MODULAR AND THRESHOLD SUBWORD COUNTING AND MATRIX REPRESENTATIONS OF FINITE MONOIDS

JORGE ALMEIDA, STUART MARGOLIS, BENJAMIN STEINBERG, AND MIKHAIL VOLKOV

1. Background and motivation

Recall that a word u over a finite alphabet Σ is said to be a *subword* of a word $v \in \Sigma^*$ if, for some $n \geq 1$, there exist words $u_1, \dots, u_n, v_0, v_1, \dots, v_n \in \Sigma^*$ such that $u = u_1 u_2 \cdots u_n$ and

$$v = v_0 u_1 v_1 u_2 v_2 \cdots u_n v_n. \tag{1.1}$$

The subword relation reveals interesting combinatorial properties and plays a prominent role in formal language theory. For instance, recall that languages consisting of all words over Σ having a given word $u \in \Sigma^*$ as a subword serve as a generating system for the Boolean algebra of so-called *piecewise testable* languages. It was a deep study of combinatorics of the subword relation that led Simon [20,21] to his elegant algebraic characterization of piecewise testable languages. Further, the natural idea to put certain rational constraints on the factors $v_0, v_1, \dots, v_n$ that may appear in a decomposition of the form (1.1) gave rise to the useful notion of a marked product of languages studied from the algebraic viewpoint by Schützenberger [18], Reutenauer [10], Straubing [23], Simon [22], amongst others.

Yet another natural idea is to count how many times a word $v \in \Sigma^*$ contains a given word u as a subword, that is, to count different decompositions of the form (1.1). Clearly, if one wants to stay within the realm of rational languages, one can only count up to a certain threshold and/or modulo a certain number. For instance, one may consider Boolean combinations of languages consisting of all words over Σ having t modulo p occurrences of a given word $u \in \Sigma^*$ (where p is a given prime number). This class of languages also admits a nice algebraic characterization, see [5, Sections VIII.9 and VIII.10] and also [25]. Combining modular counting with rational constraints led to the idea of marked products with modular counters explored, in particular, by Weil [27] and Peladeau [7].

The most natural version of threshold counting is formalized via the notion of an unambiguous marked product in which one considers words $v \in \Sigma^*$

Date: Version of June 25, 2005.

1991 *Mathematics Subject Classification.* 68R15, 68Q70, 20M35.

Key words and phrases. subword relation, marked product, modular counting, finite monoid, matrix representation.

having exactly one decomposition (1.1) with a given subword u and given rational constraints on the factors $v_0, v_1, \ldots, v_n$. Such unambiguous marked products have been investigated by Schützenberger [19], Pin [8], Pin, Straubing, and Thérien [9], amongst others.

Many known facts on marked products rely on rather difficult techniques from finite monoid theory, namely, on bilateral semidirect product decomposition results of Rhodes *et al.* [14, 16]. These results are proved using Rhodes's classification of maximal proper surmorphisms [11, 15, 6] via case-by-case analysis of the kernel categories of such maps [14, 16]. The aim of the present paper is to give easier and – we hope – more conceptual proofs of several crucial facts about marked products by using matrix representations of finite monoids as a main tool. In particular, we are able to prove the results of Peladeau and Weil in one step, without any case-by-case analysis and without using the machinery of categories. Rather we adapt Simon's analysis of the combinatorics of multiplying upper triangular matrices [22] from the case of Schützenberger products to block upper triangular matrices. We failed to obtain such a purely combinatorial argument for the case of unambiguous products; we still need to use a lemma on kernel categories. Nevertheless we have succeeded in avoiding the decomposition results and case-by-case analysis.

In Section 2 we collect a few facts from the theory of matrix representations of finite monoids. Some of these facts are new; their proofs can be found in the forthcoming paper by the authors [3]. The announced applications to marked products with modular counters and unambiguous marked products are presented in Section 3.

2. Results from Representation Theory

The reader is referred to [4, Chapter 5] and [17] for the basic results of monoid representation theory. All monoids in this paper are assumed to be finite except for the monoid of matrices over an infinite field.

Let M be a monoid and K a field. A (matrix) *representation* of M over K of *degree* n is a homomorphism $\rho : M \to M_n(K)$, where $M_n(K)$ is the monoid of all $n \times n$ matrices over K. Set $V = K^n$. Then a subspace W of V is said to be M-*invariant* if $(M\rho)W \subseteq W$. The representation ρ is said to be *irreducible* if the only M-invariant subspaces are $\{0\}$ and V.

