Recent BRICS Report Series Publications

BRICS RS-02-20 Ésik & Larsen: Regular Languages Definable by Lindström Quantifiers

BRICS

Basic Research in Computer Science

Regular Languages Definable by Lindström Quantifiers

(Preliminary Version)

Zoltán Ésik
Kim G. Larsen

BRICS Report Series RS-02-20
ISSN 0909-0878 April 2002

Regular Languages Definable by Lindström Quantifiers*
Preliminary Version

Zoltán Ésik
Dept. of Computer Science
University of Szeged
P.O.B. 652
6701 Szeged, Hungary
esik@inf.u-szeged.hu

Kim G. Larsen
Dept. of Computer Science
Aalborg University
Fredrik Bajers Vej 7E
9220 Aalborg, Denmark
kgl@cs.auc.dk

April 18, 2002

Abstract

In our main result, we establish a formal connection between Lindström quantifiers with respect to regular languages and the double semidirect product of finite monoid-generator pairs. We use this correspondence to characterize the expressive power of Lindström quantifiers associated with a class of regular languages.

1 Introduction

By the classic result of Büchi [6], Elgot [12] and Trakhtenbrot [39], the regular languages are exactly those definable by the sentences of a certain monadic second order logic over words. Moreover, Mc Naughton and Papert [22] proved that the first-order sentences of this logic define an important subclass of the regular languages, namely the star-free languages. By

*Research supported by BRICS (Basic Research in Computer Science), Aalborg, Denmark.

Schützenberger's theorem [29], the star-free languages are exactly those that can be recognized by the aperiodics, i.e., by those finite monoids containing no nontrivial groups.

Because of the limited expressive power of first-order logic on words, and in the search for characterizations of other important subclasses of regular languages in terms of formal logic, Straubing, Therien and Thomas [33, 34] introduced generalized, or modular quantifiers $\exists^{(m,r)}$, where $m \geq 2$ and $r = 0, \ldots, m-1$, with the following meaning: A word u satisfies a sentence $\exists^{(m,r)}x.\varphi$ iff the number of assignments of positions in u to the variable x satisfying φ is congruent to r modulo m. They proved that a language is definable in the logic involving, in addition to first-order quantifiers, the above modular quantifiers iff its syntactic monoid is finite and solvable, i.e., it contains only solvable groups. This class of regular languages first arose in [30, 35]. And if no first-order quantifiers are allowed, then a language is definable iff its syntactic monoid is a finite solvable group. In fact, Straubing, Therien and Thomas also studied the more general setting when the moduli of the generalized quantifiers in the logic are restricted to a finite set of (prime) numbers. See also Straubing, Therien [32] for a more recent account, and Baziramwabo, McKenzie, Therien [4] for a corresponding extension of linear temporal logic with modular counting.

In order to express regular languages having non-solvable syntactic monoids, Barrington, Immerman and Straubing [3] associated a family of quantifiers with each finite group containing a quantifier corresponding to each group element. When the group is cyclic of order m, the associated quantifiers are essentially the modular quantifiers $\exists^{(m,r)}$. They showed that a language is definable in first-order logic enriched with group quantifiers corresponding to the members of a subclass $\mathbf{G}$ of the finite groups iff the language is regular (so that its syntactic monoid is finite), and every simple group that divides the syntactic monoid of the language divides a group in $\mathbf{G}$. (See, e.g., [11] for a definition of when a group divides a group or a monoid.) Moreover, if only group quantifiers are allowed, then a language is definable iff in addition to the above conditions its syntactic monoid is a group. When $\mathbf{G}$ is empty, by Schützenberger's theorem one obtains the Mc Naughton–Papert characterization of first-order definable languages. The theorem of Barrington, Immerman and Straubing easily extends to quantifiers associated with finite monoids.

The quantifiers associated with finite monoids and groups, and thus the

modular quantifiers, are all special cases of (simple) Lindström quantifiers associated with (regular) languages, defined in [7]. (For more general treatments of Lindström quantifiers the reader is referred to Lindström [19] and Ebbinghaus and Flum [10], Chapter 12. See also the generalized quantifiers of Immerman [16] and the LNCS volume Väänänen [40].) By the results of Barrington, Immerman, Straubing [3], extended to monoids, it follows that when the Lindström quantifiers concern regular languages, then only regular languages can be defined. Moreover, when $\mathbf{G}$ is a class of finite groups (or monoids), and $\mathcal{L}_{\mathbf{G}}$ is the class of regular languages that can be recognized by the members of the class $\mathbf{G}$, then a language L is definable in first-order logic with Lindström quantifiers with respect to the languages in $\mathcal{L}_{\mathbf{G}}$ iff L is regular and every simple group divisor of the syntactic monoid of L divides a group (or monoid) in $\mathbf{G}$.

Our initial motivation for studying regular Lindström quantifiers was the question of characterizing those classes $\mathcal{L}$ of regular languages which are expressively complete in the sense that every regular language is definable by a sentence possibly involving, in addition to ordinary quantifiers, Lindström quantifiers with respect to the languages in $\mathcal{L}$. By the classic theorem of Büchi, Elgot, and Trakhtenbrot, first-order logic, enriched with Lindström quantifiers with respect to the languages in an expressively complete class $\mathcal{L}$, has the same expressive power as monadic second-order logic. Moreover, by the above-mentioned results of Barrington, Immerman and Straubing, a necessary condition of the expressive completeness of a class $\mathcal{L}$ is that $\mathcal{L}$ be group-complete, i.e., every finite (non-abelian simple) group be a divisor of the syntactic monoid of a language in $\mathcal{L}$. We will show that this condition, together with a condition involving the existence of certain cycles in the syntactic monoids of the languages in $\mathcal{L}$, which is equivalent to the expressibility of all of the one-letter languages $(a^n)^*$, $n \geq 2$, is necessary and sufficient as long as $\mathcal{L}$ is closed with respect to taking quotients, which is a natural assumption on $\mathcal{L}$. (A technical justification of this assumption is given later in the sequel.) On the other hand, we show that neither condition is sufficient by itself. But when $\mathcal{L}$ is closed with respect to taking quotients and padding, then $\mathcal{L}$ is expressively complete iff it is group-complete and at least one of the one-letter languages $(a^n)^*$, $n \geq 2$ is definable. Moreover, when $\mathcal{L}$ is closed with respect to taking quotients and arbitrary inverse homomorphic images, then $\mathcal{L}$ is expressively complete iff $\mathcal{L}$ is group-complete.

Formal logic in connection with words and languages has several general techniques such as model theoretic games and deep algebraic techniques developed in the theory of finite semigroups and automata. A general account of these methods is given in Straubing [31]. In particular, the semidirect product and the wreath product, and their symmetric versions, the double semidirect product[1] and the block product, defined by Rhodes and Tilson [28] (or the triple product of Eilenberg [11]), and the Krohn-Rhodes theorem [17] have been the fundamental tools for several of the aforementioned results. The same holds for our investigation. In our main technical result, Theorem 7.4, we make a bridge between Lindström quantifiers and the double semidirect product, or the block product. Particular instances of this correspondence appear in above cited works, see e.g. the proofs of lemmas VI.1.2, VI.1.4, VII.2.2. and VII.2.3 in Straubing [31]. In fact, we will make use of a version of the double semidirect product and the block product that concerns finite monoids with a distinguished set of generators.

For the connection between circuit complexity and generalized quantifiers, we refer to Barrington, Immerman, Straubing [3], Barrington, Compton, Straubing, Therien [2], and the last two chapters of Straubing [31]. For second-order Lindström quantifiers and their relation to leaf language definability, see Burtschick, Vollmer [7], Peichl, Vollmer [23], Galota, Vollmer [15]. For results regarding the connection between the semidirect product and the expressive power of temporal logics, see Cohen, Perrin, Pin [8], Therien, Wilke [36] and Baziramwabo, McKenzie, Therien [4]. The last paragraph of the paper McKenzie, Schwentick, Therien, Vollmer [20] contains an indication of the possibility of handling nested monoidal quantifiers in the logical framework by series connections of 2-way automata. The texts Pin [25], Straubing [31] and Thomas [37, 38] are excellent surveys of the subject.

The paper is organized as follows. In Section 2, we associate a Lindström quantifier with any (regular) language and establish some simple properties of Lindsröm quantifiers. In Section 3, we relate Lindsröm quantifiers to literal varieties of regular languages. Sections 4 and 5 are devoted to monoid-generator pairs, and to the operations of double semidirect product and block product on monoid-generator pairs. In Section 6, we define varieties of monoid-generator pairs and extend the double semidirect product and the block product to varieties. In Section 7, we establish a formal connection between Lindsröm quantifiers and the double semidirect product (block product, respectively) on varieties. In Section 8, we review the

[1]The double semidirect product is called the bilateral semidirect product in [31].

Krohn-Rhodes theorem and establish some of its consequences. In Section 9, we apply the results of Section 7 and 8 to provide characterizations of the expressive power of Lindström quantifiers.

We have tried to make the paper accessible not only for the experts but also for a larger audience.

2 Lindström quantifiers, defined

Suppose that $\mathcal{L}$ is a class of regular languages. We associate with $\mathcal{L}$ a language of formal logic $\mathbf{Lin}(\mathcal{L})$. For each alphabet (i.e., finite nonempty set) Σ, the *formulas of* $\mathbf{Lin}(\mathcal{L})$ *(over* Σ*)* are defined as follows. We assume that a fixed countable set of variables is given, and that each alphabet comes with a linear order defined on the letters of the alphabet.

- For each $a \in \Sigma$ and each variable x, $P_a(x)$ is an (atomic) formula. Moreover, when x, y are variables, $x < y$ is an (atomic) formula.
- For all formulas φ and ψ, both $\varphi \vee \psi$ and $\neg\varphi$ are formulas. Moreover, false is a formula.
- Suppose that $K \subseteq \Delta^*$ is in $\mathcal{L}$, where $\Delta = \{b_1, \ldots, b_m\}$, $m \geq 1$ is some alphabet. Then for all variables x and formulas φ_{b_i}, $b_i \in \Delta$, $i < m$,

$$Q_K x.\langle \varphi_{b_1}, \ldots, \varphi_{b_{m-1}} \rangle \tag{1}$$

is a formula.

We say that the variable x is *bound* in (1). The set of *free variables* in a formula is defined in the standard way. We identify any two formulas that differ only in the bound variables. Thus, we may assume that the bound variables of a formula are pairwise different, and different from any free variable. A formula with no free variables is called a *sentence*.

Suppose that u is a word in Σ^* of length n, say $u = u_1 \ldots u_n$, where the u_i are letters. Moreover, suppose that φ is a formula of $\mathbf{Lin}(\mathcal{L})$ over Σ whose free variables are contained in the finite set V. Given a function $\lambda : V \to [n]$, where $[n] = \{1, \ldots, n\}$, we say that (u, λ) *satisfies* φ, in notation $(u, \lambda) \models \varphi$, if

- φ is of the form $P_a(x)$ and $u_{\lambda(x)} = a$; or φ is of the form $x < y$ and $\lambda(x) < \lambda(y)$, or
- φ is of the form $\varphi_1 \vee \varphi_2$ and $(u, \lambda) \models \varphi_1$ or $(u, \lambda) \models \varphi_2$; or φ is of the form $\neg\psi$ and it is not the case that $(u, \lambda) \models \psi$, or
- φ is of the form (1), where Δ is ordered by $b_1, \ldots, b_m$, and the *characteristic word* [7] $\overline{u} = \overline{u_1} \ldots \overline{u_n}$ determined by (u, λ) and the formula belongs to K, where for each $i \in [n]$, $\overline{u_i}$ is the least b_j, $j < m$ such that we have $(u, \kappa) \models \varphi_{b_j}$ for the function $\kappa : V \cup \{x\} \to [n]$ with $\kappa(y) = \lambda(y)$, for all $y \in V$, and $\kappa(x) = i$.[2] When no such b_j exists, we define $\overline{u_i} = b_m$.

For all pairs (u, λ), relation $(u, \lambda) \models$ false does not hold.

Some notational conventions. In the sequel, in addition to the boolean connectives $\vee$ and $\neg$, we will also use the connective $\wedge$ and other boolean connectives. These are treated as abbreviations. We use **true** to denote $\neg$false. Moreover, we write $x \leq y$ for $\neg(y < x)$, $x = y$ for $(x \leq y) \wedge (y \leq x)$, etc. In quantified formulas (1), we may assume that the subformulas φ_{b_i} are *pairwise disjoint*, i.e., no pair (u, λ) satisfies two or more φ_{b_i}. Also, we may define φ_{b_m} as $\neg(\bigvee_{i<m} \varphi_{b_i})$ and write (1) as $Q_K x.\langle \varphi_{b_i} \rangle_{b_i \in \Delta}$. Note that the ordering on Δ now becomes irrelevant. *Below, when writing $Q_K x.\langle \varphi_{b_i} \rangle_{b_i \in \Delta}$, we will always assume that the φ_{b_i} satisfy the above conditions.* When φ is a sentence, we will write $u \models \varphi$ whenever $(u, \lambda) \models \varphi$ for all, or for some $\lambda : V \to [n]$.

Example 2.1

- Suppose that $K \subseteq \{b_1, b_2\}^*$ is the language $K_\exists = b_2^* b_1 (b_1 + b_2)^*$. Then $(u, \lambda) \models Q_K x.\langle \varphi \rangle$ iff there is an extension $\kappa : V \cup \{x\} \to [n]$ of $\lambda : V \to [n]$ such that $(u, \kappa) \models \varphi$. Thus, the Lindtrsöm quantifier corresponding to $K_\exists$ is the ordinary existential quantifier. When $K \subseteq \{b_1, b_2\}^*$ is the language $K_\forall = b_1^*$, the corresponding Lindström quantifier is the ordinary universal quantifier.
- Suppose that M is a set of integers > 1. Let $\mathcal{L}_M$ consist of all languages $C_m^r \subseteq \{b_1, b_2\}^*$, $m \in M$, $r = 0, \ldots, m-1$, where C_m^r is the set of all words u in $\{b_1, b_2\}^*$ such that the number of b_1s in u is congruent to r modulo m. Then $Q_{C_m^r}$ is the "modular quantifier" $\exists^{(m,r)}$ of Straubing, Therien and Thomas [34] and Straubing [31], and $\mathbf{Lin}(\mathcal{L}_M)$ is the class of all languages definable by modular quantification

[2] When the φ_{b_j} contain no free variables other than x, then the function $u \mapsto \overline{u}$ is called a *first-order translation* in [18] and [20].

with respect to moduli in M. (Note that it is sufficient to allow modular quantifiers with respect to prime moduli as in [31, 34].)

- Let L_m, where $m > 1$ is an integer, denote the language $(b_1^m)^*$, considered as a subset of $\{b_1\}^*$. Then for every alphabet Σ, $Q_{L_m} x.\langle\rangle$ is a sentence over Σ, and for every word $u \in \Sigma^*$, $u \models Q_{L_m} x.\langle\rangle$ iff the length of u is congruent to 0 mod m.
- One can express temporal modalities by Lindström quantifiers. Recall from Pnueli [26], Cohen, Perrin, Pin [8] that the formulas of propositional linear temporal logic over an alphabet Σ are generated from atomic propositions p_a, $a \in \Sigma$, by the boolean connectives $\vee$ and $\neg$, and the next and until modalities denoted X and U. For more details and the definition of semantics we refer to [26, 8]. Let K_X denote the two-letter language $(b_1 + b_2) b_1 (b_1 + b_2)^*$, and let K_U denote the three-letter language $b_2^* b_1 (b_1 + b_2 + b_3)^*$. Using these notations, we can translate each formula φ of propositional temporal logic over Σ into a formula $\tau(\varphi)$ involving ordinary quantifiers and Lindström quantifiers with respect to the languages K_X and K_U. We define:

  1. $\tau(p_a) = \exists x.((\forall y. x \leq y) \wedge P_a(x))$, for all $a \in \Sigma$.
  2. $\tau(\varphi \vee \psi) = \tau(\varphi) \vee \tau(\psi)$ and $\tau(\neg\varphi) = \neg\tau(\varphi)$.
  3. $\tau(\mathsf{X}\varphi) = Q_{K_\mathsf{X}} x.\langle \tau(\varphi)[\geq x] \rangle$ and $\tau(\varphi \mathsf{U} \psi) = Q_{K_\mathsf{U}} x.\langle \tau(\psi)[\geq x], \tau(\varphi)[\geq x] \rangle$.

  (Here, $\tau(\varphi)[\geq x]$ denotes a *relativization* of the formula $\tau(\varphi)$. See Straubing [31].) Then, for each word $u \in \Sigma^*$ and temporal logic formula φ, it holds that $u \models \varphi$ iff $u \models \tau(\varphi)$.

Given a sentence φ of $\mathbf{Lin}(\mathcal{L})$ over the alphabet Σ, we let $L_\varphi \subseteq \Sigma^*$ denote the *language defined by* φ:

$$L_\varphi = \{u \in \Sigma^* : u \models \varphi\}.$$

With some abuse of notation, we write also $\mathbf{Lin}(\mathcal{L})$ to denote the class of all languages definable by sentences of this logic. Moreover, we define $\mathbf{FO}(\mathcal{L}) = \mathbf{Lin}(\mathcal{L} \cup \{K_\exists\}) = \mathbf{Lin}(\mathcal{L} \cup \{K_\forall\})$. Thus, $\mathbf{FO}(\emptyset)$ is just the class of *first-order definable languages* of [22, 31].

Example 2.2

- Let $\mathcal{L}$ consist of all finite languages. Then $\mathbf{Lin}(\mathcal{L})$ is the class of all finite or co-finite languages.
- Let $\mathcal{L}$ consist of the languages C_m^r defined in Example 2.1. Then, as shown in Straubing, Therien, Thomas [34], $\mathbf{Lin}(\mathcal{L})$ ($\mathbf{FO}(\mathcal{L})$, respectively) is the class of all regular languages whose syntactic monoid is a solvable group (monoid, respectively). See also Section 9.

- More generally, when M denotes a set of integers > 1, $\mathbf{Lin}(\mathcal{L}_M)$ consists of those regular languages whose syntactic monoids are solvable groups of order n such that any prime divisor of n divides an integer in M. Moreover, $\mathbf{FO}(\mathcal{L}_M)$ consists of those regular languages L such that every subgroup of the syntactic monoid of L has this property. See Straubing, Therien, Thomas [34].

