Quantifier Alternation in Two-Variable First-Order Logic with Successor Is Decidable*

Manfred Kufleitner and Alexander Lauser

University of Stuttgart, FMI, Germany
`{kufleitner,lauser}@fmi.uni-stuttgart.de`

Abstract

We consider the quantifier alternation hierarchy within two-variable first-order logic $\mathrm{FO}^2[<, \mathrm{suc}]$ over finite words with linear order and binary successor predicate. We give a single identity of omega-terms for each level of this hierarchy. This shows that for a given regular language and a non-negative integer m it is decidable whether the language is definable by a formula in $\mathrm{FO}^2[<, \mathrm{suc}]$ which has at most m quantifier alternations. We also consider the alternation hierarchy of unary temporal logic $\mathrm{TL}[\mathsf{X}, \mathsf{F}, \mathsf{Y}, \mathsf{P}]$ defined by the maximal number of nested negations. This hierarchy coincides with the $\mathrm{FO}^2[<, \mathrm{suc}]$ quantifier alternation hierarchy.

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

Keywords and phrases automata theory, semigroups, regular languages, first-order logic

Digital Object Identifier 10.4230/LIPIcs.STACS.2013.305

1 Introduction

Around 1960, Büchi, Elgot and Trakhtenbrot independently showed that monadic second-order logic (MSO) over finite words defines the class of regular languages [2, 6, 33]. Since then numerous fragments of MSO have been considered. A theoretical motivation for fragments is the study of the rich structure within the regular languages. For this purpose, fragments form the basis of a descriptive complexity theory: The simpler the formula for defining a language is, the simpler this language is. From a practical point of view, simpler fragments often lead to more efficient algorithms for decision problems such as satisfiability.

The most prominent fragment of MSO is first-order logic FO. The atomic predicates of FO are the unary predicate $\lambda(x) = a$ stating that position x is labeled by the letter a, and the binary predicates $x = y$ and $x < y$ with the natural interpretation. The successor predicate $\mathrm{suc}(x, y)$ is easily definable in FO by saying that $x < y$ and that there is no position between x and y. McNaughton and Papert showed that a language is FO-definable if and only if it is star-free [18]. Combined with Schützenberger's characterization of star-free languages in terms of finite aperiodic monoids [21], it follows that a language is FO-definable if and only if its syntactic monoid is aperiodic. The latter property is decidable and one can thus effectively check whether a regular language (given *e.g.* by a nondeterministic automaton or an MSO formula) is definable in FO. The two most famous hierarchies within FO are the Straubing-Thérien hierarchy and Brzozowski's dot-depth hierarchy. The Straubing-Thérien hierarchy coincides with the quantifier alternation inside FO without the successor predicate [25, 29], and Brzozowski's dot-depth hierarchy is captured by quantifier alternation including the successor predicate [3]; see also [20, 31]. Here, quantifier alternation is defined in terms of blocks of quantifiers for formulae in prenex normal form. Note that

* The authors were supported by the German Research Foundation (DFG) under grant DI 435/5-1.

30th Symposium on Theoretical Aspects of Computer Science (STACS'13).
Editors: Natacha Portier and Thomas Wilke; pp. 305–316
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
SYMPOSIUM ON THEORETICAL ASPECTS OF COMPUTER SCIENCE

by introducing new variables, every formula is equivalent to a formula in prenex normal form. Deciding membership of level m for these hierarchies is one of the most challenging open problems in automata theory. To date only the very first levels (*i.e.*, $m = 1$) of both hierarchies are known to be decidable [9, 24].

By Kamp's Theorem, first-order logic FO^3 with only three different names for the variables and full first-order logic FO have the same expressive power [8]. However, two variables are not sufficient for defining all first-order definable languages. The fragment FO$^2[<]$ without successor predicate has a huge number of different characterizations; see *e.g.* [5, 28]. One of them is the variety **DA** of finite monoids [22]; *cf.* [30]. For quantifier alternation inside FO^2 one cannot readily rely on prenex normal forms. However, in FO^2 negations can be moved towards the atomic formulae, and hence every formula is equivalent to a negation-free counterpart. The fragment FO^2_m consists of all FO^2-formulae whose negation-free counterpart has at most m blocks of quantifiers on each path of the parse tree. Kufleitner and Weil have shown that for every $m \geq 1$ it is decidable whether a given regular language is definable in FO$^2_m[<]$ without successor predicate [16]. They have given an effective algebraic characterization in terms of levels of the Trotter-Weil hierarchy of finite monoids [34]; see also [15]. In addition, restrictions of many other characterizations of the FO$^2[<]$-definable languages admit algebraic counterparts within this hierarchy [12, 17]. The proof of Kufleitner and Weil's characterization of FO$^2_m[<]$ relies on a combinatorial tool known under the terms *ranker* [35] and *turtle program* [23]. A connection between FO$^2_m[<]$ and rankers was established by Weis and Immerman [35] and further exploited by Kufleitner and Weil [17]. Straubing has given another algebraic characterization of FO$^2_m[<]$ in terms of weakly iterated block products of $\mathcal{J}$-trivial monoids [27]. Recently, Krebs and Straubing [10] were able to use this characterization for giving identities of omega-terms for FO$^2_m[<]$, thereby obtaining another effective characterization of FO$^2_m[<]$.

In this paper, we consider the quantifier alternation hierarchy inside FO$^2[<, \mathrm{suc}]$ with successor predicate. The logic FO$^2[<, \mathrm{suc}]$ is strictly more expressive than FO$^2[<]$ without successor. Thérien and Wilke [30] have given an algebraic characterization of FO$^2[<, \mathrm{suc}]$ which, by a previous result of Almeida, is known to coincide with the decidable variety **LDA** of finite semigroups [1]; see also [4]. For every $m \geq 2$ we give a single identity of omega-terms such that a language is definable in FO$^2_m[<, \mathrm{suc}]$ if and only if its syntactic semigroup satisfies this identity. It is thus decidable whether a given regular language is FO$^2_m[<, \mathrm{suc}]$-definable.

Our proof is by induction on m with Knast's Theorem on dot-depth one [9] as base case. For $m = 1$, there is a small difference between the availability and the absence of min- and max-predicates; this is identical to the situation for dot-depth one [11]. The main ingredients of our proof are *(i)* string rewriting techniques, *(ii)* combinatorial properties of **LDA**, and *(iii)* relativization techniques for FO^2_m. As a byproduct, we show that quantifier alternation in FO$^2[<, \mathrm{suc}]$ coincides with alternation in unary temporal logic $\mathrm{TL}[\mathsf{X}, \mathsf{F}, \mathsf{Y}, \mathsf{P}]$ where the latter is based on the nesting depth of negations. This last property can also be seen using a translation from FO^2 to unary temporal logic by Etessami, Vardi, and Wilke [7].

Missing proofs can be found in the technical report [14].

