Algebraic Characterization of the Alternation Hierarchy in $FO^2[<]$ on Finite Words*

Howard Straubing

Computer Science Department, Boston College
Chestnut Hill, Massachusetts, USA 02467
straubin@cs.bc.edu

Abstract

We give an algebraic characterization of the quantifier alternation hierarchy in first-order two-variable logic on finite words. As a result, we obtain a new proof that this hierarchy is strict. We also show that the first two levels of the hierarchy have decidable membership problems, and conjecture an algebraic decision procedure for the other levels.

1998 ACM Subject Classification F.4.1 Mathematical Logic; F.4.3 Formal Languages

Keywords and phrases Automata, finite model theory

Digital Object Identifier 10.4230/LIPIcs.CSL.2011.525

1 Introduction

We study first-order sentences interpreted in finite words over a finite alphabet Σ, with the single relation $<$ on positions in the word. It is well known (Kamp [6], Immerman and Kozen [5]) that every such sentence is equivalent to one in which only three variables are used. There has been extensive study, from the standpoint of first-order and temporal logic, automata theory, and algebra, of the fragment $FO^2[<]$ of sentences that use only two variables. (See, for example, Ettesami, Vardi and Wilke [4]; Schwentick, Thérien and Vollmer [13]; Straubing and Thérien [16]. Tesson and Thérien [17] give a broad-ranging survey of the many places in which the class of languages definable in this logic arises.)

Weis and Immerman [20] examined the hierarchy within $FO^2[<]$ based on alternation of quantifiers. Using model-theoretic methods, they showed that this hierarchy is strict. Kufleitner and Weil [9] show that each level of the hierarchy defines a variety of languages. This implies, among other things, that whether a regular language $L \subseteq \Sigma^*$ can be defined by a sentence of a given level k in the hierarchy is completely determined by the syntactic monoid $M(L)$ of L. While they do not provide an explicit algebraic description of the levels, Kufleitner and Weil do show that one can effectively compute the alternation depth of a given language in $FO^2[<]$ with an error no more than 1.

Here we give an exact algebraic characterization of each level of the alternation hierarchy; that is, we give an algebraic description of sequence $\mathbf{V}_n$ of families of finite monoids with the property that L is defined by a sentence with k quantifier alternations if and only if $M(L) \in \mathbf{V}_k$. Our characterization is in terms of the two-sided semidirect product of finite monoids and of pseudovarieties of finite monoids. More precisely, we show that the k^{th} level of the hierarchy corresponds to the weakly iterated two-sided semidirect product of k copies of the pseudovariety $\mathbf{J}$ of $\mathcal{J}$-trivial monoids. While many algebraic decompositions of the pseudovariety $\mathbf{DA}$ corresponding to $FO^2[<]$ have been studied, and while it has always been

* Research partially supported by National Science Foundation Grant CCF-0915065

Computer Science Logic 2011 (CSL'11).
Editor: Marc Bezem; pp. 525–537

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

clear that **DA** is equal to the closure of **J** under two-sided semidirect product, the fact that the levels of this hierarchy have such a simple logical significance appears to be new.

This still leaves the question of whether the membership problem for each of the levels is decidable. This is precisely the kind of question that algebraic methods are best suited to answer, since it is often possible to reduce the problem to one of verifying identities in the syntactic monoid. We produce a sequence of identities, based on work of Almeida and Weil [2], that we conjecture characterizes membership in each of the levels of the hierarchy. We show that these identities are necessary conditions, and use this fact to give a new proof of the strictness of the alternation hierarchy. The identities are known to be sufficient for the first two levels, which gives algebraic decision procedures for determining whether a give regular language is definable by a two-variable sentence with one or two quantifier alternations.

We present general preliminaries in Section 2 and particulars about two-sided semidirect products in Section 3. We prove our characterization theorem in Section 4; the argument is an adaptation of one given in [16]. We apply the result to questions of strictness and decidability in Section 5.

2 Logical And Algebraic Preliminaries

We review these preliminaries briefly and somewhat informally. The books by Pin [10] and Straubing [15] are references for all the matters discussed here.

2.1 First-order logic

Let Σ be a finite alphabeuntitled foldert. We build first-order formulas from atomic formulas $x < y$ and $Q_\sigma x$, where $\sigma \in \Sigma$. These formulas are interpreted in finite words over Σ: variables are interpreted as positions, with $x < y$ indicating that position x is strictly to the left of position y, and $Q_\sigma x$ indicating that the letter in position x is σ. A *sentence* (a formula without free variables) accordingly *defines* the *language* $L \subseteq \Sigma^*$ of all words w that satisfy the sentence. For example, if $\Sigma = \{\sigma, \tau\}$, then the set of words in which both σ and τ occur, and in which there are at least two occurrences of σ to the left of the first occurrence of τ is defined by the sentence

$$\exists x(Q_\tau x \wedge \forall y(y < x \to Q_\sigma y) \wedge \exists z_1 \exists z_2(z_1 < z_2 \wedge z_2 < x)).$$

As mentioned in the introduction, every first-order sentence of this kind is equivalent to one in which there are only three variables, provided we are allowed to reuse variable symbols. Here we are concerned with the languages definable by sentences of the logic we denote by $FO^2[<]$, consisting of formulas in which only two variables are used. This logic is known to have strictly weaker expressive power than full first-order logic. Note however, that the language in the example above is definable in this restricted logic, by the sentence

$$\exists x(Q_\tau x \wedge \forall y(y < x \to Q_\sigma y) \wedge \exists y(y < x \wedge \exists x(x < y))).$$