For technical reasons, we also associate a language with formulas φ over Σ containing free variables. We follow the definitions in [31]. Let V denote a finite set of variables containing all of the free variables of φ. A *V-structure* over Σ is a word $u = u_1 \dots u_n$ in $(\Sigma \times P(V))^*$, where $P(V)$ denotes the power set of V, such that each variable in V appears exactly once in the right hand component of a letter $u_i = (a_i, X_i)$. Thus, the sets X_i are pairwise disjoint and their union is V. Note that the empty word is a V-structure iff $V = \emptyset$. Given the V-structure u, the left hand components a_i determine a word $v = a_1 \dots a_n$ in Σ^*, and the right hand components determine a function $\lambda : V \to [n]$, defined by $\lambda(x) = i$ iff $x \in X_i$. We say that u satisfies φ, denoted $u \models \varphi$, if $(v, \lambda) \models \varphi$. The language L_φ defined by φ consists of all V-structures u over Σ with $u \models \varphi$.

PROPOSITION 2.3 *For each class $\mathcal{L}$ of regular languages, it holds that $\mathcal{L} \subseteq$ $\mathbf{Lin}(\mathcal{L})$.*

Proof. Given $K \subseteq \Delta^*$ in $\mathcal{L}$, where $\Delta = \{b_1, \dots, b_k\}$, the language defined by the sentence $Q_K x.\langle P_{b_i}(x)\rangle_{b_i \in \Delta}$ is K. □

We say that a class $\mathcal{L}$ of regular languages is *closed with respect to the boolean operations* if for each alphabet Σ, the regular languages over Σ which are in $\mathcal{L}$ contain $\emptyset$ and Σ^* and form a boolean algebra. Moreover, we say that $\mathcal{L}$ is *closed with respect to inverse literal (homo)morphisms* if for all alphabets Σ, Δ and letter preserving homomorphisms $h : \Sigma^* \to \Delta^*$ (i.e., such that $h(\Sigma) \subseteq \Delta$), and for all languages $L \subseteq \Delta^*$, if L is in $\mathcal{L}$, then

$$h^{-1}(L) \quad = \quad \{u \in \Sigma^* : h(u) \in L\}$$

is also in $\mathcal{L}$.

PROPOSITION 2.4 *For each class $\mathcal{L}$ of regular languages it holds that $\mathbf{Lin}(\mathcal{L}) =$ $\mathbf{Lin}(\mathcal{L}')$, where $\mathcal{L}'$ is the least class of (regular) languages containing $\mathcal{L}$ which is closed with respect to the boolean operations and inverse literal morphisms.*

Proof. Since $\mathcal{L} \subseteq \mathcal{L}'$, it is clear that $\mathbf{Lin}(\mathcal{L}) \subseteq \mathbf{Lin}(\mathcal{L}')$. The reverse inclusion follows from the following claims.

Claim 1 Suppose that $K, K_1, K_2 \subseteq \Delta^*$ are (regular) languages and for each $b \in \Delta$, φ_b is a formula of $\mathbf{Lin}(\mathcal{L})$ over Σ with free variables in $V \cup \{x\}$. Then for each V-structure u over Σ,

$$u \models Q_{K_1 \cup K_2} x.\langle\varphi_b\rangle_{b\in\Delta} \quad\Leftrightarrow\quad u \models Q_{K_1} x.\langle\varphi_b\rangle_{b\in\Delta} \text{ or } u \models Q_{K_2} x.\langle\varphi_b\rangle_{b\in\Delta}$$
$$u \models Q_{\Delta^* - K} x.\langle\varphi_b\rangle_{b\in\Delta} \quad\Leftrightarrow\quad u \models \neg(Q_K x.\langle\varphi_b\rangle_{b\in\Delta}).$$

Claim 2 Suppose that $h : \Delta^* \to \Delta'^*$ is a literal morphism and that φ_b, $b \in \Delta$ are pairwise disjoint $\mathbf{Lin}(\mathcal{L})$ formulas over Σ with free variables in $V \cup \{x\}$ such that each $(V \cup \{x\})$-structure over Σ satisfies $\bigvee_{b\in\Delta} \varphi_b$. Let $K' \subseteq \Delta'^*$ and $K = h^{-1}(K')$. Then there exist $\mathbf{Lin}(\mathcal{L})$ formulas $\varphi_{b'}$, $b' \in \Delta'$ over Σ with free variables in $V \cup \{x\}$ such that $Q_K x.\langle\varphi_b\rangle_{b\in\Delta}$ and $Q_{K'} x.\langle\varphi_{b'}\rangle_{b'\in\Delta'}$ define the same set of V-structures over Σ. In fact, we can set

$$\varphi_{b'} \quad = \quad \bigvee_{h(b)=b'} \varphi_b,$$

for all $b' \in \Delta'$. (When b' is not in the image of h, we have $\varphi_{b'} = \mathsf{false}$.) □

REMARK 2.5 When L is a regular language, one can use any finite automaton accepting L to express the Lindström quantifier Q_L in monadic second order logic [31] and use the theorem of Büchi [6], Elgot [12] and Trakhtenbrot [39] to establish that if each φ_{b_i} defines a regular language (of $(V \cup \{x\})$-structures), then (1) defines a regular language (of V-structures). Thus, for any class $\mathcal{L}$ of regular languages, $\mathbf{Lin}(\mathcal{L})$ contains only regular languages. This result may also be seen as an instance of a general fact about Lindström quantifiers, cf. Exercise 12.1.1 in [10]. This fact also follows from Theorem 7.4. See also Barrington, Immerman, Straubing [3], Lautemann, McKenzie, Schwentick [18], and Theorem 7.4.

3 Literal varieties of regular languages

Suppose that $\mathcal{L}$ is a class of regular languages. We call $\mathcal{L}$ a *literal variety* if it is closed with respect to the boolean operations, left and right quotients and inverse literal homomorphisms. Thus, if $L, L_1, L_2 \subseteq \Sigma^*$ are in a literal variety $\mathcal{L}$ and $a \in \Sigma$, then $L_1 \cup L_2$, $\Sigma^* - L$ and $a^{-1}L, La^{-1}$ are also in $\mathcal{L}$, where

$$a^{-1}L = \{u \in \Sigma^* : au \in L\}$$
$$La^{-1} = \{u \in \Sigma^* : ua \in L\}.$$

Moreover, if h is a literal morphism $\Sigma^* \to \Delta^*$ and $L \subseteq \Delta^*$ is in $\mathcal{L}$, then $h^{-1}(L)$ is also in $\mathcal{L}$. Note that every literal variety contains, for each finite alphabet Σ, the language Σ^* and the empty language.

Literal varieties are a generalization of the $*$-*varieties* of Eilenberg [11] and Pin [24] that are closed with respect to arbitrary inverse homomorphisms.

PROPOSITION 3.1 *For each class $\mathcal{L}$ of regular languages,* **Lin**$(\mathcal{L})$ *is closed with respect to the boolean operations and inverse literal morphisms. Moreover, when $\mathcal{L}$ is a class of regular languages closed with respect to quotients, then* **Lin**$(\mathcal{L})$ *is a literal variety.*

Proof. We have remarked in Section 2 that **Lin**$(\mathcal{L})$ contains only regular languages. The closure of **Lin**$(\mathcal{L})$ under the boolean operations is straightforward. To prove that **Lin**$(\mathcal{L})$ is closed with respect to inverse literal homomorphisms, suppose that $L \subseteq \Delta^*$ is in **Lin**$(\mathcal{L})$ and h is a literal homomorphism $\Sigma^* \to \Delta^*$. Let φ denote a sentence of **Lin**$(\mathcal{L})$ over Δ defining L. Then $h^{-1}(L)$ is defined by the sentence over Σ obtained by replacing each atomic formula $P_b(x)$ in φ, where x is a variable and $b \in \Delta$, by the disjunction of all formulas $P_a(x)$ with $h(a) = b$. When b is not in the image of h, we replace $P_b(x)$ by the formula false.

To complete the proof, assume now that $\mathcal{L}$ is closed with respect to taking quotients, and suppose that φ is a formula of **Lin**$(\mathcal{L})$ over the finite alphabet Σ possibly containing free variables from the finite set V. Let L_φ denote the set of all V-structures (over Σ) defined by φ. We argue by induction on the structure of φ to prove that for each letter $a \in \Sigma$, the set of V-structures $a^{-1}L_\varphi$ is definable by some formula of **Lin**$(\mathcal{L})$ with free variables in V. Moreover, we show that for each letter $a \in \Sigma$ and variable $x \in V$, the set of $(V - \{x\})$-structures $(a, \{x\})^{-1}L_\varphi$ is definable by some formula with free variables in $V - \{x\}$. (The definability of right quotients can be established in the same way.) The basis case is obvious, including the case when φ is false, as are the cases when φ is of the form $\varphi_1 \vee \varphi_2$ or $\neg\psi$. One uses the fact that the operation of taking left quotients commutes with the boolean operations. Suppose finally that φ is of the form $Q_K x.\langle\psi_{b_j}\rangle_{b_j \in \Delta}$, where $K \subseteq \Delta^* = \{b_1, \ldots, b_m\}^*$ is a language in $\mathcal{L}$ and each ψ_{b_j} is a formula of **Lin**$(\mathcal{L})$ over Σ with free variables in $V \cup \{x\}$, where $x \notin V$. For each b_j, let L_{b_j} denote the set of all $(V \cup \{x\})$-structures defined by ψ_{b_j}. It follows by the induction hypothesis that for each b_j there is a formula ρ_{b_j} that defines the set of V-structures $(a, \{x\})^{-1}L_{b_j}$, i.e., the set of all V-structures u such that $(a, \{x\})u \models \psi_{b_j}$. Moreover, for each b_k there exists a formula ψ'_{b_k} over Σ with free variables in $V \cup \{x\}$ defining the set of $(V \cup \{x\})$-structures in $a^{-1}L_{b_k}$, i.e., the set of all $(V \cup \{x\})$-structures u such that $au \models \psi_{b_k}$. For each b_j, let

$$\alpha_{b_j} = \rho_{b_j} \wedge Q_{b_j^{-1}K} x.\langle\psi'_{b_k}\rangle_{b_k \in \Delta}.$$

Given a V-structure u, we have $u \models \alpha_{b_j}$ iff $(a, \{x\})u \models \psi_{b_j}$ and $\overline{u} \in b_j^{-1}K$, where $\overline{u}$ is the characteristic word determined by u and the formula $Q_{b_j^{-1}K} x.\langle\psi'_{b_k}\rangle_{b_k \in \Delta}$. It follows that $u \models \alpha_{b_j}$ iff the characteristic word determined by au and the formula φ starts with b_j and belongs to K. Thus, letting

$$\alpha = \bigvee_{b_j \in \Delta} \alpha_{b_j},$$

it holds that $u \models \alpha$ iff $au \models \varphi$ iff $u \in a^{-1}L_\varphi$, showing that $a^{-1}L_\varphi$ is in **Lin**$(\mathcal{L})$. The fact that $(a, \{y\})^{-1}L_\varphi$ is in **Lin**$(\mathcal{L})$, for each variable $y \in V$ and letter $a \in \Sigma$, can be established by a similar argument. □

PROPOSITION 3.2 *Suppose that $\mathcal{L}$ is a class of regular languages closed with respect to left and right quotients. Then* **Lin**$(\mathcal{L}) =$ **Lin**$(\mathcal{L}')$, *where $\mathcal{L}'$ is the literal variety generated by $\mathcal{L}$.*

Proof. We have proved that **Lin**$(\mathcal{L}) =$ **Lin**$(\mathcal{L}'')$, where $\mathcal{L}''$ is the smallest class containing $\mathcal{L}$ that is closed with respect to the boolean operations and inverse literal homomorphisms. When $\mathcal{L}$ is closed with respect to quotients, so is $\mathcal{L}''$, so that $\mathcal{L}'' = \mathcal{L}'$, the literal variety generated by $\mathcal{L}$. □

4 Monoid-generator pairs

A *monoid-generator pair* (M, A), or *mg-pair*, for short, consists of a monoid M and a set A of generators of M. When M is finite, we call (M, A) a finite mg-pair. A *morphism* $(M, A) \to (N, B)$ of mg-pairs is a monoid homomorphism $h : M \to N$ such that $h(A) \subseteq B$. It is clear that mg-pairs and their morphisms form a category with respect to function composition. When $\underline{1} = \{1\}$ denotes a trivial monoid, we have that $(\underline{1}, \emptyset)$ is initial and $(\underline{1}, \{1\})$ is a terminal object of this category. We call a morphism $h : (M, A) \to (N, B)$ *surjective* if $h(A) = B$, so that also $h(M) = N$, and *injective* if it is an injective function. Moreover, we call (N, B) a *quotient* of (M, A) if there is surjective morphism $(M, A) \to (N, B)$, and a *sub-mg pair* of (M, A) if $N \subseteq M$ and the inclusion $N \to M$ is a morphism (thus, $B = A \cap N$). Finally, we say that (M, A) *covers* (N, B), or that (N, B) *divides* (M, A), if (N, B) is a quotient of a sub-mg of (M, A). We let $<$ denote this relation. It is clear that when $(N, B) < (M, A)$, the monoid N is a quotient of a submonoid of M, i.e., $N < M$ as defined in [11]. Also, $<$ is a reflexive and transitive both on mg-pairs and on monoids.

REMARK 4.1 Suppose that there are a *subsemigroup* S of M and a subset C of A that generates S such that there is a surjective semigroup homomorphism $S \to M$ that maps C onto B. Then we have $(N, B) < (M, A)$. (Note that S may not contain the identity element of M.) The converse statement is valid whenever the identity element of N is a nonempty product over B.

REMARK 4.2 Suppose that M and N are monoids. We recall from Eilenberg [11] that a *covering* $N \to M$ is a relation $\varphi : N \to M$, viewed as a function $N \to P(M)$, such that

- $\varphi(n) \neq \emptyset$, for all $n \in N$,
- for all $n_1, n_2 \in N$, if $n_1 \neq n_2$ then $\varphi(n_1) \cap \varphi(n_2) = \emptyset$,
- $1 \in \varphi(1)$, and
- $\varphi(n_1)\varphi(n_2) \subseteq \varphi(n_1)\varphi(n_2)$, for all $n_1, n_2 \in N$.

It is known that $N < M$ iff there is a covering $N \to M$.

We may define a covering $(N, B) \to (M, A)$, where (M, A) and (N, B) are mg-pairs, as a covering $\varphi : N \to M$ such that for each $b \in B$ there exists some $a \in A$ with $a \in \varphi(b)$. (The first condition above in the definition of covering then becomes redundant.) We will return to coverings in the Appendix.

EXAMPLE 4.3 • For every monoid M, the pair (M, M) is an mg-pair. Moreover, for monoids M, N, we have that $N < M$ iff $(N, N) < (M, M)$.

- When Σ is a (finite) alphabet, (Σ^*, Σ) is an mg-pair. Given any mg-pair (M, A) and function $h : \Sigma \to A$, there is a unique morphism $(\Sigma^*, \Sigma) \to (M, A)$ extending h. (We denote this morphism by h as well.) Thus, (Σ^*, Σ) is a *free* mg-pair.
- Each *automaton* $(Q, \Sigma, \cdot)$ with transition function $\cdot : Q \times \Sigma \to Q$ gives rise to an mg-pair $(M_Q, \overline{\Sigma})$. Its monoid component M_Q is the monoid of all state transformations $Q \to Q$ induced by the words in Σ^*, and the set of generators $\overline{\Sigma}$ consists of those transformations induced by the letters in Σ.
- Each mg-pair (M, A) may be regarded as an automaton freely generated by the identity element of M whose action is given by right multiplication $(m, a) \mapsto ma$, $m \in M$, $a \in A$. In fact, the category of mg-pairs is equivalent to the category of one-generated input reduced free automata (i.e., in which different input letters induce different state transformations), whose morphisms preserve the free generator and the transitions (with a change in the alphabet).
- When (M, A) is an mg-pair, $B \subseteq A$ and $Q \subseteq M$ is closed with respect to right multiplication with the elements of B, then Q and B determine an mg-pair $(N, \overline{B})$. Here, N is the quotient of the submonoid M' of M generated by the elements in B with respect to the congruence $\sim_Q$ defined by $x \sim_Q y$ iff $qx = qy$ for all $q \in Q$. Moreover, $\overline{B}$ consists of the congruence classes of the elements of B.

Below we will identify a monoid M with the mg-pair (M, M).

Each mg-pair may be used as a recognizer. Let (M, A) denote a not necessarily finite mg-pair and let $h : (\Sigma^*, \Sigma) \to (M, A)$ be a morphism, so that h is a monoid homomorphism $\Sigma^* \to A$ with $h(\Sigma) \subseteq A$. Given a set $F \subseteq M$, the language *recognized*, or *accepted* by (M, A) with h and F is the set

$$\{u \in \Sigma^* : h(u) \in F\}.$$

It is clear that a language is regular iff it can be recognized by a finite mg-pair.

Any language can be recognized by an mg-pair. Given a language $L \subseteq \Sigma^*$, let M_L denote the syntactic monoid of L, and let $\eta_L : \Sigma^* \to M_L$ denote the syntactic homomorphism of L, cf. Eilenberg [11], Pin [24]. Then $(M_L, \eta_L(\Sigma))$ is an mg-pair, called the *syntactic mg-pair* of L. Moreover, η is a morphism $(\Sigma^*, \Sigma) \to (M_L, \eta(\Sigma))$, called the *syntactic morphism* of L.

The following fact is an adaptation of well-known results from Eilenberg [11] and Pin [24].

PROPOSITION 4.4 • *The language recognized by $(M_L, \eta_L(\Sigma))$ with the syntactic morphism η_L and the set $\eta_L(L)$ is L.*

- *Suppose that (M, A) accepts L with $h : (\Sigma^*, \Sigma) \to (M, A)$ and $F \subseteq M$. Suppose that h is surjective. Then there is a (unique) morphism $h' : (M, A) \to (M_L, \eta_L(\Sigma))$ such that*

$$\eta_L \quad = \quad (\Sigma^*, \Sigma) \xrightarrow{h} (M, A) \xrightarrow{h'} (M_L, \eta_L(\Sigma)).$$

- *A language $L \subseteq \Sigma^*$ can be recognized by an mg-pair (M, A) iff we have $(M_L, \eta_L(\Sigma_L)) < (M, A)$.*

Suppose that $\mathbf{K}$ is a class of finite mg-pairs. We define $\mathbf{Lin(K)} = \mathbf{Lin}(\mathcal{L}_{\mathbf{K}})$, where $\mathcal{L}_{\mathbf{K}}$ is the class of all (regular) languages recognizable by the members of $\mathbf{K}$. Moreover, we define $\mathbf{FO(K)} = \mathbf{FO}(\mathcal{L}_{\mathbf{K}})$.