2 Preliminaries

Throughout, A denotes a finite alphabet. The set of all finite words is A^* and the set of all finite, nonempty words is A^+. Let $u = a_1 \cdots a_n$ with $a_i \in A$. The set of *positions* of u is $\mathrm{pos}(u) = \{1, \ldots, n\}$ and its *length* is $|u| = n$. If I is an interval, then $u[I]$ denotes the factor of u covered by the interval of positions $\mathrm{pos}(u) \cap I$. If $I = [i; j]$, then $u[i; j]$ is an abbreviation for $u[I]$. In particular, if $1 \leq i \leq j \leq n$, then $u[i;j] = a_i \cdots a_j$. The *k-factor alphabet* is $\text{alph}_k(u) = \{a_i \cdots a_{i+k-1} \in A^k \mid 1 \leq i \leq n-k+1\}$

First-Order Logic. We consider first-order logic over finite words with order and successor predicates. Atomic first-order formulae are $\top$ for *true*, $\bot$ for *false*, label predicates $\lambda(x) = a$ with $a \in A$, comparisons $x = y$, $x < y$ and successor $\text{suc}(x, y)$ as well as minimum $\min(x)$ and maximum $\max(x)$. Here x and y are variables ranging over positions of a word which forms a model as a labeled, linearly ordered set of positions. Formulae can be composed by the usual Boolean connectives, *i.e.*, if φ and ψ are first-order formulae, then so are the disjunction $\varphi \vee \psi$, the conjunction $\varphi \wedge \psi$, and the negation $\neg\varphi$. Moreover, formulae can be composed by existential quantification $\exists x\, \varphi$ and universal quantification $\forall x\, \varphi$. The semantics is as usual; see *e.g.* [13, 32]. We use the notation $\varphi(x_1, \ldots, x_n)$ to indicate that at most the variables $x_1, \ldots, x_n$ occur freely in φ. We write $u \models \varphi(i_1, \ldots, i_n)$ for $u \in A^*$ and positions $i_j \in \text{pos}(u)$ if φ is true over u with x_j being interpreted by i_j. A formula without free variables is a *sentence* and in this case we simply write $u \models \varphi$. For any class $\mathcal{F}$ of first-order formulae, $\mathcal{F}[\mathcal{C}]$ is the restriction to formulae in $\mathcal{F}$ which, apart from $\top$, $\bot$, label predicates, and equality, only use predicates in $\mathcal{C} \subseteq \{<, \text{suc}, \min, \max\}$.

The fragment $\text{FO}^2 = \text{FO}^2[<, \text{suc}, \min, \max]$ of first-order logic contains all formulae which use at most two different names for variables, say x and y. For FO^2-formulae $\varphi(x)$ with free variable x we stipulate the convention that $\varphi(y)$ is the FO^2-formula obtained by interchanging x and y. Using De Morgan's laws and the usual dualities between existential and universal quantifiers, one can see that every formula in FO^2 is equivalent to a formula with negations only applied to atomic formulae. We call such formulae *negation-free* (since negations could be eliminated by adding negative predicates to an extended signature). The fragment FO^2_m consists of all formulae in FO^2 with quantifier alternation depth at most m, *i.e.*, formulae such that the negation-free counterpart has at most m blocks of quantifiers on every path of the parse tree. Therefore, if we drop the two-variable restriction, every FO^2_m-formula admits a prenex normal form with m blocks of quantifiers. In other words negation-free formulae in FO^2_m have at most $m-1$ alternations of nested existential and universal quantifiers. Note that FO^2_m is closed under negation. The fragment $\text{FO}^2_{m,n}$ contains all formulae in FO^2_m with quantifier depth at most n.

Unary Temporal Logic. Unary temporal logic $\text{TL}[\mathsf{X}, \mathsf{F}, \mathsf{Y}, \mathsf{P}]$ consists of all formulae built from $\top$ for *true*, $\bot$ for *false*, labels a with $a \in A$, compositions using Boolean connectives as in first-order logic, and temporal modalities $\mathsf{X}\, \varphi$, $\mathsf{F}\, \varphi$, $\mathsf{Y}\, \varphi$, and $\mathsf{P}\, \varphi$ for $\varphi \in \text{TL}[\mathsf{X}, \mathsf{F}, \mathsf{Y}, \mathsf{P}]$. Formulae of unary temporal logic are interpreted over a word relative to a current position. The semantics is declared by the following FO^2-formulae in one free variable: We let $a(x) \equiv \big(\lambda(x) = a\big)$ and

$$\begin{aligned} (\mathsf{X}\, \varphi)(x) &\equiv \exists y \left(\text{suc}(x, y) \wedge \varphi(y)\right), & (\mathsf{F}\, \varphi)(x) &\equiv \exists y \left(x \leq y \wedge \varphi(y)\right), \\ (\mathsf{Y}\, \varphi)(x) &\equiv \exists y \left(\text{suc}(y, x) \wedge \varphi(y)\right), & (\mathsf{P}\, \varphi)(x) &\equiv \exists y \left(y \leq x \wedge \varphi(y)\right). \end{aligned}$$

Here and in the sequel, $\equiv$ means syntactic equality. We often use this symbol instead of equality in order to avoid confusion with the symbol $=$ occurring in atomic predicates. The formulae for the remaining constructs are as usual. The modalities X (neXt) and F (Future) are called *future modalities* whereas the modalities Y (Yesterday) and P (Past) are called *past modalities*. In order to define $u \models \varphi$ without a distinguished position in u, we start evaluation in front (position 0) for future modalities and after (position $|u| + 1$) the word u

for past modalities. More formally, for a word $u \in A^*$ we define $u \not\models a$ and

$$u \models \mathsf{X}\,\varphi \text{ if and only if } u \models \varphi(1), \qquad u \models \mathsf{F}\,\varphi \text{ if and only if } u \models \mathsf{F}\,\varphi(1),$$
$$u \models \mathsf{Y}\,\varphi \text{ if and only if } u \models \varphi(|u|), \qquad u \models \mathsf{P}\,\varphi \text{ if and only if } u \models \mathsf{P}\,\varphi(|u|).$$

Boolean connectives and atomic formulae $\top$ and $\bot$ are defined as usual. For example, the formula $\mathsf{X}\,a \wedge \mathsf{Y}\,b$ defines the language aA^*b. Let $\mathrm{TL}_m[\mathsf{X},\mathsf{F},\mathsf{Y},\mathsf{P}]$ be the fragment of unary temporal logic consisting of the Boolean combinations of formulae with at most $m-1$ nested negations. Let $\mathrm{TL}_{m,n}[\mathsf{X},\mathsf{F},\mathsf{Y},\mathsf{P}]$ consist of all formulae in $\mathrm{TL}_m[\mathsf{X},\mathsf{F},\mathsf{Y},\mathsf{P}]$ with operator depth at most n, *i.e.*, there are at most n nested temporal modalities. For a formula φ in first-order logic or in unary temporal logic, let $L(\varphi) = \{u \in A^+ \mid u \models \varphi\}$ be the language defined by φ.

Algebra. Let S be a finite semigroup. An element $x \in S$ is *idempotent* if $x^2 = x$. The set of all idempotents of S is denoted $E(S)$. For every finite semigroup S there exists an integer $\omega \geq 1$ such that each ω-power is idempotent in S. *Green's relations* are an important concept in the structure theory of finite semigroups: For $x, y \in S$ let $x \leq_{\mathcal{R}} y$ if $x = y$ or $x \in yS$ and symmetrically let $x \leq_{\mathcal{L}} y$ if $x = y$ or $x \in Sy$. For $\mathcal{G} \in \{\mathcal{R}, \mathcal{L}\}$ let $x \mathrel{\mathcal{G}} y$ if both $x \leq_{\mathcal{G}} y$ and $y \leq_{\mathcal{G}} x$; and let $x <_{\mathcal{G}} y$ if $x \leq_{\mathcal{G}} y$ but not $y \leq_{\mathcal{G}} x$. We also view S as an alphabet and write $u \in S^+$ for a word with letters from S. For words $u, v \in S^+$ we say that a relation $u \mathrel{\mathcal{G}} v$ "holds in S", if the relation is satisfied after evaluating u and v in S. We use this frequently for equality and Green's relations. All semigroups in this paper are nonempty.

