Weakly Iterated Block Products of Finite Monoids

Howard Straubing[1] and Denis Thérien[2*]

[1] Boston College
straubin@cs.bc.edu
[2] McGill University
denis@cs.mcgill.ca

Abstract. The block product of monoids is a bilateral version of the better known wreath product. Unlike the wreath product, block product is not associative. All decomposition theorems based on iterated block products that have appeared until now have assumed right-to-left bracketing of the operands. We here study what happens when the bracketing is made left-to-right. This parenthesization is in general weaker than the traditional one. We show that weakly iterated block products of semilattices correspond exactly to the well-known variety **DA** of finite monoids: if groups are allowed as factors, the variety **DA*G** is obtained. These decomposition theorems allow new, simpler, proofs of recent results concerning the defining power of generalized first-order logic using two variables only.

1 Introduction

Rhodes and Tilson [3] introduced the bilateral semidirect product of monoids, and the related block product, and used them to develop the notion of the kernel of a homomorphism of monoids. The underlying idea behind such products is quite old; its precursors can be found in the "triple products" of Eilenberg [1], in the work of Schützenberger on the Schützenberger product [4] and on sequential bimachines [5]. In [3] we find the following bilateral version of the Krohn-Rhodes theorem: Every finite monoid M divides an iterated bilateral semidirect product

$$(M_n ** \cdots (M_3 ** (M_2 ** M_1)) \cdots),$$

where each M_i is either a semilattice or a simple group that divides M. In particular, if M is aperiodic, then M divides an iterated product of semilattices. While bilateral products are somewhat unwieldy to work with, they result in decompositions with simpler factors than are possible with unilateral products, making them especially suitable for some applications. For example, Straubing [9] applies

* contact author: School of Computer Science, McGill University, 3480 University, Montréal, QC, Canada, H3A 2A7. Research supported by NSERC, FCAR and the Von Humboldt Foundation.

the bilateral Krohn-Rhodes theorem to find logical characterizations of classes of regular languages.

In the present paper we consider what happens when we bracket the iterated bilateral semidirect product in the opposite direction. We find that the monoids that divide an iterated bilateral semidirect product

$$(\cdots((M_1 ** M_2) ** M_3) ** \cdots ** M_n)$$

of semilattices are precisely the members of the pseudovariety $\mathbf{DA}$, and those that divide an iterated product of groups and semilattices are precisely the members of $\mathbf{DA} * \mathbf{G}$. This constitutes a kind of inside-out Krohn-Rhodes theorem. We also give an application of our decomposition result: A new, transparent proof of theorems of Thérien, Wilke and Straubing [11, 8] on the definability of regular languages by generalized first-order sentences with two variables.

2 Algebraic Preliminaries

We suppose that the reader is familiar with the fundamental notions concerning the connections between semigroups and automata: division of monoids, recognition of regular languages by finite monoids, and the definition and basic properties of the syntactic monoid. We refer the reader to Chapter 1 of Pin [2] for an introduction to this material.

2.1 One-sided and bilateral products of monoids

Let M and N be finite monoids. Following a convention introduced by Eilenberg, we will write the product in M additively. Thus we write the identity of M as 0, and the k^{th} power of $m \in M$ as $k \cdot m$. This is done to make the notation more readable, and not to suggest that M is commutative. A *left action* of N on M associates to each pair $(n, m) \in N \times M$ an element nm of M, subject to the following laws:

$$n(m + m') = nm + nm'$$

$$(nn')m = n(n'm)$$

$$n0 = 0$$

$$1m = m$$

for all $m, m' \in M$, $n, n' \in N$. Given such a left action we define the *semidirect product* $M * N$ with respect to this left action as the monoid whose underlying set is $M \times N$ with multiplication given by

$$(m, n)(m', n') = (m + nm', nn'),$$

for all $m, m' \in M$, $n, n' \in N$. It is straightforward to verify that this multiplication is associative, and that $(0, 1)$ is the identity for this multiplication; thus

$M * N$ is indeed a monoid. There may be many different left actions of M on N, giving rise to nonisomorphic semidirect products $M * N$.

We define a *right action* of N on M analogously, and define the *reverse semidirect product* $N *_r M$ with respect to this action as the monoid structure on $N \times M$ with multiplication given by

$$(n, m)(n', m') = (nn', mn' + m').$$

Suppose we have both a left and a right action of N on M, and that these two actions satisfy

$$(nm)n' = n(mn'),$$

for all $m \in M$, $n, n' \in N$. We can then define another monoid structure on $M \times N$ with multiplication given by

$$(m, n)(m', n') = (mn' + nm', nn').$$

Once again, it is straightforward to verify that this is an associative multiplication with identity $(0, 1)$. We call the resulting monoid a *bilateral semidirect product* and denote it $M * *N$. Observe that every ordinary and reverse semidirect product is a special instance of the bilateral semidirect product, since we can define one of the two actions to be the identity map on M for all $n \in N$.

