodic. Since $T/W(T)$ divides R, there exists an injective (elementary in the terminology of [26]) and hence aperiodic relational morphism $T/W(T) \to R$. Since $S/W(S)$ divides S, $(S/W(S))^1$ divides S^1 and thus there exists also an aperiodic relational morphism $(S/W(S))^1 \to S^1$. Finally we obtain by compostion an aperiodic relational morphism $\psi : T \to S^1$. □

Corollary 2.7. *Let V be a monoidal variety of semigroups satisfying $A^{-1}V = V$. Then for every finite semigroup $S \in V$ there exists a finite prefix code P such that*

(1) *S divides $S(P^+)$.*

(2) *$S(P^+)$ is in V.*

Corollary 2.8. *Let V be a variety of monoids satisfying $A^{-1}V = V$. Then for every finite monoid $M \in V$ there exists a finite prefix code P such that*

(1) *M divides $M(P^*)$,*

(2) *$M(P^*) \in V$.*

Proof. By Theorem 2.6 there exists a finite prefix code P such that the syntactic semigroup T of P^+ satisfies

(a) M divides T.

(b) There exists an aperiodic relational morphism $\psi : T \to M$.

Now $M(P^*) = T^1$ and thus M divides T^1. Moreover ψ can be extended to an aperiodic relational morphism $T^1 \to M$ by setting $1\psi = \{1\}$. Now, since $M \in V$ and $A^{-1}V = V$, $T^1 \in V$. □

Here is another consequence of Theorem 2.3.

Theorem 2.9. *Let V be a variety of semigroups satisfying $A^{-1}V = V$ and $UV = V$. Then for every finite semigroup $S \in V$ there exists a finite prefix code P such that*

(1) *S divides $S(P^+)$,*

(2) *$S(P^+)$ is in V.*

Proof. By Theorem 2.3, there exists a finite prefix code P such that the syntactic semigroup T of P^+ satisfies

(a) S divides T.

(b) $T' = T/W(T)$ is a subsemigroup of R.

It follows that $T'E(T')T'$ is a subsemigroup of $RE(R)R$. But $R = K \cup N \cup \{0\}$ and since N is nilpotent, $E(R)$ is contained in the ideal $K^0 = K \cup \{0\}$ of R. Therefore $RE(R)R$ is a subsemigroup of K^0. Let $\psi : K^0 \to S/W(S)$ be the relation defined by

$$\begin{cases}(r,u,s)\psi = \{u\},\\ 0\psi = S/W(S).\end{cases}$$

A proof similar to the proof of Proposition 2.5 shows that ψ is an aperiodic relational morphism. Now $S \in V$ by hypothesis and thus $S/W(S) \in V$. Since $V = A^{-1}V$, K^0 is also in V and since $T'E(T')T'$ is a subsemigroup of K^0, $T'E(T')T' \in V$. Now since $V = UV$, $T/W(T) = T' \in V$ and finally $T \in V$ since there exists an aperiodic morphism $T \to T/W(T)$.

Note that Theorem 2.5 improves condition (3) of Theorem 2.2 since every local variety of semigroups V satisfies $UV = V$. Indeed assume that $S \in UV$, that is $SE(S)S \in V$. Then, for all $e \in E(S)$, $eSe = ee(Se)$ is a subsemigroup of $SE(S)$ and thus $eSe \in V$. Since V is local, $S \in V$. □

We remark also that condition (2) of Theorem 2.2 is no longer true with our new construction. However we can obtain an analoguous result which is easier to prove and sufficient for the applications. Recall that a code $P \subset A^+$ is *very pure* [3] or *circular* [1] iff for all $u, v \in A^+$, $uv \in P^+$ and $vu \in P^+$ imply $u \in P^+$ and $v \in P^+$. A language L is *strictly locally testable* if there exist four finite sets $U, V, W, F \subset A^+$ such that $L = ((UA^* \cap A^*V) \setminus A^*WA^*) \cup F$. It is shown in [3] that a language L is strictly locally testable iff there exists an integer $n > 0$ such that all words of A^+ of length $\geq n$ have rank ≤ 1 in the minimal automaton of L.

We can now state:

Theorem 2.10. *Let L be a strictly locally testable language and let $\mathscr{A}$ be the minimal automaton of L. Then there exists a finite prefix code P such that:*

(1) *P is very pure.*

(2) *$S(P^+)$ is locally idempotent and commutative.*

(3) *$S(L)$ divides $S(P^+)$.*

Proof. First note that (a) and (b) are two equivalent properties [3]. Let S be the transition semigroup of $\mathscr{A}$. By Theorem 2.3 there exists a finite prefix code P such that $T = S(P^+)$ satisfies

(a) S divides T.

(b) $T/W(T)$ divides R.