We cannot use standard constructions to write such sentences in prefix form without increasing the number of variables. Nonetheless, it is still possible to describe a different sort of normal form that will allow us to define the depth of quantifier alternation in a formula. We allow atomic formulas with $\leq$ as well as $<$, replace every occurrence of $\neg Q_\sigma x$ by

$$\bigvee_{\tau \in \Sigma \setminus \{\sigma\}} Q_\tau x,$$

and apply DeMorgan's Laws to move negations past conjunctions, disjunctions and quantifiers. We obtain as a result an equivalent formula that contains only existential and universal quantifiers, and the connectives $\wedge$ and $\vee$, with no occurrences of negation. This does not change the number of disuntitled foldertinct variable symbols. Consider the parse tree of a formula in this form. Every interior node is labeled by either $\wedge$, $\vee$, or a quantifier. Consider just the sequence of quantifiers on a path from the root to a leaf: this sequence contains alternating blocks of existential and universal quantifiers. The maximum number of such blocks over all paths in the tree is the *alternation depth* of the formula. For example, the sentence displayed above has alternation depth 2. We write $FO^2_n[<]$ for the fragment of $FO^2[<]$ consisting of formulas with alternation depth no more than n.

2.2 Finite monoids

A *monoid* is a set M together with an associative operation for which there is an identity element $1 \in M$. If Σ is an alphabet, then Σ^* is a monoid with concatenation of words as the multiplication. Σ^* is the *free monoid* on Σ: this means that every map $\alpha : \Sigma \to M$, where M is a monoid, extends in a unique fashion to a homomorphism from Σ^* into M.

Apart from free monoids, all the monoids we consider in this paper are finite. If M is a finite monoid, and $m \in M$, then there is a unique $e \in \{m^k : k > 1\}$ that is *idempotent, i.e.,* $e^2 = e$. We denote this element m^ω.

If M, N are monoids then we say M *divides* N, and write $M \prec N$, if M is a homomorphic image of a submonoid of N.

We are interested in monoids because of their connection with automata and regular languages: A *congruence* on Σ^* is an equivalence relation $\sim$ on Σ^* such that $u_1 \sim u_2$, $v_1 \sim v_2$, implies $u_1v_1 \sim u_2v_2$. The classes of $\sim$ then form a monoid $M = \Sigma^*/\sim$, and the map $u \mapsto [u]_\sigma$ sending each word to its congruence class is a homomorphism from Σ^* onto M. If $L \subseteq \Sigma^*$, then $\equiv_L$, the *syntactic congruence* of L, is the coarsest congruence for which L is a union of congruence classes. The quotient monoid $\Sigma^*/\equiv_L$ is called the *syntactic monoid* of L and is denoted $M(L)$.

We say that a monoid M *recognizes* a language $L \subseteq \Sigma^*$ if there is a homomorphism $\alpha : \Sigma^* \to M$ and a subset X of M such that $\alpha^{-1}(X) = L$. The following proposition gives the fundamental properties linking automata to finite monoids.

▶ **Proposition 1.**
Let $L \subseteq \Sigma^*$.

- *A monoid M recognizes L if and only if $M(L) \prec M$.*
- *L is a regular language if and only if $M(L)$ is finite.*

2.3 Varieties and identities

A collection **V** of finite monoids closed under finite direct products and division is called a *pseudovariety* of finite monoids. (The prefix 'pseudo' is because of the restriction to finite direct products, as the standard use of 'variety' in universal algebra does not include this requirement.)

Let Ξ be the countable alphabet $X = \{x_1, x_2, \ldots\}$. A *term* over Ξ is built from the letters by concatenation and application of a unary operation $v \mapsto v^\omega$. For example, $(x_1x_2)^\omega x_1$ is a term. We will interpret these terms in finite monoids in the obvious way, by considering a valuation $\psi : \Xi \to M$ and giving concatenation and the ω operator their usual meaning in M. For this reason, we do not distinguish between $(uv)w$ and $u(vw)$, where u, v and w are

themselves terms, nor between terms u^ω and $(u^\omega)^\omega$, as these will be equivalent under every valuation.

An *identity* is a formal equation $u = v$, where u and v are terms. We say that a monoid M *satisfies* the identity, and write $M \models (u = v)$, if u and v are equal under every valuation into M. The family of all finite monoids satisfying a given set of identities is a pseudovariety, and we say that the pseudovariety is *defined* by the set of identities. (We hasten to add that the identities we consider here are merely special instances of a much more general class of *pseudoidentities*. Under this broader definition, every pseudovariety is defined by a set of pseudoidentities. See, for instance, Almeida [1].)

We consider three particular pseudovarieties that will be of importance in this paper. First, the pseudovariety **Ap** consists of the *aperiodic* finite monoids, those that contain no nontrivial groups. **Ap** is defined by the identity $x_1^\omega = x_1 x_1^\omega$. If Σ is a finite alphabet and $L \subseteq \Sigma^*$ is a regular language, then $M(L) \in \mathbf{Ap}$ if and only if L is definable by a first-order sentence over $<$.

The pseudovariety **DA** is defined by the pair of identities

$$(x_1x_2x_3)^\omega x_2 (x_1x_2x_3)^\omega = (x_1x_2x_3)^\omega, x_1^\omega = x_1 x_1^\omega.$$

There are many equivalent characterizations of this pseudovariety in terms of other identities, the ideal structure of the monoids, and logic. For us the most important one is this: If $L \subseteq \Sigma^*$, then $M(L) \in \mathbf{DA}$ if and only if L is definable in $FO^2[<]$.

The pseudovariety **J** consists of finite monoids that satisfy the pair of identities

$$(x_1x_2)^\omega = (x_2x_1)^\omega, x_1^\omega = x_1 x_1^\omega.$$

This is equivalent to the identities

$$(x_1x_2)^\omega x_1 = x_2 (x_1x_2)^\omega = (x_1x_2)^\omega, x_1^\omega = x_1 x_1^\omega.$$

Alternatively, **J** consists of finite monoids M such that for all $s, t \in M$, $MsM = MtM$ implies $s = t$. Such monoids are said to be $\mathcal{J}$-trivial.