We denote by $K[M]$ the *monoid algebra* of M, that is, the K-algebra with basis M, whose multiplication extends the multiplication of M. Clearly, any representation $\rho : M \to M_n(K)$ uniquely extends to a K-algebra homomorphism $K[M] \to M_n(K)$. This homomorphism defines a $K[M]$-module structure on the space $V = K^n$. The representation ρ is irreducible if and only if the associated $K[M]$-module is simple. Thus, by choosing a composition series of V, considered as a $K[M]$-module, one can choose a basis for V such that $M\rho$ consists of block upper triangular matrices where the monoids formed by the diagonal blocks are images of M under certain irreducible representations. These irreducible blocks are uniquely determined by ρ and are called the *irreducible constituents* of ρ.

The *regular representation of* M is the representation $\rho_M : M \to M_{|M|}(K)$ on the vector space $K[M]$ extending the homomorphism that maps each element $m \in M$ to the left translation $\lambda_m : m' \mapsto mm'$ of the set M. This is a faithful representation (meaning ρ_M is injective). Moreover, every irreducible representation of M is an irreducible constituent of ρ_M.

If M is a monoid and K is a field, then we define the *Rhodes radical* $\mathrm{Rad}_K(M)$ to be the congruence on M associated to the direct sum of all the irreducible representations of M over K. Equivalently, it is the restriction to M of the congruence on $K[M]$ associated to the Jacobson radical. Alternatively, if we consider the regular representation, placed in block upper triangular form, then the Rhodes radical is the congruence associated to the projection to the block diagonal.

Recall that a *pseudovariety* of monoids (semigroups) is a class of finite monoids (semigroups) closed under the formation of finite direct products, submonoids (subsemigroups) and homomorphic images [1,5]. If $\mathbf{V}$ is a pseudovariety of monoids, then $\mathbf{LV}$ denotes the pseudovariety of semigroups S such that, for each idempotent $e \in S$, the monoid eSe belongs to $\mathbf{V}$. Let $\mathbf{I}$ denote the trivial pseudovariety and $\mathbf{G}_p$ denote the pseudovariety of p-groups for p prime. If $\mathbf{V}$ is a pseudovariety of semigroups, a homomorphism $\varphi : M \to N$ of monoids is called a $\mathbf{V}$-*morphism* if, for each idempotent $f \in N$, one has $f\varphi^{-1} \in \mathbf{V}$.

With this notation, Rhodes showed [12, 17] that if K has characteristic 0, then $\mathrm{Rad}_K(M)$ is the largest congruence $\equiv$ on M such that the quotient $\varphi : M \to M/{\equiv}$ is an $\mathbf{LI}$-morphism. The authors have generalized this [3] to show that if K has characteristic $p > 0$ (a prime), then $\mathrm{Rad}_K(M)$ is the largest congruence $\equiv$ on M such that the quotient $\varphi : M \to M/{\equiv}$ is an $\mathbf{LG}_p$-morphism. Two proofs of these results are given in [3]. The first proof uses the Wedderburn theory of finite dimensional algebras; the second proof uses classical semigroup representation theory and follows along the lines of [12, 17]. One of the key algebraic results used in the first proof, and that we shall use later, is the following, whose proof we include to give the flavor of things. We shall use the fact that a semigroup S is locally a group (in $\mathbf{LG}$ for $\mathbf{G}$ the pseudovariety of groups) if and only if it does not contain a copy of the two element semilattice $\{e, f \mid e = e^2 = ef = fe,\ f = f^2\}$; in this case S is a nilpotent extension of a simple semigroup. By $E(S)$ we denote the set of all idempotents of a semigroup S.

Lemma 2.1. *Let $\varphi : A \to B$ be a morphism of K-algebras with* $\ker\varphi$ *nilpotent. Let S be a finite subsemigroup of A. Then if* $\mathrm{char}K = 0$*, respectively p, then $\varphi|_S$ is an* $\mathbf{LI}$*-morphism, respectively* $\mathbf{LG}_p$*-morphism.*

Proof. Without loss of generality, we may assume that S spans A and hence that A is finite dimensional. Let $e_0 \in E(B)$ and $U = e_0\varphi|_S^{-1}$. First we show that U does not contain a copy of the two element semilattice. Indeed, suppose that $e, f \in E(U)$ and $ef = fe = e$. Then

$$(f - e)^2 = f^2 - ef - fe + e^2 = f - e.$$

Since $f - e \in \ker\varphi$, a nilpotent ideal, we conclude $f - e = 0$, that is, $f = e$. As observed before the formulation of the lemma, this means that U is locally a group.

Now let G be a maximal subgroup of U with identity e. Then $g - e \in \ker\varphi$. Since g and e commute, if the characteristic is p, then, for large enough n,

$$0 = (g - e)^{p^n} = g^{p^n} - e$$

and so G is a p-group. Thus $U \in \mathbf{LG}_p$.