PROPOSITION 4.5 *For each class $\mathbf{K}$ of finite mg-pairs, $\mathbf{Lin(K)}$ is a literal variety.*

Proof. By Proposition 2.4, we know that $\mathbf{Lin(K)}$ is closed with respect to the boolean operations and inverse literal morphisms. Since $\mathcal{L}_{\mathbf{K}}$ is closed with respect to quotients, so is $\mathbf{Lin(K)} = \mathbf{Lin}(\mathcal{L}_{\mathbf{K}})$, by Proposition 3.1. □

Let U_1 denote a two-element monoid which is not a group. Note that U_1 is isomorphic to the syntactic monoid of the language $K_\exists$, defined in Example 2.1. Moreover, every language $L \subseteq \Delta^*$ recognizable in U_1 is either the empty language, or the language Δ^*, or the inverse image of $K_\exists$ or $K_\forall$ with respect to a literal morphism $\Delta^* \to \{b_1, b_2\}^*$. Using this fact and Proposition 2.4, we immediately have:

PROPOSITION 4.6 *For each class $\mathbf{K}$ of mg-pairs, $\mathbf{FO(K)} = \mathbf{Lin}(\mathbf{K} \cup \{U_1\})$.*

Below we will use this fact without mention.

5 Double semidirect product and block product

The double semidirect product and the block product of monoids were introduced in [28]. In this section, we extend these notions to mg-pairs.

Suppose that (S, A) and (T, B) are mg-pairs. We write the monoid operation of S additively without assuming that the operation is commutative. We denote by 0 the identity element of S. A (monoidal) *left action* of T on (S, A) is a function

$$\begin{aligned} T \times S &\to S \\ (t, s) &\mapsto ts \end{aligned}$$

subject to the following conditions for all $s, s' \in S$ and $t, t' \in T$:

$$\begin{aligned} (tt')s &= t(t's) \\ t(s + s') &= ts + ts' \\ 1s &= s \\ t0 &= 0. \end{aligned}$$

Moreover, it is required that

$$ta \in A, \quad \text{for all } t \in T, a \in A.$$

A *right action* $S \times T \to S$, $(s, t) \mapsto st$ is defined symmetrically. Actions $T \times S \to S$ and $S \times T \to S$ are *compatible* if

$$(ts)t' \quad = \quad t(st'),$$

for all $t, t' \in T$ and $s \in S$. Due to the above laws, we will write just tst' for $(ts)t' = t(st')$, $tt's$ for $t(t's) = (tt')s$, etc.

Given a compatible pair of left and right actions of T on (S, A), we define the *double semidirect product* $(S, A){\star\star}(T, B)$ as follows. First, we define

$$(s, t)(s', t') \quad = \quad (st' + ts', tt'),$$

for all $s, s' \in S$ and $t, t' \in T$. It is known, cf. [28, 31] that $S \times T$, equipped with this operation, is a monoid with identity element $(0, 1)$, called the double semidirect product of S and T determined by the actions. Let R denote the submonoid of this monoid generated by the set $A \times B$. We define the double semidirect product $(S, A){\star\star}(T, B)$ to be the mg-pair $(R, A \times B)$.

Two special cases are of particular interest. The notion of *semidirect product* [11] $(S, A){\star}(T, B)$ involves only a left action of T on (S, A) and corresponds to the double semidirect product $(S, A){\star\star}(T, B)$ determined by the same left action and the trivial right action: $st = s$, for all $s \in S$ and $t \in T$. When both actions are trivial, we obtain the *direct product* $(S, A) \times (T, B)$. This is the mg-pair $(R, A \times B)$, where R is the submonoid of the usual direct product $S \times T$ generated by $A \times B$. The direct product is the categorical product in the category of mg-pairs.

REMARK 5.1 The double semidirect product of monoids is closely related to the *triple product* of Eilenberg [11], vol. B. Any double semidirect product $S{\star\star}T$ of monoids S and T embeds in a triple product $[T, S, T]$ determined by the same actions. Moreover, as shown in Rhodes, Tilson [28], any triple product $[T_1, S, T_2]$ of monoids S, T_1, T_2 equipped with a monoidal right action $S \times T_1 \to S$, $(s, t_1) \mapsto st_1$, and a monoidal left action $T_2 \times S \to S$, $(t_2, s) \mapsto t_2 s$, is isomorphic to the double semidirect product $S{\star\star}(T_1 \times T_2)$ with actions $(t_1, t_2)s = t_2 s$ and $s(t_1, t_2) = st_1$, for all $s \in S$ and $t_i \in T_i$, $i = 1, 2$.

Suppose that (S, A) and (T, B) are mg-pairs. Then $(S, A)^{T\times T}$, the $(T \times T)$-fold direct product of (S, A) with itself is an mg-pair $(R, A^{T\times T})$. Here, R is the submonoid of $S^{T\times T}$ generated by the set $A^{T\times T}$. The *block product* $(S, A)\Box(T, B)$ is the double semidirect product

$$(R, A^{T\times T}){\star\star}(T, B)$$

determined by the following compatible left and right actions:

$$(tf)(t_1, t_2) = f(t_1 t, t_2)$$
$$(ft)(t_1, t_2) = f(t_1, tt_2),$$

for all $f \in R$ and $t, t_1, t_2 \in T$. The reader should have no difficulty in verifying that $tf, ft \in R$ for all $f \in R$ and $t \in T$. Moreover, when $f \in A^{T\times T}$, we have $tf, ft \in A^{T\times T}$, for all $t \in T$. The *wreath product* $(S, A) \circ (T, B)$ is defined in a similar way. It is the semidirect product $(R, A^T \times B) = (S, A)^T \star (T, B)$ determined by the left action

$$(tf)(t_1) = f(t_1 t), \quad t_1 \in T,$$

for all $f \in R$ and $t \in T$.

PROPOSITION 5.2 *For any mg-pairs (M, A) and (N, B), every double semidirect product $(M, A \times B) = (S, A){\star\star}(T, B)$ is isomorphic to a sub mg-pair of the block product $(N, A^{T\times T} \times B) = (S, A)\Box(T, B)$.*

Proof. We follow the proof of Proposition 7.1 from Rhodes, Tilson [28]. For each $s \in S$ let $f_s : T \times T \to S$ denote the function $(t_1, t_2) \mapsto t_1 s t_2$, $t_1 \in T_1, t_2 \in T_2$. Note that when $s \in A$, then f_s maps $T \times T$ into A. It is shown in Rhodes [27] that the assignment

$$(s, t) \mapsto (f_s, t), \quad (s, t) \in S \times T$$

defines an injective morphism $S{\star\star}T \to S\Box T$. Moreover, if $(s, t) \in A \times B$, then $(f_s, t) \in A^{T\times T} \times B$. To complete the proof we still need to show that if $(s, t) \in M$, then $(f_s, t) \in N$. However, if (s, t) is an n-fold product over $A \times B$, for some $n \geq 0$, then (f_s, t) is an n-fold product over $A^{T\times T} \times B$. □

6 Varieties of mg-pairs

In the rest of the paper, other than free mg-pairs (Σ^*, Σ), we will only consider finite mg-pairs, i.e., pairs (M, A) such that M is a finite monoid. Each finite monoid M may be identified with the mg-pair (M, M).

A *(pseudo)variety of mg-pairs* is a nonempty class $\mathbf{V}$ of (finite) mg-pairs closed with respect to the direct product and division, i.e., such that

- $(S, A), (T, B) \in \mathbf{V} \Rightarrow (S, A) \times (T, B) \in \mathbf{V}$, and
- $(S, A) < (T, B),\ (T, B) \in \mathbf{V} \Rightarrow (S, A) \in \mathbf{V}$.

A *closed class of mg-pairs* is a nonempty class of mg-pairs that is closed with respect to the double semidirect product and division. Since the direct product is a special case of the double semidirect product, any closed class is a variety. Therefore we will also call closed classes as *closed varieties*. It is clear that each class of mg-pairs is contained in a least variety and in a least closed variety.

Given varieties $\mathbf{V}$ and $\mathbf{W}$, we define

- $\mathbf{V}{\star\star}\mathbf{W}$ to be the variety generated by all double semidirect products $(M, A){\star\star}(T, B)$, where $(M, A) \in \mathbf{V}$ and $(T, B) \in \mathbf{W}$,
- $\mathbf{V}\Box\mathbf{W}$ to be the variety generated by all block products $(M, A)\Box(T, B)$, where $(M, A) \in \mathbf{V}$ and $(T, B) \in \mathbf{W}$.

PROPOSITION 6.1 *For all varieties* $\mathbf{V}$ *and* $\mathbf{W}$, *it holds that* $\mathbf{V}{**}\mathbf{W} = \mathbf{V}\Box\mathbf{W}$. *Moreover, an mg-pair is in* $\mathbf{V}{**}\mathbf{W}$ *iff it is covered by a double semidirect product* $(S,A){**}(T,B)$, *or, equivalently, by a block product* $(S,A)\Box(T,B)$, *where* $(S,A) \in \mathbf{V}$ *and* $(T,B) \in \mathbf{W}$.

Proof. Since a block product $(S,A)\Box(T,B)$ is a double semidirect product $(S,A)^{B\times B}{**}(T,B)$, and since varieties are closed with respect to the direct product, it follows that $\mathbf{V}\Box\mathbf{W} \subseteq \mathbf{V}{**}\mathbf{W}$. The reverse inclusion follows from Proposition 5.2.

The proof of the second claim uses the fact that any direct product of double semidirect products is isomorphic to a double semidirect product of direct products, and similarly for the block product. The argument is quite standard. All facts formulated in Proposition 6.1 are well-known for monoid varieties. See Rhodes [27]. □

COROLLARY 6.2 *A nonempty class of mg-pairs is a closed variety iff it is closed with respect to division and the block product.*

PROPOSITION 6.3 *For all varieties* $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$, *it holds that*

$$(\mathbf{V}_1{**}\mathbf{V}_2){**}\mathbf{V}_3 \quad \subseteq \quad \mathbf{V}_1{**}(\mathbf{V}_2{**}\mathbf{V}_3).$$

This fact is known to hold for varieties of monoids, cf. Rhodes [27], p. 460. As communicated to the authors by John Rhodes, the proof uses the Kernel Theorem (Theorem 7.4) of Rhodes, Tilson [28]. In the Appendix, we extend the kernel construction to mg-pairs.

COROLLARY 6.4 *For every variety* $\mathbf{V}$, *the least closed variety containing* $\mathbf{V}$ *can be constructed as the class* $\bigcup_{n\geq 0} \mathbf{V}^{(n)}$, *where* $\mathbf{V}^{(0)}$ *is the class of all trivial (i.e., singleton) mg-pairs and* $\mathbf{V}^{(n+1)} = \mathbf{V}{**}\mathbf{V}^{(n)} = \mathbf{V}\Box\mathbf{V}^{(n)}$, *for all* $n \geq 0$.

For later use we note:

PROPOSITION 6.5 *For any mg-pairs* (S,A) *and* (T,B), *any language recognizable by* $(S,A)\times(T,B)$ *is a boolean combination of languages recognizable by* (S,A) *and* (T,B).

A version of Eilenberg's Variety Theorem [11, 24] holds. The proof is standard.

THEOREM 6.6 *The function that maps a variety* $\mathbf{V}$ *of mg-pairs to the class of (regular) languages whose syntactic mg-pair is in* $\mathbf{V}$ *is an order isomorphism from the lattice of varieties of mg-pairs onto the lattice of literal varieties of regular languages. The inverse of this function takes a literal language variety* $\mathcal{L}$ *to the variety of mg-pairs containing all mg-pairs that only accept languages in* $\mathcal{L}$.

It follows by Proposition 4.4 that the function $\mathbf{V} \mapsto \mathcal{L}$ maps a variety $\mathbf{V}$ to the class of all languages that can be recognized by the members of $\mathbf{V}$. Moreover, the inverse assignment takes a literal variety to the variety generated by the syntactic mg-pairs of the languages in $\mathcal{L}$.

The following fact is well-known.

LEMMA 6.7 *Let* $L \subseteq A^*$ *be a regular language. Then every language in* A^* *recognizable by the syntactic morphism* η_L *is a boolean combination of quotients of* L.

LEMMA 6.8 *Let* $L \subseteq A^*$ *be a regular language and let* B *denote an alphabet. Then every language in* B^* *recognizable by the syntactic mg-pair of* L *is the inverse image under a literal morphism* $B^* \to A^*$ *of a language in* A^* *which is a boolean combination of quotients of* L.

Proof. Let $h : (B^*, B) \to (M_L, \eta_L(A))$ be a morphism and let $K \subseteq B^*$ with $h^{-1}(h(K)) = K$. Since η_L is surjective and (B^*, B) is free, there exists a literal morphism $\varphi : B^* \to A^*$ such that

$$h \quad = \quad (B^*, B) \xrightarrow{\varphi} (A^*, A) \xrightarrow{\eta_L} (M_L, \eta_L(A)).$$

Thus, $K = \varphi^{-1}(K')$, where $K' = \eta_L^{-1}(h(K))$. By Lemma 6.7, K' is a boolean combination of quotients of L. □

For a class $\mathcal{L}$ of regular languages, we let $\mathbf{K}_{\mathcal{L}}$ denote the class of all syntactic mg-pairs of the languages in $\mathcal{L}$.

PROPOSITION 6.9 • *Suppose that $\mathcal{L}$ is a class of regular languages and $\mathbf{K}_{\mathcal{L}}$ is the class of all syntactic mg-pairs of the members of $\mathcal{L}$. Then $\mathbf{Lin}(\mathcal{L}) \subseteq \mathbf{Lin}(\mathbf{K}_{\mathcal{L}})$. Moreover, if $\mathcal{L}$ is closed with respect to (left and right) quotients, then $\mathbf{Lin}(\mathcal{L}) = \mathbf{Lin}(\mathbf{K}_{\mathcal{L}})$.*

- *For a nonempty class $\mathbf{K}$ of mg-pairs let $\mathbf{V}$ denote the variety of mg-pairs generated by $\mathbf{K}$. Then $\mathbf{Lin}(\mathbf{K}) = \mathbf{Lin}(\mathbf{V})$.*
- *Suppose that $\mathcal{L}$ is a class of regular languages and $\mathbf{V}$ is the variety of all mg-pairs generated by the syntactic mg-pairs of the languages in $\mathcal{L}$. We have that $\mathbf{Lin}(\mathcal{L}) =\subseteq \mathbf{Lin}(\mathbf{V})$. Moreover, when $\mathcal{L}$ is closed with respect to (left and right) quotients, then $\mathbf{Lin}(\mathcal{L}) = \mathbf{Lin}(\mathbf{V})$.*

Proof. The inclusion $\mathbf{Lin}(\mathcal{L}) \subseteq \mathbf{Lin}(\mathbf{K}_{\mathcal{L}})$ is obvious. Assume now that $\mathcal{L}$ is closed with respect to taking quotients. Then, by Lemma 6.8, every language recognizable by some mg-pair in $\mathbf{K}_{\mathcal{L}}$ is the inverse image with respect to a literal morphism of a boolean combination of quotients of a language in $\mathcal{L}$. It follows from Proposition 2.4 that $\mathbf{Lin}(\mathbf{K}_{\mathcal{L}}) \subseteq \mathbf{Lin}(\mathcal{L})$.

As for the second claim, the inclusion $\mathbf{Lin}(\mathbf{K}) \subseteq \mathbf{Lin}(\mathbf{V})$ is obvious. For the reverse inclusion, note that any language that can be recognized by an mg-pair in $\mathbf{V}$ is a boolean combination of languages recognizable by the mg-pairs in $\mathbf{K}$, (use Proposition 6.5), and then apply Proposition 2.4.

The last claim is immediate from the first two. □

7 Lindström quantifiers and the block product

In this section, we assume that $\mathbf{K}$ is a class of mg-pairs. We let $\mathbf{V}$ denote the variety of mg-pairs generated by $\mathbf{K}$. The quantification depth $qd(\varphi)$ of a formula of $\mathbf{Lin}(\mathbf{K})$ is defined to be the length of the longest chain of nested quantifiers in the formula. For each $n \geq 0$, we let $\mathbf{Lin}_n(\mathbf{K})$ denote the the class of languages definable by formulas φ of this logic with $qd(\varphi) \leq n$. By the same argument as above, we have that $\mathbf{Lin}_n(\mathbf{K}) = \mathbf{Lin}_n(\mathbf{V})$. Our main result in this section will show that language is in $\mathbf{Lin}_n(\mathbf{K})$ iff it can be recognized by an mg-pair in $\mathbf{V}^{(n)}$.

PROPOSITION 7.1 *Suppose that all languages recognizable by (T, B) belong to $\mathbf{Lin}_n(\mathbf{K})$, for some $n \geq 0$, and suppose that $(S, A) \in \mathbf{K}$. Then any language recognizable in any double semidirect product $(R, A\times B) = (S, A)**(T, B)$ belongs to $\mathbf{Lin}_{n+1}(\mathbf{K})$.*

Proof. Let Σ be an alphabet and let

$$h \ : \ \Sigma^* \to R$$

denote a monoid homomorphism with $h(\Sigma) \subseteq A\times B$, so that h is a morphism $(\Sigma^*, \Sigma) \to (R, A\times B)$. It suffices to show that $L = h^{-1}(r_0)$ is in $\mathbf{Lin}_{n+1}(\mathbf{K})$, for each $r_0 = (s_0, t_0) \in R$.