A theorem due to I. Simon [14] describes the regular languages whose syntactic monoids are in **J**. Let $w \in \Sigma^*$. We denote by $c(w)$ the *content* of w; that is, the set of letters of Σ that appear in w. We say that $v = \sigma_1 \cdots \sigma_k$, where each $\sigma_i \in \Sigma$, is a *subword* of w if

$$w = w_0 \sigma_1 w_1 \cdots \sigma_k w_k$$

for some $w_i \in \Sigma$. We denote by L_v the set of all words in Σ^* of which v is a subword. Let $k \geq 1$. We define an equivalence relation $\sim_k$ on Σ^* that identifies two words if and only if they contain the same subwords of length no more than k. (In particular, $w_1 \sim_1 w_2$ if and only if $c(w_1) = c(w_2)$.) Simon's theorem is:

▶ **Theorem 2.** *Let $L \subset \Sigma^*$ be a regular language. The following are equivalent:*

- *$M(L) \in \mathbf{J}$.*
- *L is a boolean combination of languages of the form L_u, with $u \in \Sigma^*$.*
- *L is a union of $\sim_k$-classes for some $k \geq 1$.*

The equivalence of the last two items is obvious. It is rather easy to show that $\Sigma^*/\sim_k \in \mathbf{J}$, and as a result the last two items imply that L is recognized by a monoid in **J**, and thus by Proposition 1, $M(L) \in \mathbf{J}$. The deep content of the theorem is that the first condition implies the others. The theorem can also be formulated in first-order logic: $M(L) \in \mathbf{J}$ if and only if L is defined by a boolean combination of Σ_1-sentences over $<$.

3 Two-sided Semidirect Products

In this section we describe an operation on both finite monoids and pseudovarieties, the *two-sided semidirect product.* This was given its formal description by Rhodes and Tilson [11], but it has precursors in automata theory in the work of Schützenberger on sequential bimachines [12], Krohn, Mateosian and Rhodes [8], and Eilenberg on triple products [3].

Let M and N be finite monoids. We will follow the standard practice of writing the product in N additively, and thus write its identity element as 0. This is not intended to suggest that N is commutative, but is simply a device for making the notation more readable. A *left action* of M on N is a mapping

$$(m, n) \mapsto mn \in N$$

from $M \times N$ into N that satisfies the axioms

$$\begin{aligned} m(n_1 + n_2) &= mn_1 + mn_2 \\ m_1(m_2 n) &= (m_1 m_2)n \\ m0 &= 0 \\ 1n &= n \end{aligned}$$

for all $m, m_1, m_2 \in M$, $n, n_1, n_2 \in N$.

A *right action* $(m, n) \mapsto nm$ of M on N is defined analogously. A *compatible pair* of actions consists of a left action and a right action of M on N that satisfy the additional axiom

$$m_1(nm_2) = (m_1 n)m_2,$$

for all $m_1, m_2 \in M$, $n \in N$. This justifies the notation $m_1 n m_2$ that we will henceforth use.

Given such a compatible pair, we define a monoid called the *two-sided semidirect product* $N * *M$. The underlying set is just the cartesian product $N \times M$, and the multiplication is given by

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

It is straightforward to verify that this product is associative, and that $(0, 1)$ is the identity element.

Observe that the notation $N * *M$ suppresses mention of the action pair, so in fact there may be several non-isomorphic two-sided semidirect products of N and M. There is always at least one compatible action pair: these are the actions given by $mn = nm = n$ for all m, n. In this case, the resulting two-sided semidirect product reduces to the direct product. Moreover, there is always a compatible pair of actions of M on a direct product of $|M|^2$ copies of N. If we view the latter as the set of maps $F : M \times M \to N$ with componentwise multiplication, then the actions are given by

$$(mF)(m_1, m_2) = F(m_1, mm_2)$$

$$(Fm)(m_1, m_2) = F(m_1 m, m_2).$$

The resulting two-sided semidirect product is called the *block product* of N and M. This monoid has every two-sided semidirect product $N * *M$ as a divisor.

If $\mathbf{V}$ and $\mathbf{W}$ are pseudovarieties of finite monoids then we define $\mathbf{W} * *\mathbf{V}$ to be the collection of finite monoids that divide some two-sided semidirect product $N * *M$ with

$M \in \mathbf{V}$, $N \in \mathbf{W}$. $\mathbf{W} * *\mathbf{V}$ is itself a pseudovariety. We stress that this operation on pseudovarieties is not associative.

Throughout the proof of the main theorem we will use the following description of the regular languages recognized by members of $\mathbf{W} * *\mathbf{V}$. This is adapted from Thérien [18], and is a relatively straightforward translation from the definition of the product. Let $\alpha : \Sigma^* \to M$ be a homomorphism into a finite monoid, and let $\Gamma = M \times \Sigma \times M$. We view Γ as another alphabet. We define a length-preserving map τ_α (not a homomorphism) from Σ^* to Γ^* by

$$\tau_\alpha(\sigma_1 \cdots \sigma_k) = \gamma_1 \cdots \gamma_k,$$

where

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

(If $i = 1$, we interpret the right-hand side as $(1, \sigma_1, \alpha(\sigma_2 \cdots \sigma_k))$, and similarly if $i = k$.)

▶ **Proposition 3.** *Let $L \subseteq \Sigma^*$ be a regular language. $M(L) \in \mathbf{W} * *\mathbf{V}$ if and only if there exist $M \in \mathbf{V}$, and a homomorphism $\alpha : \Sigma^* \to M$, such that L is a boolean combination of sets of the form*

$$\tau_\alpha^{-1}(K) \cap \alpha^{-1}(m),$$

where $m \in M$ and $K \subseteq \Gamma^*$ is recognized by a monoid in $\mathbf{W}$.