Classes of finite semigroups are often defined by *identities* of omega-terms. An *omega-term* over a set of variables Σ is defined inductively. Every $x \in \Sigma$ is an omega-term, and if u and v are omega-terms, then so are uv and u^ω. A finite semigroup S *satisfies* the identity $u = v$ if for each homomorphism $h : \Sigma^+ \to S$ we have $h(u) = h(v)$. Here, h is extended to omega-terms by letting $h(u^\omega)$ be the idempotent generated by $h(u)$.

For every $e \in E(S)$ the set eSe forms the so-called *local monoid* at e. A semigroup S belongs to **LDA** if every local monoid eSe satisfies $(xy)^\omega x (xy)^\omega = (xy)^\omega$. This is equivalent to saying that we have $(exeye)^\omega exe (exeye)^\omega = (exeye)^\omega$ in S for all $x, y \in S$ and all $e \in E(S)$. Note that if S is in **LDA** and if $e \in E(S)$ and $x, y \in eSe$ then, $(xy)^\omega = (xy)^{\omega-1} x (yx)^\omega y = (xy)^{\omega-1} x (yx)^\omega y (yx)^\omega y = (xy)^{2\omega} y (xy)^\omega = (xy)^\omega y (xy)^\omega$. Thus despite its asymmetric definition, **LDA** is left-right-symmetric.

A homomorphism $h : A^+ \to S$ to a finite semigroup S *recognizes* a language $L \subseteq A^+$ if $h^{-1}(h(L)) = L$. A semigroup S *recognizes* $L \subseteq A^+$ if there exists a homomorphism $h : A^+ \to S$ which recognizes L. For $u, v \in A^+$ let $u \equiv_L v$ if $puq \in L$ is equivalent to $pvq \in L$ for all $p, q \in A^*$. The relation $\equiv_L$ over A^+ is a congruence and the semigroup $A^+/{\equiv_L}$, also denoted by $\mathrm{Synt}(L)$ and called the *syntactic semigroup* of L, is the unique minimal semigroup recognizing L. Moreover, it is effectively computable (*e.g.* from an automaton for L); *cf.* [19].

3 Alternation within Two-Variable First-Order Logic with Successor

We define classes $\mathbf{W}_m$ of finite semigroups which will yield an algebraic characterization of $\mathrm{FO}^2_m[<, \mathrm{suc}]$. To this end, we inductively define sequences of omega-terms U_m, V_m with variables $e, f, x_i, y_i, s, t, p_i, q_i$. For $m = 1$ we define $U_1 = (e^\omega s f^\omega x_1 e^\omega)^\omega s (f^\omega y_1 e^\omega t f^\omega)^\omega$ and $V_1 = (e^\omega s f^\omega x_1 e^\omega)^\omega t (f^\omega y_1 e^\omega t f^\omega)^\omega$ and for $m \geq 2$

$$U_m = (p_m U_{m-1} q_m x_m)^\omega p_m U_{m-1} q_m (y_m p_m U_{m-1} q_m)^\omega,$$
$$V_m = (p_m U_{m-1} q_m x_m)^\omega p_m V_{m-1} q_m (y_m p_m U_{m-1} q_m)^\omega.$$

By definition, a semigroup is in $\mathbf{W}_m$ if it satisfies the identity $U_m = V_m$. The class $\mathbf{W}_1$ is Knast's algebraic characterization of dot-depth one [9]. The only difference between U_1 and V_1 is the central variable in U_1 being s and in V_1 being t. Intuitively, this difference is hidden more and more in U_m and V_m with increasing m.

The following result is the main contribution of this paper. The remainder of this section is dedicated to its proof.

▶ **Theorem 1.** *Let $m \geq 2$ and let $L \subseteq A^+$. The following assertions are equivalent:*
1. *L is definable in $\mathrm{FO}^2_m[<, \mathrm{suc}]$.*
2. *L is definable in $\mathrm{TL}_m[\mathsf{X}, \mathsf{F}, \mathsf{Y}, \mathsf{P}]$.*
3. *$\mathrm{Synt}(L) \in \mathbf{W}_m$.*

Before turning to the proof of Theorem 1 we record the following decidability corollary. For $m = 1$ it relies on a characterization of two-sided ideals inside dot-depth one [11].

▶ **Corollary 2.** *For every positive integer m one can decide whether a given regular language $L \subseteq A^+$ is definable in $\mathrm{FO}^2_m[<, \mathrm{suc}]$.* ◀

We start with the hard part of the proof of Theorem 1, *i.e.*, with the implication from (3) to (1). This is essentially Proposition 13 whose proof requires some preparatory work: We first show that every $\mathbf{W}_m$ is contained in $\mathbf{LDA}$ (Lemma 3) which allows us to use a combinatorial property of $\mathbf{LDA}$ (given in Lemma 6). Then a relativization technique for FO^2_m (Lemma 7) is used for defining a congruence $\approx_{m,n}$ (Definition 8) as a tool for FO^2_m. The connection between this congruence and FO^2_m is established by Lemma 10. Using a string rewriting system, a special factorization (given in Lemma 12) finally leads to an inductive scheme to prove Proposition 13.

In the proof of Theorem 1 at the very end of this section we sketch how to show the reverse implication as well as how to incorporate unary temporal logic.

▶ **Lemma 3.** *For all $m \geq 1$ we have $\mathbf{W}_m \subseteq \mathbf{LDA}$.*

Proof. Let S be a finite semigroup and let $\omega \geq 1$ be an integer such that x^ω is idempotent for all $x \in S$. Let $x, y \in S$ and let $e \in S$ be idempotent. Setting $e_1 = f_1 = s = e$, $x_1 = xey$, $y_1 = x$, $t = y$ we get $U_1 = (exeye)^\omega$ in S and $V_1 = (exeye)^\omega eye(exeye)^\omega$ in S. Setting all other variables occurring in U_m or in V_m to be e, we see $U_m = (exeye)^\omega$ in S and $V_m = (exeye)^\omega eye(exeye)^\omega$ in S. Thus if $S \in \mathbf{W}_m$ and $e \in E(S)$, then eSe satisfies the identity $(xy)^\omega = (xy)^\omega y(xy)^\omega$, *i.e.*, $S \in \mathbf{LDA}$. ◀

The next lemma is an intermediate result for Lemma 5 and Lemma 6 both of which yield important combinatorial properties of semigroups in $\mathbf{LDA}$.

▶ **Lemma 4.** *Let $S \in \mathbf{LDA}$, let $x, y, z \in S$, and let $e \in E(S)$.*
1. *If $xe \mathrel{\mathcal{R}} ye$ in S, then $xe \mathrel{\mathcal{R}} xez$ if and only if $ye \mathrel{\mathcal{R}} yez$.*
2. *If $ex \mathrel{\mathcal{L}} ey$ in S, then $ex \mathrel{\mathcal{L}} zex$ if and only if $ey \mathrel{\mathcal{L}} zey$.*

Proof. Since $\mathbf{LDA}$ is left-right symmetric, it suffices to show (1). Suppose $xe \mathrel{\mathcal{R}} xez$. Since $ye \mathrel{\mathcal{R}} xe \mathrel{\mathcal{R}} xez$ there exist s, t such that $xe = yes$ and $ye = xezt$. We get $ye = ye(esezte)$. Pumping the factor in the parentheses and using $\mathbf{LDA}$ yields $ye = ye(esezte)^\omega = ye(esezte)^\omega ezte(esezte)^\omega \in yezS$. ◀

▶ **Lemma 5.** *Let $S \in \mathbf{LDA}$, let $u, v \in S^+$, let $s, t \in S^*$ with $\mathrm{alph}_{|S|+1}(vs) = \mathrm{alph}_{|S|+1}(vt)$ and $|v| \geq |S|$.*

1. *If* $u \mathrel{\mathcal{R}} uv$ *in* S*, then* $u \mathrel{\mathcal{R}} uvs$ *in* S *if and only if* $u \mathrel{\mathcal{R}} uvt$ *in* S.
2. *If* $u \mathrel{\mathcal{L}} vu$ *in* S*, then* $u \mathrel{\mathcal{L}} svu$ *in* S *if and only if* $u \mathrel{\mathcal{L}} tvu$ *in* S.