For each $\sigma \in \Sigma$, let $s_\sigma \in A$ denote the left-hand component of $h(\sigma)$. We have

$$h(w) \ = \ \left(\sum_{w=w'\sigma w''} \pi(h(w'))s_\sigma\pi(h(w'')),\ \pi(h(w)) \right),$$

for all $w \in \Sigma^*$, where π denotes the projection $R \to T$, $\pi((s,t)) = t$, for all $(s,t) \in R$. Note that each $\pi(h(w'))s_\sigma\pi(h(w''))$ belongs to A. Since the composite of h and π is a homomorphism $\Sigma^* \to T$ with $\pi(h(\Sigma)) \subseteq B$, it follows from our assumptions that for each $t \in T$ there is a sentence α_t of $\mathbf{Lin}(\mathbf{K})$ with $qd(\alpha_t) \leq n$ such that for all words $w \in \Sigma^*$, $\pi(h(w)) = t$ iff $w \models \alpha_t$. For each $a \in A$ let $\varphi_a(x)$ be the formula in the free variable x,

$$\bigvee_{t's_\sigma t''=a} P_\sigma(x) \wedge \alpha_{t'}[< x] \wedge \alpha_{t''}[> x],$$

where $\alpha_{t'}[< x]$ and $\alpha_{t''}[> x]$ denote relativizations of $\alpha_{t'}$ and $\alpha_{t''}$, respectively, defined in the usual way, cf. [31]. Then let ψ be the sentence

$$Q_K x.\langle\varphi_a(x)\rangle_{a\in A},$$

where $K \subseteq A^*$ denotes the regular language recognized by (S, A) with the element s_0 and the morphism $(A^*, A) \to (S, A)$ which is the identity function on A. It is clear from the construction that ψ defines the set of all words $w \in \Sigma^*$ such that the left-hand component of $h(w)$ is s_0. Thus, $\alpha_{t_0} \wedge \psi$ defines L. Moreover, $qd(\alpha_{t_0} \wedge \psi) \leq n+1$. □

Suppose that Σ is a finite alphabet and V is a finite set of variables. Given a morphism $h : ((\Sigma \times P(V))^*, \Sigma \times P(V)) \to (M, A)$ and a set $F \subseteq M$, where (M, A) is an mg-pair, the language of V-structures (over Σ) recognized by

(M,A) with h and F consists of all V-structures $u \in (\Sigma \times P(V))^*$ such that $h(u) \in F$. In other words, it is the intersection of the language recognized by (M,A) with h and F with the language of all V-structures over Σ.

Suppose now that $K \subseteq \Delta^* = \{b_1, \ldots, b_k\}^*$, and consider a formula

$$\psi \quad = \quad Q_K x.(\varphi_{b_i})_{b_i \in \Delta},$$

where each $\varphi_{b_i} = \varphi_{b_i}(x, y_1, \ldots, y_k)$ is a formula of $\mathbf{Lin(K)}$ over the alphabet Σ whose free variables are among $x, y_1, \ldots, y_k$. Let (M,A) denote the syntactic mg-pair of K, or any mg-pair by which K can be recognized, and for each $b_i \in \Delta$, let (N_i, B_i) denote an mg-pair recognizing the language of $(V \cup \{x\})$-structures $L_i = L_{\varphi_i} \subseteq (\Sigma \times P(V \cup \{x\}))^*$ defined by φ_i, where $V = \{y_1, \ldots, y_k\}$.

Proposition 7.2 *The language L of V-structures over Σ defined by ψ can be recognized by the block product*

$$(M,A)\Box[(N_1,B_1) \times \ldots \times (N_k,B_k)].$$

Proof. Let (N,B) denote the product $(N_1,B_1) \times \ldots \times (N_k,B_k)$, so that $B = B_1 \times \ldots \times B_k$ and N is the submonoid generated by B in the direct product $N_1 \times \ldots \times N_k$. For each $i \in [k]$, let η_i denote a morphism

$$((\Sigma \times P(V \cup \{x\}))^*, \Sigma \times P(V \cup \{x\})) \quad \to \quad (N_i, B_i),$$

recognizing L_i, and let η_K denote the syntactic morphism of K, so that $\eta_K : (\Delta^*, \Delta) \to (M,A)$.

We define

$$\theta : ((\Sigma \times P(V))^*, \Sigma \times P(V)) \quad \to \quad (M,A)\Box(N,B)$$

by

$$\theta((a,X)) \quad = \quad (F_{(a,X)}, \eta_1((a,X)), \ldots, \eta_k((a,X))),$$

$(a,X) \in \Sigma \times P(V)$, where for all $n_1, n_1' \in N_1, \ldots, n_k, n_k' \in N_k$ such that $(n_1, \ldots, n_k), (n_1', \ldots, n_k') \in N$,

$$\begin{aligned} F_{(a,X)}((n_1, \ldots, n_k), (n_1', \ldots, n_k')) = \eta_K(b_i) \Leftrightarrow \\ \Leftrightarrow \quad n_i\eta_i((a, X \cup \{x\}))n_i' \in \eta_i(L_i) \\ \text{and } n_i\eta_j((a, X \cup \{x\}))n_j' \notin \eta_j(L_j),\ j < i. \end{aligned}$$

Note that we indeed have that

$$F_{(a,X)} \quad \in \quad A^{N \times N}$$

and

$$\theta((a,X)) \quad \in \quad A^{N \times N} \times B_1 \times \ldots \times B_k.$$

Let $w = (a_1, X_1) \ldots (a_n, X_n) \in (A \times P(V))^*$ be a V-structure and write F_i for $F_{(a_i, X_i)}$, for all $i \in [n]$. Then we have

$$\begin{aligned} \theta(w) \quad &= \quad (F_1, \eta_1((a_1,X_1)), \ldots, \eta_k((a_1,X_1))) \ldots \\ &= \quad \ldots (F_n, \eta_1((a_n,X_n)), \ldots, \eta_k((a_n,X_n))) \\ &= \quad (F, \eta_1(w), \ldots, \eta_k(w)), \end{aligned}$$

where

$$\begin{aligned} &F((1,\ldots,1),(1,\ldots,1)) = \\ &= \prod_{i=1}^{n} F_i((\eta_1((a_1,X_1)\ldots(a_{i-1},X_{i-1})), \ldots, \eta_k((a_1,X_1)\ldots(a_{i-1},X_{i-1}))), \\ &\qquad (\eta_1((a_{i+1},X_{i+1})\ldots(a_n,X_n)), \ldots, \eta_k((a_{i+1},X_{i+1})\ldots(a_n,X_n)))) \\ &= \prod_{i=1}^{n} G_i. \end{aligned}$$

Now, for each $i \in [n]$, G_i is $\eta_K(b_j)$ for the least j such that

$$\eta_j((a_1,X_1)\ldots(a_i, X_i \cup \{x\})\ldots(a_n,X_n)) \quad \in \quad \eta_j(L_j),$$

i.e., the least j with

$$(a_1,X_1)\ldots(a_i, X_i \cup \{x\})\ldots(a_n,X_n) \quad \in \quad L_j.$$

(Note that there is always such a j.) Thus, $F((1,\ldots,1),(1,\ldots,1)) = \eta_K(\overline{w})$ for the word $\overline{w} = w_1 \ldots w_n$, where for each i, $w_i = b_j$ for the least (and by agreement unique) j with $(a_1,X_1)\ldots(a_i, X_i \cup \{x\})\ldots(a_n,X_n) \models \varphi_{b_j}$. Thus, the language L defined by ψ is exactly the language recognized by θ with those elements $(F, n_1, \ldots, n_k)$ of the block product satisfying $F((1,\ldots,1),(1\ldots,1)) \in \eta_K(K)$. □

REMARK 7.3 Let π denote the projection

$$(M,A)\Box[(N_1,B_1)\times\ldots\times(N_k,B_k)] \quad\to\quad (N_1,B_1)\times\ldots\times(N_k,B_k),$$

and for each $i\in[k]$, let π_i denote the projection

$$(N_1,B_1)\times\ldots\times(N_k,B_k) \quad\to\quad (N_i,B_i).$$

The morphism θ constructed above has the property that the composite

$$\begin{aligned}(\Sigma\times P(V))^* &\xrightarrow{\theta} (M,A)\Box[(N_1,B_1)\times\ldots\times(N_k,B_k)]\\ &\xrightarrow{\pi} (N_1,B_1)\times\ldots\times(N_k,B_k)\\ &\xrightarrow{\pi_i} (N_i,B_i)\end{aligned}$$

is the morphism $(\Sigma\times P(V))^*\to(N_i,B_i)$ obtained by restricting the function $\eta_i:(\Sigma\times P(V\cup\{x\}))^*\to N_i$ to $(\Sigma\times P(V))^*$. In particular, for each $i\in[k]$, the restriction of the above composite morphism to Σ^* agrees with the restriction of η_i to Σ^*. (We regard Σ as a subset of $\Sigma\times P(V)$ which in turn is a subset of $\Sigma\times P(V\cup\{x\})$.) Thus, if each $(N_i',B_i')=(\eta_i(\Sigma^*),\eta_i(\Sigma))$ belongs to $\mathbf{W}$, for some variety $\mathbf{W}$, then $(\theta(\Sigma^*),\theta(\Sigma))$ belongs to $\mathbf{V}\Box\mathbf{W}$, since $(\theta(\Sigma^*),\theta(\Sigma))$ embeds in $(M,A)\Box[(N_1',B_1')\times\ldots\times(N_k',B_k')]$.

Recall that $\hat{\mathbf{V}}$ denotes the least variety of mg-pairs containing $\mathbf{V}$ (or $\mathbf{K}$) closed with respect to the double semidirect product, or the block product. By Corollary 6.4, we have $\hat{\mathbf{V}}=\bigcup_{n\geq 0}\mathbf{V}^{(n)}$, where $\mathbf{V}^{(0)}$ is the class of trivial mg-pairs and $\mathbf{V}^{(n+1)}=\mathbf{V}{\star\star}\mathbf{V}^{(n)}=\mathbf{V}\Box\mathbf{V}^{(n)}$.

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

THEOREM 7.4 *For every alphabet Σ and language $L\subseteq\Sigma^*$, and for every $n\geq 0$, we have $L\in\mathbf{Lin}_n(\mathbf{K})$ iff the syntactic mg-pair of L belongs to $\mathbf{V}^{(n)}$. Thus, $L\in\mathbf{Lin}(\mathbf{K})$ iff the syntactic mg-pair of L belongs to $\hat{\mathbf{K}}=\hat{\mathbf{V}}$, the least closed variety containing $\mathbf{K}$, or $\mathbf{V}$.*

Proof. Suppose first that the syntactic mg-pair (M_L,A_L) of L belongs to $\mathbf{V}^{(n)}$, for some $n\geq 0$. When $n=0$, (M_L,A_L) is trivial and thus L is either the empty set or Σ^*. In either case, L can be defined by a sentence of quantification depth 0, namely by false or true, proving that $L\in\mathbf{Lin}_n(\mathbf{K})$. We

proceed by induction on n. When $n>0$, we have that L can be recognized by a double semidirect product

$$(S,A){\star\star}(T,B),$$

where $(S,A)\in\mathbf{V}$ and $(T,B)\in\mathbf{V}^{(n-1)}$. By the induction hypothesis, every language recognizable by (T,B) is in $\mathbf{Lin}_{n-1}(\mathbf{K})$. Thus, by Proposition 7.1, we have that L belongs to $\mathbf{Lin}_n(\mathbf{K})$.

Suppose now that φ is a formula of $\mathbf{Lin}(\mathbf{K})$ over the alphabet Σ with free variables included in the finite set V. Let $qd(\varphi)=n$. We argue by induction on the structure of φ to show that L_φ can be recognized by a morphism $\theta:(\Sigma\times P(V))^*\to(M,A)$ such that $(\theta(\Sigma^*),\theta(\Sigma))$ belongs to $\mathbf{V}^{(n)}$. When φ is an atomic formula then any two words in Σ^* are equivalent with respect to the syntactic congruence of L_φ. Thus, $(\eta_{L_\varphi}(\Sigma^*),\eta_{L_\varphi}(\Sigma))$ is trivial and is thus in $\mathbf{V}^{(0)}$. Suppose that φ is $\varphi_1\vee\varphi_2$, and suppose that L_{φ_i} can be recognized by the morphism $\theta_i:(\Sigma\times P(V))^*\to(M_i,A_i)$ such that $(\theta_i(\Sigma^*),\theta_i(\Sigma))$ belongs to $\mathbf{V}^{(n)}$, $i=1,2$. Then L_φ can be recognized by the target pairing

$$\begin{aligned}\theta=\langle\theta_1,\theta_2\rangle:(\Sigma\times P(V))^* &\to (M_1,A_1)\times(M_2,A_2)\\ (\sigma,X) &\mapsto (\theta_1((\sigma,X)),\ \theta_2((\sigma,X))).\end{aligned}$$

Since

$$(\theta(\Sigma^*),\theta(\Sigma)) \quad<\quad (\theta_1(\Sigma^*),\theta_1(\Sigma))\times(\theta_2(\Sigma^*),\theta_2(\Sigma)),$$

and since varieties are closed with respect to direct product and division, it follows that $(\theta(\Sigma^*),\theta(\Sigma))\in\mathbf{V}^{(n)}$. When φ is of the form $\neg\psi$, the result follows by using that L_φ and L_ψ can be recognized by the same mg-pairs. Suppose finally that $\varphi=Q_Kx.\langle\varphi_{b_1},\ldots,\varphi_{b_k}\rangle$, where $B\subseteq\Delta^*$, $\Delta=\{b_1,\ldots,b_k\}$, is a language recognized by some mg-pair (M,A) in $\mathbf{K}$, and where each φ_{b_i} is a formula of $\mathbf{Lin}(\mathbf{K})$ of quantifier depth at most $n-1$. By the induction hypothesis, each L_{φ_i} can be recognized by a morphism

$$\theta_i:(A\times P(V\cup\{x\}))^* \quad\to\quad (N_i,B_i)$$

such that $(\theta_i(\Sigma^*),\theta_i(\Sigma))\in\mathbf{V}^{(n-1)}$. But then, by Remark 7.3, L_φ can be recognized by a morphism

$$\theta:(\Sigma\times P(V))^* \quad\to\quad (M,A)\Box[(N_1,B_1)\times\ldots\times(N_k,B_k)]$$

such that $(\theta(\Sigma^*),\theta(\Sigma)$ is in $\mathbf{V}^{(n)}$. □

Corollary 7.5 *A language belongs to* $\mathbf{FO}(\mathbf{K})$ *iff its syntactic mg-pair is in* $\hat{\mathbf{K}}_1$, *where* $\mathbf{K}_1 = \mathbf{K} \cup \{U_1\}$.

Corollary 7.6 *Suppose that* $\mathcal{L}$ *is a class of regular languages closed with respect to taking quotients. Let* $\mathbf{V}$ *denote the smallest closed variety of mg-pairs containing the syntactic mg-pairs of the languages in* $\mathcal{L}$. *Then* $\mathbf{Lin}(\mathcal{L}) = \mathbf{Lin}(\mathbf{V})$.

Proof. We know that $\mathbf{Lin}(\mathcal{L}) = \mathbf{Lin}(\mathbf{K}_{\mathcal{L}})$ (Proposition 6.9). Since $\hat{\mathbf{K}}_{\mathcal{L}} = \mathbf{V}$, by Theorem 7.4 we have $\mathbf{Lin}(\mathcal{L}) = \mathbf{Lin}(\mathbf{V})$. □

Corollary 7.7 *For any closed variety* $\mathbf{V}$, *it holds that* $\mathbf{Lin}(\mathbf{V})$ *is the literal variety corresponding to* $\mathbf{V}$ *by the Variety Theorem (Theorem 6.6). Thus, when* $\mathbf{V}_1$ *and* $\mathbf{V}_2$ *are literal varieties,* $\mathbf{V}_1 \subseteq \mathbf{V}_2$ *iff* $\mathbf{Lin}(\mathbf{V}_1) \subseteq \mathbf{Lin}(\mathbf{V}_2)$.

Proof. Let $\mathcal{V}$ denote the literal variety corresponding to $\mathbf{V}$ by the Variety Theorem. Then for all languages $\mathcal{V}$, we have $L \in \mathcal{V}$ iff the syntactic mg-pair of L is in $\mathbf{V}$ iff $L \in \mathbf{Lin}(\mathbf{V})$, since $\mathbf{V} = \hat{\mathbf{V}}$. □

Corollary 7.8 *For any class* $\mathbf{K}$ *of mg-pairs, it holds that* $\mathbf{Lin}(\mathbf{Lin}(\mathbf{K})) = \mathbf{Lin}(\mathbf{K})$.

Proof. It is clear that $\mathbf{Lin}(\mathbf{K}) \subseteq \mathbf{Lin}(\mathbf{Lin}(\mathbf{K}))$, see Proposition 2.3. As for the reverse inclusion, let $L \in \mathbf{Lin}(\mathbf{Lin}(\mathbf{K}))$. Then $L \in \mathbf{Lin}(\mathbf{K}')$, where $\mathbf{K}'$ is the class of all syntactic mg-pairs of the languages in $\mathbf{Lin}(\mathbf{K})$. But by Theorem 7.4, any such mg-pair belongs to the least closed variety $\mathbf{V}$ containing $\mathbf{K}$. Thus, $L \in \mathbf{Lin}(\mathbf{V})$. But, again by Theorem 7.4, the syntactic mg-pair of L belongs to the least closed variety containing $\mathbf{V}$, which is $\mathbf{V}$. We conclude that the syntactic mg-pair of L is in $\mathbf{V}$. Thus, again by Theorem 7.4, $L \in \mathbf{Lin}(\mathbf{K})$. □

Corollary 7.9 *The operation* $\mathcal{L} \mapsto \mathbf{Lin}(\mathcal{L})$ *is a closure operation on those classes of regular languages closed with respect to quotients.*

Proof. We have already seen that $\mathcal{L} \subseteq \mathbf{Lin}(\mathcal{L})$. It is clear that $\mathbf{Lin}(\mathcal{L}) \subseteq \mathbf{Lin}(\mathcal{L}')$ whenever $\mathcal{L} \subseteq \mathcal{L}'$. If $\mathcal{L}$ is closed with respect to quotients, then, by

Proposition 6.9, $\mathbf{Lin}(\mathcal{L}) = \mathbf{Lin}(\mathbf{K}_{\mathcal{L}})$. Thus, by Corollary 7.8, $\mathbf{Lin}(\mathbf{Lin}(\mathcal{L})) = \mathbf{Lin}(\mathbf{Lin}(\mathbf{K}_{\mathcal{L}})) = \mathbf{Lin}(\mathbf{K}_{\mathcal{L}}) = \mathbf{Lin}(\mathcal{L})$. □

In the same way, the operation $\mathcal{L} \mapsto \mathbf{FO}(\mathcal{L})$, defined on classes of regular languages $\mathcal{L}$ closed with respect to quotients, is also a closure operation.

