Around Dot-Depth One

Manfred Kufleitner* Alexander Lauser*

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

Abstract. The dot-depth hierarchy is a classification of star-free languages. It is related to the quantifier alternation hierarchy of first-order logic over finite words. We consider subclasses of languages with dot-depth 1/2 and dot-depth 1 obtained by prohibiting the specification of prefixes or suffixes. As it turns out, these language classes are in one-to-one correspondence with fragments of alternation-free first-order logic without min- or max-predicate, respectively. For all fragments, we obtain effective algebraic characterizations. Moreover, we give new proofs for the decidability of the membership problem for dot-depth 1/2 and dot-depth 1.

Keywords: automata theory; semigroups; regular languages; first-order logic; finite model theory

1 Introduction

The dot-depth hierarchy $\mathcal{B}_n$ for $n \in \mathbb{N} + \{1/2, 1\}$ has been introduced by Cohen and Brzozowski [3]. A very similar hierarchy is the Straubing-Thérien hierarchy $\mathcal{L}_n$, see [21, 24]. Both hierarchies are strict [2] and they exhaust the class of star-free languages. A classical result of McNaughton and Papert is that a language is star-free if and only if it is definable in first-order logic [12]. Thomas [26] has tightened this result by showing that there is a one-to-one correspondence between the dot-depth hierarchy (and also between the Straubing-Thérien hierarchy) and the quantifier alternation hierarchy of first-order logic. More precisely, the dot-depth hierarchy is related to the quantifier alternation hierarchy over the signature $[<, +1, \min, \max]$, whereas the Straubing-Thérien hierarchy corresponds to the quantifier alternation hierarchy over the signature $[<]$.

Schützenberger has shown that a language is star-free if and only if its syntactic semigroup is aperiodic [18]. The latter property is decidable. Together with the result of McNaughton and Papert, this yields a decision procedure for definability in first-order logic. Effectively determining the level of a language in the dot-depth hierarchy, or equivalently in the quantifier alternation hierarchy of first-order logic, is one of the most challenging open problems in automata theory. For $n \in \mathbb{N}$, Straubing has shown that membership in $\mathcal{B}_n$ is decidable if and only if membership in $\mathcal{L}_n$ is decidable [22]. This result has been extended to the half-levels

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

by Pin and Weil [17]. Simon has shown that the class of piecewise testable languages $\mathcal{L}_1$ is decidable [19]. Later, Knast [9] gave an effective algebraic characterization of $\mathcal{B}_1$. Decidability of $\mathcal{L}_{1/2}$ was shown by Pin [14], and the levels $\mathcal{B}_{1/2}$ and $\mathcal{L}_{3/2}$ are decidable by a result of Pin and Weil [16]. The decidability of $\mathcal{B}_{3/2}$ was first shown by Glaßer and Schmitz [5, 6]. Other proofs were given by Pin and Weil [17] and by Kallas *et al.* [7]. To date, no other levels are known to be decidable.

In this paper, we focus on subclasses of $\mathcal{B}_{1/2}$ and $\mathcal{B}_1$. For both $\mathcal{B}_{1/2}$ and $\mathcal{B}_1$ we give new proofs for their effective algebraic characterizations. The proof of Pin and Weil [16] for $\mathcal{B}_{1/2}$ is based on factorization forests [20], and the proof of Knast [9] as well as the simplified version of Thérien [25] for $\mathcal{B}_1$ are based on a generalization of finite monoids, so-called *finite categories* [28]. Our proof for $\mathcal{B}_1$ is a generalization of Klíma's proof [8] for $\mathcal{L}_1$. The main advantage of our proofs for $\mathcal{B}_{1/2}$ and $\mathcal{B}_1$ over previous ones is that the constants involved in finding language descriptions for given algebraic objects are more explicit (and therefore smaller).

The main original contributions of this paper are effective algebraic characterizations of fragments of alternation-free first-order logic over the signatures $[<, +1, \min]$ without max-predicate, $[<, +1, \max]$ without min, and $[<, +1]$ without min and max. These fragments also admit language characterizations in terms of subclasses of $\mathcal{B}_{1/2}$ and $\mathcal{B}_1$. The corresponding language classes are obtained by prohibiting the specification of prefixes or suffixes. A more detailed overview of our results can be found in the summary in Section 7.