Proof. Since **LDA** is left-right symmetric, it suffices to show (1). Assume $u \mathrel{\mathcal{R}} uv \mathrel{\mathcal{R}} uvs$ in S. We want to show $u \mathrel{\mathcal{R}} uvt$ in S. This is trivial if t is the empty word. Otherwise we factorize $vt = pwz$ such that $|w| < |wz| = |S| + 1$ with $w = we$ in S for some idempotent e of S. Note that every sequence $x_1, \ldots, x_{|S|} \in S$ has a prefix which admits an idempotent stabilizer, *i.e.*, there exists $i \in \{1, \ldots, |S|\}$ and $e \in E(S)$ such that $x_1 \cdots x_i = x_1 \cdots x_i e$ in S; see *e.g.* [11, Lemma 1] for a proof of this claim. Since vs and vt have the same factors of length $|S| + 1$, we find a factorization $vs = s_1 w z s_2$. Let $x = us_1 w$ and $y = upw$. By induction $u \mathrel{\mathcal{R}} y$ and thus $xe = x \mathrel{\mathcal{R}} y = ye$ in S. Moreover, $xe \mathrel{\mathcal{R}} xez$ and by Lemma 4 we see $ye \mathrel{\mathcal{R}} yez$ in S. This implies the claim. ◀

Choosing s to be the empty word and $t = a$ immediately yields the following consequence.

▶ **Lemma 6.** *Let* $S \in \mathbf{LDA}$*, let* $u, v \in S^+$*, let* $a \in S$ *and let* $|v| \geq |S|$.
1. *If* $u \mathrel{\mathcal{R}} uv >_{\mathcal{R}} uva$ *in* S*, then* $\mathrm{alph}_{|S|+1}(v) \neq \mathrm{alph}_{|S|+1}(va)$.
2. *If* $u \mathrel{\mathcal{L}} vu >_{\mathcal{L}} avu$ *in* S*, then* $\mathrm{alph}_{|S|+1}(v) \neq \mathrm{alph}_{|S|+1}(av)$. ◀

The next lemma gives the main combinatorial properties of $\mathrm{FO}^2_m[<, \mathrm{suc}]$ for our purpose, namely relativizations of formulae to certain factors of deterministic factorizations.

▶ **Lemma 7.** *Let* $\varphi \in \mathrm{FO}^2[<, \mathrm{suc}]$ *and let* $v, w \in A^+$.
1. *There exist formulae* $\langle\varphi\rangle_{<\mathsf{X}w}$ *and* $\langle\varphi\rangle_{>\mathsf{X}w}$ *such that for all* $u = u_1 w u_2$ *with a unique occurrence of the factor* w *in the prefix* $u_1 w$*:*
$$u \models \langle\varphi\rangle_{<\mathsf{X}w}(i,j) \;\text{ iff }\; u_1 \models \varphi(i,j) \;\text{ for all } 1 \leq i,j \leq |u_1|,$$
$$u \models \langle\varphi\rangle_{>\mathsf{X}w}(i,j) \;\text{ iff }\; u_2 \models \varphi(i - |u_1 w|, j - |u_1 w|) \;\text{ for all } |u_1 w| < i,j \leq |u|.$$
2. *There exist formulae* $\langle\varphi\rangle_{<\mathsf{Y}v}$ *and* $\langle\varphi\rangle_{>\mathsf{Y}v}$ *such that for all* $u = u_1 v u_2$ *with a unique occurrence of the factor* v *in the suffix* $v u_2$*:*
$$u \models \langle\varphi\rangle_{<\mathsf{Y}v}(i,j) \;\text{ iff }\; u_1 \models \varphi(i,j) \;\text{ for all } 1 \leq i,j \leq |u_1|,$$
$$u \models \langle\varphi\rangle_{>\mathsf{Y}v}(i,j) \;\text{ iff }\; u_2 \models \varphi(i - |u_1 v|, j - |u_1 v|) \;\text{ for all } |u_1 v| < i,j \leq |u|.$$
3. *There exists a formula* $\langle\varphi\rangle_{[v;w]}$ *such that for all* $u = u_1 v u_2 w u_3$ *with a unique occurrence of the factor* v *in* $v u_2 w u_3$ *and a unique occurrence of the factor* w *in* $u_1 v u_2 w$*:*
$$u \models \langle\varphi\rangle_{[v;w]}(i,j) \;\text{ iff }\; u_2 \models \varphi(i - |u_1 v|, j - |u_1 v|) \;\text{ for all } |u_1 v| < i,j \leq |u_1 v u_2|.$$

Moreover, if $\varphi \in \mathrm{FO}^2_{m,n}[<, \mathrm{suc}]$*, then*
1. $\langle\varphi\rangle_{<\mathsf{X}w} \in \mathrm{FO}^2_{m+1,n+|w|}[<, \mathrm{suc}]$ *and* $\langle\varphi\rangle_{>\mathsf{X}w} \in \mathrm{FO}^2_{m,n+|w|}[<, \mathrm{suc}]$,
2. $\langle\varphi\rangle_{<\mathsf{Y}v} \in \mathrm{FO}^2_{m,n+|v|}[<, \mathrm{suc}]$ *and* $\langle\varphi\rangle_{>\mathsf{Y}v} \in \mathrm{FO}^2_{m+1,n+|v|}[<, \mathrm{suc}]$*, and*
3. $\langle\varphi\rangle_{[v;w]} \in \mathrm{FO}^2_{m+1,n+N}[<, \mathrm{suc}]$ *for* $N = \max\{|v|, |w|\}$. ◀

The relativization of the previous lemma leads to the congruence in the following definition. This congruence is our tool for the combinatorics of FO^2_m in the subsequent proofs.

▶ **Definition 8.** Let $u, v \in A^*$. For $m, n \geq 0$ we let $u \approx_{m,0} v$ and $u \approx_{0,n} v$. For $n \geq 1$ let $u \approx_{1,n} v$ if u and v are contained in the same monomials $w_1 A^+ w_2 \cdots A^+ w_\ell$ with $w_i \in A^+$ and $|w_1 \cdots w_\ell| \leq n$. For $m \geq 2$ and $n \geq 1$ let $u \approx_{m,n} v$ if $\mathrm{alph}_k(u) = \mathrm{alph}_k(v)$ and $\mathrm{pref}_k(u) = \mathrm{pref}_k(v)$ and $\mathrm{suff}_k(u) = \mathrm{suff}_k(v)$ for all $k \leq n$, and all of the following hold:

1. if $u = u_1 w u_2$ and $v = v_1 w v_2$ with $1 \leq |w| \leq n$ such that the factor w has a unique occurrence in the prefixes $u_1 w$ and $v_1 w$, then $u_1 \approx_{m-1,n-|w|} v_1$ and $u_2 \approx_{m,n-|w|} v_2$,
2. if $u = u_1 w u_2$ and $v = v_1 w v_2$ with $1 \leq |w| \leq n$ such that the factor w has a unique occurrence in the suffixes $w u_2$ and $w v_2$, then $u_1 \approx_{m,n-|w|} v_1$ and $u_2 \approx_{m-1,n-|w|} v_2$,
3. if $u = u_1 w u_2 w' u_3$ and $v = v_1 w v_2 w' v_3$ with $|ww'| \leq n$ such that the factor w has a unique occurrence in the suffixes $w u_2 w' u_3$ and $w v_2 w' v_3$ and such that the factor w' has a unique occurrence in the prefixes $u_1 w u_2 w'$ and $v_1 w v_2 w'$, then $u_2 \approx_{m-1,n-|ww'|} v_2$. ◀

An elementary verification shows that $\approx_{m,n}$ is a congruence. Since this fact is not used in this paper, we do not record it as lemma. The following is also straightforward.