4 The Main Theorem

We define a sequence $\mathbf{V}_n$ of pseudovarieties of finite monoids as follows: $\mathbf{V}_1 = \mathbf{J}$, and, for $n \geq 1$, $\mathbf{V}_{n+1} = \mathbf{V}_n * *\mathbf{J}$.

▶ **Theorem 4.** *Let Σ be a finite alphabet, and let $L \subseteq \Sigma^*$. Let $n \geq 1$. $L \in FO_n^2[<]$ if and only if $M(L) \in \mathbf{V}_n$.*

The remainder of this section is devoted to the proof of Theorem 4.

We first prove that if L is recognized by a monoid in $\mathbf{V}_n$, then L is defined by a sentence of $FO_n^2[<]$. We show this by induction on n. For the case $n = 1$, Theorem 2 says that L is a finite boolean combination of languages of the form L_u, where $u \in \Sigma^*$. Each L_u is defined by a two-variable sentence with alternation depth 1 in an obvious way: For example, $L_{\sigma\tau\tau}$ is defined by the sentence

$$\exists x(Q_\sigma x \wedge \exists y(x < y \wedge Q_\tau y \wedge \exists x(y < x \wedge Q_\tau x))).$$

Now suppose $n > 1$. Then L is recognized by a monoid in $\mathbf{V}_{n-1} * *\mathbf{J}$. There are accordingly monoids $M \in \mathbf{J}$, $N \in \mathbf{V}_{n-1}$, and a morphism $\alpha : \Sigma^* \to M$ as in Proposition 3 above. We need to show that there is a formula of alternation depth no more than n defining each language of the form

$$\alpha^{-1}(m) \cap \tau_\alpha^{-1}(K),$$

where $m \in M$ and $K \subseteq \Gamma^* = (M \times \Sigma \times M)^*$ is recognized by N.

By Theorem 2, $\alpha^{-1}(m)$ is a boolean combination of languages of the form L_u, and so, as we saw above, is definable in alternation depth 1. So we turn to $\tau_\alpha^{-1}(K)$. By the inductive hypothesis, K is defined by a sentence ψ of alternation depth no more than $n-1$. The trick is to rewrite ψ to obtain a defining sentence for $\tau_\alpha^{-1}(K)$ while increasing the alternation depth by at most 1. This is accomplished simply by taking each of the atomic formulas $Q_{(m,\sigma,m')}x$ occurring in ψ and replacing it by a formula with x free and with quantifier depth 1. What should this formula say? It must assert that the letter in position x is σ, that the prefix

$v \in \Sigma^*$ of letters preceding this position satisfies $\alpha(v) = m$, and, similarly, that the suffix v' following this position satisfies $\alpha(v') = m'$. The first of these conditions is given by $Q_\sigma x$. The second, is, by Theorem 2, equivalent to a boolean combination of formulas asserting that v contains $\sigma_1, \ldots, \sigma_r$ as a subword, which is expressed by

$$\exists y(y < x \wedge Q_{\sigma_r} y \wedge \exists x(x < y \wedge Q_{\sigma_{r-1}} x \wedge \cdots)),$$

and the third by a boolean combination of analogous formulas. We accordingly replace $Q_{(m,\sigma,m')}x$ by a boolean combination of formulas with alternation depth no more than 1 to obtain the defining sentence for $\tau_\alpha^{-1}(K)$.

We now prove the converse: if L is defined by a sentence of $FO_n^2[<]$, then it is recognized by a monoid in $\mathbf{V}_n$.

Suppose $n > 0$, and let ϕ be a two-variable defining sentence for L. We write this in our standard form described earlier. Let us look at a quantified subformula ψ of ϕ that has quantifier alternation 1 and that is maximal for this property. We call ψ an *innermost block* of ϕ. In terms of the parse tree of ϕ, we are looking for nodes of minimal depth that are labeled by a quantifier, and such that every quantifier in the subtree rooted at this node is of the same type. The innermost blocks of ϕ are the formulas given by these subtrees.

If ϕ itself has quantifier alternation 1, then each innermost block ψ is a sentence, and ϕ is obtained from these blocks by disjunction and conjunction. Otherwise ψ has one free variable. Let's say this free variable is x. Suppose the quantifier in ψ is $\exists$. (If the quantifier in ψ is $\forall$ then its negation is a formula in which the only quantifier is $\exists$; we apply the transformations described below to this existential formula.) Since there are no negations in ψ we can, in the standard way (but introducing new variables in the process) rewrite ψ in prefix form as an ordinary Σ_1 formula

$$\exists y_1 \exists y_2 \cdots \exists y_r \theta(x, y_1, \ldots, y_r).$$

where θ is quantifier-free.

If $n = 1$ then the free variable x does not appear. Thus we can further rewrite ψ as a disjunction of sentences of the form

$$\exists y_1 \exists y_2 \cdots \exists y_r \Big(\bigwedge_{i=1}^{r} Q_{\sigma_i} y_i \wedge \rho(y_1, \ldots, y_r) \Big),$$

where ρ uniquely specifies the ordering among the y_i. (For example, with $r = 3$, ρ might have the form $y_1 = y_3 < y_2$.) Seen this way, ψ simply asserts the presence of certain subwords (and, had we begun with a universal quantifier, the absence of certain subwords.) In this case ϕ defines a boolean combination of languages of the form L_u, which by Theorem 2, is recognized by a monoid in $\mathbf{J}$. This is the base of our induction.