Call a class $\mathcal{L}$ of regular languages *Lindström closed* if $\mathcal{L}$ is closed with respect to quotients and $\mathbf{Lin}(\mathcal{L}) = \mathcal{L}$ (or equivalently, $\mathbf{Lin}(\mathcal{L}) \subseteq \mathcal{L}$).

The above results can be summarized in a single statement.

Corollary 7.10 *The assignment* $\mathbf{V} \mapsto \mathcal{V} = \mathbf{Lin}(\mathbf{V})$ *defines an order isomorphism between closed varieties of mg-pairs and Lindsröm closed classes* $\mathcal{V}$ *of regular languages. Moreover, this map is just the restriction of the bijection provided by the Variety Theorem to closed varieties of mg-pairs that maps a variety* $\mathbf{V}$ *to the class of those regular languages whose syntactic monoid is in* $\mathbf{V}$.

Note that the inverse assignment maps a Lindström closed class $\mathcal{L}$ to the closed variety generated by $\mathbf{K}_{\mathcal{L}}$.

Corollary 7.11 *The assignment* $\mathbf{V} \mapsto \mathbf{FO}(\mathbf{V})$ *defines an order isomorphism between closed varieties of mg-pairs containing* U_1 *and Lindsröm closed classes of regular languages containing* $\mathbf{FO}$. *Moreover, this map is just the restriction of the bijection provided by the Variety Theorem to closed varieties of mg-pairs containing* U_1.

The least closed variety containing (the mg-pair corresponding to) U_1 is the variety of aperiodic mg-pairs. See below.

8 The Krohn-Rhodes theorem

In this section we first review a version of the fundamental theorem of Krohn and Rhodes [17, 1] which involves the double semidirect product (block product). The original formulation of the theorem involved the wreath product, and its automata theoretic equivalent, the cascade product. Our presentation follows Straubing [31], Appendix A. Then we review some results from Dömösi, Ésik [9] and Ésik [13] and apply them in conjunction with the Krohn-Rhodes Theorem to obtain descriptions of the class of mg-pairs that can be generated from a given class of mg-pairs by the double semidirect product and division. The results of this section will be applied in Section 9 in the characterization of the expressive power of Lindström quantifiers with respect to regular languages.

All monoids considered in this section are assumed to be finite. As before, we will identify a monoid M with the mg-pair (M, M). For a class $\mathbf{K}$ of monoids (mg-pairs), we let $\hat{\mathbf{K}}$ denote the least class of monoids (mg-pairs) containing $\mathbf{K}$ which is closed with respect to the double semidirect product (block product) and division. It is a simple matter to show that when $\mathbf{K}$ is a class of monoids and $\mathbf{K}_1$ is the corresponding class of mg-pairs (M, M), $M \in \mathbf{K}$, then an mg-pair (S, A) belongs to $\hat{\mathbf{K}}_1$ iff S belongs to $\hat{\mathbf{K}}$. Moreover, $\hat{\mathbf{K}}_1$ is the class of all mg-pairs (M, A) such that $M \in \hat{\mathbf{K}}$. When $\mathbf{K}$ denotes a class of mg-pairs, then we let $\overline{\mathbf{K}}$ denote the class of all monoid components of the mg-pairs in $\mathbf{K}$. When $\mathbf{K}$ is a class of monoids and M is a monoid, $M < \mathbf{K}$ means that there is a monoid $S \in \mathbf{K}$ with $M < S$. Recall that U_1 denotes a two-element monoid which is not a group. Moreover, recall that a (finite) group G is called *simple* if it is nontrivial and has no nontrivial normal subgroup. It is a simple matter to show that a (finite) monoid M is group iff $U_1 < M$ does not hold. In fact, if M is not a group, then it contains an idempotent e other than the identity element 1. Then $\{1, e\}$ is a submonoid of M which is isomorphic to U_1.

THEOREM 8.1 Krohn–Rhodes Theorem

- *Part 1. The following two conditions are equivalent for a nontrivial (finite) monoid M.*
  1. *M is a simple group or isomorphic to U_1.*
  2. *For every class $\mathbf{K}$ of monoids, if $M \in \hat{\mathbf{K}}$ then $M < \mathbf{K}$.*
- *Part 2. Suppose that M is a monoid and $\mathbf{K}$ is a class of monoids containing at least one monoid which covers U_1. Then $M \in \hat{\mathbf{K}}$ iff for every simple group G, if $G < M$ then $G < \mathbf{K}$. Moreover, when $\mathbf{K}$ is a class of groups, so that $U_1 \not< \mathbf{K}$, then $M \in \mathbf{K}$ iff M is a group and for every simple group G, if $G < M$ then $G < \mathbf{K}$.*

REMARK 8.2 If one defines $\hat{\mathbf{K}}$ as the closure of $\mathbf{K}$ with respect to the semidirect product (or wreath product) and division, then the result remains true provided that in Part 1 both U_1 and the three element monoid U_2 with two right zero elements are allowed, and if U_1 is replaced by U_2 in Part 2. In fact, the original formulations of the Krohn-Rhodes Theorem (Krohn, Rhodes [17], Arbib [1]) used the wreath product and/or the corresponding automata theoretic notion of cascade composition.

Recall that a monoid M is called *aperiodic* Eilenberg [11], Pin [24], if it contains no nontrivial group, or equivalently, if no nontrivial group (or simple group) divides M. Moreover, recall that M *solvable*, cf. Pin [24], Straubing [31], if every group included in M is solvable. (Such a group does not necessarily contain the identity element of M.) We denote the class of all aperiodics and the class of all groups by $\mathbf{A}$ and $\mathbf{G}$, respectively. Moreover, we denote by $\mathbf{GSol}$ the class of solvable groups and by $\mathbf{MSol}$ the class of solvable monoids, i.e., those monoids that contain only solvable groups. Moreover, when P is a set of prime numbers, we denote by $\mathbf{GSol}_P$ the subclass of $\mathbf{GSol}$ determined by those solvable groups whose order is a product of primes in P. The variety $\mathbf{MSol}_P$ is defined likewise. Note that when P is empty, $\mathbf{MSol}_P = \mathbf{A}$, and when P is the set of all prime numbers, then $\mathbf{GSol}_P = \mathbf{GSol}$ and $\mathbf{MSol}_P = \mathbf{MSol}$. More generally, when $\mathbf{S}$ denotes a class of simple groups closed with respect to division, we let $\mathbf{G_S}$ denote the class of groups all of whose simple group divisors lie in $\mathbf{S}$. The class $\mathbf{M_S}$ is the class of all those monoids which only contain groups in $\mathbf{G_S}$. When $\mathbf{S}$ is empty, $\mathbf{M_S}$ is the class of all aperiodic monoids. Moreover, when $\mathbf{S}$ is the class of cyclic simple groups, then $\mathbf{G_S} = \mathbf{GSol}$ and $\mathbf{M_S} = \mathbf{MSol}$. And when $\mathbf{S}$ contains all simple groups, $\mathbf{G_S}$ is the class $\mathbf{G}$ of all groups, and $\mathbf{M_S} = \mathbf{M}$ is the class of all monoids. By the Krohn-Rhodes theorem, the above classes are all closed varieties (of monoids), i.e., they are closed with respect to the double semidirect product (block product) and division.

COROLLARY 8.3 *Let $\mathbf{K}$ denote a class of monoids and let $\mathbf{S}$ denote a class of simple groups closed with respect to division.*

- *$\mathbf{G_S} \subseteq \hat{\mathbf{K}}$ iff $G < \mathbf{K}$ holds for all $G \in \mathbf{S}$. Moreover, $\hat{\mathbf{K}} = \mathbf{G_S}$ iff $\mathbf{K} \subseteq \mathbf{G_S}$ and $G < \mathbf{K}$ holds for all $G \in \mathbf{S}$ iff $U_1 \not< \mathbf{K}$ and for all simple groups G it holds that $G < \mathbf{K}$ iff $G \in \mathbf{S}$.*

- $\mathbf{M_S} \subseteq \hat{\mathbf{K}}$ *iff* $U_1 < \mathbf{K}$ *and* $G < \mathbf{K}$, *for all* $G \in \mathbf{S}$. *Moreover,* $\hat{\mathbf{K}} = \mathbf{M_S}$ *iff* $\mathbf{K} \subseteq \mathbf{M_S}$ *and* $U_1 < \mathbf{K}$ *and* $G < \mathbf{K}$, *for all* $G \in \mathbf{S}$ *iff* $U_1 < \mathbf{K}$ *and for all simple groups* G *it holds that* $G < \mathbf{K}$ *iff* $G \in \mathbf{S}$.

In particular, we obtain:

- $\hat{\mathbf{K}} = \mathbf{M}$ iff the monoid U_1 as well as each (non-abelian) simple group is covered by some monoid in $\mathbf{K}$.
- $\mathbf{G} \subseteq \hat{\mathbf{K}}$ iff $G < \mathbf{K}$ holds for all simple groups G. Moreover, $\hat{\mathbf{K}} = \mathbf{G}$ iff $\hat{\mathbf{K}} \subseteq \mathbf{G}$ and $G < \mathbf{K}$ holds for all simple groups G iff $U_1 \not< \mathbf{K}$ and $G < \mathbf{K}$ holds for all simple groups G.
- $\mathbf{A} \subseteq \hat{\mathbf{K}}$ iff $U_1 < \mathbf{K}$. Moreover, $\hat{\mathbf{K}} = \mathbf{A}$ iff $\mathbf{K} \subseteq \mathbf{A}$ and $U_1 < \mathbf{K}$ iff no (simple) group divides $\mathbf{K}$ and $U_1 < \mathbf{K}$.
- $\mathbf{GSol_P} \subseteq \hat{\mathbf{K}}$ iff $G < \mathbf{K}$ holds for all cyclic groups Z_p of prime order $p \in P$. Moreover, $\hat{\mathbf{K}} = \mathbf{GSol_P}$ iff $\mathbf{K} \subseteq \mathbf{GSol_P}$ and $G < \mathbf{K}$ holds for all cyclic groups Z_p with $p \in P$ iff $U_1 \not< \mathbf{K}$ an for each simple group G we have $G < \mathbf{K}$ iff G is cyclic with order in P.

We now turn our attention to mg-pairs. Below we identify any class $\mathbf{K}$ of monoids with the class of all mg-pairs (M, A) such that $M \in \mathbf{K}$. Thus, for example, $\mathbf{M}$ also denotes the class of all mg-pairs, $\mathbf{MSol}$ the class all mg-pairs whose monoid component is solvable, etc.

Suppose that M is a monoid, (S, A) is an mg-pair, and $n \geq 1$. Following Dömösi, Ésik [9], we say that M *divides* (S, A) *in length* $n \geq 1$, denoted $M|^{(n)}(S, A)$, if S contains a subsemigroup T that maps homomorphically onto M under a homomorphism $h : T \to M$ such that each set $h^{-1}(m)$, $m \in M$ contains an n-fold product of elements in A (i.e., an element in A^n). We define $M|(S, A)$ iff there is some n with $M|^{(n)}(S, A)$. The following fact is clear.

PROPOSITION 8.4 *Let* T *denote the submonoid generated by* A^n *in* S. *Then* $M|^{(n)}(S, A)$ *iff* $M < (T, A^n)$.

PROPOSITION 8.5 *If* $M|^{(n)}(S, A)$ *then there is a multiple* m *of* n *and a subsemigroup* T *of* S *contained in* A^m *such that* M *is a homomorphic image of* T.

This is shown in Ésik [13], cf. Lemma 3.3. Since M is a monoid, T can be chosen to be a monoid as well. However, T may not contain the identity element of S. Also, when M is a group, T can be assumed to be a group as well.

PROPOSITION 8.6 *Suppose that* (S, A) *is an mg-pair and* M *is a non-abelian simple group or the monoid* U_1. *If* $M < S$ *then* $M|(S, A)$.

This is a very particular case of Proposition 3.5 in Ésik [13]. See also Maurer, Rhodes [21]. The case $M = U_1$ is obvious.

PROPOSITION 8.7 *Suppose that* M *is* U_1 *or a simple group, moreover, suppose that* $M|(S, A)\star\star(T, B)$, *where* (S, A) *and* (T, B) *are mg-pairs. Then either* $M|(S, A)$ *or* $M|(T, B)$.

Proof. When M is U_1, or a non-abelian simple group, this follows from the first part of the Krohn-Rhodes theorem and Proposition 8.6. Thus, to complete the proof, it suffices to establish the claim for (cyclic) groups of prime order. So suppose that G is a cyclic group with prime order p such that $G|(S, A)\star\star(T, B) = (M, A \times B)$. By Proposition 8.5, G is a homomorphic image of a group in M all of whose members are n-fold products of the elements in $A \times B$, for some $n \geq 1$. By the first Sylow theorem, this group in turn contains a cyclic subgroup H of order p. Let (f, e) denote the identity element of H and let (s, t) denote any element of H different from (f, e). If $t \neq e$ then clearly t generates a cyclic subgroup of order p in T (whose identity element is e), all of whose elements are n-fold products over B. We conclude that $G|^{(n)}(T, B)$. So suppose now that $t = e$. Then the right-hand component of each element of H is e. It follows as in Straubing [31], p. 64, or Eilenberg [11], v. B, p. 143, that the function $(s, e) \mapsto ese$, $(s, e) \in H$ is an injective homomorphism $H \to S$. Since each $(s, e) \in H$ is an n-fold product over $A \times B$, it follows that each element ese, $(s, e) \in H$ is an n-fold product over A. Thus, we have $G|^{(n)}(S, A)$. □

Suppose that M is a monoid and $\mathbf{K}$ is a class of mg-pairs, and suppose that $n \geq 1$. Below we will write $M|^{(n)}\mathbf{K}$ ($M|\mathbf{K}$, respectively) to denote that there exists an mg-pair $(S, A) \in \mathbf{K}$ such that $M|^{(n)}(S, A)$ ($M|(S, A)$, respectively).

COROLLARY 8.8 *Let* M *be* U_1 *or a simple group, and let* $\mathbf{K}$ *denote a class of mg-pairs. If* $M|\hat{\mathbf{K}}$ *then* $M|\mathbf{K}$.

A *counter* of length n is an mg-pair consisting of a cyclic group Z_n of order n and a singleton generating set. We let $(Z_n, \{a\})$ denote a counter of order n. A nontrivial counter is a counter of length > 1. Given a monoid M and an element $a \in M$, the *period* of a is the least positive integer p such that there exists some m with $a^m = a^{m+p}$, i.e., the period of the cyclic semigroup generated by a.

LEMMA 8.9 *Suppose that a nontrivial counter divides a double semidirect product $(S, A \times B) = (M, A){\star\star}(N, B)$. Then there is a nontrivial counter which divides (M, A) or (N, B).*

Proof. By assumption, there is some $(a, b) \in A \times B \subseteq S$ with period $n > 1$. If the period of b is > 1, then we are done. So suppose that the period of b is 1, i.e., $b^k = b^{k+1}$, for some k. Consider the sequence

$$(a, b), (a, b)^2, \ldots$$

which is, by assumption, ultimately periodic with period n. But for all $\ell \geq 0$,

$$\begin{aligned}(a,b)^{2k+\ell} &= (ab^{2k+\ell-1} + bab^{2k+\ell-2} + \ldots + b^{2k+\ell-1}a,\ b^{2k+\ell}) \\ &= (ab^k + bab^k + \ldots + \overbrace{b^kab^k + \ldots + b^kab^k}^{\ell \text{ times}} + b^kab^{k-1} + \ldots + b^ka,\ b^k),\end{aligned}$$

showing that n must be the same as the period of $b^kab^k \in A$ in M. It follows that a counter of length n divides (M, A). □

COROLLARY 8.10 *Given a class* $\mathbf{K}$ *of mg-pairs,* $\hat{\mathbf{K}}$ *contains a nontrivial counter iff a nontrivial counter divides an mg-pair in* $\mathbf{K}$.

REMARK 8.11 The proof of Lemma 8.9 can easily be modified to show that a counter of length $n > 1$ divides a double semidirect product $(M, A){\star\star}(N, B)$ iff there exist integers p, q such that n divides pq, moreover, a counter of length p divides (N, B), and a counter of length q divides (M', A^p), where M' is the submonoid of M generated by A^p.

It is shown in Dömösi, Ésik [9] that if $M|^n(S, A)$, then M divides a wreath product

$$(S, A) \circ (R, B) \circ (Z_n, \{a\}), \qquad (2)$$

where R is aperiodic. (Actually this fact is shown in [9] for finite automata and the cascade product, moreover, only a special type of aperiodic automata, namely *definite automata* are needed in the construction. The wreath product is associative, this is why no parentheses appear in (2).) Thus, by the Krohn-Rhodes Theorem and Proposition 6.3, we have:

PROPOSITION 8.12 *Suppose that $M|^n(S, A)$. Then $M \in \hat{\mathbf{K}}$, where* $\mathbf{K}$ *consists of U_1, a counter of length n, and the mg-pair (S, A).*

COROLLARY 8.13 *Suppose that* $\mathbf{K}$ *is a class of mg-pairs such that $U_1 \in \hat{\mathbf{K}}$ and such that for each simple group G with $G < \overline{\mathbf{K}}$ there exists some $n \geq 1$ with $G|^{(n)}\mathbf{K}$ and $(Z_n, \{a\}) \in \hat{\mathbf{K}}$. Then an mg-pair (M, A) belongs to $\hat{\mathbf{K}}$ iff for every simple group G, if $G < M$ then $G < \overline{\mathbf{K}}$.*

Proof. One direction is immediate from the Krohn–Rhodes Theorem. The other direction follows from the Krohn–Rhodes Theorem and Proposition 8.12. Indeed, assume that every simple group divisor of M divides the underlying monoid of an mg-pair in $\mathbf{K}$. Let $G_1, \ldots, G_k$ denote, up to isomorphism, all of the simple group divisors of M. By assumption, for each i there exists n_i with $G_i|^{(n_i)}\mathbf{K}$ and $(Z_{n_i}, \{a\}) \in \hat{\mathbf{K}}$. Since also $U_1 \in \hat{\mathbf{K}}$, it follows from Proposition 8.12 that $G_i \in \hat{\mathbf{K}}$. Since this holds for all $i \in [n]$, thus, by the Krohn–Rhodes Theorem, $(M, A) \in \hat{\mathbf{K}}$. □

COROLLARY 8.14 *Suppose that* $\mathbf{K}$ *is a class of mg-pairs such that $\hat{\mathbf{K}}$ contains U_1 as well as all the counters. Moreover, suppose that for all simple groups G, if $G < \overline{\mathbf{K}}$ then $G|\mathbf{K}$. Then an mg-pair (M, A) belongs to $\hat{\mathbf{K}}$ iff for every simple group G, if $G < M$ then $G|\mathbf{K}$.*

Call a class $\mathbf{K}$ of mg-pairs *group-complete* if every group divides some monoid in $\overline{\mathbf{K}}$. Since every group embeds in a (non-abelian) simple group, by Proposition 8.6 we have that $\mathbf{K}$ is group-complete iff every (non-abelian) simple group divides in equal lengths some mg-pair in $\mathbf{K}$.