We now describe two related products. Once again, let M and N be finite monoids. We return to using the standard multiplicative notation for the product in M. The *wreath product* $M \circ N$ is a monoid structure on $M^N \times N$, with multiplication given by

$$(F, n)(F', n') = (G, nn'),$$

where for all $n'' \in N$,

$$G(n'') = F(n'')F'(n''n).$$

The *block product* $M \Box N$ is a monoid structure on $M^{N \times N} \times N$, with multiplication given by

$$(F, n)(F', n') = (G, nn'),$$

where for all $(n_1, n_2) \in N \times N$,

$$G(n_1, n_2) = F(n_1, n'n_2)F'(n_1 n, n_2).$$

The following proposition summarizes the essential facts about these products. The proofs (which are all quite simple) can be found in Eilenberg [1] or in Rhodes and Tilson [3]. In what follows, if M is a finite monoid, then M^r denotes the reversed monoid.

Proposition 1. *Let M, N be finite monoids.*
(a) If M and N are groups, then every semidirect product $M * N$ is a group.
(b) If $M * *N$ is a bilateral semidirect product with N a group, then $M * *N$ is isomorphic to a semidirect product $M * N$.

(c) Given a bilateral semidirect product $M**N$ there exist a left action of N on M and a right action of N on the resulting semidirect product $M*N$ such that $M**N$ is isomorphic to a submonoid of $N*_r(M*N)$.
(d) For every bilateral semidirect product $M**N$ there is a bilateral semidirect product M^r**N^r such that $(M**N)^r$ is isomorphic to M^r**N^r.
(e) $M \circ N$ is isomorphic to a semidirect product $M' * N$, where M' is the direct product of $|N|$ copies of M.
(f) Every semidirect product $M*N$ divides $M \circ N$.
(g) $M \Box N$ is isomorphic to a bilateral semidirect product $M''**N$, where M'' denotes the direct product of $|N|^2$ copies of M.
(h) Every bilateral semidirect product $M**N$ divides $M\Box N$.
(i)If $M_1 \prec M_2$, $N_1 \prec N_2$, then $M_1 \circ N_1 \prec M_2 \circ N_2$, and $M_1 \Box N_1 \prec M_2 \Box N_2$.
(j) Let 1 denote the trivial monoid. Then $M \circ 1$, $1 \circ M$, $M\Box 1$, and $1 \Box M$ are all isomorphic to M.
(k) $M \circ N \prec M \Box N$.

2.2 Product pseudovarieties

A *pseudovariety* of finite monoids is a collection of finite monoids that contains all divisors of its members, and the direct products of any two of its members. We use standard names for certain important pseudovarieties: $\mathbf{J}_1$ denotes the pseudovariety of finite semilattices (*i.e.,* idempotent and commutative monoids), $\mathbf{R}$ the pseudovariety of $\mathcal{R}$-trivial monoids, $\mathbf{G}$ the pseudovariety of finite groups, and $\mathbf{1}$ the pseudovariety whose only member is the trivial monoid. Another important pseudovariety, denoted $\mathbf{DA}$, will be discussed at length below.

If $\mathbf{V}$ and $\mathbf{W}$ are pseudovarieties, then $\mathbf{V} * \mathbf{W}$ denotes the family of finite monoids that divide a semidirect product $M*N$, where $M \in \mathbf{V}$ and $N \in \mathbf{W}$. We define $\mathbf{V} *_r \mathbf{W}$ analogously. We denote by $\mathbf{V}\Box\mathbf{W}$ the family of finite monoids that divide a bilateral semidirect product $M**N$, with $M \in \mathbf{V}$ and $N \in \mathbf{W}$. By Proposition 1, $\mathbf{V} * \mathbf{W}$ is also the family of divisors of wreath products $M \circ N$, with $M \in \mathbf{V}$ and $N \in \mathbf{W}$, and similarly $\mathbf{V}\Box\mathbf{W}$ is the family of divisors of block products $M\Box N$. $\mathbf{V} * \mathbf{W}$, $\mathbf{V} *_r \mathbf{W}$, and $\mathbf{V}\Box\mathbf{W}$ are themselves all pseudovarieties.

If $\mathbf{V}$ is a pseudovariety, we set $\mathbf{V}^r = \{M^r : M \in \mathbf{V}\}$. This, too, is a pseudovariety.

The following proposition summarizes properties of these product varieties. Most of these are direct consequences of Proposition 1. See [1] or [3] for the proofs.

Proposition 2. *Let* $\mathbf{U}$,$\mathbf{V}$,$\mathbf{W}$ *be pseudovarieties of finite monoids.*
(a) If $\mathbf{V}, \mathbf{W} \subseteq \mathbf{G}$, *then* $\mathbf{V} * \mathbf{W} \subseteq \mathbf{G}$.
(b) If $\mathbf{W} \subseteq \mathbf{G}$, *then* $\mathbf{V}\Box\mathbf{W} = \mathbf{V} * \mathbf{W}$.
(c) $\mathbf{V}\Box\mathbf{W} \subseteq \mathbf{W} *_r (\mathbf{V} * \mathbf{W})$.
(d) $\mathbf{V}^r\Box\mathbf{W}^r = (\mathbf{V}\Box\mathbf{W})^r$.
(e)$(\mathbf{U} * \mathbf{V}) * \mathbf{W} = \mathbf{U} * (\mathbf{V} * \mathbf{W})$.