▶ **Lemma 9.** *If $m, n \geq 1$ and $u, v \in A^*$ with $u \approx_{m,n} v$, then $u \approx_{m-1,n} v$ and $u \approx_{m,n-1} v$.* ◀

The next lemma connects $\mathrm{FO}^2_{m,n}$ with the combinatorial properties captured by $\approx_{m,n}$. For $u, v \in A^*$ let $u \equiv_{1,n} v$ if u and v model the same formulae in $\mathrm{FO}^2_{1,n}[<, \mathrm{suc}, \min, \max]$. For $m \geq 2$ and $u, v \in A^*$ let $u \equiv_{m,n} v$ if u and v model the same formulae in $\mathrm{FO}^2_{m,n}[<, \mathrm{suc}]$. We have to include min and max predicates at level 1 for technical reasons.

▶ **Lemma 10.** *If $m, n \geq 0$ and $u, v \in A^*$ with $u \equiv_{m,n+1} v$, then $u \approx_{m,n} v$.* ◀

In other words the previous lemma shows that $\equiv_{m,n+1}$ is a refinement of $\approx_{m,n}$. In particular, $\approx_{m,n}$ has finite index. The next lemma is an auxiliary statement used in the proof of Lemma 12. It says that $\approx_{1,n}$ equivalence of u and v allows order comparison for certain factors in the words u and v.

▶ **Lemma 11.** *Let $u, v \in A^+$ and consider factorizations $u = x_1 u_1 \cdots x_k u_k = u'_1 y_1 \cdots u'_\ell y_\ell$ and $v = x_1 v_1 \cdots x_k v_k = v'_1 y_1 \cdots v'_\ell y_\ell$ with $k, \ell \geq 1$ and $u'_1, v'_1, u_k, v_k \in A^*$ and $x_i, y_i \in A^+$ such that*

- *$x_1 u_1 \cdots x_k$ is the shortest prefix of u contained in $x_1 A^+ x_2 \cdots A^+ x_k$ and $x_1 v_1 \cdots x_k$ is the shortest prefix of v contained in $x_1 A^+ x_2 \cdots A^+ x_k$,*
- *$y_1 \cdots u'_\ell y_\ell$ is the shortest suffix of u contained in $y_1 A^+ y_2 \cdots A^+ y_\ell$ and $y_1 \cdots v'_\ell y_\ell$ is the shortest suffix of v contained in $y_1 A^+ y_2 \cdots A^+ y_\ell$.*

Let $\Delta_u = |u| - |x_1 u_1 \cdots u_{k-1}| - |u'_2 \cdots u'_\ell y_\ell|$ and let $\Delta_v = |v| - |x_1 v_1 \cdots v_{k-1}| - |v'_2 \cdots v'_\ell y_\ell|$. If $u \approx_{1,n} v$ for $n = |x_1 \cdots x_k| + |y_1 \cdots y_\ell|$, then the relative order of the occurrences of x_k and y_1 is the same in u and v, i.e., one of the following conditions applies:

1. *$\Delta_u > |x_k y_1|$ and $\Delta_v > |x_k y_1|$.*
2. *$\Delta_u < 0$ and $\Delta_v < 0$.*
3. *$\Delta_u = \Delta_v$.* ◀

The main combinatorial ingredient for the implication from $\mathbf{W}_m$ to FO^2_m is the factorization in the following lemma. It combines properties of **LDA** and $\approx_{m,n}$.

▶ **Lemma 12.** *Let $S \in \mathbf{LDA}$, let $m \geq 2$, let $N = 2\,|S|^2$ and let $u, v \in S^+$ such that $u \approx_{m,n+N} v$. Then there exist factorizations $u = w_0 s_1 w_1 \cdots s_\ell w_\ell$ and $v = w_0 t_1 w_1 \cdots t_\ell w_\ell$ with $w_i, s_i, t_i \in S^+$ and $|w_0 \cdots w_\ell| \leq N$ such that for all $1 \leq i \leq \ell$ the following hold:*

1. *$s_i \approx_{m-1,n} t_i$,*
2. *$w_0 s_1 \cdots w_{i-1} \mathrel{\mathcal{R}} w_0 s_1 \cdots w_{i-1} s_i$ in S,*
3. *$w_i \cdots t_\ell w_\ell \mathrel{\mathcal{L}} t_i w_i \cdots t_\ell w_\ell$ in S.*

Proof. Let $X' = \{1\} \cup \{i \in \mathrm{pos}(u) \mid 1 < i \leq |u|\,,\, u[1;i-1] >_{\mathcal{R}} u[1;i] \text{ in } S\}$ be the set of positions of u which cause an $\mathcal{R}$-descent when reading u from left to right. Let X be the set of positions j such that there exists $i \in X'$ with $0 \leq i - j \leq |S|$, *i.e.*, we include all $|S|$ positions to the left of each $i \in X'$. Let Y' and Y be defined left-right symmetrically on v, *i.e.*, $Y' = \{|v|\} \cup \{i \in \text{pos}(v) \mid 1 < i \leq |v|, v[i-1;|v|] >_{\mathcal{L}} u[i;|v|] \text{ in } S\}$ and Y is the set of positions j such that $0 \leq j - i \leq |S|$ for some $i \in Y'$. Let $X = X_1 \cup \cdots \cup X_k$ with $X_i \neq \emptyset$ being maximal subsets of consecutive positions of X such that all positions of X_i are smaller than all positions of X_{i+1}. Symmetrically, let $Y = Y_1 \cup \cdots \cup Y_{k'}$ with $Y_i \neq \emptyset$ being maximal subsets of consecutive positions of Y such that all positions of Y_i are smaller than all positions of Y_{i+1}.

Let $x_i = u[X_i]$ and $y_i = u[Y_i]$ be the factors of u and v covered by the positions of X_i and Y_i, respectively. By construction and Lemma 6 (1), we see that $u[1;\max(X_i)]$ is the shortest prefix of u which is contained in $x_1 S^+ x_2 \cdots S^+ x_i$. Symmetrically, $v[\min(Y_i);|v|]$ is the shortest suffix of v which is contained in $y_i S^+ y_{i+1} \cdots S^+ y_{k'}$ by Lemma 6 (2). We use these properties to transfer the positions of X to v and the positions of Y to u. Specifically we let $Y'' = Y''_1 \cup \cdots \cup Y''_{k'}$ be such that each Y''_i is an interval of positions of u with $u[Y''_i] = y_i$ and $u[\min(Y''_i);|u|]$ is the shortest suffix of u which is contained in $y_i S^+ y_{i+1} \cdots S^+ y_{k'}$. And we let $X = X''_1 \cup \cdots \cup X''_k$ be such that each X''_i is an interval of positions of v with $v[X''_i] = x_i$ and $v[1;\max(X''_i)]$ is the shortest prefix of v which is contained in $x_1 S^+ x_2 \cdots S^+ x_i$. Note that $u \in S^* y_1 S^+ y_2 \cdots S^+ y_{k'}$ and $v \in x_1 S^+ x_2 \cdots S^+ x_k S^*$ because $u \approx_{m,n+N} v$.