If the characteristic is 0, then we observe that $(g - e)^n = 0$ for some n. So by taking the regular representation ρ of G, we see that $g\rho$ is a matrix with minimal polynomial of the form $(x - 1)^n$; that is $g\rho$ is unipotent. A quick consideration of the Jordan canonical form for such $g\rho$ shows that if $g\rho$ is not the identity matrix, then it has infinite order. It follows that $g = e$ and so G is trivial. Thus $U \in \mathbf{LI}$. □

We remark that if A is an algebra of block upper triangular matrices, B is the diagonal block algebra, and φ is the projection to the diagonal block, then the kernel is contained in the algebra of upper triangular matrices with zero diagonal; this algebra is nilpotent and so Lemma 2.1 applies in this context.

We recall that if $\mathbf{V}$ is a pseudovariety of semigroups and $\mathbf{W}$ is a pseudovariety of monoids, then the *Malcev product* $\mathbf{V} ⓜ \mathbf{W}$ is the pseudovariety generated by all monoids M with a $\mathbf{V}$-morphism to a monoid in $\mathbf{W}$. Given our description of the Rhodes radical, it follows from results of Rhodes and Tilson [6, 13, 26] that $M \in \mathbf{LI} ⓜ \mathbf{W}$ if and only if $M/\mathrm{Rad}_{\mathbb{Q}}(M) \in \mathbf{W}$ and $M \in \mathbf{LG}_p ⓜ \mathbf{W}$ if and only if $M/\mathrm{Rad}_{\mathbb{F}_p}(M) \in \mathbf{W}$, where $\mathbb{F}_p$ is the finite field of order p.

3. Applications to Marked Products

In this section we present two applications of representation theory to studying marked products. More can be found in [3].

Recall that Eilenberg established [5, Vol.B, Chap. VII] a correspondence between pseudovarieties of monoids and so-called varieties of languages. If $\mathbf{V}$ is a pseudovariety of monoids and Σ a finite alphabet, then $\mathcal{V}(\Sigma^*)$ denotes the set of all languages over Σ that can be recognized by monoids in $\mathbf{V}$. (Such languages are often referred to as $\mathbf{V}$-*languages*.) The operator $\mathcal{V}$ that assigns each free monoid Σ^* the set $\mathcal{V}(\Sigma^*)$ is said to be *the variety of languages associated to* $\mathbf{V}$. The syntactic monoid [5, loc. cit.] of a rational language L will be denoted M_L. It is known that L is a $\mathbf{V}$-language if and only if $M_L \in \mathbf{V}$.

3.1. **Products with Counter.** Our first application is to prove the results of Peladeau and Weil [7, 27] on products with counter.

Let $L_0, \ldots, L_m \subseteq \Sigma^*$, $a_1, \ldots, a_m \in \Sigma$ and let n be an integer. Then the *marked product with modulo n counter* $L = (L_0a_1L_1 \cdots a_mL_m)_{r,n}$ is the language of all words $w \in \Sigma^*$ with r factorizations modulo n of the form $w = u_0a_1u_1 \cdots a_mu_m$ with each $u_i \in L_i$. One can show that L is rational [27] (see also the proof of Theorem 3.2 below). Using a decomposition result of Rhodes and Tilson [14] (see also [16]) based on case-by-case analysis of kernel categories of maximal proper surmorphisms (see [11, 15, 6]), Weil characterized the closure of a variety $\mathcal{V}$ under marked products with modulo p counter. This required iterated usage of the so-called "block product" principle. But Weil missed that the Boolean algebra generated by $\mathcal{V}(\Sigma^*)$ and marked products with modulo p counters of members $\mathcal{V}(\Sigma^*)$ is already closed under marked products with modulo p counters; this was later observed by Peladeau [7]. The difficulty arises because it is not so clear how to combine marked products with modulo p counters into new marked products with modulo p counters.

We use representation theory to prove the result in one fell swoop. Our approach is inspired by a paper of Simon [22] dealing with marked products and the Schützenberger product of finite monoids.

Lemma 3.1. *Let $\mathbf{V}$ be a pseudovariety of monoids, $\varphi : \Sigma^* \to M$ be a morphism with M finite. Let K be a field of characteristic p and suppose that M can be represented faithfully by block upper triangular matrices over K so that the monoids formed by the diagonal blocks of the matrices in the image of M all belong to $\mathbf{V}$. Let $F \subseteq M$. Then $L = F\varphi^{-1}$ is a Boolean combination of members of $\mathcal{V}(\Sigma^*)$ and of marked products with modulo p counter $(L_0a_1L_1 \cdots a_nL_n)_{r,p}$ with the $L_i \in \mathcal{V}(\Sigma^*)$.*