Since L is strictly locally testable, there exists an integer $n > 0$ such that every word of length $\geq n$ has rank ≤ 1 in $\mathscr{A}$. Therefore $S^n \subset W(S)$ and $S/W(S)$ is nilpotent. It follows that R is nilpotent. Indeed let $f = 0$ be an idempotent of R. Then $f = (r,u,r)$ for some idempotent $u \in S \setminus W(S)$, a contradiction. Thus by (b), $T/W(T)$ is nilpotent. Let e be an idempotent of T. Then $e \in W(T)$ and therefore $eTe = e(eTe)e \subset eW(T)e = \{e, 0\}$. Thus T is locally idempotent and commutative.

3. Varieties described by their finite codes

Let V be a variety of semigroups (resp. monoids) and let $\mathscr{V}$ be the corresponding variety of languages. By definition $\mathscr{V}$ is *described* by a class $\mathscr{C}$ of codes if $\mathscr{V}$ is the smallest variety which contains the language of the form C^+ (resp. C^*) where $C \in \mathscr{C}$. Similarly, $\mathscr{V}$ is described by its finite prefix codes, if $\mathscr{V}$ is the smallest variety which contains all finite prefix codes P such that $S(P^+)$ (resp. $M(P^*)$) is in V.

The main result of this section solves a conjecture of [13].

Theorem 3.1. *Every $*$-variety closed under product is described by its finite prefix codes.*

Proof. Let $\mathscr{V}$ be a $*$-variety and let V be the corresponding variety of monoids. Let W be the variety of monoids generated by all monoids of V of the form $M(P^*)$ for some finite prefix code P. Clearly $W \subset V$. Conversely let $M \in V$. By Theorem 2.6, there exists a finite prefix code P such that M divides $M(P^*)$ and such that there exists an aperiodic relational morphism. $\psi : M(P^*) \to M^1 = M$. Thus $M(P^*) \in A^{-1}V$ and since $\mathscr{V}$ is closed under product, $A^{-1}V = V$ by Straubing's theorem. Therefore $M(P^*) \in V$ and hence $M(P^*) \in W$ by definition. Now since M divides $M(P^*)$, $M \in W$ and thus $V = W$. It follows by Eilenberg's theorem that $\mathscr{V}$ is described by its finite prefix codes. □

The corresponding result for $+$-varieties is more involved.

Theorem 3.2. *Let $\mathscr{V}$ be a $+$-variety closed under product and let V be the corresponding variety of semigroups. If V is monoidal, or if $UV = V$, then $\mathscr{V}$* is described by its finite prefix codes.

The proof is the same except that we use Theorems 2.8 and 2.9 instead of Theorem 2.6.

It is an open problem to know if Theorem 3.2 still holds without the conditions 'monoidal' or '$UV = V$'. In particular, we don't know if the condition $V = A^{-1}V$ implies that V is monoidal.

Here are some explicit examples of applications of Theorem 3.2. Let V_n be the sequence of varieties of semigroups defined inductively as follows:

$$V_0 = A,$$

$$V_{n+1} = V_n * G * A.$$

The Krohn–Rhodes theorem (see [4]) implies that $\bigcup_{n \geq 0} V_n = S$, the variety of all semigroups. Given a semigroup S the smallest integer n such that $S \in V_n$ is called the *complexity* of S. It is known that $V_{n+1} \setminus V_n \neq \emptyset$ for all $n \geq 0$. Furthermore, for every $n \geq 0$, $A^{-1}V_n = V_n$ and V_n is monoidal [26]. Therefore:

Theorem 3.3. *The +-varieties corresponding to semigroups of complexity $\leq n$ are described by their finite prefix codes.*

More generally we have the following result.

Theorem 3.4. *Let $H_1, H_2, \ldots, H_n$ be a sequence of varieties of groups and let $\mathscr{V}$ be the +-variety corresponding to the variety of semigroups $V = A * H_1 * A \cdots * A * H_n * A$. Then $\mathscr{V}$* is described by its finite prefix codes.

Proof. For this proof we will need some results of [26]. First of all, the proof of Proposition 2.1 of [26, p. 321] shows that V is a monoidal variety. Next we show that $A^{-1}V = V$ by induction on n. This is trivial for $n = 0$. In the general case it is sufficient to show that if $\psi : S \to T$ is a surjective aperiodic morphism such that $T \in V$, then $S \in V$. By (10.5) of [26, p. 364] we have $S < \hat{S} < B \circ \hat{T}$ where $B \in A$ and $\hat{S}$ denotes the Rhodes expansion of S [26, p. 361]. Now since $T \in V$, $T < A * (H * R)$ where $R \in V' = A * H_2 * A * \cdots * H_n * A$, $H \in H_1$ and $A \in A$. Now by (10.6) of [26, p. 364], $\hat{T} < A' * (H \hat{*} R)$ for some $A' \in A$ and by (13.1) of [26, p. 376], $(H \hat{*} R) < A'' * (H' * \hat{R})$ where H' is a direct product of copies of H and $A'' \in A$. Now by (9.4) of [26, p. 362] there exists an aperiodic morphism $\nu : \hat{R} \to R$. Therefore $\hat{R} \in A^{-1}V'$ and by the induction hypothesis $A^{-1}V' = V'$. It follows that $(H \hat{*} R) \in A * H_1 * V'$ and thus $\hat{T} \in A * (A * H_1 * V') = A * H_1 * V' = V$ and finally $S \in V$ since $S < B \circ \hat{T}$. Theorem 3.4 now follows from Theorem 3.2. □