If $n > 1$, then we rewrite ψ as a disjunction of formulas of the form

$$\exists z_1 \cdots \exists z_t \exists z'_1 \cdots \exists z'_t \theta,$$

where θ is

$$\bigwedge_{i=1}^{t} (Q_{\sigma_i} z_i \wedge (z_i < x)) \wedge \rho_1(z_1, \cdots, z_t) \wedge \bigwedge_{j=1}^{t'} (Q_{\sigma'_j} z'_j \wedge (z'_j > x)) \wedge \rho_2(z'_1, \cdots, z'_t).$$

Let's denote this formula, which has x free, by

$$\zeta[x, u, v, \rho_1, \rho_2],$$

where $u = (\sigma_1, \ldots, \sigma_t)$ $v = (\sigma'_1, \ldots, \sigma'_{t'})$. If we started with an innermost block beginning with a universal quantifer, then this procedure produces the negation of a disjunction of these formulas. So we suppose ϕ has been transformed so that all its innermost blocks have been replaced by boolean combinations of such ζ.

Let s be the maximum of all the t, t' that occur in these formulas. Let $M = \Sigma^*/\sim_s$, as defined in Section 2. Let $\alpha : \Sigma^* \to M$ be the homomorphism that maps each word to its $\sim_s$-class. Recall that $M \in \mathbf{J}$.

We now rewrite ϕ and replace it by a new sentence ϕ' over the alphabet $\Gamma = M \times \Sigma \times M$. The idea is simply to express properties of a word $w \in \Sigma^*$ in terms of properties of $\tau_\alpha(w) \in \Gamma^*$ This is easy to do, because the two words have the same set of positions, and because the letters of $\tau_\alpha(w)$ encode additional information about each position. The subformula $\zeta[x, u, v, \rho_1, \rho_2]$ states that the prefix of w consisting of positions to the left of the position x contains a certain subword w_1 of length no more than s, and that the suffix consisting of positions to the right of x contains another such subword w_2. Equivalently, the letter of $\tau_\alpha(w)$ in position x is (m, σ, m'), where the $\sim_s$-class m contains w_1 as a subword, and the $\sim_s$-class m' contains w'_2. We thus replace each ζ by a disjunction of the atomic formulas $Q_{(m,\sigma,m')}x$ over all such m, m'. The result is that all the innermost blocks have now been eliminated and replaced by a boolean combination of these atomic formulas, which can in turn be written as a disjunction of such formulas.

We also replace each $Q_\sigma x$ that occurs outside an innermost block by the disjunction of the $Q_{(m,\sigma,m')}x$ over all $m, m' \in M$. The resulting sentence ϕ' is a two-variable sentence of quantifier depth $n-1$. Thus, by the induction hypothesis, the language $K \subseteq \Gamma^*$ defined by ϕ' is recognized by a monoid N in $\mathbf{V}_{n-1}$. We have constructed ϕ' so that $w \models \phi$ if and only if $\tau_\alpha(w) \models \phi'$. Thus, by Proposition 3, L is recognized by a monoid in $\mathbf{V}_{n-1} * *\mathbf{J} = \mathbf{V}_k$.

5 Strictness and Decidability

Here we use our main theorem to give a new proof that the alternation hierarchy is strict. This was first shown in [20]. We also discuss the question of decidability of the levels of the hierarchy.

We define two sequences of terms $\{u_n\}_{n\geq 1}$, $\{v_n\}_{n\geq 1}$ as follows.

$$u_1 = (x_1x_2)^\omega, v_1 = (x_2x_1)^\omega.$$

If $n \geq 1$, we set

$$u_{n+1} = (x_1 \cdots x_{2n}x_{2n+1})^\omega u_n (x_{2n+2}x_1 \cdots x_{2n})^\omega$$

$$v_{n+1} = (x_1 \cdots x_{2n}x_{2n+1})^\omega v_n (x_{2n+2}x_1 \cdots x_{2n})^\omega$$

▶ **Proposition 5.** *Let $n \geq 1$. If $M \in \mathbf{V}_n$, then $M \models (u_n = v_n)$, and $M \models (x_1^\omega = x_1x_1^\omega)$.*

Proof. For $n = 1$, the Proposition follows from the identities defining $\mathbf{J}$ that were given in Section 2. For the inductive step we will make repeated use of the following identities that also hold in $\mathbf{J}$, and that are direct consequences of the ones we gave earlier:

$$(x_1x_2x_3)^\omega x_2 = (x_1x_2x_3)^\omega = x_2(x_1x_2x_3)^\omega \tag{1}$$

Suppose then that $n \geq 1$, and that the Proposition holds for n. It is well known—and in any case follows easily from the kind of argument we give below—that the two-sided

semidirect product preserves aperiodicity. So we will only show $M \models (u_{n+1} = v_{n+1})$ for $M \in \mathbf{V}_{n+1}$. Since satisfaction of identities is preserved under division, we only need to show this in the case where M is a two-sided semidirect product $T * *K$, with $T \in \mathbf{V}_n$ and $K \in \mathbf{J}$. Consider a map ϕ from the set $\{x_1, x_2, \ldots\}$ into M with

$$\phi(x_i) = m_i = (t_i, k_i) \in T * *K$$

for all $i \geq 1$. Now suppose $x_{i_1} \cdots x_{i_p}$ is a term formed just by concatenating variables (*i.e.*, without using ω). Then

$$\phi(x_{i_1} \cdots x_{i_p}) = \left(\sum_{j=1}^{p} k_{i_1} \cdots k_{i_{j-1}} t_{i_j} k_{i_{j+1}} \cdots k_{i_p}, k_{i_1} \cdots k_{i_p} \right). \tag{2}$$