Now, consider the factorization $u = w_0 s_1 w_1 \cdots s_\ell w_\ell$ with $s_i \in S^+$ such that the w_i are the factors covered by maximal subsets of consecutive positions in $X \cup Y''$. Intuitively, this means that we merge overlapping and adjacent factors x_i and y_j in u. Lemma 11 shows that the relative order of those concrete occurrences of x_i and y_j is the same in v as in u. Therefore, if we consider the factorization of v which is covered by maximal subsets of consecutive positions in $X'' \cup Y$, then we end up with the same factors in the same order, *i.e.*, we have $v = w_0 t_1 w_1 \cdots t_\ell w_\ell$ for some $t_i \in S^+$. Since the $\mathcal{R}$-class and the $\mathcal{L}$-class can descend at most $|S| - 1$ times, we have $|X' \cup Y'| \leq 2|S|$ and thus $|w_0 \cdots w_\ell| \leq |X \cup Y''| \leq 2|S|^2$. Moreover, by construction every $\mathcal{R}$-descent when reading prefixes of u as well as every $\mathcal{L}$-descent when reading suffixes of v is covered by some factor w_i showing (2) and (3).

It remains to show $s_i \approx_{m-1,n} t_i$ for all i. An intermediate step is the following claim.

Claim. *If $s_k w_k \cdots s_\ell w_\ell \approx_{m,n+N} t_k w_k \cdots t_\ell w_\ell$ for some $N \geq |w_k \cdots w_\ell|$, then $s_i \approx_{m-1,n} t_i$ for all $i \in \{k, \ldots, \ell\}$.*

The proof of this claim is by induction on $\ell - k$. Every w_i either arises from some x_j or some y_j or both. Therefore, the w_i's inherit the properties of the corresponding x_j's and y_j's of being the first occurrence (respectively being the last occurrence). If there is no w_i arising from an x_j, then every w_i has a unique occurrence in $w_i s_{i+1}$ as well as in $w_i t_{i+1}$. Thus $s_i \approx_{m-1,n} t_i$ for all i by an $(\ell - k)$-fold application of condition (2) in the definition of $\approx_{m,n}$ (from right to left). For $i = k$ this uses Lemma 9.

Fix the first w_i which arises from an x_j. We have $s_j \approx_{m-1,n} t_j$ for all $j > i$ by condition (1) in the definition of $\approx_{m,n}$ and induction. If $i = k$, then $s_k \approx_{m-1,n} t_k$ again by condition (1) in the definition of $\approx_{m,n}$. Assume therefore $i > k$ in the sequel. Let $h \geq i$ be minimal such that w_h arises from some y_j; note that w_ℓ arises from $y_{k'}$. By a repeated application of condition (2) in the definition of $\approx_{m,n}$ we get that $s_k w_k \cdots s_h \approx_{m,n+N'} t_k w_k \cdots t_h$ for $N' = |w_k \cdots w_{h-1}|$. Now w_{i-1} has a unique occurrence in each of the words $w_{i-1} s_i \cdots s_h$ and $w_{i-1} t_i \cdots t_h$. Therefore, by repeatedly applying condition (2) in the definition of $\approx_{m,n}$ we see that $s_j \approx_{m-1,n} t_j$ for all $k \leq j < i$. If $h > i$, then by condition (3) in the definition of $\approx_{m,n}$ we see that $s_i \approx_{m-1,n} t_i$; and if $h = i$, then this follows from condition (2) in the definition of $\approx_{m,n}$. This concludes the proof of the claim.

Now by condition (1) in the definition of $\approx_{m,n}$, we see $s_1 w_1 \cdots s_\ell w_\ell \approx_{m,n+N'} t_1 w_1 \cdots t_\ell w_\ell$

for $N' = N - |w_0|$ and the above claim yields $s_j \approx_{m-1,n} t_j$ for all $1 \leq j \leq \ell$. ◀

The following proposition essentially shows how to pass from $\mathbf{W}_m$ to $\mathrm{FO}^2_m[<, \mathrm{suc}]$. The key to its proof is a string rewriting system which enables induction on the parameter m. Intuitively we consider the maximal quotient of a semigroup in $\mathbf{W}_m$ contained in $\mathbf{W}_{m-1}$. Since the latter is given by an omega-identity, this quotient can be described by a string rewriting system. A single rewriting step of this system corresponds to one application of the omega-identity for $\mathbf{W}_{m-1}$ and can be lifted to $\mathbf{W}_m$ relatively easily.

▶ **Proposition 13.** *For every $S \in \mathbf{W}_m$ with $m \geq 1$ there exists $n \geq 1$ such that $u \approx_{m,n} v$ implies $u = v$ in S for all $u, v \in S^+$.*

Proof. We perform an induction on m. By Knast's Theorem [9], if L is recognized by a semigroup $S \in \mathbf{W}_1$, then the language L is a Boolean combination of monomials $w_1 A^+ w_2 \cdots A^+ w_\ell$. Choosing $n \geq 1$ such that for all these monomials we have $|w_1 \cdots w_\ell| \leq n$ yields the claim for $m = 1$.

Let $\omega > |S|$ be an integer such x^ω is idempotent in S for all $x \in S$. Consider the relation $\to$ on S^+ given by $s \to t$ if $s = t$ in S or if $s = p u_{m-1} q$ and $t = p v_{m-1} q$ for some $p, q \in S^*$ and some $x_i, e, y_i, f, p_i, q_i, z, z' \in S^+$ such that $u_1 = (e^\omega z f^\omega x_1 e^\omega)^\omega z (f^\omega y_1 e^\omega z' f^\omega)^\omega$ and $v_1 = (e^\omega z f^\omega x_1 e^\omega)^\omega z' (f^\omega y_1 e^\omega z' f^\omega)^\omega$ and for $i \geq 2$ we have

$$u_i = (p_i u_{i-1} q_i x_i)^\omega p_i u_{i-1} q_i (y_i p_i u_{i-1} q_i)^\omega, \quad v_i = (p_i u_{i-1} q_i x_i)^\omega p_i v_{i-1} q_i (y_i p_i u_{i-1} q_i)^\omega.$$

Let $\overset{*}{\leftrightarrow}$ be the reflexive, symmetric and transitive closure of $\to$. The relation $\overset{*}{\leftrightarrow}$ is a congruence of finite index (since $S^+/\overset{*}{\leftrightarrow}$ is a quotient of S). Moreover $x^\omega \overset{*}{\leftrightarrow} x^{2\omega}$ for all $x \in S^+$ and $S^+/\overset{*}{\leftrightarrow} \in \mathbf{W}_{m-1}$.

Claim 1. *Let $u, s, t \in S^+$. If $s \to t$, then $u \mathrel{\mathcal{R}} us$ in S if and only if $u \mathrel{\mathcal{R}} ut$ in S.*

Assume without restriction that $s \neq t$ in S. We have $\mathrm{alph}_{|S|+1}(s) = \mathrm{alph}_{|S|+1}(t)$ by construction of u_{m-1} and v_{m-1}. Note that by choice of ω, in particular both words have the same prefix and the same suffix of length $|S| + 1$. Lemma 5 yields Claim 1.