Other examples, already given in [13] include the variety $\bar{H}$ of all semigroups whose groups are in a given variety of groups H and, in particular, the variety S of all semigroups and the variety A of aperiodic semigroups.

Finally let ***Inv*** be the variety of semigroups generated by all inverse semigroups. The following result was proved in [11] by the same methods.

Theorem 3.5. *The +-variety $\mathscr{G}_{\mathscr{W}}$ corresponding to **Inv** is described by its finite biprefix codes.*

We conclude this section with two negative results. The first result concerns the variety ***DA*** of all semigroups whose regular $\mathscr{J}$-classes are idempotent semigroups. Let us recall a useful property of a semigroup S in ***DA*** [8]. If e is an idempotent of S and if $s_1, \ldots, s_n$ are elements of S such that $e \leq_{\mathscr{J}} s_1, \ldots, e \leq_{\mathscr{J}} s_n$, then $es_1 \cdots s_n e = e$. Then we have:

Theorem 3.6. *Let $C \subset A^+$ be a finite code such that $S(C^+) \in \mathbf{DA}$. Then $C \subset A$. Therefore the +-variety corresponding to **DA** is not described by its finite codes.*

Proof. Let $S = S(C^+)$ and let $\pi : A^+ \to S$ be the syntactic morphism. For each $u \in C$ there exists $n > 0$ such that $(u\pi)^n = e$ is idempotent. Let a be a letter of the word u.

Then $e \leq_{\mathcal{J}} a\pi$ and since $S \in \boldsymbol{DA}$ it follows that $(u^n a^m u^n)\pi = e$ for all $m > 0$. Therefore for all $m > 0$, $u^n a^m u^n \in C^+ \pi\pi^{-1} = C^+$. If we choose m greater than twice the maximum length of the words in C, the decomposition of the word $u^n a^m u^n$ over C contains a factor of the form a^k for some $k > 0$. Moreover, since $S(C^+) \in \boldsymbol{DA}$, $S(C^+)$ is aperiodic and thus so is $S(C^+ \cap a^+) = S((a^k)^+)$. Since $S((a^k)^+)$ contains the cyclic group Z_k, it follows that $k = 1$. Therefore for each letter a occurring in a word of C we have $a \in C$. It follows that $C \subset A$ since C is a code. □

The second negative result concerns the variety $\boldsymbol{LR}$ of all semigroups S such that for all idempotent $e \in S$, eSe is $\mathcal{R}$-trivial.

Theorem 3.7. *Let P be a finite prefix code such that $S(P^+) \in \boldsymbol{LR}$. Then $S(P^+)$ is locally idempotent and commutative and P is very pure. In particular, the $+$-variety corresponding to $\boldsymbol{LR}$ is not described by its finite prefix codes.*

Proof. Let $\mathcal{A} = (Q, A)$ be the minimal automaton of P^+. This automaton defines a transformation semigroup $X = (Q, S)$ where $S = S(P^+)$ is the syntactic semigroup of P^+. Moreover, since P is a prefix code, it is known that X is transitive, that is, for all states $q_1, q_2 \in Q$, there exists $s \in S$ such that $q_1 s = q_2$. Let e be an idempotent of S and let q_1, q_2 be two states fixed by e. We claim that $q_1 = q_2$. Indeed, since X is transitive, there exist $s, t \in s$ such that $q_1 s = q_2$ and $q_2 t = q_1$. It follows that $q_1(ese) = q_1 se = q_2 e = q_2$ and similarly $q_2(qte) = q_1$. Choose $n > 0$ such that $(esete)^n$ is idempotent. Then we have $q_1(esete)^n = q_1$ and $q_1(esete)^n se = q_2$. But $(esete)^n \mathcal{R} (esete)^n se$ and since eSe is $\mathcal{R}$-trivial it follows that $(esete)^n = (esete)^n se$. Thus $q_1 = q_2$, proving the claim. Therefore $Qe = \{qe \mid q \in Q\}$ is a singleton and every idempotent of S has rank ≤ 1 in $\mathcal{A}$. It follows [3] that P is very pure and that $S(P^+)$ is locally idempotent and commutative. □

Following Brzozowski [4], a language $L \subset A^+$ has *dot-depth one* if it is in the boolean algebra generated by languages of the form uA^*, A^*v or $A^*u_1A^*u_2 \cdots A^*u_nA^*$ when $n > 0$ and $u, v, u_1, \ldots, u_n \in A^+$. It is known that the languages of dot-depth one form a $+$-variety. The corresponding variety of semigroups $\boldsymbol{B}_1$ has been recently characterized by Knast [5].