There is an integer q such that $m^\omega = m^q$ for all $m \in T, K$ or M. Thus

$$\begin{aligned} \phi(u_{n+1}) &= \phi((x_1 x_2 \cdots x_{2n+1})^q u_n (x_{2n+2} x_1 \cdots x_{2n})^q) \\ &= (t, \gamma(u_n)), \end{aligned}$$

where t is a sum of the form displayed in Equation 2, and $\gamma(x_i) = k_i$ for all i. Let us analyze the summands of t. Let $s = q \cdot (2n+1)$. If $j \leq s$, then the j^{th} summand is

$$k_{i_j} = k_{i_1} \cdots k_{i_{j-1}} t_{i_j} k_R,$$

where

$$k_R = (k_{2n+2} k_1 \cdots k_{2n})^q.$$

This follows from the absorbing property given in Equation 1 of the $\mathcal{J}$-trivial monoid K. Similarly, if $s \geq p - j$, the j^{th} summand is

$$k_L t_{i_j} k_{i_{j+1}} \cdots k_{i_p},$$

where

$$k_L = (k_1 \cdots k_{2n} k_{2n+1})^q.$$

If $s < j < n - s$, then the j^{th} summand is $k_L t_{i_j} k_R$, so that the sum of these middle terms is

$$\sum_{j=s+1}^{n-s} k_L t_{i_j} k_R = k_L \left(\sum_{j=s+1}^{n-s} t_{i_j} \right) k_R = k_L \psi(u_n) k_R,$$

where $\psi(x_i) = t_i$ for all i. We thus can write $\phi(u_{n+1})$ in the form

$$\phi(u_{n+1}) = (t_L + k_L \psi(u_n) k_R + t_R, \gamma(u_{n+1})).$$

When we compute $\phi(v_{n+1})$, the values of t_L and t_R are unchanged, and we have

$$\phi(v_{n+1}) = (t_L + k_L \psi(v_n) k_R + t_R, \gamma(v_{n+1})).$$

From the identities for $\mathbf{J}$ in Equation 1 we have $K \models (u_{n+1} = v_{n+1})$, and thus $\gamma(u_{n+1}) = \gamma(v_{n+1})$. From the inductive hypothesis we have $T \models (u_n = v_n)$, so $\psi(u_n) = \psi(v_n)$, and thus $\phi(u_{n+1}) = \phi(v_{n+1})$, as required.

◀

We will use these identities to show that the alternation hierarchy is strict. We begin by defining, for each finite alphabet Σ, an equivalence relation $\equiv_\Sigma$ on Σ^*. (In fact, $\equiv_\Sigma$ will be a congruence on Σ^* whose quotient is the *free idempotent monoid* on Σ. This construction is very well known; see, for, example, Eilenberg [3].)

If $|\Sigma| = 1$, then $\equiv_\Sigma$ identifies two distinct words over Σ if and only if they are both nonempty (so that there are two equivalence classes, one containing the empty word, and the other containing all the nonempty words). Now suppose $|\Sigma| > 1$, and that $\equiv_\Gamma$ has been defined for all proper subalphabets Γ of Σ. Let $w_1, w_2 \in \Sigma^*$. If the set of distinct letters $\Gamma = c(w_1)$ appearing in w_1 is a proper subset Γ of Σ, then we set $w_1 \equiv_\Sigma w_2$ if and only if $c(w_2) = \Gamma$, and $w_1 \equiv_\Gamma w_2$. Otherwise, $c(w_1) = c(w_2) = \Sigma$. Let u_i denote the maximal prefix of w_i such that $c(u_i) \neq \Sigma$, and similarly let v_i denote the maximal suffix of w_i such that $c(v_i) \neq \Sigma$. We can then write

$$w_i = u_i\sigma_i y_i = z_i\tau_i v_i,$$

where $\sigma_i, \tau_i \in \Sigma$. We define $w_1 \equiv_\Sigma w_2$ if and only if $\sigma_1 = \sigma_2$, $\tau_1 = \tau_2$, $u_1 \equiv_{c(u_1)} u_2$, and $v_1 \equiv_{c(v_1)} v_2$.

Easily, $\equiv_\Sigma$ is a congruence of finite index on Σ^*. We denote the $\equiv_\Sigma$-class of $w \in \Sigma^*$ by $[w]_\equiv$. The language $[w]_\equiv$ is regular; moreover, for every word $u \in \Sigma^*$, $u \equiv_\Sigma u^2$, which implies that $m^2 = m$, or, equivalently $m^\omega = m$, for every $m \in M([w]_\Sigma)$.

▶ **Lemma 6.** *Let $|\Sigma| = n$. Every class of $\equiv_\Sigma$ is definable by a sentence of $FO^2_n[<]$.*

Proof. We prove this by induction on n. For $n = 1$, we have $\Sigma = \{\sigma\}$, and the two classes are defined by the sentences

$$\exists x Q_\sigma x$$

and

$$\forall x(x < x).$$

(Note that we allow our formulas to be interpreted in the empty word, which satisfies every universally quantified sentence.)

Assume now that $n > 1$, and that the claim is true for all subalphabets of Σ. Let $w \in \Sigma^*$. If $c(w) \neq \Sigma$, then $[w]_\Sigma = [w]_\Gamma$ for some proper subalphabet Γ of Σ. The inductive hypothesis implies that this class is defined by a sentence of $FO^2_{n-1}[<]$. So we assume $c(w) = \Sigma$, and write $w = u\sigma x = y\tau v$, where u, v are, respectively, the maximal prefix and suffix of w that do not contain all the letters of Σ. To express the property that every letter except σ occurs in the prefix w we use the sentence

$$\exists x(Q_\sigma x \wedge \bigwedge_{\sigma' \neq \sigma} \exists y(y < x \wedge Q_{\sigma'} y) \wedge \forall y(y \geq x \vee \bigvee_{\sigma' \neq \sigma} Q_{\sigma'} y))).$$

Note that this sentence has alternation depth $2 \leq n$.

To express the property that the prefix preceding the first position containing σ belongs to a particular $\equiv_\Gamma$-class, where $\Gamma = c(u)$, we apply the inductive hypothesis: There is a sentence ψ of alternation depth less than n defining $[u]_\Gamma$. We modify ψ by replacing each existentially quantified subformula $\exists x\zeta$ by

$$\exists x(\zeta \wedge \forall y(y \geq x \vee \bigvee_{\sigma' \neq \sigma} Q_{\sigma'} y)),$$

and each universally quantified subformula $\forall x\zeta$ by

$$\forall x(\zeta \vee \exists y(y \leq x \wedge Q_\sigma y)).$$

The resulting sentence has alternation depth no more than n and defines the set of strings such that the maximal prefix that does not contain σ is in $[u]_\Gamma$. The conjunction of these two sentences, along with the analogues for the suffix, defines $[w]_\Sigma$.