Let $\mathcal{C}$ be a collection of finite monoids, and let $\mathbf{V}$ be the smallest pseudovariety containing $\mathcal{C}$. We denote by $pc(\mathcal{C})$ the union of the pseudovarieties $\mathbf{V}$, $\mathbf{V} * \mathbf{V}$, $\mathbf{V} * \mathbf{V} * \mathbf{V}$, etc. Note that we are implicitly using part (e) of the above proposition in this definition. $pc(\mathcal{C})$ is the smallest pseudovariety containing $\mathcal{C}$ that is closed under *; pc stands for "product closure".

For pseudovarieties $\mathbf{U}, \mathbf{V}, \mathbf{W}$, we have

$$(\mathbf{U} \Box \mathbf{V}) \Box \mathbf{W} \subseteq \mathbf{U} \Box (\mathbf{V} \Box \mathbf{W}),$$

however the converse inclusion is in general false. Because of this non-associativity, we must be careful about how we define iterated product varieties for the block product. We denote by $wb^k(\mathbf{V})$ the pseudovariety

$$(\cdots((\mathbf{V} \Box \mathbf{V}) \Box \mathbf{V}) \Box \cdots \Box \mathbf{V}),$$

with k occurrences of $\mathbf{V}$. We denote by $wbpc(\mathbf{V})$ the union of the $wb^k(\mathbf{V})$ over all $k \geq 0$. We also denote (not without some abuse of notation) by $wbpc(\mathbf{V} \cup \mathbf{W})$ the union of all the pseudovarieties

$$[\mathbf{V}_1, \ldots, \mathbf{V}_k] = (\cdots((\mathbf{V_1} \Box \mathbf{V_2}) \Box \mathbf{V_3}) \Box \cdots \Box \mathbf{V_k}),$$

where for each i, $\mathbf{V}_i = \mathbf{V}$ or $\mathbf{V}_i = \mathbf{W}$. It may not be obvious that this is a pseudovariety. To see that it is, observe that parts (i) and (j) of Proposition 1 can be used to show that the pseudovarieties $[\mathbf{V}_1, \ldots, \mathbf{V}_k]$ and $[\mathbf{W}_1 \ldots, \mathbf{W}_m]$ are both contained in the pseudovariety $[\mathbf{V}_1, \ldots, \mathbf{V}_k, \mathbf{W}_1, \ldots, \mathbf{W}_m]$, and thus $wbpc(\mathbf{V} \cup \mathbf{W})$ contains the direct product of any two of its members. $wbpc$ stands for "weak block product closure".

2.3 The pseudovariety DA

If M is a finite monoid and $m \in M$, then we denote by m^ω the unique power of m that is idempotent. **DA** consists of all finite monoids M such that for all $x, y, z \in M$,

$$(xyz)^\omega y (xyz)^\omega = (xyz)^\omega.$$

It follows directly from this characterization that **DA** is a pseudovariety. If we take $x = z = 1$, we find that every monoid in **DA** satisfies the identity

$$y^\omega y = y^\omega,$$

and thus is aperiodic (i.e., contains no nontrivial groups). **DA** was introduced by Schützenberger [6].

We give an equivalent characterization of **DA** in terms of congruences on finitely generated free monoids. This is due to Thérien and Wilke [11]. Let Σ be a finite alphabet, with $|\Sigma| = n$. We define a family of equivalence relations $\sim_{n,k}$, $k \geq 0$, on Σ^* as follows. If either $n = 0$ or $k = 0$, $\sim_{n,k}$ is the trivial congruence that identifies all words of Σ^*. Now suppose n and k are both positive. We will define $\sim_{n,k}$ inductively: If $w \in \Sigma^*$ then we denote by $c(w)$ the set of letters appearing in w. If $\sigma \in c(w)$, then w has a unique factorization

$$w = w_0 \sigma w_1,$$

where $\sigma \notin c(w_0)$. We define

$$l_\sigma^{n,k}(w) = ([w_0]_{n-1,k}, [w_1]_{n,k-1}),$$

where $[v]_{m,r}$ denotes the $\sim_{m,r}$-equivalence class of v. Observe that the definition above makes sense, because w_0 is a word over the $(n-1)$-letter alphabet $\Sigma - \{\sigma\}$. Similarly, there is a unique factorization

$$w = w_0' \sigma w_1',$$

such that $\sigma \notin c(w_1')$. We set

$$r_\sigma^{n,k}(w) = ([w_0']_{n-1,k}, [w_1']_{n,k-1}).$$

We now set, for $v, w \in \Sigma^*$, $v \sim_{n,k} w$ if and only if $c(v) = c(w)$, and for all $\sigma \in c(v)$,

$$l_\sigma^{n,k}(v) = l_\sigma^{n,k}(w), r_\sigma^{n,k}(v) = r_\sigma^{n,k}(w).$$

It is easy to verify that every $\sim_{n,k}$ is a congruence of finite index on Σ^*. Thérien and Wilke show that the quotient monoids $\Sigma^*/\sim_{n,k}$ are all in $\mathbf{DA}$, and that every monoid in $\mathbf{DA}$ is a homomorphic image of some $\Sigma^*/\sim_{n,k}$.