Corollary 3.8. *The variety of languages of dot-depth one is not described by its finite prefix codes.*

Proof. It is known [5] that if $S \in \boldsymbol{B}_1$, then for every idempotent $e \in S$, eSe is $\mathcal{J}$-trivial. In particular $\boldsymbol{B}_1 \subset \boldsymbol{LR}$. Consequently, if P is a finite prefix code such that P^+ has dot-depth one, then P is very pure. Now the variety described by finite very pure prefix codes is the variety of locally testable languages [3], and this variety is strictly contained in in the variety of languages of dot-depth one. □

4. Languages and product of varieties

As we pointed out in the introduction the product of two varieties of semigroups or monoids is one of the most important operations on varieties. However, little is known in general about this operation. The aim of this section is to describe the operation on languages corresponding to the operation $V \mapsto V * W$ for some suitable choices of W. Analogous results for $W = LI$ have been obtained recently by Straubing [24]. We first need a study of the operation $L \mapsto aL$ where L is a language of A^* and a is a letter. More precisely, given a monoid recognizing L we want to describe a monoid recognizing aL.

Proposition 4.1. *Let $L \subset A^*$ be a language recognized by a monoid M and let a be a letter. Then aL is recognized by $(M \times U_1) \circ U_1$.*

Proof. We give a self-contained proof although it follows from the general results of [16]. Since L is recognized by M, there exists a morphism $\gamma : A^* \to M$ and a subset P of M such that $L = P\gamma^{-1}$. Define a morphism $\psi : A^* \to (M \times U_1) \circ U_1$ by setting, for all letters $b \in A$

$$b\psi = (f, 0) \quad \text{where } 1f = (1, 1) \text{ and } 0f = (a\gamma, 1) \quad \text{if } b = a,$$

$$b\psi = (f, 0) \quad \text{where } 1f = (1, 0) \text{ and } 0f = (b\gamma, 1) \quad \text{if } b \neq a.$$

Let Q be the subset of $(M \times U_1) \circ U_1$ defined by

$$Q = \{(f, 0) \mid 1f \in P \times \{1\}\}.$$

Then we have $Q\psi^{-1} = \{u \in A^* \mid u\psi \in Q\}$. Set $u = a_1 \cdots a_n$ and $a_i\psi = (f_i, 0)$. Then we have $(a_1 \cdots a_n)\psi = (f, 0)$ where

$$1f = 1f_1 0f_2 \cdots 0f_n = \begin{cases} ((a_2 \cdots a_n)\gamma, 0) & \text{if } a_1 \neq a, \\ ((a_2 \cdots a_n)\gamma, 1) & \text{if } a_1 = a. \end{cases}$$

Therefore $Q\psi^{-1} = \{a_1 \cdots a_n \in A^* \mid a_1 = a \text{ and } (a_2 \cdots a_n)\gamma \in P\} = aL$. Thus aL is recognized by $(M \times U_1) \circ U_1$. □

Let W be a variety of monoids and let $\mathscr{W}$ be the corresponding variety of languages. A morphism $\alpha : A^* \to B^*$ is called a (prefix) coding if it is injective and if $A\alpha$ is a (prefix) code. It is called a (prefix) W-coding if $(A\alpha)^* \in B^*\mathscr{W}$. In particular, an A-coding is called a pure coding because in this case $A\alpha$ is a pure code. Then we have:

Proposition 4.2. *Let W be a variety of monoids such that $A^{-1}W = W$ and let $\alpha : A^* \to B^*$ be a prefix W-coding. Then there exists a monoid $N \in W$ such that if $L \subset A^*$ is recognized by M, $L\alpha$ is recognized by $M \circ N$.*

Proof. It is proved in [16] that $L\alpha$ is recognized by $M \circ N$ where N is the so-called 'petal monoid' of the code $A\alpha$. Moreover, it is shown in [9] that there exists an aperiodic morphism from N onto the syntactic monoid of $(A\alpha)^*$. Since α is a $\mathbf{W}$-coding, $M((A\alpha)^*) \in \mathbf{W}$ and thus $N \in \mathbf{W}$ since $A^{-1}\mathbf{W} = \mathbf{W}$.

Corollary 4.3. *Let $\alpha : A^* \to B^*$ be a pure prefix coding. Then there exists an aperiodic monoid N such that if $L \subset A^*$ is recognized by M, $L\alpha$ is recognized by $M \circ N$.*

We are now ready to state the main result of this section.