◀

▶ **Lemma 7.** *Let $n \geq 1$, and let $|\Sigma| = 2n$. There is a word $w \in \Sigma^*$ such that $M([w]_\Sigma)$ does not satisfy $(u_n = v_n)$.*

Proof. Let u'_n and v'_n be the terms that result from removing all occurrences of the operator ω from u_n and v_n, respectively. Let $\Sigma = \{\sigma_1, \cdots, \sigma_{2n}\}$, and let $w_1^{(n)}, w_2^{(n)} \in \Sigma^*$ be the respective words that result when each occurrence of a variable x_i in u_n or v_n is replaced by σ_i. It is enough to show that $w_1^{(n)} \not\equiv_\Sigma w_2^{(n)}$. The reason is this: We can take $M = M([w_1^{(n)}]_\Sigma)$. Since $M \models (x^\omega = x)$, if we had $M \models (u_n = v_n)$ then $M \models (u'_n = v'_n)$. But that case we would obtain $w_1^{(n)} \equiv_\Sigma w_2^{(n)}$.

We prove that $w_1^{(n)} \not\equiv_\Sigma w_2^{(n)}$ by induction on n. For $n = 1$ we have

$$w_1^{(1)} = \sigma_1\sigma_2 \not\equiv_\Sigma \sigma_2\sigma_1 = w_2^{(1)}.$$

For $n > 1$ we have

$$w_j^{(n)} = \sigma_1 \cdots \sigma_{2n-1} w_j^{(n-1)} \sigma_{2n}\sigma_1 \cdots \sigma_{2n-2},$$

for $j = 1, 2$. The maximal prefix of $w_j^{(n)}$ that does not contain all the letters of Σ is $z_j = \sigma_1 \cdots \sigma_{2n-1} w_j^{(n-1)}$, and the maximal suffix of z_j not containing all the letters of $c(z_j)$ is $w_j^{(n-1)}$. By the inductive hypothesis, $w_1^{(n-1)} \not\equiv_\Gamma w_2^{(n-1)}$, where $\Gamma = \{\sigma_1, \ldots, \sigma_{2n-2}\}$, so we cannot have $w_1^{(n)} \equiv_\Sigma w_2^{(n)}$.

◀

We get the strictness of the hierarchy as a consequence of these two lemmas:

▶ **Theorem 8.** *For every $n > 1$ there is a language definable in $FO^2_n[<]$ that is not definable in $FO^2_{n-1}[<]$.*

Proof. For every $n > 1$, we must have $\mathbf{V}_{n-1} \subsetneq \mathbf{V}_n$, since equality at one level would imply equality at all higher levels, and we would have, in particular, $\mathbf{V}_n = \mathbf{V}_{2n}$ for some $n \geq 1$. But the two Lemmas, coupled with Theorem 4 and Proposition 5, provide an example of a language whose syntactic monoid is in $\mathbf{V}_{2n} \backslash \mathbf{V}_n$. Thus $\mathbf{V}_{n-1} \subsetneq \mathbf{V}_n$. Since every pseudovariety is generated by the syntactic monoids it contains, Theorem 4 gives the result. ◀

We now discuss the problem of *decidabilty:* Suppose we are given a regular language $L \subseteq \Sigma^*$, either by an automaton that recognizes L, or in terms of some other representation, such as a regular expression, from which we can effectively construct an automaton. Is there an algorithm for determining whether L can be defined by a sentence of $FO^2_n[<]$ for a fixed n? Of course, this begs the question of whether L can be defined by a sentence of $FO^2[<]$ at all, but this problem is solved by earlier work: Compute the syntactic monoid of $M(L)$ and determine whether $M(L)$ is in **DA**, by verifying the identities for **DA**. (Note that in verifying the identities in a particular monoid M, the symbol ω in these identities can be replaced by $|M|$.)

Algebraic methods provide a powerful tool for answering such decision questions (and, more generally, for proving that a given language cannot be defined in a logic, as we did in Lemma 7 above), since the multiplication table of the syntactic monoid of L can be effectively computed from any reasonable representation of L. However, in order to apply this method

here, we need an algorithm for determining whether a given finite monoid belongs to $\mathbf{V}_n$, for any given n. Identities defining these pseudovarieties would provide us with precisely such an algorithm, but the identities $(u_n = v_n, x_1^\omega = x_1 x_1^\omega)$ that we have exhibited have only been proved to be necessary conditions for membership in $\mathbf{V}_n$.

In fact, these identities are from a paper by by Almeida and Weil [2], where they appear as part of a general scheme for obtaining identities for pseudovarieties of the form $\mathbf{V} * *\mathbf{J}$ when $\mathbf{V} \subseteq \mathbf{DA}$ and identities for $\mathbf{V}$ are known. The result stated there would imply that satisfaction of $(u_n = v_n, x_1^\omega = x_1 x_1^\omega)$ is also a sufficient condition for membership in $\mathbf{V}_n$, and thus resolve the decidability question. However, this paper is known to contain an error. A second paper, by Weil [19], explains the nature of the problem: The proof that the identities are sufficient requires a particular finite rank property for categories that are globally covered by members of $\mathbf{V}_{n-1}$. (Even defining these terms would take us too far afield; the interested reader is referred to [2] and [19] and the many references given there.)