3 The Decomposition Theorem

Our main result is:

Theorem 1. *Let* $\mathbf{H}$ *be any pseudovariety of finite groups. Then*

$$wbpc(\mathbf{J}_1 \cup \mathbf{H}) = \mathbf{DA} * pc(\mathbf{H}).$$

In particular, taking $\mathbf{H} = \mathbf{1}$, we have

$$wbpc(\mathbf{J}_1) = \mathbf{DA}.$$

We give the proof in the next two subsections.

3.1 Proof that $wbpc(\mathbf{J}_1 \cup \mathbf{H}) \subseteq \mathbf{DA} * pc(\mathbf{H})$

We first show:

Lemma 1. $\mathbf{DA} \Box \mathbf{J}_1 \subseteq \mathbf{DA}$.

Proof. By Proposition 2,

$$\mathbf{DA}\Box\mathbf{J}_1 \subseteq \mathbf{J}_1 *_r (\mathbf{DA} * \mathbf{J}_1).$$

Since **DA** is closed under reversal, it suffices, again by applying Proposition 2 to prove the lemma with $\Box$ replaced by $*$. We show this using the defining identity for **DA**:

$$(xyz)^\omega y(xyz)^\omega = (xyz)^\omega.$$

That is, we will show that if M is a monoid satisfying this identity, and $N \in \mathbf{J}_1$, then any semidirect product $M*N$ satisfies the same identity. Let $x, y, z \in M*N$, with

$$x = (x_1, x_2), y = (y_1, y_2), z = (z_1, z_2).$$

Since M and N are both aperiodic, there exists $k > 0$ such that

$$u^\omega = u^k = u^{k+1}$$

for all $u \in M**N$, and (using additive notation in M)

$$k \cdot v = (k+1) \cdot v$$

for all $v \in M$. Since $N \in \mathbf{J}_1$ we have

$$(x_2y_2z_2)^2 = x_2y_2z_2x_2 = x_2y_2z_2y_2 = x_2y_2z_2z_2 = x_2y_2z_2,$$

so that

$$\begin{aligned}(xyz)^\omega &= (xyz)^{k+1} \\ &= (x_1 + x_2y_1 + x_2y_2z_1 + x_2y_2z_2 \cdot k \cdot (x_1 + y_1 + z_1), x_2y_2z_2)\end{aligned}$$

Thus

$$\begin{aligned}(xyz)^\omega y(xyz)^\omega = (x_1 + x_2y_1 + x_2y_2z_1 + x_2y_2z_2 \cdot (k \cdot (x_1 + y_1 + z_1) + \\ y_1 + (k+1) \cdot (x_1 + y_1 + z_1)), x_2y_2z_2)\end{aligned}$$

Since M satisfies the identity for **DA**,

$$k \cdot (x_1 + y_1 + z_1) + y_1 + (k+1) \cdot (x_1 + y_1 + z_1) = k \cdot (x_1 + y_1 + z_1),$$

from which it follows that

$$(xyz)^\omega y(xyz)^\omega = (xyz)^\omega.$$

To complete the proof that $wbpc(\mathbf{J}_1 \cup \mathbf{H}) \subseteq \mathbf{DA} * pc(\mathbf{H})$, it is enough to show

$$(\mathbf{DA} * pc(\mathbf{H}))\Box\mathbf{H} \subseteq \mathbf{DA} * pc(\mathbf{H}),$$

and

$$(\mathbf{DA} * pc(\mathbf{H}))\Box\mathbf{J}_1 \subseteq \mathbf{DA} * pc(\mathbf{H}).$$

The first inclusion follows from Proposition 2.

$$\begin{aligned}(\mathbf{DA} * pc(\mathbf{H}))\Box\mathbf{H} &= (\mathbf{DA} * pc(\mathbf{H})) * \mathbf{H}\\ &= \mathbf{DA} * (pc(\mathbf{H}) * \mathbf{H})\\ &= \mathbf{DA} * pc(\mathbf{H}).\end{aligned}$$

For the second, we argue as in the proof of Lemma 1: It is enough to show this inclusion when $\Box$ is replaced by $*$. We now use the following two facts, first proved in Stiffler [7]:

$$\mathbf{R} = pc(\mathbf{J}_1),$$

and

$$\mathbf{H} * \mathbf{J}_1 \subseteq \mathbf{R} * \mathbf{H},$$

for any pseudovariety of groups $\mathbf{H}$. This gives

$$\begin{aligned}(\mathbf{DA} * pc(\mathbf{H})) * \mathbf{J}_1 &= \mathbf{DA} * (pc(\mathbf{H}) * \mathbf{J}_1)\\ &\subseteq \mathbf{DA} * (\mathbf{R} * pc(\mathbf{H}))\\ &= (\mathbf{DA} * pc(\mathbf{J}_1)) * pc(\mathbf{H})\\ &\subseteq \mathbf{DA} * pc(\mathbf{H}),\end{aligned}$$

by Proposition 2 and Lemma 1.