Claim 2. *Let $u, v, s, t \in S^+$ with $s \overset{*}{\leftrightarrow} t$. If $u \mathrel{\mathcal{R}} us$ and $v \mathrel{\mathcal{L}} tv$ in S, then $usv = utv$ in S.*

Since $s \overset{*}{\leftrightarrow} t$, there exists $k \geq 0$ and $w_0, \ldots, w_k \in S^+$ such that $s = w_0$ and $w_k = t$ and such that either $w_{i-1} \to w_i$ or $w_i \to w_{i-1}$ for each $1 \leq i \leq k$. Claim 1 and its left-right dual, yield that $u \mathrel{\mathcal{R}} uw_i$ and $v \mathrel{\mathcal{L}} w_i v$ in S for all i. It therefore suffices to show the claim for $s \to t$. The claim is trivial if $s = t$ in S. Otherwise suppose $s = p_m u_{m-1} q_m$ and $t = p_m v_{m-1} q_m$. Since $u \mathrel{\mathcal{R}} us$ in S, there exists $x_m \in S$ such that $u = usx_m$ in S. Since $v \mathrel{\mathcal{L}} tv$ in S, the left-right dual of Claim 1 implies $v \mathrel{\mathcal{L}} sv$ in S. Hence, there exists $y_m \in S$ such that $v = y_m sv$ in S. Now $u = u(p_m u_{m-1} q_m x_m)^\omega$ in S and $v = (y_m p_m u_{m-1} q_m)^\omega v$ in S and with $S \in \mathbf{W}_m$ we see

$$\begin{aligned} usv &= u(p_m u_{m-1} q_m x_m)^\omega p_m u_{m-1} q_m (y_m p_m u_{m-1} q_m)^\omega v \\ &= u(p_m u_{m-1} q_m x_m)^\omega p_m v_{m-1} q_m (y_m p_m u_{m-1} q_m)^\omega v = utv \text{ in } S, \end{aligned}$$

thus establishing Claim 2.

Since $S^+/\overset{*}{\leftrightarrow} \in \mathbf{W}_{m-1}$, by induction there exists $n \geq 1$ such that $s \approx_{m-1,n} t$ implies $s \overset{*}{\leftrightarrow} t$ for all $s, t \in S^+$. Let $u, v \in S^+$ and suppose $u \approx_{m,n+N} v$ for $N = 2|S|^2$. Let $u = w_0 s_1 w_1 \cdots s_\ell w_\ell$ and $v = w_0 t_1 w_1 \cdots t_\ell w_\ell$ be the factorizations given by Lemma 12; in particular $s_i \approx_{m-1,n} t_i$ and $w_0 s_1 \cdots w_{i-1} \mathrel{\mathcal{R}} w_0 s_1 \cdots w_{i-1} s_i$ in S and $w_i \cdots t_\ell w_\ell \mathrel{\mathcal{L}} t_i w_i \cdots t_\ell w_\ell$. By choice of n we have $s_i \overset{*}{\leftrightarrow} t_i$ for all i and repeated application of Claim 2 yields the following chain of identities valid in S:

$$\begin{aligned}
v &= w_0 t_1 w_1 t_2 \cdots t_{\ell-1} w_{\ell-1} t_\ell w_\ell \\
&= w_0 s_1 w_1 t_2 \cdots t_{\ell-1} w_{\ell-1} t_\ell w_\ell \\
&= w_0 s_1 w_1 s_2 \cdots t_{\ell-1} w_{\ell-1} t_\ell w_\ell \\
&\vdots \\
&= w_0 s_1 w_1 s_2 \cdots s_{\ell-1} w_{\ell-1} t_\ell w_\ell \\
&= w_0 s_1 w_1 s_2 \cdots s_{\ell-1} w_{\ell-1} s_\ell w_\ell = u.
\end{aligned}$$

This concludes the proof. ◄

We are now ready to prove Theorem 1.

Proof of Theorem 1. We shall first show "(3) ⇒ (1)". Afterwards we sketch the proof for the implications "(1) ⇒ (3)" and "(1) ⇒ (2)"; note that the implication "(2) ⇒ (1)" is trivial because the semantics of temporal logic formulae is given by two-variable first-order formulae with quantifier alternations originating in negations. We refer to the technical report [14] for full proofs.

"(3) ⇒ (1)": Suppose $S \in \mathbf{W}_m$ and the homomorphism $h : A^+ \to S$ recognizes $L \subseteq A^+$. Combining Proposition 13 and Lemma 10, we see that there exists an integer $n \geq 1$ such that $u \equiv_{m,n} v$ for $u, v \in S^+$ implies $u = v$ in S. Now if $u \equiv_{m,n} v$ for $u, v \in A^+$, then $h(u) = h(v)$. Thus, by specifying the $\equiv_{m,n}$-classes of A^+ which are contained in L, we obtain a formula $\varphi \in \mathrm{FO}^2_{m,n}[<, \mathrm{suc}]$ such that $L(\varphi) = h^{-1}(h(L)) = L$. Note that the syntactic semigroup of L recognizes L.

Sketch of "(1) ⇒ (3)": The overall proof scheme is reminiscent of a recent proof of Straubing [27] which shows that $\mathrm{FO}^2_m[<]$-definable languages are recognized by a monoids in the so-called weakly iterated two-sided semidirect product $((\mathbf{J} ** \mathbf{J}) ** \mathbf{J}) \cdots ** \mathbf{J}$ where $\mathbf{J}$ appears n times. To avoid technical notation our formulation is not in terms of semidirect products, however. More concretely we show that formulae in FO^2_m up to a certain quantifier depth are unable to disprove the defining identity of $\mathbf{W}_m$; this yields a recognizing semigroup of L and thus the claim since the syntactic semigroup is a divisor of any semigroup recognizing L. To this end, an extended alphabet is used to annotate every position by certain information about sequences of factors occurring in the prefix ending and the suffix starting at this position. This allows to reduce the alternation depth of formulae by replacing so-called *innermost quantified blocks* by alphabet information. Induction then yields the claim. The most technical part of this step is to enable induction by showing that certain central factors of the annotated identities for $\mathbf{W}_m$ are obtained from the identities for $\mathbf{W}_{m-1}$ over the extended alphabet.

Sketch of "(1) ⇒ (2)": This can be seen using the construction in [7, proof of Theorem 1] by means of which Etessami, Vardi, and Wilke showed that FO^2 coincides with $\mathrm{TL}[\mathsf{X}, \mathsf{F}, \mathsf{Y}, \mathsf{P}]$; their statement does not involve the alternation depth explicitly, though. Roughly speaking, for every formula in FO^2_m with one free variable an equivalent formula in $\mathrm{TL}_m[\mathsf{X}, \mathsf{F}, \mathsf{Y}, \mathsf{P}]$ is constructed. The idea is to split up quantifier with respect to the order type. For example, the quantifier $\exists x\, \varphi$ is equivalent to the disjunction

$$(\exists x < y - 1 \colon \varphi) \vee (\exists x = y - 1 \colon \varphi) \vee (\exists x = y \colon \varphi) \vee (\exists x = y + 1 \colon \varphi) \vee (\exists x > y + 1 \colon \varphi).$$

In addition, we make explicit the label of the variable x and use syntactic bookkeeping to keep track of the label and the order type. Under the condition that these information be correct, induction yields temporal logic formula for the subformula φ. Now this presupposition can be ensured using the modalities X, F, Y, and P. For example, the subformula in the first parentheses would be YYP φ' where φ' is the formula for φ with respect to the label and the order type $x < y - 1$ which is obtained by induction. ◀

Conclusion

We showed that quantifier alternation for the logic $\mathrm{FO}^2[<, \mathrm{suc}]$ is decidable by giving a single identity of omega-terms for each level $\mathrm{FO}^2_m[<, \mathrm{suc}]$. The key ingredient in our proof is a rewriting technique which allows us to apply induction on m.