COROLLARY 8.15 *Let* $\mathbf{K}$ *be a class of mg-pairs. Then $\hat{\mathbf{K}}$ is the class of all mg-pairs iff the following hold:*

1. *$\hat{\mathbf{K}}$ contains U_1 and all counters.*

2. $\mathbf{K}$ *is group-complete.*

REMARK 8.16 It is clear that $\hat{\mathbf{K}}$ contains all counters iff it contains all counters of prime power length.

EXAMPLE 8.17 For each $n \geq 1$, let S_n denote the *symmetric group* of all permutations of the set $[n]$. When $n \geq 3$, S_n is generated by the cyclic permutation $\rho = (12\ldots n)$ and the transposition $\pi = (12)$. Hence, $(S_n, \{\rho, \pi\})$ is an mg-pair. Let $\mathbf{K}$ consist of (U_1, U_1) and all the mg-pairs $(S_n, \{\rho, \pi\})$, $n \geq 3$. Then both conditions of Corollary 8.15 are satisfied, so that $\hat{\mathbf{K}}$ is the class of all mg-pairs.

Note that $(S_n, \{\rho, \pi\})$ is just the mg-pair of the automaton whose states are the integers in the set $[n]$ which has two input letters that induce the permutations ρ and π, respectively.

EXAMPLE 8.18 We modify the previous example to show that $U_1 \in \mathbf{K}$ and group-completeness of $\mathbf{K}$ do not imply that $\hat{\mathbf{K}}$ contains any counter. So let $\mathbf{K}$ consist of U_1 and, for each $n \geq 3$, the mg-pair of the following automaton Q_n with $2n + 1$ states. The state set of Q_n consists of the integers $1, 2, \ldots, 2n$ and the state $2'$, and there are four input letters, a, b, c, d. For each state q and letter x, $qx = q$, except for the following cases.

$$\begin{aligned}
(2i-1)a &= 2i, \quad i \in [n] \\
(2i)b &= 2i+1, \quad i \in [n-1] \\
(2n)b &= 1 \\
1c &= 2 \\
2d &= 3 \\
3c &= 2' \\
2'd &= 1.
\end{aligned}$$

Thus, on the set of odd integers, the word ab induces the cyclic permutation $(13...(2n-1))$ and cd induces the transposition (13). Thus, $\mathbf{K}$ is group-complete and contains U_1. However, no non-trivial counter divides any mg-pair in $\mathbf{K}$, since for each Q_n, any letter $x \in \{a, b, c, d\}$ induces the same function as x^2, and similarly for U_1. (See Corollary 8.10).

This example can be modified to show that there is a class $\mathbf{K}$ of mg-pairs which is group-complete, contains U_1 as well as each counter whose length is not a multiple of a given prime number p, but such that no counter of length p belongs to $\hat{\mathbf{K}}$.

COROLLARY 8.19 *Let* $\mathbf{K}$ *be a class of mg-pairs. Then* $\hat{\mathbf{K}} \supseteq \mathbf{MSol}$ *iff the following hold:*

1. $\hat{\mathbf{K}}$ *contains* U_1 *and all counters.*

2. For each (cyclic) group G *of prime order it holds that* $G|\mathbf{K}$.

Moreover, $\hat{\mathbf{K}} = \mathbf{MSol}$ *iff the above conditions hold and* $\mathbf{K} \subseteq \mathbf{MSol}$.

More generally, we have:

COROLLARY 8.20 *Let* $\mathbf{K}$ *be a class of mg-pairs and let* $\mathbf{S}$ *be a class of simple groups containing the cyclic groups of prime order and closed with respect to division. Then* $\hat{\mathbf{K}} \supseteq \mathbf{M_S}$ *iff the following hold:*

1. $\hat{\mathbf{K}}$ *contains* U_1 *and all counters.*

2. For each $G \in \mathbf{S}$ *it holds that* $G|\mathbf{K}$.

Moreover, $\hat{\mathbf{K}} = \mathbf{M_S}$ *iff the above conditions hold and* $\mathbf{K} \subseteq \mathbf{M_S}$.

8.1 Monoid-generator pairs with identity

Given an integer $n > 1$, the mg-pair $(Z_n, \{a, 1\})$ consists of a cyclic group Z_n of order n and the set $\{a, 1\}$, where a is a generator of Z_n and 1 is the identity element of Z_n.

PROPOSITION 8.21 *For each* n*, it holds that* Z_n *divides a direct power of* $(Z_n, \{a, 1\})$.

Proof. Map each $(n-1)$-tuple $(a^{k_1}, \ldots, a^{k_{n-1}})$ in the direct power

$$(Z_n, \{a, 1\})^{n-1} = (Z_n^{n-1}, \{a, 1\}^{n-1})$$

to the element

$$a^{k_1} a^{2k_2} \ldots a^{(n-1)k_{n-1}}$$

in Z_n. □

PROPOSITION 8.22 *Suppose that* $\mathbf{K}$ *is a class of mg-pairs and* P *is a set of prime numbers. Then* $\hat{\mathbf{K}} \supseteq \mathbf{GSol}_P$ *iff for each prime number* $p \in P$ *it holds that* $(Z_p, \{a, 1\}) \in \hat{\mathbf{K}}$. *Moreover,* $\hat{\mathbf{K}} = \mathbf{GSol}_P$ *iff the above condition holds and* $\mathbf{K} \subseteq \mathbf{GSol}_P$.

Proof. This follows from Proposition 8.21 and Corollary 8.3. □

The mg-pairs $(Z_n, \{a, 1\})$ have the property that the identity element appears in the generator set. We call such mg-pairs as *mg-pairs with identity*, or *mgi-pairs*, for short.

LEMMA 8.23 *Suppose that S is a monoid and (M, A) is an mgi-pair. If $S < M$ then there exists some n_0 such that $S|^{(n)}(M, A)$ holds for all $n \geq n_0$. In particular, $S|(M, A)$.*

Proof. Since A is a set of generators for M, each element of M can be written as a product of elements of A. Let n_0 be the maximum number of factors in such a representation for each element of M. Since the identity element is in A, it follows that each $m \in M$ is the product of n generators, for every $n \geq n_0$. Thus, $M|^{(n)}(M, A)$. It follows that $S|^{(n)}(M, A)$ for all monoids S with $S < M$. □

LEMMA 8.24 *Let $\mathbf{K}$ be a class of mgi-pairs. Then $U_1 \in \hat{\mathbf{K}}$ iff $\overline{\mathbf{K}}$ contains a monoid which is not a group.*

Proof. If $U_1 \in \hat{\mathbf{K}}$, then, by the Krohn-Rhodes Theorem, it holds that $U_1 < \overline{\mathbf{K}}$. But this is possible only if $\overline{\mathbf{K}}$ contains a monoid which is not a group.

Let (M, A) be an mg-pair in $\mathbf{K}$ which is not a group. Since (M, A) is an mgi-pair, A contains the identity element 1. Moreover, since M is not a group and A generates M, there exists some $a \in A$ such that $a^k \neq 1$, for all $k \geq 1$. Thus, U_1 is a homomorphic image of the submonoid M' of M generated by a. It follows that (U_1, U_1) is a morphic image of $(M', \{a, 1\})$. This proves that $U_1 < \mathbf{K}$, so that $U_1 \in \hat{\mathbf{K}}$. □

LEMMA 8.25 *Let $\mathbf{K}$ denote a class of mgi-pairs. The following conditions are equivalent.*

1. *There is a nontrivial counter $(Z_n, \{a\})$ with $(Z_n, \{a\}) < \mathbf{K}$.*

2. *$\hat{\mathbf{K}}$ contains a nontrivial counter.*

3. *$\hat{\mathbf{K}}$ contains an infinite number of non-isomorphic counters.*

Proof. We already know that the first and second conditions are equivalent (Corollary 8.10). The third condition clearly implies the second. To complete the proof we show that the first condition implies the third. Given that $(Z_n, \{a\}) < \mathbf{K}$, where $n > 1$, also $(Z_n, \{a, 1\}) < \mathbf{K}$. Thus, by Proposition 8.22, $(Z_m, \{a\}) \in \hat{\mathbf{K}}$ for all powers m such that every prime divisor of m divides n. □

PROPOSITION 8.26 *Suppose that $\mathbf{K}$ is a class of mgi-pairs such that $\overline{\mathbf{K}}$ contains a monoid which is not a group and there is a nontrivial counter that divides an mg-pair in $\mathbf{K}$. Then $\hat{\mathbf{K}}$ contains and mg-pair (M, A) iff every simple group divisor of M divides a monoid in $\overline{\mathbf{K}}$.*

Proof. By the Krohn-Rhodes Theorem, $\hat{\mathbf{K}}$ contains at most those mg-pairs (M, A) such that every simple group divisor of M divides a monoid $\overline{\mathbf{K}}$. In the rest of the proof, we show that every such mg-pair is in indeed in $\hat{\mathbf{K}}$.

By Lemma 8.24 we have $U_1 \in \hat{\mathbf{K}}$. Consider now an arbitrary monoid M such that $M < \overline{\mathbf{K}}$. By Lemma 8.23 there exists some n_0 such that $M|^{(n)}(S, A)$ for all $n \geq n_0$. Also, by Lemma 8.25, there exist an infinite number of counters in $\hat{\mathbf{K}}$ of pairwise different length. We conclude that for some n, both $M|^{(n)}(S, A)$ and $(Z_n, \{a\}) \in \hat{\mathbf{K}}$ hold. Thus, by Proposition 8.12 and Proposition 6.3 we have that $(M, M) \in \hat{\mathbf{K}}$. In particular, it follows that whenever G is a simple group with $G < \overline{\mathbf{K}}$, then $(G, G) \in \hat{\mathbf{K}}$. Since also $U_1 \in \hat{\mathbf{K}}$, it follows from the Krohn-Rhodes Theorem that $\hat{\mathbf{K}}$ contains every mg-pair (M, A) such that every simple group divisor of M divides a monoid in $\overline{\mathbf{K}}$. □

COROLLARY 8.27 *Suppose that $\mathbf{K}$ is a class of mgi-pairs. Then $\hat{\mathbf{K}}$ is the class of all mg-pairs iff the following hold:*

1. *$\overline{\mathbf{K}}$ contains a monoid which is not a group.*

2. *There is a nontrivial counter which divides an mg-pair in $\mathbf{K}$.*

3. *$\mathbf{K}$ is group-complete.*

COROLLARY 8.28 *Suppose that $\mathbf{S}$ is a nonempty class of simple groups closed with respect to division. Let $\mathbf{K}$ be a class of mgi-pairs. Then $\hat{\mathbf{K}} \supseteq \mathbf{M_S}$ iff the following hold:*

1. $\overline{\mathbf{K}}$ *contains a monoid which is not a group.*

2. There is a nontrivial counter which divides an mg-pair in $\mathbf{K}$*.*

3. $G < \mathbf{K}$ *holds for each* $G \in \mathbf{S}$*.*

Moreover, $\hat{\mathbf{K}} = \mathbf{M_S}$ *iff the above conditions hold and* $\mathbf{K} \subseteq \mathbf{M_S}$*.*

COROLLARY 8.29 *For a class* $\mathbf{K}$ *of mgi-pairs,* $\hat{\mathbf{K}} \supseteq \mathbf{A}$ *iff* $\mathbf{K}$ *contains an mgi-pairs whose underlying monoid is not a group. Moreover,* $\hat{\mathbf{K}} = \mathbf{A}$ *if this condition holds and* $\mathbf{K} \subseteq \mathbf{A}$*.*

9 Completeness

Call a class $\mathbf{K}$ of mg-pairs *Lindström-complete* if $\mathbf{Lin}(\mathbf{K})$ is the class of all regular languages, and *expressively complete* if $\mathbf{FO}(\mathbf{K})$ is the class of all regular languages. In this section we combine results from Section 7 and Section 8 to obtain characterizations of Lindström-complete and expressively complete classes. By Proposition 6.9, this also gives a characterization of those classes $\mathcal{L}$ of regular languages, closed with respect to quotients, for which $\mathbf{Lin}(\mathcal{L})$ ($\mathbf{FO}(\mathcal{L})$, respectively) is the class of all regular languages. We call such classes of regular languages Lindström complete (expressively complete, respectively) as well. We will also include relative completeness results.

In the following propositions, $\mathbf{K}$ denotes a class of mg-pairs.

PROPOSITION 9.1 $\mathbf{K}$ *is Lindström-complete iff* $\mathbf{K}$ *is group-complete and* $\hat{\mathbf{K}}$ *contains* U_1 *and all counters. Moreover,* $\mathbf{K}$ *is expressively complete iff* $\mathbf{K}$ *is group-complete and* $\hat{\mathbf{K}}$ *contains all counters.*

Proof. By Corollary 7.10, $\mathbf{Lin}(\mathbf{K})$ is the class of all regular languages iff the corresponding closed variety, $\hat{\mathbf{K}}$ is the class of all mg-pairs. By Corollary 8.15, $\hat{\mathbf{K}}$ is the class of all mg-pairs iff $\mathbf{K}$ is group-complete and $\hat{\mathbf{K}}$ contains U_1 and all counters, proving the first claim.

The second claim follows in the same way by applying Corollary 7.11 and Corollary 8.15. □

EXAMPLE 9.2 The class $\mathbf{K}$ presented in Example 8.17 is Lindström complete. Thus there exists a Lindström complete class of mg-pairs with two generators. On the other hand, no class of mg-pairs with a single generator is group-complete, hence no such class is Lindström complete, or expressively complete.

COROLLARY 9.3 *There exists no finite Lindström complete class of mg-pairs. Each Lindström complete class contains an infinite number of mg-pairs with 2 or more generators.*

This corollary is related to results proved in [5]. The following fact is a variant of a result proved in [3].

PROPOSITION 9.4 *Suppose that* $\mathbf{S}$ *is a class of simple groups closed with respect to division.*

1. $\mathbf{Lin}(\mathbf{K})$ *contains all regular languages whose syntactic monoid is in* $\mathbf{M_S}$ *iff* $\hat{\mathbf{K}}$ *contains (the mg-pairs corresponding to)* U_1 *and the simple groups in* $\mathbf{S}$*. Moreover,* $\mathbf{Lin}(\mathbf{K})$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{M_S}$ *iff each simple group divisor of the monoid component of any mg-pair in* $\mathbf{K}$ *is in* $\mathbf{S}$ *and* $\hat{\mathbf{K}}$ *contains (the mg-pairs) corresponding to* U_1 *and the simple groups in* $\mathbf{S}$*.*
2. $\mathbf{FO}(\mathbf{K})$ *contains all regular languages whose syntactic monoid is in* $\mathbf{M_S}$ *iff* $\hat{\mathbf{K}}$ *contains (the mg-pairs corresponding to) the simple groups in* $\mathbf{S}$*. Moreover,* $\mathbf{FO}(\mathbf{K})$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{M_S}$ *iff each simple group divisor of the monoid component of any mg-pair in* $\mathbf{K}$ *is in* $\mathbf{S}$ *and* $\hat{\mathbf{K}}$ *contains (the mg-pairs corresponding to) the simple groups in* $\mathbf{S}$*.*

Proof. By Corollary 7.10, $\mathbf{Lin}(\mathbf{K})$ is the literal variety corresponding to $\hat{\mathbf{K}}$, the least closed variety containing $\mathbf{K}$. Thus, the first claim follows by Corollary 8.3. For the second claim, apply Corollary 7.11 and Corollary 8.3. □

PROPOSITION 9.5 *Suppose that* $\hat{\mathbf{K}}$ *contains* U_1*, the counters and has the following property: For every simple group* G*, if* $G < \overline{\mathbf{K}}$ *then* $G|\mathbf{K}$*. Then a language* L *is in* $\mathbf{Lin}(\mathbf{K})$ *iff every simple group divisor of the syntactic monoid of* L *divides* $\mathbf{K}$*.*

Proof. By Corollary 8.14, $\hat{\mathbf{K}} = \mathbf{M}_\mathbf{S}$, where $\mathbf{S}$ is the class of all simple groups G with $G < \mathbf{K}$. Thus, the result follows from Theorem 7.4. □

PROPOSITION 9.6 *Suppose that $\mathbf{S}$ is a class of simple groups closed with respect to division containing all cyclic groups of prime order.*

1. $\mathbf{Lin(K)}$ *contains all regular languages whose syntactic monoid is in $\mathbf{M}_\mathbf{S}$ iff $\hat{\mathbf{K}}$ contains U_1 and the counters, and for each simple group G in $\mathbf{S}$ it holds that $G|\mathbf{K}$. Moreover,* $\mathbf{Lin(K)}$ *is the class of all regular languages whose syntactic monoid is in $\mathbf{M}_\mathbf{S}$ iff the above condition holds and $\mathbf{K} \subseteq \mathbf{M}_\mathbf{S}$.*
2. $\mathbf{FO(K)}$ *contains all regular languages whose syntactic monoid is in $\mathbf{M}_\mathbf{S}$ iff $\hat{\mathbf{K}}$ contains the counters and for each simple group G in $\mathbf{S}$ it holds that $G|\mathbf{K}$. Moreover,* $\mathbf{FO(K)}$ *is the class of all regular languages whose syntactic monoid is in $\mathbf{M}_\mathbf{S}$ iff the previous condition holds and $\mathbf{K} \subseteq \mathbf{M}_\mathbf{S}$.*

Proof. The first claim follows from Theorem 7.4 and Corollary 8.20. The second claim follows from the first applied to the class $\mathbf{K} \cup \{U_1\}$, and by noting that a counter belongs to the least closed variety containing $\mathbf{K} \cup \{U_1\}$ iff it belongs to the least closed variety containing $\mathbf{K}$. □

PROPOSITION 9.7 *Suppose that $\mathbf{S}$ is a class of simple groups closed with respect to division.* $\mathbf{Lin(K)}$ *contains all regular languages whose syntactic monoid is in $\mathbf{G}_\mathbf{S}$ iff $\hat{\mathbf{K}}$ contains (the mg-pairs corresponding to) the simple groups in $\mathbf{S}$. Moreover,* $\mathbf{Lin(K)}$ *is the class of all regular languages whose syntactic monoid is in $\mathbf{G}_\mathbf{S}$ iff $\hat{\mathbf{K}}$ contains (the mg-pairs corresponding to) the simple groups in $\mathbf{S}$ and $\mathbf{K} \subseteq \mathbf{G}_\mathbf{S}$.*

Proof. By Corollary 7.10 and Corollary 8.3. □

By taking $\mathbf{S}$ to be the class of all cyclic groups of prime order, from Proposition 9.6 we obtain:

PROPOSITION 9.8

1. $\mathbf{Lin(K)}$ *contains all regular languages whose syntactic monoid is in* $\mathbf{MSol}$ *iff $\hat{\mathbf{K}}$ contains U_1 and the counters, and for every prime number p it holds that $Z_p|\mathbf{K}$. Moreover,* $\mathbf{Lin(K)}$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{MSol}$ *iff $\mathbf{K} \subseteq \mathbf{MSol}$, $\hat{\mathbf{K}}$ contains U_1 and the counters, and for every prime number p it holds that $Z_p|\mathbf{K}$.*
2. $\mathbf{FO(K)}$ *contains all regular languages whose syntactic monoid is in* $\mathbf{MSol}$ *iff $\hat{\mathbf{K}}$ contains the counters, and for every prime number p it holds that $Z_p|\mathbf{K}$. Moreover,* $\mathbf{FO(K)}$ *is the class of all regular languages whose syntactic monoid is solvable iff $\mathbf{K} \subseteq \mathbf{MSol}$, $\hat{\mathbf{K}}$ contains the counters, and for every prime number p it holds that $Z_p|\mathbf{K}$.*

PROPOSITION 9.9 *Let P denote a set of prime numbers.* $\mathbf{Lin(K)}$ *contains all regular languages whose syntactic monoid is in $\mathbf{GSol}_P$ iff for every prime number $p \in P$ it holds that $(Z_p, \{a, 1\}) \in \hat{\mathbf{K}})$. Moreover,* $\mathbf{Lin(K)}$ *is the class of all regular languages whose syntactic monoid is in $\mathbf{GSol}_P$ iff $\mathbf{K} \subseteq \mathbf{GSol}_P$ and for every prime number $p \in P$ it holds that $(Z_p, \{a, 1\}) \in \hat{\mathbf{K}})$.*

Proof. This follows from Proposition 9.7 and Proposition 8.22. □

We also have:

PROPOSITION 9.10 *Suppose that P is a set of prime numbers.*

1. $\mathbf{Lin(K)}$ *is the class of all regular languages whose syntactic monoid is in $\mathbf{MSol}_P$ iff $\mathbf{K} \subseteq \mathbf{MSol}_P$, moreover, $U_1 \in \hat{\mathbf{K}}$ and $(Z_p, \{a, 1\}) \in \hat{\mathbf{K}}$, for all $p \in P$.*
2. $\mathbf{FO(K)}$ *is the class of all regular languages whose syntactic monoid is in $\mathbf{GSol}_P$ iff $\mathbf{K} \subseteq \mathbf{GSol}_P$, moreover, $(Z_p, \{a, 1\}) \in \hat{\mathbf{K}}$, for all $p \in P$.*

In particular, when P is empty, we have:

PROPOSITION 9.11 $\mathbf{Lin(K)} = \mathbf{FO}$ *iff $U_1 \in \hat{\mathbf{K}}$.*

We now translate some of the above results to classes of regular languages closed with respect to quotients. Let $\mathcal{L}$ denote such a class.

COROLLARY 9.12 *$\mathcal{L}$ is Lindström-complete iff every (non-abelian simple) group divides the syntactic monoid of some language in $\mathcal{L}$, moreover, $\mathbf{Lin}(\mathcal{L})$ contains $K_\exists$ and all of the one-letter languages $(a^m)^*$, $m \geq 2$.*

COROLLARY 9.13 *$\mathcal{L}$ is expressively complete iff every (non-abelian simple) group divides the syntactic monoid of some language in $\mathcal{L}$, moreover,* $\mathbf{Lin}(\mathcal{L})$ *contains all of the one-letter languages $(a^m)^*$, $m \geq 2$.*

COROLLARY 9.14 *Suppose that P is a set of prime numbers.*

1. $\mathbf{Lin}(\mathcal{L})$ *contains all regular languages whose syntactic monoid is in* $\mathbf{MSol}_P$ *iff* $\mathbf{Lin}(\mathcal{L})$ *contains $K_\exists$ and all languages*

$$C_p^0 \quad = \quad \{u \in \{a,b\}^* : |u|_a \equiv 0 \bmod p\},$$

where p is any prime in P. Moreover, $\mathbf{Lin}(\mathcal{L})$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{MSol}_P$ *iff the above conditions hold and the syntactic monoid of each language in $\mathcal{L}$ belongs to* $\mathbf{MSol}_P$.

2. $\mathbf{FO}(\mathcal{L})$ *contains all regular languages whose syntactic monoid is in* $\mathbf{MSol}_P$ *iff* $\mathbf{FO}(\mathcal{L})$ *contains the languages C_p^0, where p is any prime in P. Moreover,* $\mathbf{FO}(\mathcal{L})$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{MSol}_P$ *iff the above condition holds and the syntactic monoid of each language in $\mathcal{L}$ belongs to* $\mathbf{MSol}_P$.

COROLLARY 9.15 *Suppose that P is a set of prime numbers. Then* $\mathbf{Lin}(\mathcal{L})$ *contains all regular languages whose syntactic monoid is in* $\mathbf{GSol}_P$ *iff* $\mathbf{Lin}(\mathcal{L})$ *contains the languages C_p^0, where p is any prime in P. Moreover,* $\mathbf{Lin}(\mathcal{L})$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{GSol}_P$ *iff the above condition holds and the syntactic monoid of each language in $\mathcal{L}$ belongs to* $\mathbf{GSol}_P$.

COROLLARY 9.16 $\mathbf{Lin}(\mathcal{L}) = \mathbf{FO}$ *iff* $K_\exists \in \mathbf{Lin}(\mathcal{L})$.

COROLLARY 9.17

1. $\mathbf{Lin}(\mathcal{L})$ *contains all regular languages whose syntactic monoid is in* $\mathbf{MSol}$ *iff*
   (a) *every cyclic group of prime order divides in equal length the syntactic mg-pair of a language in $\mathcal{L}$,*
   (b) $\mathbf{Lin}(\mathcal{L})$ *contains $K_\exists$ and all of the one-letter languages $(a^m)^*$, $m \geq 2$.*

Moreover, $\mathbf{Lin}(\mathcal{L})$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{MSol}$ *iff the above conditions hold and the syntactic mg-pair of each language in $\mathcal{L}$ belongs to* $\mathbf{MSol}$.

2. $\mathbf{FO}(\mathcal{L})$ *contains all regular languages whose syntactic monoid is in* $\mathbf{MSol}$ *iff*
   (a) *every cyclic group of prime order divides in equal length the syntactic monoid of a language in $\mathcal{L}$,*
   (b) $\mathbf{Lin}(\mathcal{L})$, *or* $\mathbf{FO}(\mathcal{L})$ *contains all of the one-letter languages $(a^m)^*$, $m \geq 2$.*

Moreover, $\mathbf{FO}(\mathcal{L})$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{MSol}$ *iff the above conditions hold and the syntactic monoid of each language in $\mathcal{L}$ belongs to* $\mathbf{MSol}$.

9.1 Completeness and padding

Our characterizations become simpler when $\mathbf{K}$ is a class of mgi-pairs that we assume in the rest of this section. We only present three results and skip the proofs that use Corollaries 8.27, 8.28 and 8.29.

So let $\mathbf{K}$ denote a class of mg-pairs.

PROPOSITION 9.18

1. $\mathbf{K}$ *is Lindström complete iff* $\mathbf{K}$ *is group-complete, contains an mgi-pair whose underlying monoid is not a group, moreover, there exists some $n > 1$ with $(Z_n, \{a\}) < \mathbf{K}$.*
2. $\mathbf{K}$ *is expressively complete iff* $\mathbf{K}$ *is group-complete and there exists some $n > 1$ with $(Z_n, \{a\}) < \mathbf{K}$.*

PROPOSITION 9.19 *Suppose that* $\mathbf{S}$ *is a nonempty class of simple groups closed with respect to division.*

1. $\mathbf{Lin}(\mathbf{K})$ *contains the regular languages whose syntactic monoids are in* $\mathbf{M_S}$ *iff for each $G \in \mathbf{S}$ it holds that $G < \mathbf{K}$, moreover,* $\mathbf{K}$ *contains an mgi-pair whose underlying monoid is not a group and there exists some*

$n > 1$ *with* $(Z_n, \{a\}) < \mathbf{K}$. *Further,* $\mathbf{Lin}(\mathbf{K})$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{M_S}$ *iff the above conditions hold and* $\mathbf{K} \subseteq \mathbf{M_S}$.

2. $\mathbf{FO}(\mathbf{K})$ *contains the regular languages whose syntactic monoids are in* $\mathbf{M_S}$ *iff for each* $G \in \mathbf{S}$ *it holds that* $G < \mathbf{K}$ *and there exists some* $n > 1$ *with* $(Z_n, \{a\}) < \mathbf{K}$. *Further,* $\mathbf{Lin}(\mathbf{K})$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{M_S}$ *iff the above conditions hold and* $\mathbf{K} \subseteq \mathbf{M_S}$.

PROPOSITION 9.20 $\mathbf{Lin}(\mathbf{K}) \supseteq \mathbf{FO}$ *iff* $\mathbf{K}$ *contains an mgi-pair whose monoid component is not a group.*

Say that a class $\mathcal{L}$ of regular languages *admits padding* if for each $L \in \mathcal{L}$, $L \subseteq B^*$ there exists some $b_0 \in B$ such that for all words $u \in B^*$ we have $u \in L$ iff $h(u) \in L$, where $h : B^* \to B^*$ is the homomorphism such that $h(b_0)$ is the empty word and $h(b) = b$, for all $b \neq b_0$.

In the next three corollaries, we assume that $\mathcal{L}$ is closed with respect to quotients and admits padding.

COROLLARY 9.21

1. *$\mathcal{L}$ is Lindström-complete iff every (non-abelian simple) group divides the syntactic monoid of some language in $\mathcal{L}$, moreover, $\mathcal{L}$ contains a language whose syntactic monoid is not a group, and there is some $n > 1$ such that the one-letter language $(a^n)^*$ is the inverse image of a language in $\mathcal{L}$ under a literal homomorphism.*

2. *$\mathcal{L}$ is expressively complete iff every (non-abelian simple) group divides the syntactic monoid of some language in $\mathcal{L}$, moreover, there is some $n > 1$ such that the one-letter language $(a^n)^*$ is the inverse image of a language in $\mathcal{L}$ under a literal homomorphism.*

COROLLARY 9.22 *Suppose that* $\mathbf{S}$ *is a nonempty class of simple groups closed with respect to division.*

1. $\mathbf{Lin}(\mathcal{L})$ *contains the regular languages whose syntactic monoids are in* $\mathbf{M_S}$ *iff for each* $G \in \mathbf{S}$ *it holds that* $G < \mathbf{K}_\mathcal{L}$, *moreover,* $\mathbf{K}_\mathcal{L}$ *contains an*

mgi-pair whose underlying monoid is not a group and there exists some $n > 1$ *such that* $(a^n)^*$ *is the inverse image of a language in* $\mathcal{L}$ *under a literal homomorphism. Further,* $\mathbf{Lin}(\mathcal{L})$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{M_S}$ *iff the above conditions hold and* $\mathbf{K}_\mathcal{L} \subseteq \mathbf{M_S}$.

2. $\mathbf{FO}(\mathcal{L})$ *contains the regular languages whose syntactic monoids are in* $\mathbf{M_S}$ *iff for each* $G \in \mathbf{S}$ *it holds that* $G < \mathbf{K}_\mathcal{L}$ *and there exists some* $n > 1$ *such that* $(a^n)^*$ *is the inverse image of a language in* $\mathcal{L}$ *under a literal homomorphism. Further,* $\mathbf{Lin}(\mathcal{L})$ *is the class of all regular languages whose syntactic monoid is in* $\mathbf{M_S}$ *iff the above conditions hold and* $\mathbf{K}_\mathcal{L} \subseteq \mathbf{M_S}$.

COROLLARY 9.23 $\mathbf{Lin}(\mathcal{L}) \supseteq \mathbf{FO}$ *iff* $\mathcal{L}$ *contains a language whose syntactic monoid is not a group.*

10 Further work

A future version of this paper will contain more applications, including. e.g., some results presented in [14], and some generalizations of that results. Extensions of the main result to other structures will be considered in subsequent papers.

11 Acknowledgment

The results of this paper were obtained during the first author's visit at BRICS and the Department of Computer Science of Aalborg University. He is indebted to members of these institutions for their hospitality.

Both authors would like to thank Prof. John L. Rhodes for communicating to us an outline of the argument that proves Proposition 6.3 for monoid varieties. In the Appendix, we have modified this argument for varieties of mg-pairs.

References

[1] M. A. Arbib, Ed.: *Algebraic Theory of machines, languages, and Semigroups*, With a major contribution by Kenneth Krohn and John L. Rhodes, Academic Press, 1968.

[2] D. A. Mix Barrington, K. Compton, H. Straubing and D. Therien: Regular languages in NC^1, *J. Computer and System Sciences*, 44(1992), no. 3, 478–499.

[3] D. Barrington, N. Immerman and H. Straubing: On uniformity within NC^1, *J. Computer System Science*, 41(1990), 274–306.

[4] A. Baziramwabo, P. McKenzie and D. Therien: Modular temporal logic, in: *14th Annual IEEE Symposium on Logic in Computer Science, Trento, 1999*, IEEE Computer Society, 344–351.

[5] D. Beauquier and A. Rabinovitch: Monadic logic of order over naturals has no finite base, *Journal of Logic and Computation*, to appear.

[6] J. R. Büchi: Weak second order logic and finite automata, *Zeit. Math. Logik Grund. Math.*, 6(1960), 66–92.

[7] H.-J. Burtschick and H. Vollmer: Lindström quantifiers and leaf language definability, *International J. of Foundations of Computer Science*, 9(1998), 277-294.

[8] J. Cohen, D. Perrin, J.-E. Pin: On the expressive power of temporal logic, *J. Computer and System Sciences*, 46(1993), 271–294.

[9] P. Dömösi and Z. Ésik: Critical classes for the α_0-product, *Theoret. Comput. Sci.*, 61(1988), 17–24.

[10] H.-J. Ebbinghaus and J. Flum: *Finite Model Theory*, 2nd. ed., Springer, 1999.

[11] S. Eilenberg: *Automata, Languages, and Machines*, vol. A and B, Academic Press, 1974, 1976.

[12] C. Elgot: Decision problems of finite automata design and related arithmetics, *Trans. Amer. Math. Soc.*, 98(1961), 21–51.

[13] Z. Ésik: Results on homomorphic realization of automata by α_0-products, *Theoret. Comput. Sci.*, 87(1991), 229–249.

[14] Z. Ésik and M. Ito: Temporal logic with cyclic counting and the degree of aperiodicity of finite automata, *BRICS Report Series*, RS-01-53, 2001.

[15] M. Gałota, H. Vollmer: A generalization of the Büchi–Elgot–Trakhtenbrot theorem, in: *Computer Science Logic, 15th International Workshop, CSL 2001, Paris, 2001*, LNCS 2142, Springer, 355–368.

[16] N. Immerman: *Descriptive Complexity*, Graduate Texts in Computer Science, Springer-Verlag, New York, 1999.

[17] K. Krohn and J. Rhodes: Algebraic theory of machines. I. Prime decomposition theorem for finite semigroups and machines, *Trans. Amer. Math. Soc.*, 116(1965), 450–464.

[18] C. Lautemann, P. McKenzie and Th. Schwentick: The descriptive complexity approach to LOGCFL, *J. Computer System Science*, 62(2001), 629–652.

[19] P. Lindström: First order predicate logic with generalized quantifiers. *Theoria*, 32(1966), 186–195.

[20] P. McKenzie, Th. Schwentick, D. Therien, H. Vollmer: The many faces of a translation, in: *27th Automata, Languages and Programming, 27th International Colloquium, ICALP'00* LNCS 1853, 890–901.

[21] W. D. Maurer and J. L. Rhodes: A property of finite simple non-abelian groups, *Proc. Amer. Math. Soc.*, 16(1965), 552–554.

[22] R. McNaughton and S. Papert: *Counter-Free Automata*, MIT Press, 1971.

[23] T. Peichl and H. Vollmer: Finite automata with generalized acceptance criteria, in: *26th Automata, Languages and Programming, 26th International Colloquium, ICALP'99, Prague*, LNCS 1644, Springer, 605-614.

[24] J.-E. Pin: *Varieties of Formal Languages*, Plenum, 1986.

[25] J.-E. Pin: Logic, semigroups and automata on words, *Ann. Math. Artificial Intelligence*, 16(1996), 343–384.

[26] A. Pnueli: The temporal logic of programs, in: *18th IEEE Symp. Foundations of Computer Science, Providence, 1977*, 46–57.

[27] J. Rhodes: Undecidability, automata, and pseudovarieties of finite semigroups, *Int. J. of Algebra nd Computation,*, 9(1999), 455–473.

[28] J. Rhodes and B. Tilson: The kernel of monoid morphisms, *J. Pure and Appl. Alg.*, 62(1989), 227–268.

[29] M. P. Schützenberger: On finite monoids having only trivial subgroups, *Information and Control*, 8(1965), 190–194.

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

[31] H. Straubing: *Finite Automata, Formal Logic, and Circuit Complexity*, Birkhauser, 1994.

[32] H. Straubing, D. Therien: Regular languages defined with a bounded number of variables, in: *STACS 2001*, LNCS 2010, Springer, 2001, 555–562.

[33] H. Straubing, D. Therien and W. Thomas: Regular languages defined with generalized quantifiers, *Automata, languages and programming (Tampere, 1988)*, Lecture Notes in Comput. Sci., 317, Springer, Berlin, 1988, 561–575.