3.2 Proof that $\mathbf{DA} * pc(\mathbf{H}) \subseteq wbpc(\mathbf{J}_1 \cup \mathbf{H})$.

We first note that it is sufficient to show $\mathbf{DA} \subseteq wbpc(\mathbf{J}_1)$. For suppose that this is true. If $M \in \mathbf{DA} * pc(\mathbf{H})$ we have

$$M \prec N \circ G$$

for some $N \in \mathbf{DA}$, $G \in \mathbf{H}$. Our assumption and Proposition 1 then give:

$$\begin{aligned}M &\prec N \circ H_k \circ \cdots \circ H_1\\ &\prec (\cdots((N\Box H_k)\Box H_{k-1})\Box\cdots\Box H_1)\\ &\prec (\cdots((V_r\Box V_{r-1})\Box V_{r-2})\Box\cdots\Box V_1)\Box H_k\Box\cdots\Box H_1),\end{aligned}$$

where $H_1, \ldots, H_k \in \mathbf{H}$, and $V_1, \ldots, V_r \in \mathbf{J}_1$. Thus $M \in wbpc(\mathbf{J}_1 \cup \mathbf{H})$.

To prove $\mathbf{DA} \subseteq wbpc(\mathbf{J}_1)$, we use the generating family of congruences for $\mathbf{DA}$, defined in 2.3. We thus need to show that for every finite alphabet Σ, with $|\Sigma| = n$, and every $k \geq 0$, $\Sigma^*/\sim_{n,k} \in wbpc(\mathbf{J}_1)$. This is trivially true if $n = 0$ or $k = 0$. If n and k are both positive, then we claim

$$\Sigma^*/\sim_{n,k} \in wb^{n+k-1}(\mathbf{J}_1).$$

We prove this by induction on $n + k - 1$. It is enough to show that for each $w \in \Sigma^*$ and each $\sigma \in c(w)$, the languages

$$L_1 = \{v \in \Sigma^* : c(v) = c(w)\},$$

$$L_2 = \{v \in \Sigma^* : l_\sigma^{n,k}(v) = l_\sigma^{n,k}(w)\},$$
$$L_2 = \{v \in \Sigma^* : r_\sigma^{n,k}(v) = r_\sigma^{n,k}(w)\},$$

are all recognized by monoids in $wb^{n+k-1}(\mathbf{J}_1)$, since each class of $\sim_{n,k}$ is a boolean combination of these languages.

L_1 is obviously recognized by the monoid whose elements are the subsets of Σ, with union as multiplication—this is in $\mathbf{J}_1$. We now need only show that L_2 is recognized by a monoid in $wb^{n+k-1}(\mathbf{J}_1)$, since the result for L_3 follows from the reversal closure of the block product of pseudovarieties (Proposition 2). We will now show that L_2 is recognized by the monoid

$$M = ((\Sigma - \{\sigma\})^*/\sim_{n-1,k} \quad \times \quad \Sigma^*/\sim_{n,k-1}) \circ U_1,$$

which, coupled with the inductive hypothesis, gives the desired result. To show recognition, we define a map $\phi : \Sigma \to M$ as follows: We set $\phi(\sigma) = (G_\sigma, 0)$, and $\phi(\tau) = (F_\tau, 1)$ for $\tau \neq \sigma$, where

$$F_\tau(1) = ([\tau]_{n-1,k}, 1),$$
$$G_\sigma(1) = (1, 1),$$
$$F_\tau(0) = (1, [\tau]_{n,k-1}),$$
$$G_\sigma(0) = (1, [\sigma]_{n,k-1}).$$

The map ϕ extends to a unique homomophism from Σ^* into M. Let $w \in \Sigma^*$, and let $\phi_1(w)$ and $\phi_2(w)$ denote, respectively, the left and right co-ordinates of $\phi(w)$. It follows readily that if $w \in \Sigma^*$ and $\sigma \notin c(w)$, then $\phi_2(w) = 1$, and $(\phi_1(w))(1) = ([w]_{n-1,k}, 1)$. Otherwise, $\phi_2(w) = 0$, and

$$(\phi_1(w))(1) = ([w_0]_{n-1,k}, [w_1]_{n,k-1}),$$

where $w = w_0 \sigma w_1$ is the unique factorization with $\sigma \notin c(w_0)$.Thus $L_2 = \phi^{-1}(X)$, where

$$X = \{(G, 0) : G(1) = ([w_0]_{n-1,k}, [w_1]_{n,k-1})\}.$$

Thus M recognizes L_2, as required.

4 An Application to Logic

4.1 Regular languages and generalized first-order logic

Regular languages can be defined by sentences of first-order logic, using the following scheme: Variables in a sentence denote positions in a word over the underlying input alphabet Σ. There are two kinds of atomic formulas: $x < y$, which is interpreted to mean that position x is to the left of position y, and $Q_\sigma x$, which means that the letter in position x is σ. A sentence such as

$$\exists x (\forall y (\neg y < x) \wedge Q_\sigma x)$$

is satisfied by all words having at least one letter, and whose first letter is σ. Thus the sentence *defines* the regular language $\sigma\Sigma^*$, consisting of all the words that satisfy the sentence. We will allow our sentences to contain, in addition to the usual existential and universal quantifiers, *modular* quantifiers $\exists^{r \bmod n}$, where $0 \leq r < n$. A formula of the form

$$\exists^{r \bmod n} x\phi(x)$$

is interpreted to mean that the number of positions x for which $\phi(x)$ holds is congruent to r modulo n.