As we have already mentioned, for $n = 1$ the identities $(u_n = v_n, x_1^\omega = x_1 x_1^\omega)$ are known to define $\mathbf{J}$. The finite rank property, thanks to a theorem of Knast [7], is known to hold for $\mathbf{J}$, and therefore the identities for $n = 2$ do indeed define $\mathbf{J} * *\mathbf{J}$. As a consequence, we have:

▶ **Theorem 9.** *It is decidable whether a given regular language is definable in* $FO_1^2[<]$, *or in* $FO_2^2[<]$.

. For level 1, the answer is again membership of the syntactic monoid in $\mathbf{J}$; for the second level the answer is unknown. We suspect that the problem of alternation depth in $FO^2[<]$, while still challenging, will turn out to be easier.

The finite rank property is not known to hold for $\mathbf{V}_{n-1}$, if $n > 2$. Thus the decidability problem remains open for higher levels of the hierarchy. This does not rule out the possibility that the identities might be proved sufficient even without the assumption of finite rank.

▶ **Conjecture 10.** *Let* $n \geq 1$. $M \in \mathbf{V}_n$ *if and only if* $M \models (u_n = v_n)$, *and* $M \models (x_1^\omega = x_1 x_1^\omega)$. *In particular, it is decidable whether a given regular language is definable in* $FO_n^2[<]$.

. For level 1, the answer is again membership of the syntactic monoid in $\mathbf{J}$; for the second level the answer is unknown. We suspect that the problem of alternation depth in $FO^2[<]$, while still challenging, will turn out to be easier.

Kufleitner and Weil [9] also study the alternation hierarchy algebraically, and introduce a very different-looking sequence of pseudovarieties $\mathbf{W}_n$ with the property that the syntactic monoid of every language with alternation depth exactly n is between $\mathbf{W}_n$ and $\mathbf{W}_{n+1}$. These pseudovarieties are known to have decidable membership problems. Kufleitner and Weil conjecture that in fact this sequence of pseudovarieties exactly captures the alternation hierarchy. This conjecture would settle the decision problem for alternation depth (and would also, coupled with our results, imply $\mathbf{W}_n = \mathbf{V}_n$ for all n.) It would be interesting to try to establish containments between $\mathbf{W}_n$ and $\mathbf{V}_n$.

Finally, we mention the similar-looking problem of *dot-depth.* Full first-order logic over $<$ interpreted in finite words defines all the languages with syntactic monoids in $\mathbf{Ap}$. The problem of determining the exact alternation depth of a language in this setting–the problem of calculating the so-called *dot-depth* of a language–has been open for nearly forty years. For level 1, the answer is again membership of the syntactic monoid in $\mathbf{J}$; for the second level the answer is unknown. We suspect that the problem of alternation depth in $FO^2[<]$, while still challenging, will turn out to be easier.

Acknowledgements. Many thanks to Phillip Weis, Neil Immerman, Jorge Almeida and Pascal Weil for sharing and discussing their work with me.

References

1. J. Almeida, *Finite Semigroups and Universal Algebra,* World Scientific, Singapore, 1994.
2. J. Almeida, P. Weil. "Profinite Categories and Semidirect Products, *J. Pure and Applied Algebra* **123** (1998) 1 - 50.
3. S. Eilenberg, *Automata, Languages and Machines,* vol. B, Academic Press, New York, 1976.
4. K. Etessami, M. Vardi, and T. Wilke, "First-Order Logic with Two Variables and Unary Temporal Logic", *Proceedings, 12th IEEE Symposium on Logic in Computer Science,* 228-235 (1996).
5. N. Immerman and D. Kozen, "Definability with a Bounded Number of Bound Variables", *Information and Computation,* **83**, 121-139 (1989).
6. J. Kamp, *Tense Logic and the Theory of Linear Order,* Ph.D. thesis, UCLA (1968).
7. R. Knast, "Some Theorems on Graph Congruences", *Informatique Théorique et Applications,* 331-342 (1983).
8. K. Krohn, R. Mateosian, and J. Rhodes, "Methods of the Algebraic Theory of Machines. I: Decomposition Theorem for Generalized Machines: Properties Preserved under Series and Parallel Compositions of Machines", *J. Comput. Syst. Sci,* 55-85 (1967).
9. M. Kufleitner, P. Weil, "On FO2 Quantifier Alternation over Words", MFCS 2009, Lecture Notes in Computer Science **5734** (Springer, 2009), pp. 513-524.
10. J. E. Pin, *Varieties of Formal Languages,* Plenum, London, 1986.
11. J. Rhodes and B. Tilson, "The Kernel of Monoid Morphisms", *J. Pure and Applied Algebra* **62** (1989) 227–268.
12. "A remark on finite transducers", *Information and Control* **4** (1961), 185-196.
13. T. Schwentick, D. Thérien and H. Vollmer. "Partially-ordered Two-way Automata: A New Characterization of **DA**', In *Proc. 5th Developments in Language Theory (DLT 2001),* (2001) 239-250.
14. I. Simon. "Piecewise testable events", in *Automata Theory and Formal Languages,* (1975), 214-222.
15. H. Straubing, *Finite Automata, Formal Logic and Circuit Complexity,* Birkhäuser, Boston, 1994.
16. H. Straubing and D. Thérien, "Weakly Iterated Block Products of Finite Monoids", *LATIN 2002, Lecture Notes in Computer Science,* **2286** (Springer, 2002),91-104.
17. P. Tesson and D. Thérien, "Diamonds are Forever: The Variety **DA**", in *Semigroups, Algorithms, Automata and Languages*, World Scientific, Singapore (2002), 475-500.
18. D. Thérien, "Two-sided wreath products of categories", *J. Pure and Applied Algebra* **74** (1991) 307-315.
19. P. Weil. "Profinite Methods in Semigroup Theory, *Intern. J. Algebra and Computation* **12** (2002) 137-178.
20. P. Weis and N. Immerman, "Structure Theorem and Strict Alternation Hierarchy for FO^2 on Words", *Logical Methods in Computer Science,* **5** (2009).