[34] H. Straubing, D. Therien and W. Thomas: Regular languages defined with generalized quantifiers, *Information and Computation*, 118(1995), 289–301.

[35] D. Therien: Classification of finite monoids: The language approach, *Theoretical Computer Science*, 14(1981), 195-208.

[36] D. Therien and Th. Wilke: Temporal logic and semidirect products: An effective characterization of the until hierarchy, in: *37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington*, IEEE Computer Society, 1996, 256-263.

[37] W. Thomas: Automata on infinite objects, in: *Handbook of Theoretical Computer Science*, Vol. B, Elsevier, Amsterdam, 1990, 133–191.

[38] W. Thomas: Languages, automata, and logic, in: *Handbook of Formal Languages*, Vol. 3, Springer, 1997, 389–455.

[39] B. A. Trakhtenbrot: Finite automata and logic of monadic predicates, *Doklady Akademii Nauk SSSR*, 140(1961), 326–329.

[40] J. Väänänen (Ed.): *Generalized Quantifiers and Computation*, LNCS 1754, Springer, 1997.

Appendix

In this appendix, we prove Proposition 6.3. The argument is an adaptation of the proof of the corresponding fact for monoid varieties, communicated to the authors by John Rhodes. In this section, by a monoid or category we will always mean a finite monoid, or category, respectively.

First, we recall from Rhodes, Tilson [28] the notion of the *kernel* K_φ of a monoid morphism $\varphi : M \to N$. It is a category constructed as a quotient of category W_φ defined as follows. The *objects* of W_φ are all ordered pairs $\mathbf{n} = (n_L, n_R)$ of elements n_L, n_R in the image $\varphi(M)$ of M. (We will follow the convention of [28] that if a boldface letter $\mathbf{x}$ denotes a pair of elements of a monoid, then x_L and x_R are the left and right hand components of this pair.) An *arrow* $\mathbf{n} \to \mathbf{n}'$ of W_φ takes the form $(n_L, (m,n), n'_R)$, where $m \in M$, $n \in N$ with $n = \varphi(m)$ are such that $n_L n = n'_L$ and $nn'_R = n_R$. (Thus, we could as well just write (n_L, m, n'_R), but we want to keep the notation consistent with that of [28]. The reason for the more complex notation of [28] is due to the fact that the kernel construction also applies to relational morphisms φ of monoids, whereas in this paper we only consider the particular case when φ is a function.) Note that n_R and n'_L and thus $\mathbf{n}$ and $\mathbf{n}'$ can be recovered from the notation $(n_L, (m,n), n'_R)$. Below we will sometimes just write (m,n) for $(n_L, (m,n), n'_R)$ when there is no danger of confusion. The composite of consecutive arrows $(n_L, (m,n), n'_R) : \mathbf{n} \to \mathbf{n}'$ and $(n'_L, (m',n'), n''_R) : \mathbf{n}' \to \mathbf{n}''$ is defined as $(n_L, (mm', nn'), n''_R)$. Note that the identity arrows of W_φ take the form $(n_L, (m,1), n_R)$ where $\varphi(m) = 1$. Each arrow $(n_L, (m,n), n'_R)$ induces a function

$$\begin{aligned} [n_L, (m,n), n'_R] : \varphi^{-1}(n_L) \times \varphi^{-1}(n'_R) &\to M \\ (m_L, m_R) &\mapsto m_L m m_R. \end{aligned}$$

The relation that identifies any two parallel arrows inducing the same function is shown to be a (category) congruence in [28]. The kernel K_φ is then defined as the quotient of W_φ with respect to this congruence. Following [28], we will denote a morphism of K_φ as $[n_L, (m,n), n'_R]$, or just $[m,n]$.

Suppose now that (M,A) and (N,B) are mg-pairs and φ is a morphism $(M,A) \to (N,B)$, so that φ is also a monoid homomorphism $M \to N$. Then we define the kernel of φ to be the pair (K_φ, A_φ), where K_φ is the category constructed above, and where A_φ is a distinguished collection of morphisms of K_φ: it consists of those morphisms $[a,b]$ of K_φ with $a \in A$. Since $b = \varphi(a)$, and since φ is a morphism of mg-pairs, it then follows that $b \in B$. Since A is a generating set of M and φ preserves the generators, it follows that (K_φ, A_φ) is a *category-generator pair*, or *cg-pair*: each arrow of K_φ is either an identity arrow or the composite of some arrows in A_φ.

Suppose that K is a category, N is a monoid, and φ is a relation from the arrows of K to N, viewed as a function from the arrows of K to the set of all subsets of N. We say that φ is a *covering* $K \to N$ if the following conditions hold:

- $\varphi(mm') \subseteq \varphi(m)\varphi(m')$, for all composable arrows m, m'.
- For all identity arrows e it holds that $1 \in \varphi(e)$.
- For all arrows m it holds that $\varphi(m) \neq \emptyset$.
- For all m, m', if $m \neq m'$ then $\varphi(m) \cap \varphi(m') = \emptyset$.

When K and N are equipped with generators, i.e., when (K,A) is a cg-pair and (N,B) is an mg-pair, then a covering $\varphi : (K,A) \to (N,B)$ also satisfies that for each $a \in A$ there is some $b \in B$ with $b \in \varphi(a)$. Since A is a set of generators, the third condition above becomes redundant. Note that each covering $\varphi : (K,A) \to (N,B)$ contains a covering φ' such that whenever $n \in \varphi'(m)$, it holds that either m is an identity arrow and $n = 1$, or there exist $a_1, \ldots, a_k \in A$, $b_1, \ldots, b_k \in B$, $k \geq 1$ with $b_1 \in \varphi(a_1), \ldots, b_k \in \varphi(a_k)$ such that m is the composite $a_1 \ldots a_k$ and n is $b_1 \ldots b_k$. The above definition also applies to one object categories K which may conveniently be identified with their hom-sets. In that case the concept reduces to the notion of covering defined earlier in Section 4: if (M,A) and (N,B) are mg-pairs and φ is a covering $(M,A) \to (N,B)$, then (M,A) is a morphic image of a sub mg-pair of (N,B). Note that each injective morphism $(M,A) \to (N,B)$ is a covering $(M,A) \to (N,B)$, moreover, the relational inverse of each

surjective morphism $(M,A) \to (N,B)$ is a covering $(N,B) \to (M,A)$, i.e., in the opposite direction.

The notion of covering can be generalized to a pair of categories, and in fact to cg-pairs. Given categories K and K', a covering $\varphi : K \to K'$ assigns an object to each object of K, a set $\varphi(m)$ of morphisms of K' to each morphism m of K, compatible with the object map, such that the obvious analogies of the above conditions hold. A covering $\varphi : (K,A) \to (K',B)$ between cg-pairs (K,A) and (K',B) also satisfies that for each arrow $a \in A$ there is an arrow $b \in B$ with $b \in \varphi(a)$. The composite of two coverings is defined in the expected way.

LEMMA 11.1 *The composite of coverings $\psi : K \to K'$ and $\psi' : K' \to K''$ is a covering $K \to K''$. Similarly, if $\psi : (K,A) \to (K',A')$ and $\psi' : (K',A') \to (K'',A'')$ are coverings, then the composite of ψ with ψ' is a covering $K \to K''$.*

The notion of covering is related to the double semidirect product by the Kernel Theorem of Rhodes, Tilson [28], also known as the Covering Lemma. We need a version of this result.

THEOREM 11.2 *Let $\varphi : (M,A) \to (N,B)$ be a morphism of mg-pairs, and let (V,C) be an mg-pair satisfying $(K_\varphi, A_\varphi) < (V,C)$. Then $(M,A) < (V,C)\Box(N,B)$.*

Proof. We follow the argument given in the proof of the Kernel Theorem (Theorem 7.4) in Rhodes, Tilson [28]. Let $\psi : (K_\varphi, A_\varphi) \to (V,C)$ be a covering. For each pair $m \in M$, $n \in N$ with $n \in \varphi(m)$, define

$$F(m,n) \quad = \quad \{f \in V^{N \times N} : f(n_1,n_2) \in \psi([n_1,(m,n),n_2]),\ n_1,n_2 \in \varphi(M)\}.$$

Then let the relation $\theta : M \to V\Box N$ be defined by

$$\theta(m) \quad = \quad \{(f,n) : n \in \varphi(m),\ f \in F(m,n)\}.$$

It is shown in [28] that θ is a covering $M \to V\Box N$. For each $m \in M$ let $\theta'(m) = \theta(m) \cap W$, where W denotes the monoid component of $(V,C)\Box(N,B)$, i.e., $(V,C)\Box(N,B) = (W, C^{N\times N} \times B)$. If we can show that for each $a \in A$ there is some $b \in B$ and $f \in C^{N\times N}$ with $(f,b) \in \theta(a)$, then, using the fact that W is a submonoid of $V\Box N$, it follows that θ' is a covering $(M,A) \to (V,C)\Box(N,B)$. But for a given a let $b = \varphi(a)$. Since ψ is a covering $(K_\varphi, A_\varphi) \to (V,C)$, for each $n_1, n_2 \in \varphi(M)$ there is some $c \in C$ with $c \in \psi([n_1,(a,b),n_2])$. So let f map each pair $(n_1,n_2) \in \varphi(M)^2$ to such a c, and let the $f(n_1,n_2)$ be an arbitrary element of C if n_1 or n_2 is not in $\varphi(M)$. □

Suppose now that $(M,A){\star\star}(T,C)$ and $(N,B){\star\star}(T,C)$ are double semidirect products so that T acts on M and on N on the left and on the right. Following Rhodes Tilson [28], we say that these actions are compatible with a morphism $\varphi : (M,A) \to (N,B)$ if for all $m \in M$, $n \in N$ and $t \in T$, if $\varphi(m) = n$ then $\varphi(tm) = tn$ and $\varphi(mt) = nt$. In this case we define a morphism

$$\begin{aligned} \varphi{\star\star}(T,C) : (M,A){\star\star}(T,C) &\to (N,B){\star\star}(T,C) \\ (m,t) &\mapsto (\varphi(m),t). \end{aligned}$$

The reader should have no difficulty to check that $\varphi{\star\star}(T,C)$ is indeed a morphism. In the same way, we define

$$\begin{aligned} \varphi{\star\star}T : M{\star\star}T &\to N{\star\star}T \\ (m,t) &\mapsto (\varphi(m),t), \end{aligned}$$

where $M{\star\star}T$ and $N{\star\star}T$ are the double semidirect products of M and N with T determined by the actions.

PROPOSITION 11.3 *Under the previous assumptions, if the actions of T on M and on N are compatible with φ, then*

$$(K_{\varphi\star\star(T,C)}, A_{\varphi\star\star(T,C)}) < (K_\varphi, A_\varphi).$$

Proof. In the proof of Rhodes and Tilson [28], Theorem 6.2, it is shown that $K_{\varphi\star\star T} < K_\varphi$. This is achieved by mapping each object (n,t) of $K_{\varphi\star\star T}$ to the object $(n_L t_R, t_L n_R)$ of K_φ, and by relating each arrow

$$[(m,t),(n,t)] : (\mathrm{n},\mathrm{t}) \to (\mathrm{n}',\mathrm{t}') \tag{3}$$

in $K_{\varphi\star\star T}$ to

$$[t_L m t'_R, t_L n t'_R] : (n_L t_R, t_L n_R) \to (n'_L t'_R, t'_L n'_R). \tag{4}$$

Note that when $m \in A$ (and thus by $\varphi(m) = n$ also $n \in B$), then $t_L m t'_R \in A$ and $t_L n t'_R \in B$. Let ψ denote this covering. By Lemma 11.4 and its proof, there is a covering

$$\rho : K_{\varphi\star\star(T,C)} \to K_{\varphi\star\star T}$$

that is the identity on objects and relates a morphism $[(a,c),(b,c)]$ in $K_{\varphi\star\star(T,C)}$, where $a \in A$, $b \in B$ and $c \in C$ with the corresponding morphism $[(a,c),(b,c)]$ in $K_{\varphi\star\star T}$. The composite of the two coverings ρ and ψ is the required covering. □

LEMMA 11.4 *Suppose that φ is a homomorphism $M \to N$, M' is a submonoid of M, and N' is a submonoid of N such that the restriction of φ to M is a homomorphism $M' \to N'$. Then $K_{\varphi'} < K_\varphi$.*

This is proved in Rhodes and Tilson [28], Corollary 5.4. It is clear that every object of $K_{\varphi'}$ is an object of K_φ. It is shown in [28] that the relation that is the identity function on objects and relates each morphism $[(m,n)]$ in $K_{\varphi'}$ with the morphism $[(m,n)]$ in K_φ is a covering.

We let $\underline{1}$ denote a trivial monoid $\underline{1} = \{1\}$. Thus, $(\underline{1},\{1\})$ is a trivial mg-pair.

PROPOSITION 11.5 *Let (M,A) denote an mg-pair and let φ denote the unique (collapsing) morphism $(M,A) \to (\underline{1},\{1\})$. Then*

$$(K_\varphi, A_\varphi) < (M,A).$$

Proof. The relation that relates each arrow $[1,(m,1),1]$ with m is a covering. □

COROLLARY 11.6 *Suppose that $(M,A){\star\star}(N,B)$ is a double semidirect product. Let π denote the projection $(M,A){\star\star}(N,B) \to (N,B)$, $(m,n) \mapsto n$. Then $(K_\pi, A_\pi) < (M,A)$.*

Proof. The projection π is (essentially) $\varphi{\star\star}(N,C)$, where φ denotes the collapsing morphism $(M,A) \to (\underline{1},\{1\})$. (Note that φ is compatible with any actions.) Thus, the result follows from Propositions 11.5 and Proposition 11.3. □

Given a double semidirect product $((M,A){\star\star}(N,B)){\star\star}(T,C)$, where $(M,A){\star\star}(N,B) = (V, A\times B)$, we say that the actions of T (on V) are *pointwise* if there exist left and right actions of T on M and N such that

$$\begin{aligned} t(m,n) &= (tm,tn) \\ (m,n)t &= (mt,nt), \end{aligned}$$

for all $(m,n) \in V$ and $t \in T$. It follows that the left and right actions of T on N are compatible and determine a double semidirect product $(N,B){\star\star}(T,C)$.

LEMMA 11.7 *Suppose that $((M,A){\star\star}(N,B)){\star\star}(T,C)$ is a double semidirect product such that the actions of T are pointwise and thus determine a double semidirect product $(N,B){\star\star}(T,C)$. Then the actions of T on $(M,A){\star\star}(N,B)$ and on (N,B) are compatible with the projection morphism $\pi : (M,A){\star\star}(N,B) \to (N,B)$.*

Proof. Immediate from the definitions. □

PROPOSITION 11.8 *Suppose that $\mathbf{V}_1, \mathbf{V}_2$ and $\mathbf{V}_3$ are varieties of mg-pairs. Then an mg-pair is in $(\mathbf{V}_1{\star\star}\mathbf{V}_2){\star\star}\mathbf{V}_3$ iff it divides a semidirect product*

$$((M,A){\star\star}(N,B)){\star\star}(T,C)$$

such that $(M,A) \in \mathbf{V}_1$, $(N,B) \in \mathbf{V}_2$, $(T,C) \in \mathbf{V}_3$ and the actions of T are pointwise.

Proof. One direction is trivial. Suppose now that (S,D) is in $(\mathbf{V}_1{\star\star}\mathbf{V}_2){\star\star}\mathbf{V}_3$. Then, by Proposition 6.1, (S,D) divides a block product

$$((M,A)\square(N,B))\square(T,C)$$

which is a double semidirect product

$$((M,A)^{N\times N}{\star\star}(N,B))^{T\times T}{\star\star}(T,C)$$

with suitable actions. This double semidirect product is in turn isomorphic to a double semidirect product

$$((M,A)^{N\times N\times T\times T}{\star\star}(N,B)^{T\times T}){\star\star}(T,C)$$

where the actions of T are given by

$$\begin{aligned} t(f,g) &= (f',g') \\ (f,g)t &= (f'',g''), \end{aligned}$$

where

$$\begin{aligned} f'(n_1,n_2,t_1,t_2) &= f(n_1,n_2,t_1t,t_2) \\ g'(t_1,t_2) &= g(t_1t,t_2), \end{aligned}$$

and similarly for f'' and g''. Since f' does not depend on g and g' does not depend on f, the left action is pointwise. Likewise the right action. Now let

$$\begin{aligned} (M',A') &= (M,A)^{N\times N\times T\times T} \\ (N',B') &= (N,B)^{T\times T}. \end{aligned}$$

We have that (S,D) divides a double semidirect product

$$((M',A'){\star\star}(N',B')){\star\star}(T,C)$$

such that the actions of T are pointwise. Since varieties are closed with respect to the direct product, we also have $(M',A') \in \mathbf{V}_1$ and $(N',B') \in \mathbf{V}_2$. □

We now complete the proof of Proposition 6.3. We want to prove that for all varieties of mg-pairs $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$,

$$(\mathbf{V}_1{\star\star}\mathbf{V}_2){\star\star}\mathbf{V}_3 \subseteq \mathbf{V}_1{\star\star}(\mathbf{V}_2{\star\star}\mathbf{V}_3).$$

By Proposition 6.1, we only need to show that each mg-pair

$$((M,A){\star\star}(N,B)){\star\star}(T,C)$$

such that $(M,A) \in \mathbf{V}_1$, $(N,B) \in \mathbf{V}_2$ and $(T,C) \in \mathbf{V}_3$ is in $\mathbf{V}_1{\star\star}(\mathbf{V}_2{\star\star}\mathbf{V}_3)$. Moreover, by Proposition 11.8, we may assume that the actions of T are pointwise. But then, by Lemma 11.7, the projection $\pi : (M,A){\star\star}(N,B) \to (N,B)$ is compatible with the actions of T, and moreover, by Corollary 11.6, it holds that

$$(K_{\pi{\star\star}(T,C)}, A_{\pi{\star\star}(T,C)}) < (M,A).$$

Thus, by Theorem 11.2 applied to the morphism

$$\pi{\star\star}(T,C) : ((M,A){\star\star}(N,B)){\star\star}(T,C) \to (N,B){\star\star}(T,C),$$

we have

$$((M,A)\star\star(N,B))\star\star(T,C) \quad < \quad (M,A)\Box((N,B)\star\star(T,C)),$$

proving that $((M,A)\star\star(N,B))\star\star(T,C)$ is in $\mathbf{V}_1\star\star(\mathbf{V}_2\star\star\mathbf{V}_3)$. □