See Straubing [9] for an extensive treatment of this method of defining formal languages with formulas of logic. In practice, it has been found that classes of languages defined by "natural"-looking boolean-closed classes of sentences can usually be characterized in terms of the syntactic morphisms and syntactic monoids of the languages.

In this section we apply Theorem 1 to give a new proof of the following result, due to Thérien and Wilke [11] for the case $n = 1$ (no modular quantifiers) and to Straubing and Thérien [8] for the general case. Let $\mathbf{G}_{com}^{(n)}$ denote the pseudovariety of finite abelian groups whose exponents divide n.

Theorem 2. *Let Σ be a finite alphabet, and let $n \geq 1$. A language $L \subseteq \Sigma^*$ is definable by a sentence that uses only two variables, ordinary quantifiers and modular quantifiers of modulus n if and only if $M(L) \in \mathbf{DA} * pc(\mathbf{G}_{com}^{(n)})$.*

In the case $n = 1$, the modular quantifier is superfluous, so the theorem says that L is definable by a two-variable sentence if and only if $M(L) \in \mathbf{DA}$.

4.2 Languages recognized by bilateral semidirect products

Let Σ be a finite alphabet, M a finite monoid, and $\alpha : \Sigma^* \to M$ a homomorphism. We set $\Gamma = M \times \Sigma \times M$, and define a length-preserving map $\tau_\alpha : \Sigma^* \to \Gamma^*$ by

$$\tau_\alpha(\sigma_1 \cdots \sigma_n) = \gamma_1 \cdots \gamma_n,$$

where $\sigma_i \in \Sigma$ and

$$\gamma_i = (\alpha(\sigma_1 \cdots \sigma_{i-1}), \sigma_i, \alpha(\sigma_{i+1} \cdots \sigma_n)) \in \Gamma,$$

for $i = 1, \ldots, n$. (In the above equation, we take the left component of γ_1 and the right component of γ_n to be the identity of M.) The following facts about homomorphisms from finitely-generated free monoids into bilateral semidirect products are from Thérien [10], where they are stated in terms of congruences on finitely generated free monoids.

Lemma 2. *(a) Let Σ, M, Γ be as above. Let N be a finite monoid, and let $\beta : \Sigma^* \to N ** M$ be a homomorphism into a bilateral semidirect product. Let*

$(n,m) \in N * * M$. Then there is a homomorphism $\alpha : \Sigma^* \to M$ and a language $L \subseteq \Gamma^*$ recognized by N such that

$$\beta^{-1}(n,m) = \alpha^{-1}(m) \cap \tau_\alpha^{-1}(L).$$

(b) Let $\alpha, \Sigma, M, \Gamma$ be as above. Let $L \subseteq \Gamma^*$ be a regular language recognized by a finite monoid N. Then $\tau_\alpha^{-1}(L)$ is recognized by $N \Box M$.

.

4.3 Construction of two-variable sentences

We now prove Theorem 2 in detail. We begin by showing that every language whose syntactic monoid is in $\mathbf{DA} * pc(\mathbf{G}_{com}^{(n)})$ is definable by a two-variable sentence of the required kind. Let $L \subseteq \Sigma^*$, with $M(L) \in \mathbf{DA} * pc(\mathbf{G}_{com}^{(n)})$. By Theorem 1, there exists a homomorphism

$$\phi : \Sigma^* \to N = (\cdots((M_1 \Box M_2) \Box M_3) \Box \cdots \Box M_k,$$

where each M_i is either a semilattice or an abelian group with exponent dividing n, such that $L = \phi^{-1}(X)$ for some $X \subseteq N$. We construct a defining sentence for L by induction on k. First, suppose that M_k is a semilattice. By Lemma 2, L is a boolean combination of languages of the form $\alpha^{-1}(m)$ and $\tau_\alpha^{-1}(K)$, where $\alpha : \Sigma^* \to M_k$ is a homomorphism, $m \in M_k$, and $K \subseteq (M_k \times \Sigma \times M_k)^*$ is recognized by $(\cdots((M_1 \Box M_2) \Box M_3) \Box \cdots \Box M_{k-1}$. We need to show that both $\alpha^{-1}(m)$ and $\tau_\alpha^{-1}(K)$ are two-variable definable.

Observe that the value of $\alpha(w)$ depends only on $c(w)$. Thus if $\Sigma' \subseteq \Sigma$ we can write $\alpha(\Sigma')$ to denote the image under α of any w such that $c(w) = \Sigma'$. Consequently $\alpha^{-1}(m)$ is defined by a boolean combination of *one*-variable sentences of the form $\exists x Q_\sigma x$.