Proof. Suppose $M \leq M_t(K)$ and $t = t_1 + \cdots + t_k$ is the partition of t giving rise to the block upper triangular form. Let M_i be the monoid formed by the $t_i \times t_i$ matrices over K arising as the i^{th} diagonal blocks of the matrices in the image of M. Given $w \in \Sigma^*$ and $i, j \in \{1, \ldots, k\}$, define $\varphi_{i,j} : \Sigma^* \to M_{t_i,t_j}(K)$ by setting $w\varphi_{i,j}$ to be the $t_i \times t_j$ matrix that is the i, j-block of the block upper triangular form. So in particular $w\varphi_{i,j} = 0$ for $j < i$. Also $\varphi_{i,i}$ is a morphism $\varphi_{i,i} : \Sigma^* \to M_i$ for all i.

First we observe that we may take F to be a singleton $\{u\varphi\}$. For each $1 \leq i \leq j \leq k$, let

$$L_{i,j} = \{w \in \Sigma^* \mid w\varphi_{i,j} = u\varphi_{i,j}\}.$$

Then clearly

$$u\varphi\varphi^{-1} = \bigcap_{1 \leq i \leq j \leq k} L_{i,j}.$$

Since $L_{i,i}$ is recognized by M_i, it suffices to show $L_{i,j}$, where $1 \leq i < j \leq k$, can be written as a Boolean combination of marked products with modulo

p counter of languages recognized by the M_l. Changing notation, it suffices to show that if $1 \leq i < j \leq k$ and $C \in M_{t_i,t_j}(K)$, then

$$L(C) = \{w \in \Sigma^* \mid w\varphi_{i,j} = C\} \tag{3.1}$$

is a Boolean combination of marked products with modulo p counter of languages recognized by the M_i.

The following definitions are inspired by [22], though what Simon terms an "object", we term a "walk". A *walk* from i to j is a sequence

$$\mathfrak{w} = (i_0, m_0, a_1, i_1, m_1, \ldots, a_r, i_r, m_r) \tag{3.2}$$

where $i = i_0 < i_1 < \cdots < i_r = j$, $a_l \in \Sigma$ and $m_l \in M_{i_l}$. There are only finitely many walks. The set of walks will be denoted $\mathfrak{W}$. Given a walk $\mathfrak{w}$, we define its *value* to be

$$\mathsf{v}(\mathfrak{w}) = m_0(a_1\varphi_{i_0,i_1})m_1 \cdots (a_r\varphi_{i_{r-1},i_r})m_r \in M_{t_i,t_j}(K).$$

If $\mathfrak{w}$ is a walk, we define the *language* of $\mathfrak{w}$ to be the marked product

$$L(\mathfrak{w}) = (m_0\varphi^{-1}_{i_0,i_0})a_1(m_1\varphi^{-1}_{i_1,i_1}) \cdots a_r(m_r\varphi^{-1}_{i_r,i_r}).$$

If $w \in \Sigma^*$ and $\mathfrak{w}$ is a walk of the form (3.2), we define $w(\mathfrak{w})$ to be the *multiplicity* of w in $L(\mathfrak{w})$, that is, the number of factorizations $w = u_0a_1u_1 \cdots a_ru_r$ with $u_l\varphi_{i_l,i_l} = m_l$; this number is taken to be 0 if there are no such factorizations. If $0 \leq n < p$, we establish the shorthand

$$L(\mathfrak{w})_{n,p} = \Big((m_0\varphi^{-1}_{i_0,i_0})a_1(m_1\varphi^{-1}_{i_1,i_1}) \cdots (a_rm_r\varphi^{-1}_{i_r,i_r})\Big)_{n,p}.$$

Notice that $L(\mathfrak{w})_{n,p}$ consists of all words w with $w(\mathfrak{w}) \equiv n \bmod p$ and is a marked product with modulo p counter of $\mathcal{V}(\Sigma^*)$ languages.

The following is a variant of [22, Lemma 7].

Claim. *Let $w \in \Sigma^*$. Then*

$$w\varphi_{i,j} = \sum_{\mathfrak{w}\in\mathfrak{W}} w(\mathfrak{w})\mathsf{v}(\mathfrak{w}). \tag{3.3}$$

Proof. Let $w = b_1 \cdots b_r$ be the factorization of w in letters. Then the formula for matrix multiplication gives

$$w\varphi_{i,j} = \sum (b_1\varphi_{i_0,i_1})(b_2\varphi_{i_1,i_2}) \cdots (b_r\varphi_{i_{r-1},i_r}) \tag{3.4}$$

where the sum extends over all i_l such that $i_0 = i$, $i_r = j$ and $i_l \in \{1, \ldots, k\}$ for $0 < l < r$. Since $v\varphi_{l,n} = 0$ for $l > n$, it suffices to consider sequences such that $i = i_0 \leq i_1 \leq \cdots \leq i_r = j$. For such a sequence, we may group together neighboring indices that are equal. Then using that the $\varphi_{n,n}$ are morphisms, we see that each summand in (3.4) is the value of a walk $\mathfrak{w}$ and that $\mathfrak{w}$ appears exactly $w(\mathfrak{w})$ times in the sum. $\square$

To complete the proof, we observe that $L(C)$ (defined in (3.1)) is a Boolean combination of languages of the form $L(\mathfrak{w})_{n,p}$. Let X be the set of all functions $f : \mathfrak{W} \to \{0, \dots, p-1\}$ such that

$$\sum_{\mathfrak{w}\in\mathfrak{W}} f(\mathfrak{w})\mathsf{v}(\mathfrak{w}) = C.$$

It is then immediate from (3.3) and $\mathrm{char}K = p$ that

$$L(C) = \bigcup_{f\in X} \bigcap_{\mathfrak{w}\in\mathfrak{W}} L(\mathfrak{w})_{f(\mathfrak{w}),p}$$

completing the proof. □

Theorem 3.2. *Let $L \subseteq \Sigma^*$ be a rational language, $\mathbf{V}$ be a pseudovariety of monoids and K be a field of characteristic p. Then the following are equivalent.*

(1) $M_L \in \mathbf{LG}_p \ⓜ \mathbf{V}$;
(2) $M_L/\mathrm{Rad}_K(M_L) \in \mathbf{V}$;
(3) *M_L can be faithfully represented by block upper triangular matrices over K so that the monoids formed by the diagonal blocks of the matrices in the image of M_L all belong to $\mathbf{V}$;*
(4) *L is a Boolean combination of members of $\mathcal{V}(\Sigma^*)$ and languages $(L_0a_1L_1 \cdots a_nL_n)_{r,p}$ with the $L_i \in \mathcal{V}(\Sigma^*)$.*

Proof. The equivalence of (1) and (2) follows from the results of [3] cited in Section 2.

For (2) implies (3), take a composition series for the regular representation of M_L over K: it is then in block upper triangular form and, by (2) and the comments from Section 2, the monoids formed by diagonal blocks of matrices in the image of M_L all belong to $\mathbf{V}$.

(3) implies (4) is immediate from Lemma 3.1.

For (4) implies (1), it suffices to deal with a marked product with counter $L = (L_0a_1L_1 \cdots a_nL_n)_{r,p}$. Let $\mathcal{A}_i$ be the minimal deterministic automaton for L_i. Let $\mathcal{A}$ be the non-deterministic automaton obtained from the disjoint union of the $\mathcal{A}_i$ by attaching an edge labelled a_i from each final state of $\mathcal{A}_{i-1}$ to the initial state of $\mathcal{A}_i$. To each letter $a \in \Sigma$, we associate the matrix $a\varphi$ of the relation that a induces on the states. Since $a\varphi$ is a $0,1$-matrix, we can view it as a matrix over $\mathbb{F}_p$. In this way we obtain a morphism $\varphi : \Sigma^* \to M_k(\mathbb{F}_p)$ where k is the number of states of $\mathcal{A}$. Let $M = \Sigma^*\varphi$. Trivially, M is finite. We observe that M is block upper triangular with diagonal blocks the syntactic monoids M_{L_i} (the partition of k arises from taking the states of each $\mathcal{A}_i$). Notice that M recognizes L, since L consists of all words w such that $(w\varphi)_{s,f} = r$ where s is the start state of $\mathcal{A}_0$ and f is a final state of $\mathcal{A}_n$. Applying Lemma 2.1 to the projection to the diagonal blocks gives that M and its quotient M_L belong to $\mathbf{LG}_p \ⓜ \mathbf{V}$. □

The proof of (4) implies (1) gives a fairly easy argument that marked products of rational languages with mod p counter are rational.

Since the operator $\mathbf{LG}_p \textcircled{m} (\ )$ is idempotent, we immediately obtain the following result of [7, 27].

Corollary 3.3. *Let* $\mathbf{V}$ *be a pseudovariety of monoids and* $\mathbf{W} = \mathbf{LG}_p \textcircled{m} \mathbf{V}$. *Then*

(1) $\mathcal{W}(\Sigma^*)$ *is the smallest class of languages containing* $\mathcal{V}(\Sigma^*)$, *which is closed under Boolean operations and formation of marked products with modulo* p *counters.*
(2) $\mathcal{W}(\Sigma^*)$ *consists of all Boolean combinations of elements of* $\mathcal{V}(\Sigma^*)$ *and marked products with modulo* p *counters of elements of* $\mathcal{V}(\Sigma^*)$.

Some special cases are the following. If $\mathbf{V}$ is the trivial variety of monoids, then $\mathbf{LG}_p \textcircled{m} \mathbf{V} = \mathbf{G}_p$ and we obtain Eilenberg's result [5, Section VIII.10] that the $\mathbf{G}_p$ languages consist of the Boolean combinations of languages of the form $(\Sigma^* a_1 \Sigma^* \cdots a_n \Sigma^*)_{r,p}$. Notice that $\mathbf{G}_p$ consists of the groups unitriangularizable over characteristic p. The languages over Σ^* associated to $\mathbf{LG}_p \textcircled{m} \mathbf{Sl}$ (as observed in [2], this pseudovariety consists of the unitriangularizable monoids over characteristic p) are the Boolean combinations of languages of the forms

$$\Sigma^* a \Sigma^* \quad \text{and} \quad (\Sigma_0^* a_1 \Sigma_1^* \cdots a_n \Sigma_n^*)_{r,p}$$

where $\Sigma_i \subseteq \Sigma$.

We remark that Weil shows [27] that closing $\mathcal{V}(\Sigma^*)$ under marked products with modulo p^n counters, for $n > 1$, does not take you out of the $\mathbf{LG}_p \textcircled{m} \mathbf{V}$-languages.

3.2. Unambiguous Products.

Our next application is to recover results of Schützenberger, Pin, Straubing, and Thérien concerning unambiguous products. Our proof of one direction is along the lines of [9] but our usage of representation theory allows us to avoid using results relying on case-by-case analysis of maximal proper surmorphisms and the block product principle.

Let Σ be a finite alphabet, $L_0, \ldots, L_n \subseteq \Sigma^*$ be rational languages and $a_1, \ldots, a_n \in \Sigma$. Then the *marked product* $L = L_0 a_1 L_1 \cdots a_n L_n$ is called *unambiguous* if each word $w \in L$ has exactly one factorization of the form $u_0 a_1 u_1 \cdots a_n u_n$, where each $u_i \in L_i$. We also allow the degenerate case $n = 0$.

We shall need to use a well-known and straightforward consequence of the distributivity of concatenation over union (cf. [9]), namely, that if $L_0, \ldots, L_n$ are disjoint unions of unambiguous marked products of elements of $\mathcal{V}(\Sigma^*)$, then the same is true for any unambiguous product $L_0 a_1 L_1 \cdots a_n L_n$. We also need a lemma about languages recognized by finite monoids of block upper triangular matrices in characteristic 0.

Lemma 3.4. *Let* $\mathbf{V}$ *be a pseudovariety of monoids,* $\varphi : \Sigma^* \to M$ *be a morphism with* M *finite. Let* K *be a field of characteristic* 0 *and suppose that* M *can be represented faithfully by block upper triangular matrices over*

K so that the monoids $M_1, \ldots, M_k$ formed by diagonal blocks of matrices in the image of M all belong to $\mathbf{V}$. Let $F \subseteq M$. Then $L = F\varphi^{-1}$ is a disjoint union of unambiguous marked products $L_0a_1L_1 \cdots a_nL_n$ with the $L_i \in \mathcal{V}(\Sigma^*)$.

Proof. We induct on the number k of diagonal blocks. If there is only one block we are done.

Now let $k > 1$. We can repartition n into two blocks, one corresponding to the union of the first $k-1$ of our original blocks and the other corresponding to the last block. The first diagonal block, call it N, is block upper triangular with diagonal blocks $M_1, \ldots, M_{k-1}$; the second is just M_k. By induction, any language recognized by N is a disjoint union of unambiguous marked products $L_0a_1L_1 \cdots a_rL_r$ with the $L_i \in \mathcal{V}(\Sigma^*)$. Thus to prove the result, it suffices to show that L is a disjoint union of unambiguous marked products $L_0a_1L_1 \cdots a_nL_n$ with the L_i recognized by $N \times M_k$. It is shown in [3] that the projection from M to $N \times M_k$ has locally trivial kernel category (see [14] for the definition). Then [9, Proposition 2.2] shows us that L is a disjoint union of such unambiguous marked products. □

We ask whether there is a simple combinatorial proof of this lemma that avoids the use of [9, Proposition 2.2] along the lines of the proof of Lemma 3.1.

Theorem 3.5. *Let $L \subseteq \Sigma^*$ be a rational language, $\mathbf{V}$ be a pseudovariety of monoids and K a field of characteristic 0. Then the following are equivalent.*

(1) $M_L \in \mathbf{LI} \textcircled{m} \mathbf{V}$;
(2) $M_L/\mathrm{Rad}_K(M_L) \in \mathbf{V}$;
(3) *M_L can be faithfully represented by block upper triangular matrices over K so that the monoids formed by the diagonal blocks of the matrices in the image of M_L all belong to $\mathbf{V}$.*
(4) *L is a disjoint union of unambiguous products $L_0a_1L_1 \cdots a_nL_n$ with the $L_i \in \mathcal{V}(\Sigma^*)$.*

Proof. The equivalence of (1) and (2) follows from the results of [3] quoted in Section 2.

For (2) implies (3), take a composition series for the regular representation of M_L over K: it is then in block upper triangular form and by (2) monoids formed by diagonal blocks of matrices in the image of M_L all belong to $\mathbf{V}$.

(3) implies (4) is immediate from Lemma 3.4.

For (4) implies (1), it suffices to deal with a single unambiguous marked product $L = L_0a_1L_1 \cdots a_nL_n$. Let $\mathcal{A}_i$ be the minimal trim [5] deterministic automaton for L_i and let $\mathcal{A}$ be the non-deterministic automaton obtained from the disjoint union of the L_i by attaching an edge labelled a_i from each final state of $\mathcal{A}_{i-1}$ to the initial state of $\mathcal{A}_i$. To each letter $a \in A$, we associate the matrix $a\varphi$ of the relation that a induces on the states. In this way we obtain a morphism $\varphi : \Sigma^* \to M_k(\mathbb{Q})$ where k is the number of states of $\mathcal{A}$. Let $M = \Sigma^*\varphi$. We observe that M is block upper triangular with diagonal blocks the syntactic monoids M_{L_i} (the partition of k arises from taking the states of each $\mathcal{A}_i$). Notice that M recognizes L, since L consists of all words w such that $(w\varphi)_{s,f} > 0$ where s is the start state of $\mathcal{A}_0$ and f is a final state of $\mathcal{A}_n$. First we show that M is finite. In fact, we claim M contains only $0,1$-matrices (and hence must be finite). Indeed, suppose $(w\varphi)_{i,j} > 1$ some i, j. Since each M_{L_i} consists of $0,1$-matrices, we must have that i is a state of some $\mathcal{A}_l$ and j a state of some $\mathcal{A}_r$ with $l < r$. But $(w\varphi)_{i,j}$ is the number of paths labelled by w from i to j in $\mathcal{A}$. Thus if u, v are words reading respectively from the start state of $\mathcal{A}_0$ to i and from j to a final state of $\mathcal{A}_n$ (such exist since the $\mathcal{A}_i$ are trim), then uwv has at least two factorizations witnessing membership in L, contradicting that L was unambiguous. Since the collection of all block upper triangular matrices is an algebra over $\mathbb{Q}$, as is the collection of block diagonal matrices, an application of Lemma 2.1 to the projection to the diagonal blocks gives that $M \in \mathbf{LI} \,ⓜ\, \mathbf{V}$ and so, since $M \twoheadrightarrow M_L$, we have $M_L \in \mathbf{LI} \,ⓜ\, \mathbf{V}$. □

Since the operator $\mathbf{LI} \,ⓜ(\ )$ is idempotent, we immediately obtain the following result of [8, 9].

Corollary 3.6. *Let* $\mathbf{V}$ *be a pseudovariety of monoids and* $\mathbf{W} = \mathbf{LI} \,ⓜ\, \mathbf{V}$. *Then*

(1) $\mathcal{W}(\Sigma^*)$ *is the smallest class of languages containing* $\mathcal{V}(\Sigma^*)$, *which is closed under Boolean operations and formation of unambiguous marked products.*

(2) $\mathcal{W}(\Sigma^*)$ *consists of all finite disjoint unions of unambiguous marked products of elements of* $\mathcal{V}(\Sigma^*)$.

Recall that the Malcev product of the pseudovariety **LI** with the pseudovariety **Sl** of semilattices (idempotent-commutative monoids) is equal to the famous pseudovariety **DA** of all finite monoids whose regular $\mathcal{D}$-classes are idempotent subsemigroups (see [24] for a nice survey of combinatorial, logical and automata-theoretic characterizations of **DA**). Applying the above corollary, one obtains the classical result of Schützenberger [19] that $\mathcal{DA}(\Sigma^*)$ consists of disjoint unions of unambiguous products of the form $\Sigma_0^* a_1 \Sigma_1^* \cdots a_n \Sigma_n^*$ with $\Sigma_i \subseteq \Sigma$ for all i. It is shown in [3], using representation theory, that **DA** consists of precisely those monoids that can be faithfully represented by upper triangular matrices with zeroes and ones on the diagonal over $\mathbb{Q}$.

References

1. J. Almeida, Finite Semigroups and Universal Algebra, World Scientific, Singapore, 1994.
2. J. Almeida, S. W. Margolis and M. V. Volkov, *The pseudovariety of semigroups of triangular matrices over a finite field*, RAIRO - Inf. Theor. Appl. **39** (2005), 31–48.
3. J. Almeida, S. W. Margolis, B. Steinberg and M. V. Volkov, *Representation theory of finite semigroups, semigroup radicals and formal language theory*, in preparation.

4. A. H. Clifford and G. B. Preston, The Algebraic Theory of Semigroups, Mathematical Surveys No. 7, AMS, Providence, RI, Vol. 1, 1961.
5. S. Eilenberg, Automata, Languages and Machines, Academic Press, New York, Vol A, 1974; Vol B, 1976.
6. K. Krohn, J. Rhodes and B. Tilson, *Lectures on the algebraic theory of finite semigroups and finite-state machines*, Chapters 1, 5-9 (Chapter 6 with M. A. Arbib) of The Algebraic Theory of Machines, Languages, and Semigroups, (M. A. Arbib, ed.), Academic Press, New York, 1968.
7. P. Peladeau, *Sur le produit avec compteur modulo un nombre premier*, RAIRO - Inf. Theor. Appl. **26** (1992), 553–564.
8. J.-E. Pin, *Propriétés syntactiques du produit non ambigu*, 7^{th} ICALP, Lect. Notes Comp. Sci. **85**, Springer Verlag, Berlin, Heidelberg, New York, (1980), 483–499.
9. J.-E. Pin, H. Straubing and D. Thérien, *Locally trivial categories and unambiguous concatenation*, J. Pure Applied Algebra **52** (1988), 297–311.
10. C. Reutenauer, *Sur les variétés de langages et de monoïdes*, Proc. GI Conf. [Lect. Notes Comp. Sci. **67**], Springer-Verlag, 1979, 260–265.
11. J. Rhodes, *A homomorphism theorem for finite semigroups*, Math. Systems Theory **1** (1967), 289–304.
12. J. Rhodes, *Characters and complexity of finite semigroups*, J. Combinatorial Theory **6**, (1969), 67–85.
13. J. Rhodes, *Algebraic theory of finite semigroups: Structure numbers and structure theorems for finite semigroups*, in: Semigroups, ed. K. Folley, Academic Press, New York, 1969, 125–162.
14. J. Rhodes and B. Tilson, *The kernel of monoid morphisms*, J. Pure Appl. Algebra **62** (1989), 227–268.
15. J. Rhodes and P. Weil, *Decomposition techniques for finite semigroups using categories, I*, J. Pure Applied Algebra **62** (1989), 269 – 284.
16. J. Rhodes and P. Weil, *Decomposition techniques for finite semigroups using categories, II*, J. Pure Applied Algebra **62** (1989), 285 – 312.
17. J. Rhodes and Y. Zalcstein, *Elementary representation and character theory of finite semigroups and its application* in: Monoids and semigroups with applications (Berkeley, CA, 1989), 334–367, World Sci. Publishing, River Edge, NJ, 1991.
18. M. P. Schützenberger, *On finite monoids having only trivial subgroups*, Inf. Control **8** (1965), 190–194.
19. M. P. Schützenberger, *Sur le produit de concatenation non ambigu*, Semigroup Forum **13** (1976), 47–75.
20. I. Simon, Hierarchies of Events of Dot-Depth One, Ph. D. Thesis, University of Waterloo, 1972.
21. I. Simon, *Piecewise testable events*, Proc. 2nd GI Conf. [Lect. Notes Comp. Sci. **33**], Springer-Verlag, 1975, 214–222.
22. I. Simon, *The product of rational languages*, in: "Automata, Languages and Programming," Andrzej Lingas, Rolf Karlsson, and Svante Carlsson (eds.), Berlin, 1993. Springer-Verlag. Lecture Notes in Computer Science **700**, 430-444.
23. H. Straubing, *A generalization of the Schützenberger product of finite monoids*, Theor. Comp. Sci. **13** (1981), 137–150.
24. P. Tesson and D. Thérien, *Diamonds are forever: the variety* **DA**, in: Semigroups, Algorithms, Automata and Languages, eds. G. M. S. Gomes, J. É. Pin and P. V. Silva, World Scientific, Singapore, 2002, 475–499.
25. D. Thérien, *Subword counting and nilpotent groups*, in: Combinatorics on Words, Progress and Perspectives, ed. L. J. Cummings, Academic Press, New York, 1983, 297–305.
26. B. Tilson, *Appendix to "Algebraic theory of finite semigroups: Structure numbers and structure theorems for finite semigroups": On the p length of p-solvable semigroups:*

Preliminary results, in: Semigroups, ed. K. Folley, Academic Press, New York, 1969, 163–208.

27. P. Weil, *Closure of varieties of languages under products with counter*, J. Comput. System Sci. **45** (1992), 316–339.