There is an algebraic construction $\mathbf{V} \mapsto \mathbf{V} * \mathbf{D}$ in terms of wreath products, see *e.g.* [26]. For most logical fragments $\mathcal{F}$, whenever $\mathcal{F}$ corresponds to a variety of finite monoids $\mathbf{V}$, then the fragment $\mathcal{F}'$ obtained from $\mathcal{F}$ by adding successor predicates corresponds to the semigroup variety $\mathbf{V} * \mathbf{D}$. This is also the case for $\mathrm{FO}^2_m[<]$ and $\mathrm{FO}^2_m[<, \mathrm{suc}]$. Therefore, if $\mathbf{V}_m$ is the variety of finite monoids corresponding to $\mathrm{FO}^2_m[<]$, then our result implies $\mathbf{V}_m * \mathbf{D} = \mathbf{W}_m$.

In general, decidability of $\mathbf{V}$ is not preserved by the operation $\mathbf{V} \mapsto \mathbf{V} * \mathbf{D}$, but a particularly nice situation occurs if $\mathbf{V} * \mathbf{D} = \mathbf{LV}$. Here, a semigroup S is in $\mathbf{LV}$ if all local monoids of S are in $\mathbf{V}$. For example the variety $\mathbf{DA}$ satisfies $\mathbf{DA} * \mathbf{D} = \mathbf{LDA}$, see [1, 4]. For $\mathbf{W}_1$ however, Knast has given an example showing $\mathbf{V}_1 * \mathbf{D} \neq \mathbf{LV}_1$. In view of this example, we conjecture that $\mathbf{V}_m * \mathbf{D} \neq \mathbf{LV}_m$ for all $m \geq 1$.

Acknowledgments. We thank the anonymous referees for their suggestions which helped to improve the presentation of the paper.

References

1. J. Almeida. A syntactical proof of locality of **DA**. *Int. J. Algebra Comput.*, 6(2):165–177, 1996.
2. J. R. Büchi. Weak second-order arithmetic and finite automata. *Z. Math. Logik Grundlagen Math.*, 6:66–92, 1960.
3. R. S. Cohen and J. A. Brzozowski. Dot-depth of star-free events. *J. Comput. Syst. Sci.*, 5(1):1–16, 1971.
4. A. Costa and A. P. Escada. Some operators that preserve the locality of a pseudovariety of semigroups. Technical Report 11–37 DMUC, University of Coimbra, 2011.
5. V. Diekert, P. Gastin, and M. Kufleitner. A survey on small fragments of first-order logic over finite words. *Int. J. Found. Comput. Sci.*, 19(3):513–548, 2008. Special issue DLT 2007.
6. C. C. Elgot. Decision problems of finite automata design and related arithmetics. *Trans. Amer. Math. Soc.*, 98:21–51, 1961.
7. K. Etessami, M. Y. Vardi, and Th. Wilke. First-order logic with two variables and unary temporal logic. *Inf. Comput.*, 179(2):279–295, 2002.
8. J. A. W. Kamp. *Tense Logic and the Theory of Linear Order.* PhD thesis, University of California, 1968.
9. R. Knast. A semigroup characterization of dot-depth one languages. *RAIRO, Inf. Théor.*, 17(4):321–330, 1983.
10. A. Krebs and H. Straubing. An effective characterization of the alternation hierarchy in two-variable logic. In *FSTTCS 2012, Proceedings*, volume 18 of *LIPIcs*, pages 86–98. Dagstuhl Publishing, 2012.

11 M. Kufleitner and A. Lauser. Around dot-depth one. *Int. J. Found. Comput. Sci.*, 23(6):1323–1339, 2012.
12 M. Kufleitner and A. Lauser. The join levels of the Trotter-Weil hierarchy are decidable. In *MFCS 2012, Proceedings*, volume 7464 of *LNCS*, pages 603–614. Springer, 2012.
13 M. Kufleitner and A. Lauser. Lattices of logical fragments over words. In *ICALP 2012, Proceedings Part II*, volume 7392 of *LNCS*, pages 275–286. Springer, 2012.
14 M. Kufleitner and A. Lauser. Quantifier alternation in two-variable first-order logic with successor is decidable. *CoRR*, arXiv:1212.6500 [cs.LO], 2012.
15 M. Kufleitner and P. Weil. On the lattice of sub-pseudovarieties of **DA**. *Semigroup Forum*, 81:243–254, 2010.
16 M. Kufleitner and P. Weil. The FO^2 alternation hierarchy is decidable. In *CSL 2012, Proceedings*, volume 16 of *LIPIcs*, pages 426–439. Dagstuhl Publishing, 2012.
17 M. Kufleitner and P. Weil. On logical hierarchies within FO^2-definable languages. *Log. Methods Comput. Sci.*, 8:1–30, 2012.
18 R. McNaughton and S. Papert. *Counter-Free Automata.* The MIT Press, 1971.
19 J.-É. Pin. *Varieties of Formal Languages.* North Oxford Academic, 1986.
20 J.-É. Pin and P. Weil. Polynomial closure and unambiguous product. *Theory Comput. Syst.*, 30(4):383–422, 1997.
21 M. P. Schützenberger. On finite monoids having only trivial subgroups. *Inf. Control*, 8:190–194, 1965.
22 M. P. Schützenberger. Sur le produit de concaténation non ambigu. *Semigroup Forum*, 13:47–75, 1976.
23 Th. Schwentick, D. Thérien, and H. Vollmer. Partially-ordered two-way automata: A new characterization of DA. In *DLT 2001, Proceedings*, volume 2295 of *LNCS*, pages 239–250. Springer, 2002.
24 I. Simon. Piecewise testable events. In *Autom. Theor. Form. Lang., 2nd GI Conf.*, volume 33 of *LNCS*, pages 214–222. Springer, 1975.
25 H. Straubing. A generalization of the Schützenberger product of finite monoids. *Theor. Comput. Sci.*, 13:137–150, 1981.
26 H. Straubing. Finite semigroup varieties of the form $\mathbf{V} * \mathbf{D}$. *J. Pure Appl. Algebra*, 36(1):53–94, 1985.
27 H. Straubing. Algebraic characterization of the alternation hierarchy in $FO^2[<]$ on finite words. In *CSL 2011, Proceedings*, volume 12 of *LIPIcs*, pages 525–537. Dagstuhl Publishing, 2011.
28 P. Tesson and D. Thérien. Diamonds are forever: The variety DA. In *Semigroups, Algorithms, Automata and Languages 2001, Proceedings*, pages 475–500. World Scientific, 2002.
29 D. Thérien. Classification of finite monoids: The language approach. *Theor. Comput. Sci.*, 14(2):195–208, 1981.
30 D. Thérien and Th. Wilke. Over words, two variables are as powerful as one quantifier alternation. In *STOC 1998, Proceedings*, pages 234–240. ACM Press, 1998.
31 W. Thomas. Classifying regular events in symbolic logic. *J. Comput. Syst. Sci.*, 25:360–376, 1982.
32 W. Thomas. Languages, automata and logic. In *Handbook of Formal Languages*, volume 3, pages 389–455. Springer, 1997.
33 B. A. Trakhtenbrot. Finite automata and logic of monadic predicates (in Russian). *Dokl. Akad. Nauk SSSR*, 140:326–329, 1961.
34 P. Trotter and P. Weil. The lattice of pseudovarieties of idempotent semigroups and a non-regular analogue. *Algebra Univers.*, 37(4):491–526, 1997.
35 Ph. Weis and N. Immerman. Structure theorem and strict alternation hierarchy for FO^2 on words. *Log. Methods Comput. Sci.*, 5:1–23, 2009.