2 Preliminaries

Words and languages Let A be a finite nonempty alphabet. The set of finite words is A^*. By 1 we denote the empty word and $A^+ = A^* \setminus \{1\}$ is the set of finite nonempty words. A word $v \in A^*$ is a *prefix* (resp. *suffix*, resp. *factor*) of u if $u \in vA^*$ (resp. $u \in A^*v$, resp. $u \in A^*vA^*$). The length of a word $u \in A^*$ is $|u|$ and its alphabet is $\mathrm{alph}(u) = \{a \in A \mid u \in A^*aA^*\}$. Similarly, $\mathrm{alph}_k(u) = \left\{v \in A^k \;\middle|\; u \in A^*vA^*\right\}$ is the set of all factors of u of length k. A *quotient* of $L \subseteq A^+$ is a language of the form $u^{-1}L = \{v \in A^+ \mid uv \in L\}$ or $Lu^{-1} = \{v \in A^+ \mid vu \in L\}$ for $u \in A^*$. A language L is a *monomial* of degree m if $L = w_1A^*w_2 \cdots A^*w_n$ for some $w_1, \ldots, w_n \in A^*$ with $|w_1 \cdots w_n| = m$. A language has *dot-depth one* if it is a Boolean combination of monomials. Throughout this paper, Boolean operations are complementation, *finite* union, and *finite* intersection. Positive Boolean operations are finite union and finite intersection.

First-order logic over words We consider the first-order logic $\mathrm{FO} = \mathrm{FO}[<, +1, \min, \max]$ over nonempty finite words. We view words as sequences of labeled positions which are linearly ordered by $<$. Variables are interpreted as positions of a word. For variables x, y we have the following atomic formulas: $x < y$ says that x is a position smaller than y; and $x = y + 1$ is true if x is the immediate successor of y; the formula $\min(x)$ (resp. $\max(x)$) holds if x is the first (resp. last) position. Moreover, we always assume that we have an atomic formula $\top$ (for *true*), equality of positions $x = y$, and a predicate $\lambda(x) = a$ specifying that position x is labeled by $a \in A$. Formulas can be composed using Boolean operations, existential quantification, and universal quantification. Their semantics is as usual, see *e.g.* [4, 23, 27]. A sentence is a formula without free variables. For a sentence φ of FO we write $u \models \varphi$ if u is a model of φ and the language defined by φ is $L(\varphi) = \{u \in A^+ \mid u \models \varphi\}$.

The fragment Σ_1 consists of all FO-formulas in prenex normal form with only one block of quantifiers and these quantifiers are existential. Let $\mathcal{C} \subseteq \{<, +1, \min, \max\}$. By $\Sigma_1[\mathcal{C}]$ we denote the class of formulas in Σ_1 which only use predicates in $\mathcal{C}$, equality, and the label predicate. The fragment of alternation-free formulas over the signature $\mathcal{C}$ is $\mathbb{B}\Sigma_1[\mathcal{C}]$; it comprises all Boolean combinations of formulas in $\Sigma_1[\mathcal{C}]$.

Finite semigroups and recognizable languages Let S be a semigroup. We always assume that S is nonempty. The set of *idempotents* is $E(S) = \{e \in S \mid e^2 = e\}$. For every finite semigroup S there exists a number $\omega \geq 1$ such that for every $x \in S$, the power x^ω is the unique idempotent element generated by x. Frequently, we consider words $u, v \in S^*$ where the alphabet is a semigroup. We write "$u = v$ in S" if either $u = 1 = v$ or $u, v \in S^+$ evaluate to the same element of S.

Lemma 1 *Let S be a finite semigroup. For all $x_1, \ldots, x_{|S|} \in S$ there exist an index $i \in \{1, \ldots, |S|\}$ and an idempotent $e \in E(S)$ such that $x_1 \cdots x_i = x_1 \cdots x_i e$ in S.*

Proof: Choose some arbitrary element $x_{|S|+1} \in S$. By the pigeonhole principle there exist $i < j \leq |S| + 1$ such that $x_1 \cdots x_i = x_1 \cdots x_j$ in S. In particular $i \leq |S|$. Hence, $x_1 \cdots x_i = x_1 \cdots x_i e$ in S with $e = (x_{i+1} \cdots x_j)^\omega$. ☐

A subset $I \subseteq S$ is an *ideal* (resp. *right ideal*, resp. *left ideal*) if $S^1 I S^1 \subseteq I$ (resp. $IS^1 \subseteq I$, resp. $S^1 I \subseteq I$). Here, the monoid $S^1 = S \cup \{1\}$ is obtained by adjoining a new neutral element. Green's relations are an important tool in the study of semigroups. They are defined as follows. Let $x \leq_{\mathcal{J}} y$ (resp. $x \leq_{\mathcal{R}} y$, resp. $x \leq_{\mathcal{L}} y$) if there exist $s, t \in S^1$ such that $x = syt$ in S (resp. $x = yt$ in S, resp. $x = sy$ in S). