

First-order Fragments with Successor over Infinite Words*

Jakub Kallas¹, Manfred Kufleitner², and Alexander Lauser²

1 ENS Cachan, France

jakub.kallas@gmail.com

2 University of Stuttgart, FMI, Germany

kufleitner@fmi.uni-stuttgart.de

lauser@fmi.uni-stuttgart.de

Abstract

We consider fragments of first-order logic and as models we allow finite and infinite words simultaneously. The only binary relations apart from equality are order comparison $<$ and the successor predicate $+1$. We give characterizations of the fragments $\Sigma_2 = \Sigma_2[<, +1]$ and $\text{FO}^2 = \text{FO}^2[<, +1]$ in terms of algebraic and topological properties. To this end we introduce the factor topology over infinite words. It turns out that a language L is in $\text{FO}^2 \cap \Sigma_2$ if and only if L is the interior of an FO^2 language. Symmetrically, a language is in $\text{FO}^2 \cap \Pi_2$ if and only if it is the topological closure of an FO^2 language. The fragment $\Delta_2 = \Sigma_2 \cap \Pi_2$ contains exactly the clopen languages in FO^2 . In particular, over infinite words Δ_2 is a strict subclass of FO^2 . Our characterizations yield decidability of the membership problem for all these fragments over finite and infinite words; and as a corollary we also obtain decidability for infinite words. Moreover, we give a new decidable algebraic characterization of dot-depth $3/2$ over finite words.

Decidability of dot-depth $3/2$ over finite words was first shown by Glaßer and Schmitz in STACS 2000, and decidability of the membership problem for FO^2 over infinite words was shown 1998 by Wilke in his habilitation thesis whereas decidability of Σ_2 over infinite words is new.

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

Keywords and phrases infinite words, regular languages, first-order logic, automata theory, semi-groups, topology

Digital Object Identifier 10.4230/LIPIcs.STACS.2011.356

1 Introduction

The dot-depth hierarchy of star-free languages \mathcal{B}_n for $n \in \mathbb{N} + \{1/2, 1\}$ over finite words has been introduced by Brzozowski and Cohen [5]. Later, the Straubing-Thérien \mathcal{L}_n hierarchy has been considered [20, 23] and a tight connection in terms of so-called wreath products was discovered [19, 21]. It is known that both hierarchies are strict [4] and that they have very natural closure properties [5, 18]. Effectively determining the level n of a language in the dot-depth hierarchy or the Straubing-Thérien hierarchy is one of the most challenging open problems in automata theory. So far, the only decidable classes are \mathcal{B}_n and \mathcal{L}_n for $n \in \{1/2, 1, 3/2\}$, see e.g. [17] for an overview and [10] for level $\mathcal{B}_{3/2}$.

Thomas showed that there is a one-to-one correspondence between the quantifier alternation hierarchy of first-order logic and the dot-depth hierarchy [25]. This correspondence

* The last two authors acknowledge the support by the German Research Foundation (DFG), grant DI 435/5-1.



© J. Kallas, M. Kufleitner, and A. Lauser;

licensed under Creative Commons License NC-ND

28th Symposium on Theoretical Aspects of Computer Science (STACS'11).

Editors: Thomas Schwentick, Christoph Dürr; pp. 356–367

Leibniz International Proceedings in Informatics



LIPIcs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



holds if one allows $[\langle, +1, \min, \max]$ as a signature (we always assume that we have equality and predicates for labels of positions; in order to simplify notation, these symbols are omitted here). The same correspondence between the Straubing-Thérien hierarchy and the quantifier alternation hierarchy holds, if we restrict the signature to $[\langle]$, cf. [18]. In particular, all decidability results for the dot-depth hierarchy and the Straubing-Thérien hierarchy yield decidability of the membership problem for the respective levels of the quantifier alternation hierarchy.

The intersection $\Delta_2[\langle] = \Sigma_2[\langle] \cap \Pi_2[\langle]$ of the language classes $\Sigma_2[\langle]$ and $\Pi_2[\langle]$ of the quantifier alternation hierarchy over finite words has a huge number of different characterizations, see [22] for an overview. One of them turns out to be the first-order fragment $\text{FO}^2[\langle]$ where one can use (and reuse) only two variables [24]. The fragment $\text{FO}^2[\langle]$ is a natural restriction since three variables are already sufficient to express any first-order language over finite and infinite words [12]. Using the wreath product principle [21], one can extend $\Delta_2[\langle] = \text{FO}^2[\langle]$ to $\Delta_2[\langle, +1] = \text{FO}^2[\langle, +1]$, see e.g. [14]. Decidability of $\text{FO}^2[\langle]$ follows from the decidability of $\Sigma_2[\langle]$, but there is also a more direct effective characterization: A language over finite words is definable in $\text{FO}^2[\langle]$ if and only if its syntactic monoid is in the variety \mathbf{DA} , and the latter property is decidable. The wreath product principle yields $\mathbf{DA} * \mathbf{D}$ as an algebraic characterization of $\text{FO}^2[\langle, +1]$, but this does not immediately help with decidability. Almeida [1] has shown that $\mathbf{DA} * \mathbf{D} = \mathbf{LDA}$. Now, since \mathbf{LDA} is decidable, membership in $\text{FO}^2[\langle, +1]$ is decidable. Note that \min and \max do not yield additional expressive power for $\Delta_2[\langle]$ and $\text{FO}^2[\langle]$.

Some of the characterizations and decidability results for the quantifier alternation hierarchy and for $\text{FO}^2[\langle]$ have been extended to infinite words. Decidability of $\Sigma_1[\langle]$ and its Boolean closure $\mathbb{B}\Sigma_1[\langle]$ over infinite words is due to Perrin and Pin [15]; decidability of $\Sigma_2[\langle]$ over infinite words was shown by Bojańczyk [3]. The fragments $\Delta_2[\langle]$ and $\text{FO}^2[\langle]$ do not coincide for infinite words. In particular, decidability of $\text{FO}^2[\langle]$ does not follow from the respective result for $\Delta_2[\langle]$. Decidability of $\text{FO}^2[\langle]$ over infinite words was first shown by Wilke [27].

Over infinite words, using a conjunction of algebraic and topological properties yields further effective characterizations of the fragments $\Sigma_2[\langle]$ and $\text{FO}^2[\langle]$, cf. [7]. The key ingredient is the alphabetic topology which is a refinement of the usual Cantor topology. In addition, languages in $\text{FO}^2[\langle] \cap \Sigma_2[\langle]$ can be characterized using topological notions; namely, a language L over infinite words is in $\text{FO}^2[\langle] \cap \Sigma_2[\langle]$ if and only if L is the interior of a language in $\text{FO}^2[\langle]$ with respect to the alphabetic topology. By complementation, a language is in $\text{FO}^2[\langle] \cap \Pi_2[\langle]$ if and only if it is the topological closure of a language in $\text{FO}^2[\langle]$. This shows that topology reveals natural properties of first-order fragments over infinite words. In this paper, we continue this line of work.

Outline We combine algebraic and topological properties in order to give effective characterizations of $\Sigma_2[\langle, +1]$ (Theorem 3.1) and $\text{FO}^2[\langle, +1]$ (Theorem 4.1) over finite and infinite words. The key ingredient is a generalization of the alphabetic topology which we call the *factor topology*. As a byproduct, we give a new effective characterization of $\Sigma_2[\langle, +1]$ over finite words (Theorem 3.2), i.e., of the level 3/2 of the dot-depth hierarchy. Dually, we get a characterization of $\Pi_2[\langle, +1]$ over infinite words (Theorem 3.4). Moreover, we also obtain decidability results for the respective fragments over infinite words (in contrast to finite and infinite words simultaneously; Corollary 3.3 and Corollary 4.2). Concerning the intersection of fragments, we show that L is in $\text{FO}^2[\langle, +1] \cap \Sigma_2[\langle, +1]$ if and only if L is the interior of a language in $\text{FO}^2[\langle, +1]$ with respect to the factor topology (Theorem 6.1) and dually, L is

definable in $\text{FO}^2[<, +1] \cap \Pi_2[<, +1]$ if and only if L is the topological closure of a language in $\text{FO}^2[<, +1]$ with respect to the factor topology (Theorem 6.2). Finally, we show that $\Delta_2[<, +1]$ is a strict subclass of $\text{FO}^2[<, +1]$ and that a language L is in $\Delta_2[<, +1]$ if and only if L is in $\text{FO}^2[<, +1]$ and clopen in the factor topology (Theorem 5.1).

Due to lack of space, some proofs are omitted. For complete proofs, we refer to the full version of this paper [11].

2 Preliminaries

Words Throughout, Γ is a finite alphabet and unless stated otherwise u, v, w are finite words, and α, β, γ are finite or infinite words over the alphabet Γ . The set of all finite words is Γ^* and the set of all infinite words is Γ^ω . The empty word is denoted by 1 . We write Γ^∞ for the set of all finite and infinite words $\Gamma^* \cup \Gamma^\omega$. As usual, Γ^+ is the set of all non-empty finite words $\Gamma^* \setminus \{1\}$. If L is a subset of a monoid, then L^* is the submonoid generated by L . For $L \subseteq \Gamma^*$ we let $L^\omega = \{u_1 u_2 \cdots \mid u_i \in L \text{ for all } i \geq 1\}$ be the set of infinite products. We also let $L^\infty = L^* \cup L^\omega$. The infinite product of the empty word is empty, i.e., we have $1^\omega = 1$. Thus, $L^\infty = L^\omega$ if and only if $1 \in L$. The *length* of a word $w \in \Gamma^*$ is denoted by $|w|$. We write Γ^k for all words of length k and $\Gamma^{\geq k}$ is the set of finite words of length at least k ; similarly, $\Gamma^{< k}$ consist of all words of length less than k . By $\text{alph}_k(\alpha)$ we denote the factors of length k of α , i.e.,

$$\text{alph}_k(\alpha) = \{w \in \Gamma^k \mid \alpha = vw\beta \text{ for some } v \in \Gamma^*, \beta \in \Gamma^\infty\}.$$

As a special case, we have that $\text{alph}_1(\alpha) = \text{alph}(\alpha)$ is the *alphabet* (also called *content*) of α . We write $\text{im}_k(\alpha)$ for those factors in $\text{alph}_k(\alpha)$ which have infinitely many occurrences in α . The notation $\text{im}_k(\alpha)$ comes from “*imaginary*”.

Languages We introduce a non-standard composition \circ for sufficiently long words. Let $k \geq 1$. For $u \in \Gamma^*$ and $\alpha \in \Gamma^\infty$ define $w \circ_k \alpha$ by

$$w \circ_k \alpha = vx\beta \quad \text{if there exists } x \in \Gamma^{k-1} \text{ such that } w = vx \text{ and } \alpha = x\beta.$$

Furthermore $w \circ_k 1 = w$ and $1 \circ_k \alpha = \alpha$. In all other cases $w \circ_k \alpha$ is undefined. Note that $\text{alph}_k(u \circ_k \alpha) = \text{alph}_k(u) \cup \text{alph}_k(\alpha)$, if $u \circ_k \alpha$ is defined. In particular, the operation \circ_k does not introduce new factors of length k . For $A \subseteq \Gamma^k$ we define

$$\begin{aligned} A^{*k} &= \{w_1 \circ_k \cdots \circ_k w_n \mid n \geq 0, w_i \in A\}, \\ A^{\omega k} &= \{w_1 \circ_k w_2 \circ_k \cdots \mid w_i \in A\}, \\ A^{\infty k} &= A^{*k} \cup A^{\omega k}, \\ A^{\text{im}_k} &= \{\alpha \in \Gamma^\infty \mid \text{im}_k(\alpha) = A\}. \end{aligned}$$

If k is clear from the context, then we write $w \circ \alpha$ instead of $w \circ_k \alpha$, we write A^\circledast instead of A^{*k} , we write A^\circledcirc instead of $A^{\infty k}$, and we write $A^{\circledcirc\circledast}$ instead of A^{im_k} . Note that $\Gamma^* = \emptyset^{\circledcirc\circledast}$.

A *k-factor monomial* is a language of the form

$$P = A_1^{\circledast} \circ u_1 \circ \cdots \circ A_s^{\circledast} \circ u_s \circ A_{s+1}^{\circledcirc\circledast}$$

for $u_i \in \Gamma^{\geq k}$ and $A_i \subseteq \Gamma^k$. The *degree* of P is the length of the word $u_1 \cdots u_s$. A *k-factor polynomial* is a finite union of *k-factor monomials* and of words of length less than k . A language L is a *factor polynomial* (resp. *monomial*) if there is a number k such that L is a *k-factor polynomial* (resp. *monomial*).

Fragments of First-order Logic We think of words as labeled linear orders, and we write $x < y$, if position x comes before position y . Similarly, $x = y + 1$ means that x is the successor of y . A position x of a word α is an a -*position*, if the label of x in α is the letter a .

We denote by FO the first-order logic over words. Atomic formulas in FO are \top (for *true*), unary predicates $\lambda(x) = a$ for $a \in \Gamma$, and binary predicates $x < y$ and $x = y + 1$ for variables x and y . Variables range over positions in \mathbb{N} and $\lambda(x) = a$ means that x is an a -position. Formulas may be composed using Boolean connectives as well as existential quantification $\exists x: \varphi$ and universal quantification $\forall x: \varphi$ for $\varphi \in \text{FO}$. The semantics is as usual. A *sentence* in FO is a formula without free variables. Let $\varphi \in \text{FO}$ be a sentence. We write $\alpha \models \varphi$ if α models φ . The *language defined by* φ is $L(\varphi) = \{\alpha \in \Gamma^\infty \mid \alpha \models \varphi\}$.

The fragment $\Sigma_n[\mathcal{C}]$ of FO for $\mathcal{C} \subseteq \{<, +1\}$ consists of all sentences in prenex normal form with n blocks of quantifiers starting with a block of existential quantifiers. In addition, only binary predicates in \mathcal{C} are allowed. The fragment $\Pi_n[\mathcal{C}]$ consists of negations of formulas in $\Sigma_n[\mathcal{C}]$. We frequently identify first-order fragments with the classes of languages they define. For example, $\Delta_n[\mathcal{C}] = \Sigma_n[\mathcal{C}] \cap \Pi_n[\mathcal{C}]$ is the class of all languages which are definable in both $\Sigma_n[\mathcal{C}]$ and $\Pi_n[\mathcal{C}]$. Another important fragment is $\text{FO}^2[\mathcal{C}]$. It consists of all sentences using (and reusing) only two different names for the variables, say x and y , and where only binary predicates from \mathcal{C} are allowed. Let \mathcal{F} be a fragment of first-order logic. We say that L is \mathcal{F} -*definable over some subset* $K \subseteq \Gamma^\infty$, if there exists some formula $\varphi \in \mathcal{F}$ with $L = \{\alpha \in K \mid \alpha \models \varphi\}$. We frequently use this notion for either $K = \Gamma^*$ or $K = \Gamma^\omega$.

Finite Monoids We repeat some basic notions and properties concerning finite monoids. For further details we refer to standard textbooks such as [16]. Let M be a finite monoid. For every such monoid there exists a number $n \geq 1$ such that $a^n = a^{2n}$ for all $a \in M$, i.e., a^n is the unique idempotent power of a . The set of all idempotents of M is denoted by $E(M)$. An important tool in the study of finite monoids are *Green's relations*. At this point, we only introduce their ordered versions. We have $a \leq_{\mathcal{R}} b$ if and only if $aM \subseteq bM$, we have $a \leq_{\mathcal{L}} b$ if and only if $Ma \subseteq Mb$, and we have $a \leq_{\mathcal{J}} b$ if and only if $MaM \subseteq MbM$.

An *ordered monoid* M is equipped with a partial order \leq which is compatible with multiplication, i.e., $a \leq b$ and $c \leq d$ implies $ac \leq bd$. We can always assume that M is ordered, since equality is a compatible partial order.

The theory of first-order fragments over finite non-empty words is presented more concisely in the context of semigroups instead of monoids. In this paper however, we want to incorporate finite and infinite words in a uniform model, and our approach is heavily based on allowing words to be empty. In order to state “semigroup conditions” for monoids, we have to use surjective homomorphisms $h : \Gamma^* \rightarrow M$ instead of monoids M only.

Let $h : \Gamma^* \rightarrow M$ be a surjective homomorphism and let $e \in M$ be an idempotent. The set P_e consists of all products of the form $x_0 f_1 \cdots x_{m-1} f_m x_m$ with elements $x_0, \dots, x_m \in M$ and idempotents $f_1, \dots, f_m \in h(\Gamma^+) \subseteq M$ satisfying the following three conditions

$$\begin{aligned} e &\leq_{\mathcal{R}} x_0 f_1, \\ e &\leq_{\mathcal{J}} f_i x_i f_{i+1} \quad \text{for all } 1 \leq i \leq m-1, \\ e &\leq_{\mathcal{L}} f_m x_m. \end{aligned}$$

If $e \notin h(\Gamma^+)$, then we set $P_e = \{1\}$. Note that in this case we necessarily have $e = 1$ in M . The notation P_e is for *paths in* e . An idempotent e is said to be *locally path-top* with respect to h if $eP_e e \leq e$. Symmetrically, it is *locally path-bottom* with respect to h if $eP_e e \geq e$. If the underlying homomorphism is clear from the context, we omit the reference to it. The

homomorphism h is *locally path-top* (resp. *locally path-bottom*) if all idempotents in M are locally path-top (resp. locally path-bottom).

► **Lemma 2.1.** *Let $h : \Gamma^* \rightarrow M$ be a surjective homomorphism onto a finite monoid M . It is decidable whether M is locally path-top.*

Proof. We give an algorithm computing P_e for a given idempotent e . We define a composition on triples $T = E(M) \times M \times E(M)$ by $(f_1, x_1, f_2)(f_3, x_2, f_4) = (f_1, x_1 f_2 x_2, f_4)$ if $f_2 = f_3$; otherwise the composition is undefined. Compute the fixed point P of the equation $P = P \cup P T_e$ with $T_e = \{(f_1, x_1, f_2) \in T \mid f_1, f_2 \in h(\Gamma^+), e \leq_{\mathcal{J}} f_1 x_1 f_2\}$ and initial value $P = T_e$. This requires at most $|M|^3$ iterations. Then P_e is the set of all $x_0 f_1 x f_2 x_2$ where $(f_1, x, f_2) \in P$, $e \leq_{\mathcal{R}} x_0 f_1$ and $e \leq_{\mathcal{L}} f_2 x_2$. ◀

Let $h : \Gamma^* \rightarrow M$ be a surjective homomorphism and let $n \in \mathbb{N}$ such that a^n is idempotent for all $a \in M$. The homomorphism $h : \Gamma^* \rightarrow M$ is in **LDA** if

$$(eaebe)^n eae (eaebe)^n = (eaebe)^n$$

for all idempotents $e \in h(\Gamma^+)$ and for all $a, b \in M$. If the reference to the homomorphism is clear from the context, then we say “ $M \in \mathcal{P}$ ” for some property \mathcal{P} meaning that “ $h \in \mathcal{P}$ ”.

Recognizability A language $L \subseteq \Gamma^\infty$ is *regular* if it is recognized by some extended Büchi automaton, see e.g. [6], or equivalently, if it is definable in monadic second order logic [26]. Below, we present a more algebraic framework for recognition of $L \subseteq \Gamma^\infty$. The *syntactic preorder* \leq_L over Γ^* is defined as follows. We let $s \leq_L t$ if for all $u, v, w \in \Gamma^*$ we have the following two implications:

$$utvw^\omega \in L \Rightarrow usvw^\omega \in L \quad \text{and} \quad u(tv)^\omega \in L \Rightarrow u(sv)^\omega \in L.$$

Remember that $1^\omega = 1$. Two words $s, t \in \Gamma^*$ are syntactically equivalent, written as $s \equiv_L t$, if both $s \leq_L t$ and $t \leq_L s$. This is a congruence and the congruence classes $[s]_L = \{t \in \Gamma^* \mid s \equiv_L t\}$ form the *syntactic monoid* $\text{Synt}(L)$ of L . The preorder \leq_L on words induces a partial order \leq_L on congruence classes, and $(\text{Synt}(L), \leq_L)$ becomes an ordered monoid. It is a well-known classical result that the syntactic monoid of a regular language $L \subseteq \Gamma^\infty$ is finite, see e.g. [15, 26]. Moreover, in this case L can be written as a finite union of languages of the form $[s]_L [t]_L^\omega$ with $s, t \in \Gamma^*$ and $st \equiv_L s$ and $t^2 \equiv_L t$.

Now, let $h : \Gamma^* \rightarrow M$ be any surjective homomorphism onto a finite ordered monoid M and let $L \subseteq \Gamma^\infty$. If the reference to h is clear from the context, then we denote by $[s]$ the set of finite words $h^{-1}(s)$ for $s \in M$. The following notations are used:

- $(s, e) \in M \times M$ is a *linked pair*, if $se = s$ and $e^2 = e$.
- h *weakly recognizes* L , if

$$L = \bigcup \{[s][e]^\omega \mid (s, e) \text{ is a linked pair and } [s][e]^\omega \subseteq L\}.$$

- h *strongly recognizes* L (or simply *recognizes* L), if

$$L = \bigcup \{[s][e]^\omega \mid (s, e) \text{ is a linked pair and } [s][e]^\omega \cap L \neq \emptyset\}.$$

- L is *downward closed (on finite prefixes)* for h , if $[s][e]^\omega \subseteq L$ implies $[t][e]^\omega \subseteq L$ for all $s, t, e \in M$ where $t \leq s$.

Using Ramsey’s Theorem, one can show that for every word $\alpha \in \Gamma^\infty$ there exists a linked pair (s, e) such that $\alpha \in [s][e]^\omega$. On the other hand, two different languages of the form $[s][e]^\omega$ are not necessarily disjoint. Therefore, if L is weakly recognized by h , then there could exist some linked pair (s, e) such that $[s][e]^\omega$ and L are incomparable. If L is strongly recognized by h , then for every linked pair we have either $[s][e]^\omega \subseteq L$ or $[s][e]^\omega \cap L = \emptyset$. In particular, whenever L is strongly recognized by h , then $\Gamma^\infty \setminus L$ is also strongly recognized by h . Every regular language L is strongly recognized by its syntactic homomorphism $h_L : \Gamma^* \rightarrow \text{Synt}(L)$; $s \mapsto [s]_L$. Moreover, L is downward closed for h_L .

2.1 The factor topology

Topological properties play a crucial role in this paper. Very often a combination of algebraic and topological properties yields a decidable characterization of the fragments. Moreover, topology can be used to describe the relation between the fragments. This section introduces the topology matching the fragments $\Sigma_2[<, +1]$ and $\Pi_2[<, +1]$.

We define the k -factor topology by its basis. All sets of the form $u \circ A^\otimes$ for $u \in \Gamma^*$ and $A \subseteq \Gamma^k$ are open. Therefore, singleton sets $\{u\}$ for $u \in \Gamma^*$ are open in the k -factor topology since $\{u\} = u \circ \emptyset^\otimes$. A language is said to be *factor open* (resp. *factor closed*) if there is a natural number k such that L is open (resp. closed) in the k -factor topology.

► **Proposition 2.2.** *Let $L \subseteq \Gamma^\infty$ be a regular language. Then L is factor open if and only if L is open in the $(2|\text{Synt}(L)|)$ -factor topology.*

► **Proposition 2.3.** *It is decidable whether a regular language $L \subseteq \Gamma^\infty$ is factor open.*

3 The first-order fragment Σ_2

One of our main results is a decidable characterization of the fragment $\Sigma_2[<, +1]$ over finite and infinite words. It is a combination of a decidable algebraic and a decidable topological property. For finite words only, this yields a new decidable algebraic characterization for dot-depth $3/2$, which in turn coincides with $\Sigma_2[<, +1]$ over finite words [25].

► **Theorem 3.1.** *Let $L \subseteq \Gamma^\infty$ be a regular language. The following are equivalent:*

1. L is $\Sigma_2[<, +1]$ -definable.
2. L is a factor polynomial.
3. L is factor open and there exists a surjective locally path-top homomorphism $h : \Gamma^* \rightarrow M$ which weakly recognizes L such that L is downward closed for h .
4. L is factor open and $\text{Synt}(L)$ is locally path-top.

Next, we give a counterpart of the preceding theorem for finite words, which in turn yields a new decidable characterization of dot-depth $3/2$. The first decidable characterization was discovered by Glaßer and Schmitz [9, 10]. It is based on so-called forbidden patterns. Later, a decidable algebraic characterization was given by Pin and Weil [19].

► **Theorem 3.2.** *Let $L \subseteq \Gamma^*$ be a language. The following are equivalent over finite words:*

1. L is $\Sigma_2[<, +1]$ -definable over finite words.
2. L is a factor polynomial.
3. $\text{Synt}(L)$ is finite and locally path-top.

Proof. The language Γ^* of finite words is definable in $\Sigma_2[<]$ by stating that there is a position such that all other positions are smaller. Hence, if $L = \{w \in \Gamma^* \mid w \models \varphi\}$ for some

$\varphi \in \Sigma_2[<, +1]$, then there also exists some $\varphi' \in \Sigma_2[<, +1]$ such that $L = \{\alpha \in \Gamma^\infty \mid \alpha \models \varphi'\}$. Using Theorem 3.1, this shows “1 \Rightarrow 2”. Trivially, “2 \Rightarrow 3” follows from the same theorem. Finally, “3 \Rightarrow 1” uses the fact that every language over finite words is factor open. \blacktriangleleft

The equivalence of (1) and (2) in Theorem 3.2 was also shown by Glaßer and Schmitz using different techniques and with another formalism for defining factor polynomials [10]. As a corollary of Theorem 3.1 and Theorem 3.2 we obtain the following decidability results.

► **Corollary 3.3.** *Let L be a regular language.*

1. *For $L \subseteq \Gamma^\infty$ it is decidable, whether L is $\Sigma_2[<, +1]$ -definable.*
2. *For $L \subseteq \Gamma^*$ it is decidable, whether L is $\Sigma_2[<, +1]$ -definable over finite words.*
3. *For $L \subseteq \Gamma^\omega$ it is decidable, whether L is $\Sigma_2[<, +1]$ -definable over infinite words.*

Proof. For “1” we note that the syntactic monoid is effectively computable. Therefore, Theorem 3.1 (4) can be verified effectively by Lemma 2.1 and Proposition 2.3. Similarly, “2” follows from the decidability of Theorem 3.2 (3). The set Γ^* is definable in $\Sigma_2[<, +1]$ over Γ^∞ . Hence, $L \subseteq \Gamma^\omega$ is $\Sigma_2[<, +1]$ -definable over Γ^ω if and only if $L \cup \Gamma^*$ is $\Sigma_2[<, +1]$ -definable over Γ^∞ , and the latter condition is decidable by “1”. Therefore, assertion “3” holds. \blacktriangleleft

By duality, the properties of $\Sigma_2[<, +1]$ in Theorem 3.1 yield a decidable characterization of $\Pi_2[<, +1]$, which we state here for completeness.

► **Theorem 3.4.** *Let $L \subseteq \Gamma^\infty$ be a regular language. The following are equivalent:*

1. *L is $\Pi_2[<, +1]$ -definable.*
2. *L is factor closed and $\text{Synt}(L)$ is locally path-bottom.*

4 First-order logic with two variables

In this section, we consider two-variable first-order logic with order $<$ and successor $+1$ over finite and infinite words. The fragment $\text{FO}^2[<, +1]$ admits a temporal logic counterpart having the same expressive power [8]. It is based on unary modalities only. Wilke [27] has shown that membership is decidable for $\text{FO}^2[<, +1]$. We complement these results by giving a simple algebraic characterization of this fragment. An important concept in our proof is a refinement of the factor topology. A set of the form A^{fin} is definable in $\text{FO}^2[<, +1]$ but it is neither open nor closed in the factor topology. This observation leads to the *strict k -factor topology*. A basis of this topology is given by all sets of the form $u \circ A^{\text{fin}} \cap A^{\text{fin}}$ for $u \in \Gamma^*$ and $A \subseteq \Gamma^k$. We do not use this topology outside this section.

► **Theorem 4.1.** *Let $L \subseteq \Gamma^\infty$ be a regular language. The following are equivalent:*

1. *L is $\text{FO}^2[<, +1]$ -definable.*
2. *L is weakly recognized by some homomorphism $h : \Gamma^* \rightarrow M \in \mathbf{LDA}$ and closed in the strict $(2|M|)$ -factor topology.*
3. *$\text{Synt}(L) \in \mathbf{LDA}$.*

The syntactic monoid of a regular language is effectively computable. Hence, one can verify whether property (3) in Theorem 4.1 holds. Since both Γ^* and Γ^ω are $\text{FO}^2[<, +1]$ -definable over Γ^∞ , this gives us the following corollary.

► **Corollary 4.2.** *Let L be a regular language.*

1. *For $L \subseteq \Gamma^\infty$ it is decidable, whether L is $\text{FO}^2[<, +1]$ -definable.*
2. *For $L \subseteq \Gamma^*$ it is decidable, whether L is $\text{FO}^2[<, +1]$ -definable over finite words.*
3. *For $L \subseteq \Gamma^\omega$ it is decidable, whether L is $\text{FO}^2[<, +1]$ -definable over infinite words.*

The following proposition relates monoids in **LDA** with monoids which are simultaneously locally path-top and locally path-bottom. It is a useful tool in the proof of Theorem 4.1. Moreover, it immediately follows that $\Delta_2[<, +1]$ is a subclass of $\text{FO}^2[<, +1]$. We will further explore the relation between these two fragments in the next section.

► **Proposition 4.3.** *Let M be finite and let $h : \Gamma^* \rightarrow M$ be a surjective homomorphism. The following are equivalent:*

1. $h : \Gamma^* \rightarrow M \in \mathbf{LDA}$.
2. $eP_e e = e$ for all idempotents e of M .

► **Example 4.4.** Let $\Gamma = \{a, b, c\}$. Consider the language $L_1 = \Gamma^* ab^* a \Gamma^\infty$ consisting of all words such that there are two a 's that only contain b 's in between. It is easy to see that L_1 is $\Sigma_2[<]$ -definable. Next, we will show that L_1 is not $\text{FO}^2[<, +1]$ -definable. Choose $n \in \mathbb{N}$ such that s^n is idempotent for every $s \in \text{Synt}(L_1)$. Then $(b^n ab^n cb^n)^n \notin L_1$ whereas $(b^n ab^n cb^n)^n b^n ab^n (b^n ab^n cb^n)^n \in L_1$. This shows that $\text{Synt}(L_1)$ is not in **LDA**. By Theorem 4.1 we conclude that L_1 is not $\text{FO}^2[<, +1]$ -definable. Similarly, $L_2 = \Gamma^\infty \setminus L_1$ is definable in $\Pi_2[<]$ but not in $\text{FO}^2[<, +1]$. ◀

5 The first-order fragment Δ_2

Over finite words, the first-order fragments $\text{FO}^2[<, +1]$ and $\Delta_2[<, +1]$ have the same expressive power [14, 24]. This is not true for infinite words. Here, it turns out that $\Delta_2[<, +1]$ is a strict subclass of $\text{FO}^2[<, +1]$ and that the $\Delta_2[<, +1]$ -languages are exactly the clopen languages in $\text{FO}^2[<, +1]$.

► **Theorem 5.1.** *Let $L \subseteq \Gamma^\infty$ be a language. The following are equivalent:*

1. L is $\Delta_2[<, +1]$ -definable.
2. L is $\text{FO}^2[<, +1]$ -definable and clopen in the factor topology.

A consequence of Theorem 5.1 is that $\Delta_2[<, +1]$ is a strict subclass of $\text{FO}^2[<, +1]$. In fact, it is a strict subclass of the intersection for the fragments $\text{FO}^2[<, +1]$ and $\Sigma_2[<, +1]$.

► **Corollary 5.2.** *Over Γ^∞ , the fragment $\Delta_2[<, +1]$ is a strict subclass of the fragment $\text{FO}^2[<, +1] \cap \Sigma_2[<, +1]$ and also of the fragment $\text{FO}^2[<, +1] \cap \Pi_2[<, +1]$.*

Proof. The set of non-empty finite words Γ^+ is defined by the sentence

$$\exists x \forall y: y \leq x$$

in $\text{FO}^2[<] \cap \Sigma_2[<]$. We have to show that Γ^+ is not definable in $\Pi_2[<, +1]$. By Theorem 3.4 it suffices to show that Γ^+ is not factor closed. Let $a \in \Gamma$, and consider the word $\alpha = a^\omega \notin \Gamma^+$. Every factor open set containing α also contains some finite word $a^m \in \Gamma^+$. Hence, the complement of Γ^+ is not factor open, and therefore, Γ^+ is not factor closed. By complementation, we see that Γ^ω is definable in $\text{FO}^2[<] \cap \Pi_2[<]$ but not in $\Delta_2[<, +1]$. ◀

► **Example 5.3.** We consider another language which is definable in $\text{FO}^2[<] \cap \Sigma_2[<]$ but not in $\Delta_2[<, +1]$. Let $\Gamma = \{a, b\}$ and $L_3 = \Gamma^* ab^\infty$. The language L_3 is $\text{FO}^2[<] \cap \Sigma_2[<]$ -definable:

$$\exists x \forall y: \lambda(x) = a \wedge (\lambda(y) = a \Rightarrow y \leq x).$$

In order to show that L_3 is not definable in $\Pi_2[<, +1]$, it suffices to show that L_3 is not factor closed (Theorem 3.4). Let $k \in \mathbb{N}$. Every open set containing the word $(b^k a)^\omega \notin L_3$

also contains some word $(b^k a)^m b^\omega \in L_3$. Hence, the complement of L_3 is not k -factor open, and therefore, there is no k such that L_3 is closed in the k -factor topology.

The same reasoning also works over Γ^ω , since the language of all infinite words is definable in $\Pi_2[<, +1]$. Hence, $L'_3 = \Gamma^* a b^\omega$ is definable in $\Sigma_2[<]$ over infinite words and in $\text{FO}^2[<]$ but not in $\Delta_2[<, +1]$ over infinite words. The language L'_3 is the standard example of a language which cannot be recognized by a deterministic Büchi automaton [26, Example 4.2]. In particular, none of the fragments $\text{FO}^2[<, +1]$ or $\Sigma_2[<, +1]$ contains only deterministic languages. \triangleleft

► **Example 5.4.** Let $\Gamma = \{a, b, c\}$ and consider the language $L_4 = (\Gamma^2 \setminus \{bb\})^\circ \circ aa \circ (\Gamma^2)^\circ$ consisting of all words such that there is no factor bb before the first factor aa . The language L_4 is defined by the $\Sigma_2[<, +1]$ -sentence

$$\exists x \forall y < x : \lambda(x) = aa \wedge \lambda(y) \neq bb.$$

Here, $\lambda(x) = w$ is a shortcut saying that a factor w starts at position x . A word α is in L_4 if and only if aa is a factor of α and for every factor bb there is a factor aa to the left. These properties are $\Pi_2[<, +1]$ -definable and hence $L_4 \in \Delta_2[<, +1]$. The language L_4 is not definable in any of the fragments $\text{FO}^2[<]$, $\Sigma_2[<]$, or $\Pi_2[<]$ without successor, since its syntactic monoid is neither locally top nor locally bottom, cf. [7]. The language $L_4 \cap \Gamma^*$ has been used as an example of a language not definable in the Boolean closure of $\Sigma_2[<]$ over finite words by Almeida and Klíma [2, Proposition 6.1] as well as by Lodaya, Pandya, and Shah [14, Theorem 4]. The Boolean closure of $\Sigma_2[<]$ over finite words coincides with the second level of the Straubing-Thérien hierarchy, cf. [18, 25]. \triangleleft

6 The first-order fragments $\text{FO}^2 \cap \Sigma_2$ and $\text{FO}^2 \cap \Pi_2$

In this section, we show that topological concepts can not only be used as an ingredient for characterizing first-order fragments, but also for describing some relations between fragments. More precisely, we relate languages definable in both $\Sigma_2[<, +1]$ and $\text{FO}^2[<, +1]$ with the interiors of $\text{FO}^2[<, +1]$ -languages with respect to the factor topology. Dually, the languages in the fragment $\text{FO}^2[<, +1] \cap \Pi_2[<, +1]$ are precisely the topological closures of $\text{FO}^2[<, +1]$ -languages. Remember that for a language L , its *closure* \bar{L} is the intersection of all closed sets containing L . It can be “computed” as

$$\bar{L} = \{\alpha \in \Gamma^\infty \mid \forall U \subseteq \Gamma^\infty \text{ open with } \alpha \in U : U \cap L \neq \emptyset\}.$$

The *interior* of L is the union of all open sets contained in L . The interior of a language is the complement of the closure of its complement.

► **Theorem 6.1.** *Let $L \subseteq \Gamma^\infty$ be a regular language. The following are equivalent:*

1. $L \in \text{FO}^2[<, +1] \cap \Sigma_2[<, +1]$.
2. $L \in \text{FO}^2[<, +1]$ and L is open in the factor topology.
3. L is the factor interior of some $\text{FO}^2[<, +1]$ -definable language.

The equivalence of (1) and (2) is an immediate consequence of Theorems 3.1 and 4.1. The surprising property is (3); for example, it is not obvious that the factor interior of an $\text{FO}^2[<, +1]$ -definable language is again in $\text{FO}^2[<, +1]$. The following theorem is an immediate consequence of Theorem 6.1, obtained by complementation. In fact, the actual proof is slightly easier the other way round—proving Theorem 6.2 and then obtaining Theorem 6.1 by complementation—since the closure of a language is easier to “compute”.

► **Theorem 6.2.** *Let $L \subseteq \Gamma^\infty$ be a regular language. The following are equivalent:*

1. $L \in \text{FO}^2[<, +1] \cap \Pi_2[<, +1]$.
2. $L \in \text{FO}^2[<, +1]$ and L is closed in the factor topology.
3. L is the factor closure of some $\text{FO}^2[<, +1]$ -definable language.

The language L_3 from Example 5.3 is definable in $\text{FO}^2[<, +1] \cap \Sigma_2[<, +1]$. For any k , the k -factor closure of L_3 is Γ^∞ . Hence, one could conjecture that the factor closure of every language in $\text{FO}^2[<, +1] \cap \Sigma_2[<, +1]$ is again in $\text{FO}^2[<, +1] \cap \Sigma_2[<, +1]$ and therefore in $\Delta_2[<, +1]$. Among other things, the following example shows that this is not the case.

► **Example 6.3.** Let $\Gamma = \{a, b, c\}$. We consider the factor closure of the language L_5 defined by the $\text{FO}^2[<, +1]$ -sentence

$$\exists x: \lambda(x) = aba \wedge (\exists y > x: \lambda(y) = aba) \wedge (\neg \exists y > x: \lambda(y) = bab)$$

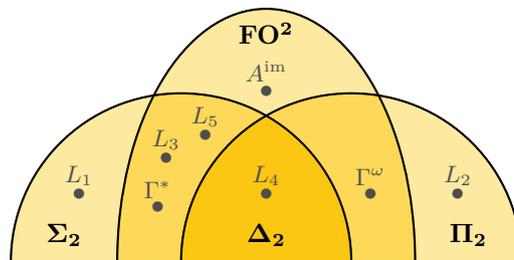
with $\lambda(x) = w$ being a macro for “an occurrence of the factor $w \in \Gamma^+$ starts at position x ”. The language L_5 contains all words of the form $u \cdot aba \cdot v \cdot aba \cdot \beta$ with $u, v \in \Gamma^*$, $\beta \in \Gamma^\infty$, and $bab \notin \text{alph}_3(aba \cdot v \cdot aba \cdot \beta)$.

The $(k+1)$ -factor closure of a language is always contained in its k -factor closure. For L_5 , we show that this inclusion is strict for $k \in \{1, 2\}$. The word $(ab)^\omega$ is in the 1-factor closure of L_5 , but not in its 2-factor closure. The word $(abacbab)^\omega$ is in the 2-factor closure of L_5 , but not in its 3-factor closure. The 1-factor closure of L_5 is $L_5 \cup \{\alpha \in \Gamma^\omega \mid a, b \in \text{im}_1(\alpha)\}$. The 2-factor closure of L_5 consists of L_5 and all words α with either $\{ab, ba, aa\} \subseteq \text{im}_2(\alpha)$ or $\{ab, ba, ac, ca\} \subseteq \text{im}_2(\alpha)$. The 3-factor closure of L_5 is similar, but it requires more case distinctions.

There is no k such that L_5 is k -factor closed. By Theorem 3.4, we see that L_5 is not $\Pi_2[<, +1]$ -definable. On the other hand, Theorem 6.2 says that every k -factor closure of L_5 is $\Pi_2[<, +1]$ -definable. Moreover, almost the same sentence as above yields $\Sigma_2[<, +1]$ -definability of L_5 . Since the 1-factor closure of L_5 is not factor open, the fragment $\text{FO}^2[<, +1] \cap \Sigma_2[<, +1]$ is not closed under taking factor closures, whereas $\text{FO}^2[<, +1]$ has this closure property by Theorem 6.2. ◁

7 Summary

We considered fragments of first-order logic over finite and infinite words. As binary predicates we allow order comparison $x < y$ and the successor predicate $x = y + 1$. Figure 1 depicts the relation between the fragments $\Sigma_2[<, +1]$, $\Pi_2[<, +1]$, and $\text{FO}^2[<, +1]$. Moreover, the languages L_1, L_2, L_3, L_4 , and L_5 from Examples 4.4, 5.3, 5.4, and 6.3 are included. For the other languages, we fix $\Gamma = \{a, b, c\}$ and $\emptyset \neq A \subsetneq \Gamma$.



■ **Figure 1** The fragments $\Sigma_2[<, +1]$, $\Pi_2[<, +1]$, and $\text{FO}^2[<, +1]$ over Γ^∞ .

The central notion for presenting our results is a partially defined composition $u \circ_k v = u'xv'$ where $u = u'x$, $v = xv'$, and $|x| = k - 1$. Using this composition, one can show that the languages definable in $\Sigma_2[<, +1]$ is exactly the class of factor polynomials. Moreover, the composition \circ_k leads to the k -factor topology, which we use in further characterizations of the successor fragments. A set is *factor open* if there exists some number k such that L is k -factor open. For every regular language L , Proposition 2.2 gives a bound k such that L is factor open if and only if L is k -factor open. Then, in Proposition 2.3, we essentially show that for a given number k it is decidable whether a regular language L is k -factor open. Altogether, in order to check whether L is factor open, we can check whether L is k -factor open, with k being the bound given by Proposition 2.2. Hence, the topological properties, which we use in the characterizations of the fragments, are decidable. Together with the decidable algebraic properties, this gives a decision procedure for deciding whether a given regular language $L \subseteq \Gamma^\infty$ or $L \subseteq \Gamma^\omega$ is definable in one of the fragments under consideration. In Table 1 we summarize our main results. All fragments are using binary predicates $[<, +1]$. The first decidable characterization of $\text{FO}^2[<, +1]$ is due to Wilke [27]. Decidability for $\Sigma_2[<, +1]$ over infinite words is new (Corollary 3.3).

Logic	Algebra	+	Topology	Languages	
Σ_2	$ePe e \leq e$	+	factor open	factor polynomials	Thm. 3.1
Π_2	$ePe e \geq e$	+	factor closed		Thm. 3.4
FO^2	LDA weak LDA	+	strictly factor closed		Thm. 4.1
Δ_2	LDA	+	factor clopen		Thm. 5.1
$\text{FO}^2 \cap \Sigma_2$	LDA	+	factor open	factor interior of FO^2	Thm. 6.1
$\text{FO}^2 \cap \Pi_2$	LDA	+	factor closed	factor closure of FO^2	Thm. 6.2

■ **Table 1** Main characterizations of some first-order fragments

Open problems The fragment $\Sigma_2[<, +1]$ has a language description in terms of factor polynomials. Without the successor predicate similar characterizations in terms of so-called unambiguous polynomials exist for the fragments $\text{FO}^2[<]$, for $\text{FO}^2[<] \cap \Sigma_2[<]$, and for $\Delta_2[<]$, cf. [7]. It is open whether these fragments admit similar characterizations if we allow the successor predicate.

Moreover, for the fragment $\Delta_2[<, +1]$ we only have an implicit decidable characterization based on the decidability of $\Sigma_2[<, +1]$ and $\Pi_2[<, +1]$ (or alternatively, based on the decidability of $\text{FO}^2[<, +1]$ and being clopen). A more direct characterization of this fragment remains open. For $\Delta_2[<]$ without successor, such a characterization shows that all languages in $\Delta_2[<]$ over infinite words are recognizable by deterministic Büchi automata.

Another important fragment is $\mathbb{B}\Sigma_1$, the Boolean closure of Σ_1 . A result of Knast [13] shows that, over finite words, it is decidable whether a regular language is definable in $\mathbb{B}\Sigma_1[<, +1, \min, \max]$, which over finite words corresponds to the first level of the dot-depth hierarchy. A similar result over infinite words is still missing.

Acknowledgments We would like to thank the anonymous referees for their comments which helped to improve the presentation of this paper.

References

- 1 J. Almeida. A syntactical proof of locality of **DA**. *Int. J. Algebra Comput.*, 6(2):165–177, 1996.
- 2 J. Almeida and O. Klíma. New decidable upper bound of the second level in the Straubing-Thérien concatenation hierarchy of star-free languages. *Discrete Math. Theor. Comput. Sci.*, 12(4):41–58, 2010.
- 3 M. Bojańczyk. The common fragment of ACTL and LTL. In *FoSSaCS 2008, Proceedings*, vol. 4962 of *LNCS*, pp. 172–185. Springer, 2008.
- 4 J. A. Brzozowski and R. Knast. The dot-depth hierarchy of star-free languages is infinite. *J. Comput. Syst. Sci.*, 16(1):37–55, 1978.
- 5 R. S. Cohen and J. A. Brzozowski. Dot-depth of star-free events. *J. Comput. Syst. Sci.*, 5(1):1–16, 1971.
- 6 V. Diekert and P. Gastin. First-order definable languages. In *Logic and Automata: History and Perspectives*, Texts in Logic and Games, pp. 261–306. Amsterdam University Press, 2008.
- 7 V. Diekert and M. Kufleitner. Fragments of first-order logic over infinite words. To appear in *Theory Comput. Syst.*, Special issue STACS 2009.
- 8 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.
- 9 Ch. Glaßer and H. Schmitz. Languages of dot-depth $3/2$. In *STACS 2000, Proceedings*, vol. 1770 of *LNCS*, pp. 555–566. Springer, 2000.
- 10 Ch. Glaßer and H. Schmitz. Languages of dot-depth $3/2$. *Theory Comput. Syst.*, 42(2):256–286, 2008.
- 11 J. Kallas, M. Kufleitner, and A. Lauser. First-order fragments with successor over infinite words. Tech. Rep. no. 2010/08, Computer Science, University of Stuttgart (Germany), 2010.
- 12 J. A. W. Kamp. *Tense Logic and the Theory of Linear Order*. PhD thesis, University of California, Los Angeles (California), 1968.
- 13 R. Knast. A semigroup characterization of dot-depth one languages. *RAIRO, Inform. Théor. Appl.*, 17(4):321–330, 1983.
- 14 K. Lodaya, P. K. Pandya, and S. S. Shah. Around dot depth two. In *DLT 2010, Proceedings*, vol. 6224 of *LNCS*, pp. 305–316. Springer, 2010.
- 15 D. Perrin and J.-É. Pin. *Infinite words*, vol. 141 of *Pure and Appl. Mathematics*. Elsevier, 2004.
- 16 J.-É. Pin. *Varieties of Formal Languages*. North Oxford Academic, 1986.
- 17 J.-É. Pin. Syntactic semigroups. In *Handbook of Formal Languages, Vol. 1*, pp. 679–746. Springer, 1997.
- 18 J.-É. Pin and P. Weil. Polynomial closure and unambiguous product. *Theory Comput. Syst.*, 30(4):383–422, 1997.
- 19 J.-É. Pin and P. Weil. The wreath product principle for ordered semigroups. *Commun. Algebra*, 30(12):5677–5713, 2002.
- 20 H. Straubing. A generalization of the Schützenberger product of finite monoids. *Theor. Comput. Sci.*, 13:137–150, 1981.
- 21 H. Straubing. Finite semigroup varieties of the form $\mathbf{V}*\mathbf{D}$. *J. Pure Appl. Algebra*, 36(1):53–94, 1985.
- 22 P. Tesson and D. Thérien. Diamonds are Forever: The Variety **DA**. In *Semigroups, Algorithms, Automata and Languages, 2001*, pp. 475–500. World Scientific, 2002.
- 23 D. Thérien. Classification of finite monoids: The language approach. *Theor. Comput. Sci.*, 14(2):195–208, 1981.
- 24 D. Thérien and Th. Wilke. Over words, two variables are as powerful as one quantifier alternation. In *STOC 1998, Proceedings*, pp. 234–240. ACM Press, 1998.
- 25 W. Thomas. Classifying regular events in symbolic logic. *J. Comput. Syst. Sci.*, 25:360–376, 1982.
- 26 W. Thomas. Automata on infinite objects. In *Handbook of Theoretical Computer Science*, pp. 133–191. Elsevier, 1990.
- 27 Th. Wilke. *Classifying Discrete Temporal Properties*. Habilitationsschrift, Universität Kiel, April 1998.