Theorem 4.4. *Let $\mathbf{V}$ and $\mathbf{W}$ be varieties of monoids and let $\mathscr{V}$ and $\mathscr{W}$ be the corresponding varieties of languages. If $\mathscr{W}$ is closed under product, that is if $A^{-1}\mathbf{W} = \mathbf{W}$, then the variety of languages corresponding to $\mathbf{V} * \mathbf{W}$ is the smallest variety $\mathscr{U}$ satisfying the following conditions*

(1) *For every alphabet A, $A^*\mathscr{U}$ contains the languages of the form $L\alpha$ where $\alpha : B^* \to A^*$ is a prefix $\mathbf{W}$-coding and $L \in B^*\mathscr{V}$.*

(2) *For every alphabet A, and for all $a \in A$, $L \in A^*\mathscr{U}$ implies $aL \in A^*\mathscr{U}$.*

Proof. Let $\mathscr{U}$ be the smallest variety of languages satisfying (1) and (2) and let $\mathscr{X}$ be the variety corresponding to $\mathbf{V} * \mathbf{W}$. To show that $\mathscr{U} \subset \mathscr{X}$ it is sufficient to prove that $\mathscr{X}$ satisfies (1) and (2).

Let $\alpha : B^* \to A^*$ be a prefix $\mathbf{W}$-coding and let $L \in B^*\mathscr{U}$. Then $M(L) \in \mathbf{V}$ and by Proposition 4.2, $M(L\alpha) \in \mathbf{V} * \mathbf{W}$. Thus $L \in A^*\mathscr{X}$. Let now $L \in A^*\mathscr{X}$. Then Proposition 4.1 shows that $M(aL) < (M(L) \times U_1) \circ U_1$ and therefore $M(aL) \in (\mathbf{V} * \mathbf{W}) * \mathbf{R}) * \mathbf{R} = \mathbf{V} * \mathbf{W} * \mathbf{R}$ since $\mathbf{R} * \mathbf{R} = \mathbf{R}$. We claim that $\mathbf{W} * \mathbf{R} = \mathbf{W}$. Indeed it is sufficient to show that any semidirect product of the form $M * U_1$ where $M \in \mathbf{W}$ is also in $\mathbf{W}$. By a result of [20, p. 164], there exists an aperiodic morphism $\psi : M * U_1 \to M$ and thus $M \in A^{-1}\mathbf{W} = \mathbf{W}$. Therefore the claim holds and $M(aL) \in \mathbf{V} * \mathbf{W}$. Consequently $aL \in A^*\mathscr{X}$ and $\mathscr{X}$ satisfies (1) and (2).

The opposite inclusion $\mathscr{X} \subset \mathscr{U}$ is more difficult to establish. The first step is the following lemma.

Lemma 4.5. *$\mathscr{W}$ is contained in $\mathscr{U}$.*

Let $P \subset A^*$ be a finite prefix code such that $P^* \in A^*\mathscr{W}$ and let $\alpha : B^* \to A^*$ be an injective morphism such that $B\alpha = P$. Then α is a prefix $\mathbf{W}$-coding and since $B^* \in B^*\mathscr{V}$, $B^*\alpha = P^* \in A^*\mathscr{U}$. Thus $\mathscr{U}$ contains all languages of the form P^* where P is a finite prefix code such that $P^* \in A^*\mathscr{W}$. But $\mathscr{W}$ is closed under product and hence described by its finite prefix codes by Theorem 3.1. Thus $\mathscr{W} \subset \mathscr{U}$.

Let now $K \in A^*\mathscr{X}$. Then there exist a semidirect product $M * N$ where $M \in \mathbf{V}$ and $N \in \mathbf{W}$ and a morphism $\gamma : A^* \to M * N$ which recognizes K. Let $\pi : M * N \to N$ be the natural projection and let $\psi : A^* \to N$ be the morphism $\psi = \gamma\pi$. Then it is proved in [15] that K is union of languages of the form $X \cap Y\sigma^{-1}$ where $X \subset A^*$ is recognized

by N, $Y \subset B^*$ is recognized by M, $B = N \times A$ and where σ is the sequential function $A^* \to B^*$ defined by

$$1\sigma = 1,$$

$$(a_1 \cdots a_n)\sigma = (1, a_1)(a_1\psi, a_2) \cdots ((a_1 \cdots a_{n-1})\psi, a_n).$$

Now since $N \in \mathbf{W}$, $X \in A^*\mathbb{W}$ and thus $X \in A^*\mathbb{W}$ by the previous lemma. Moreover, $M \in \mathbf{V}$ and thus $Y \in B^*\mathbb{W}$. Therefore since a variety is closed under boolean operations it is sufficient to show that if $Y \in B^*\mathbb{V}$, then $Y\sigma^{-1} \in A^*\mathbb{W}$.