By the inductive hypothesis, K is defined by a two-variable sentence over $\Gamma = M_k \times \Sigma \times M_k$. We obtain a sentence for $\tau_\alpha^{-1}(K)$ by replacing each subformula $Q_{(m_1,\sigma,m_2)}x$ of the sentence by a disjunction of formulas that say 'the letter in position x is σ, the set of letters in positions to the left of x is Σ_1, and the set of letters in positions to the right of x is Σ_2', where the disjunction is over all subsets Σ_1, Σ_2 of Σ such that $\alpha(\Sigma_1) = m_1$ and $\alpha(\Sigma_2) = m_2$. Such a formula is

$$\begin{aligned} Q_\sigma x \wedge \forall y (y < x \to \bigvee_{\tau \in \Sigma_1} Q_\tau y) \\ \wedge \bigwedge_{\tau \in \Sigma_1} \exists y (y < x \wedge Q_\tau y) \\ \wedge \forall y (y > x \to \bigvee_{\tau \in \Sigma_2} Q_\tau y) \\ \wedge \bigwedge_{\tau \in \Sigma_2} \exists y (y > x \wedge Q_\tau y). \end{aligned}$$

The resulting defining sentence for $\tau_\alpha^{-1}(K)$ still uses only two variables, since the new variable y that we introduced is used only within the scopes of the new quantifiers; thus we are able to re-use a variable of the same name already occurring in the defining sentence for K.

We proceed very similarly in the case where $M_k \in \mathbf{G}_{com}^{(n)}$. In this case, the value of a homomorphism $\alpha : \Sigma^* \to M_k$ on a word w is determined by the number of times, modulo n, that each letter $\tau \in \Sigma$ appears in w; that is, by a $|\Sigma|$-tuple $(n_\sigma)_{\sigma\in\Sigma}$ of elements of $\mathbf{Z}_n$. Thus $\alpha^{-1}(m)$ is defined by a boolean combination of one-variable sentences of the form $\exists^{r \bmod n} x Q_\sigma x$. We obtain a two-variable sentence for $\tau_\alpha^{-1}(K)$ from a defining sentence for K upon replacing each $Q_{(m_1,\sigma,m_2)}$ by a disjunction of sentences of the form

$$\begin{aligned} Q_\sigma x \wedge \bigwedge_{\tau\in\Sigma} \exists^{n_\tau \bmod n}(y < x \wedge Q_\tau y) \\ \wedge \bigwedge_{\tau\in\Sigma} \exists^{n'_\tau \bmod n}(y < x \wedge Q_\tau y) \end{aligned}$$

for some $|\Sigma|$-tuples $(n_\sigma)_{\sigma\in\Sigma}$ and $(n'_\sigma)_{\sigma\in\Sigma}$.

4.4 The syntactic monoid of two-variable definable languages

We complete the proof of Theorem 2 by showing that if $L \subseteq \Sigma^*$ is definable by a two-variable sentence θ all of whose modular quantifiers are of modulus n, then $M(L) \in \mathbf{DA} * pc(\mathbf{G}_{com}^{(n)})$. We first describe a kind of normal form for such sentences. Let $\mathcal{Q}$ be an innermost quantifier symbol in θ; that is, a quantifier symbol such that no other quantifier appears within its scope. We may assume that $\mathcal{Q}$ is either an ordinary existential quantifer or a modular quantifier, since the universal quantifier can be defined in terms of the existential quantifier. Let us suppose that the quantifier $\mathcal{Q}$ quantifies the variable x. We can write the subformula appearing after the quantifier as the disjunction of mutually exclusive formulas of the form

$$x\mathcal{R}y \wedge Q_\sigma x \wedge Q_\tau y,$$

where $\mathcal{R}$ is one of $<$, $>$, or $=$. We can then rewrite the entire quantified subformula as a boolean combination of formulas of the form

$$\mathcal{Q}x(x\mathcal{R}y \wedge Q_\sigma x \wedge Q_\tau y).$$

This is clear if $\mathcal{Q}$ is existential, since the existential quantifier commutes with disjunction. If $\mathcal{Q}$ is modular, then we note that

$$\exists^{r \bmod n} x(\theta_1 \vee \cdots \vee \theta_s),$$

where the θ_i are mutually exclusive, is equivalent to

$$\bigvee \bigwedge_{j=1}^{s} \exists^{r_i \bmod n} x\theta_i,$$

where the disjunction is over all s-tuples $(r_1, \ldots, r_s)$ of elements of $\mathbf{Z}_n$ whose sum is r. Now note that we can move the atomic formula $Q_\tau y$ from within the scope of the quantifer; that is,

$$\mathcal{Q}x(x\mathcal{R}y \wedge Q_\sigma x \wedge Q_\tau y)$$

is equivalent to

$$Q_\tau y \wedge \mathcal{Q}x(x\mathcal{R}y \wedge Q_\sigma x),$$

unless $\mathcal{Q}$ is the modular quantifier $\exists^{0 \bmod n}$, in which case the formula is equivalent to

$$\neg Q_\tau y \vee \neg\exists x(x\mathcal{R}y \wedge Q_\sigma x) \\ \vee (Q_\tau y \wedge \exists^{0 \bmod n} x(x\mathcal{R}y \wedge Q_\sigma x)).$$

Finally, we note that $\mathcal{Q}x(x = y \wedge Q_\sigma x)$ is equivalent to $Q_\sigma y$ if $\mathcal{Q}$ is either an existential quantifier or the modular quantifier $\exists^{1 \bmod n}$, and is never satisfied otherwise. We may thus assume that in θ, every innermost quantified subformula has the form

$$\mathcal{Q}x(x < y \wedge Q_\sigma x)$$

or

$$\mathcal{Q}x(x > y \wedge Q_\sigma x).$$

This is our normal form.

We prove $M(L) \in \mathbf{DA} * pc(\mathbf{G}_{com}^{(n)})$ by induction on the depth of nesting of the quantifiers and by the number of innermost quantifiers at this depth. At each step we will either decrease the number of quantifiers at the maximal depth, or decrease the depth. The base case is depth 0, in which the sentence simply says TRUE or FALSE, and $M(L)$ is the trivial monoid. Let us accordingly pick a quantifier at maximal depth. If the quantifier is an existential quantifier, then we set M to be the quotient of Σ^* by the congruence that identifies words v and w if and only if $c(v) = c(w)$. If the quantifier is modular, then M is the quotient of Σ^* by the congruence that identifies v and w if and only if, for all $\sigma \in \Sigma$,

$$|v|_\sigma \equiv |w|_\sigma \pmod n,$$

where $|v|_\sigma$ denotes the number of occurrences of σ in v. In either case we let $\alpha : \Sigma^* \to M$ denote the projection onto the quotient by the congruence. In the former case, elements of M can be identified with subsets of Σ, and $M \in \mathbf{J}_1$. In the latter case, elements of M are $|\Sigma|$-tuples of elements of $\mathbf{Z}_n$, and $M \in \mathbf{G}_{com}^{(n)}$.

We now transform θ into a sentence over $\Gamma = M \times \Sigma \times M$. If the selected innermost quantified formula is

$$\exists x(x < y \wedge Q_\sigma x),$$

we replace it by

$$\bigvee Q_{(\Sigma_1, \tau, \Sigma_2)} x,$$

where the disjunction is over all $\Sigma_1 \subset \Sigma$ such that $\sigma \in \Sigma_1$, and all $\tau \in \Sigma$, $\Sigma_2 \subseteq \Sigma$. In case the selected quantified subformula is

$$\exists^{r \bmod n} x(x < y \wedge Q_\sigma x),$$

we replace it by the disjunction of all

$$Q_{((n_\tau)_{\tau\in\Sigma},\tau,(n'_\tau)_{\tau\in\Sigma})}x,$$

such that $n_\sigma = r$. We replace all other subformulas of θ of the form $Q_\sigma x$ by the disjunction of all $Q_{(\Sigma_1,\sigma,\Sigma_2)}x$, where $\Sigma_1, \Sigma_2 \subseteq \Sigma$. We proceed analogously if the quantified subformula contains $x > y$ instead of $x < y$.

The result is a sentence η over Γ that has either smaller depth than θ, or fewer quantifiers at maximal depth. By the inductive hypothesis, the language $K \subseteq \Gamma^*$ defined by η has its syntactic monoid in $\mathbf{DA} * pc(\mathbf{G}_{com}^{(n)})$. We have $L = \tau_\alpha^{-1}(K)$, so $M(L) \prec M(K)\Box M$, by Lemma 2. Thus by Theorem 1, $M(L) \in \mathbf{DA} * pc(\mathbf{G}_{com}^{(n)})$.

4.5 Concluding remarks

If we take the union over all moduli $n > 0$, we obtain Theorem 2 in the form originally stated in [8]: A language is definable by a two-variable sentence if and only if it belongs to $\mathbf{DA} * \mathbf{G}_{sol}$, where $\mathbf{G}_{sol}$ is the pseudovariety of finite solvable groups. In [8] we discuss the open decision problem for $\mathbf{DA} * \mathbf{G}_{sol}$.

It is interesting to compare our proof of Theorem 2 with the logical applications of the bilateral Krohn-Rhodes theorem in Straubing [9]. In the latter, we build up our logical formulas by quantifying over existing formulas. In the present paper, however, we construct formulas by replacing atomic formulas with quantified formulas of depth 1, and that's why we need the inside-out Krohn-Rhodes theorem!

References

1. S. Eilenberg, *Automata, Languages and Machines,* vol. B, Academic Press, New York, 1976.
2. J. E. Pin, *Varieties of Formal Languages,* Plenum, London, 1986.
3. J. Rhodes and B. Tilson, "The Kernel of Monoid Morphisms", *J. Pure and Applied Algebra* **62** (1989) 227–268.
4. M. P. Schützenberger, "On finite monoids having only trivial subgroups", *Information and Control* **8** (1965) 190-194.
5. M. P. Schützenberger, "A remark on finite transducers", *Information and Control* **4** (1961), 185-196.
6. M. P. Schützenberger, "Sur le Produit de Concatenation Non-ambigu", *Semigroup Forum* **13** (1976), 47-76.
7. P. Stiffler, "Extensions of the Fundamental Theorem of Finite Semigroups", *Advances in Mathematics,* **11** 159-209 (1973).
8. H. Straubing and D. Thérien, "Regular languages defined by generalized first-order formulas with a bounded number of bound variables", *Proc. 18th Symposium on Theoretical Aspects of Computer Science* 551-562 (2001).

9. H. Straubing, *Finite Automata, Formal Logic and Circuit Complexity,* Birkhäuser, Boston, 1994.
10. D. Thérien, "Two-sided wreath products of categories", *J. Pure and Applied Algebra* **74** (1991) 307-315.
11. D. Thérien and T. Wilke, "Over Words, Two Variables are as Powerful as One Quantifier Alternation," *Proc. 30th ACM Symposium on the Theory of Computing* 256-263 (1998).