The next step consists in decomposing the transduction σ^{-1}. Let $N = \{z_1, \ldots, z_n\}$ where $z_1 = 1$ is the unit of N. Define an action of A on $\{1, \ldots, n\}$ by setting for $1 \leq i \leq n$, $i \cdot a = j$ if $z_j = z_i(a\psi)$. Thus we have an automaton $\mathscr{A} = (\{1, \ldots, n\}, A)$ whose transition monoid is N. Let c be a new letter and let $C = A \cup \{c\}$. Then the results of the previous section show that the morphism $\alpha : B^* \to C^*$ defined by

$$(z_k, a)\alpha = c^{k\tau} a c^{n\tau - k a \tau}$$

where $k\tau = 2^k - 2$ is a prefix $\mathbf{W}$-coding.

Now define a morphism $\kappa : A^* \to C^*$ by setting $a\kappa = c^{n\tau}a$. Then we obtain the desired decomposition of σ^{-1}.

Lemma 4.6. *For every language $Y \subset B^*$, we have*

$$Y\sigma^{-1} = (c^{n\tau}((y\alpha)(c^*)^{-1}))\kappa^{-1}.$$

Proof. Let $u = (z_{i_1}, a_1) \cdots (z_{i_1}, a_r)$ be a word of B^*. Then

$$u\alpha = c^{i_1\tau} a_1 c^{n\tau - i_1 a_1 \tau} c^{i_2 \tau} a_2 \cdots c^{i_r \tau} a_r c^{n\tau - i_r a_r \tau}.$$

Thus by setting $L = c^{n\tau}((u\alpha)(c^*)^{-1})$, we have

$$L = \{c^{n\tau + i_1\tau} a_1 c^{n\tau - i_1 a_1 \tau} c^{i_2\tau} a_2 \cdots c^{i_r \tau} a_r c^j \mid 0 \leq j \leq n\tau - i_r a_r \tau\}.$$

Therefore $L\kappa^{-1}$ is empty except in the case where $i_1\tau = 0$, $i_1 a_1 \tau = i_2 \tau, \ldots, i_{r-1} a_{r-1} \tau = i_r \tau$, that is, $i_1 = 1$, $i_2 = i_1 a_1, \ldots, i_r = i_{r-1} a_{r-1}$ and in this last case we have $L\kappa^{-1} = a_1 \cdots a_r$.

On the other hand, $u\sigma^{-1}$ is empty except in the case where $u = (1, a_1)(a_1\psi, a_2) \cdots ((a_1 \cdots a_{r-1})\psi, a_r)$, that is if $i_1 = 1$, $i_2 = i_1(a_1\psi), \ldots, i_r = i_{r-1}(a_{r-1}(a_{r-1}\psi)$ or equivalently $i_1 = 1$, $i_2 = i_1 a_1, \ldots, i_r = i_{r-1} a_{r-1}$. In this last case $u\sigma^{-1} = a_1 \cdots a_r = L\kappa^{-1}$ which proves the lemma. □

We can now conclude the proof of Theorem 4.4. Since $Y \in B^*\mathbb{V}$ and since α is a prefix $\mathbf{W}$-coding, $Y \in C^*\mathbb{W}$ by condition (1). Consequently $(Y\alpha)(c^*)^{-1} \in C^*\mathbb{W}$ because a variety of languages is closed under quotient. Now condition (2) implies that $L = c^{n\tau}(Y\alpha)(c^*)^{-1} \in C^*\mathbb{W}$. Finally, $LK^{-1} = Y\sigma^{-1} \in A^*\mathbb{W}$ since a variety of languages is closed under inverse morphism. Therefore $\mathbb{V} \subset \mathbb{W}$ and hence $\mathbb{V} = \mathbb{W}$. □

In the case $W=A$ we obtain the following corollary.

Corollary 4.7. *Let V be a variety of monoids and let $\mathscr{V}$ be the corresponding variety of languages. Then the variety corresponding to $V*A$ is the smallest variety containing $\mathscr{V}$ and closed under prefix pure coding and under left concatenation with letters.*

Proof. In view of Theorem 4.3, we only have to verify that the variety $\mathscr{W}$ corresponding to $V*A$ is closed under prefix pure coding. Indeed if $L\in B^*\mathscr{W}$ and if $\alpha: A^*\to B^*$ is an injective morphism such that $A\alpha$ is a prefix pure code, then $M(L\alpha)\in V*A$ and by Corollary 4.3, $M(L\alpha)\in V*A*A=V*A$ since $A*A=A$. □

We will now give a more precise description of the variety of languages $\mathscr{V}_1$ corresponding to V_1, the variety of monoids of complexity ≤ 1. Recall that a language is a group language iff its syntactic monoid is a group. Then we have

Theorem 4.8. *$\mathscr{V}_1$ is the smallest variety of languages containing the group languages and which is closed under product and under prefix pure coding.*

Proof. Since $G\subset V_1$, $\mathscr{V}_1$ contains the group languages and since $A^{-1}V_1=V_1$ [26], $\mathscr{V}_1$ is closed under product by Straubing's theorem. Moreover, Proposition 4.2 shows that $\mathscr{V}_1$ is closed under pure prefix coding. Thus $\mathscr{V}_1$ contains the smallest $*$-variety $\mathscr{V}$ containing the group languages and which is closed under product and under prefix pure coding. Conversely, the variety of monoids V corresponding to $\mathscr{V}$ contains G and hence $A^{-1}G=A*G$ since $\mathscr{V}$ is closed under product. Thus $\mathscr{V}$ contains the letters, since for each letter $a\in A$, $M(\{a\})$ is aperiodic and therefore $\mathscr{V}$ is closed under left concatenation with letters. Thus by Corollary 4.7, $\mathscr{V}$ contains $\mathscr{V}_1$ and hence $\mathscr{V}=\mathscr{V}_1$. □

One can improve the previous result by replacing the group languages by an explicitly given family of languages. As is well known the symmetric group on n elements S_n is generated, for $n\geq 2$, by the two permutations σ and τ where $\sigma=(1\cdots n)$ and $\tau=(12)$. Thus let $\Sigma=\{\sigma,\tau\}$ and let $\mathscr{A}_n=(\{1,\ldots,n\},\Sigma)$ be the automaton defined by the permutations σ and τ. The construction given in Section 2 shows that if $A=\{a,b,c\}$, then the code

$$C_n=\{a^{2^i-2}ba^{2^n-2^{i+1}} \mid 1\leq i\leq n-1\}\cup\{a^{2^n-2}ba^{2^n-2}\}$$
$$\cup\{ca^{2^n-4},a^2ca^{2^n-2}\}\cup\{a^{2^i-2}ca^{2^n-2^i} \mid 3\leq i\leq n\}$$

satisfies the two conditions

(1) $S_n<M(C_n^*)$.

(2) There exists an aperiodic relational morphism $M(C_n^*)\to S_n$.

Therefore, $M(C_n^*)\in A^{-1}G=A*G$.

Let $\alpha: \{a,b,c\}^* \to \{a,b\}^*$ be the prefix coding defined by

$$a\alpha = a, \qquad b\alpha = ba, \qquad c\alpha = b^2a.$$

Then letting $P_n = C_n\alpha$ we have

$$P_n = \{a^{2^i-2}ba^{2^n-2^{i+1}+1} \mid 1 \le i \le n-1\} \cup \{a^{2^n-2}ba^{2^n-1}\}$$
$$\cup \{b^2a^{2^n-3}, a^2b^2a^{2^n-1}\} \cup \{a^{2^i-2}b^2a^{2^n-2^i+1} \mid 3 \le i \le n\}.$$

Since $\{a, ba, b^2a\}$ is a prefix pure code, Proposition 4.2 shows that $M(P_n^*) \in (A * V) * A = V_1$. Furthermore, since $P_n^*\alpha^{-1} = C_n^*$ we have $S_n < M(C_n^*) < M(P_n^*)$. Therefore

Corollary 4.9. *$\mathscr{V}_1$ is the smallest variety of languages containing the languages P_n^* for $n \ge 2$ and which is closed under product and under prefix pure coding.*

Proof. Since $M(P_n^*) \in V_1$, $\mathscr{V}_1$ contains the languages P_n^* for $n \ge 2$. Moreover, by Theorem 4.8, $\mathscr{V}_1$ is closed under product and under prefix pure coding. Conversely, if a variety $\mathscr{V}$ contains P_n^* for $n \ge 2$, then $\mathscr{V}$ contains the group languages since for each group G there exists $n > 0$ such that $G < S_n < M(P_n^*)$. Now if $\mathscr{V}$ is closed under product and under prefix pure coding, it contains $\mathscr{V}_1$ by Theorem 4.8. □

Here is another application of Corollary 4.7. Let ***Gcom*** be the variety of all commutative groups. A description of the variety of languages $\mathscr{V}$ corresponding to $A * \mathbf{Gcom} * A$ was given in [21].

For every language $L \subset A^*$, let $(L, r, n) = \{u \in A^* \mid \mathrm{Card}(Lu^{-1}) \equiv r \bmod n\}$. Then for each alphabet A, $A^*\mathscr{V}$ is the smallest boolean algebra closed under product and containing all star-free languages and all the languages of the form (L, r, n) where L is star-free and $0 \le r < n$ are integers. Here is a different description of this variety.

Theorem 4.10. *Let $\mathscr{V}$ be the variety of languages corresponding to $A * \mathbf{Gcom} * A$. Then $\mathscr{V}$ is the smallest variety of languages such that, for all alphabets A and for all $n > 0$, $(A^n)^* \in A^*\mathscr{V}$ and which is closed under product and under prefix coding.*

Proof. Let $\mathscr{W}$ be the smallest variety such that $(A^n)^* \in A^*\mathscr{W}$ for all $n > 0$ and for all alphabets A and which is closed under product and under pure prefix coding. We first show $\mathscr{W} \subset \mathscr{V}$. Indeed, since $M(A^n)^* = \mathbb{Z}_n$ is a commutative group, we have $(A^n)^* \in A^*\mathscr{V}$ for all $n > 0$. Moreover, $\mathscr{V}$ is closed under prefix pure coding by Corollary 4.7 and $\mathscr{V}$ is closed under product by Theorem 6.2 of [26].

Conversely let W be the variety of monoids corresponding to $\mathscr{W}$. Since $M(A^n)^* = \mathbb{Z}_n$, W contains all cyclic groups and hence all commutative groups. Since $\mathscr{W}$ is closed under product $A^{-1}\mathbf{Gcom} = A * \mathbf{Gcom}$ is contained in W by Straubing's theorem and finally Corollary 4.7 shows that W contains $A * \mathbf{Gcom} * A$. Thus $W = A * \mathbf{Gcom} * A$ and this concludes the proof. □

References

[1] J. Berstel and D. Perrin, Circular codes, in: Combinatorics on Words, Progress and Perspectives (Academic Press, New York, 1983) 133–165.

[2] J. Berstel, and D. Perrin, Theory of Codes, to appear.

[3] A. de Luca, On some properties of the syntactic semigroup of very pure subsemigroups, RAIRO Informatique Théorique 14 (1980) 39–56.

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

[5] R. Knast, A semigroup characterization of dot-depth one languages, RAIRO Informatique Théorique 17 (1983) 321–330.

[6] G. Lallement, Regular semigroups with $D=R$ as syntactic monoids of finite prefix codes, Theor. Comput. Sci. 3 (1977) 35–49.

[7] G. Lallement, Cyclotomic polynomics and unions of groups, Discrete Math. 24 (1978) 19–36.

[8] G. Lallement, Semigroups and Combinatorial Applications (Interscience, New York, 1979).

[9] E. Le Rest and M. Le Rest, Sur le calcul du monoide syntactique d'un sous-monoïde finiment engendré, Semigroup Forum 21 (1980) 173–185.

[10] E. Le Rest and S.W. Margolis, On the group complexity of a finite language, in: Proc. 10th ICALP, Lecture Notes in Computer Science 154 (Springer, Berlin, 1983) 433–444.

[11] S.W. Margolis, On the syntactic transformation semigroup of a language generated by a finite biprefix code, Theoret. Comput. Sci. 21 (1982) 225–230.

[12] J.-F. Perrot, On the theory of syntactic monoids for rational languages, in: Fundamentals of Computation Theory, Lecture Notes in Computer Science, 56 (Springer, Berlin, 1979) 540–558.

[13] J.-E. Pin, On varieties of rational languages and variable-length codes, J. Pure Appl. Algebra 23 (1982) 169–196.

[14] J.-E. Pin, Langages reconnaissables et codage préfixe pur, in: Proc. 8th ICALP, Lecture Notes in Computer Science 104 (Springer, Berlin, 1981) 78–90.

[15] J.-E. Pin, Hierarchies de concaténation, PAIRO Informatique Théorique 18 (1984) 23–46.

[16] J.-E. Pin and J. Sakarovitch, Operations and transductions that preserve rationality, in: Proc. 6th GI Conference, Lecture Notes in Computer Science 145 (Springer, Berlin, 1983) 277–288,

[17] J.-E. Pin, Variétés de Langages Formels (Masson, Paris, 1984).

[18] A. Restivo, Codes an aperiodic languages, in: Proc. 1st GI Conference, Lecture Notes in Computer Science 67 (Springer, Berlin, 1979) 260–265.

[19] A. Restivo, On a question of McNaughton and Papert, Inform. and Control 25 (1974) 93–101.

[20] J. Rhodes and B. Tilson, Local complexity of finite semigroups, in: Algebra, Topology and Category Theory, a collection of papers in Honor of Samuel Eilenberg (Academic Press, New York, 1976) 149–168.

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

[22] H. Straubing, Aperiodic homomorphisms and the concatenation product of recognizable sets, J. Pure Appl. Algebra 15 (1979) 319–327.

[23] H. Straubing, Relational morphisms and operations on recognizable sets, RAIRO Informatique Théorique 15 (1981) 137–150.

[24] H. Straubing, Finite semigroup varieties of the form $V*D$, J. Pure Appl. Algebra 36 (1985) 53–94.

[25] B. Tilson, Depth decomposition theorem, Chapter 11 in [2].

[26] B. Tilson, Complexity of Semigroups and Morphisms, Chapter 12 in [4].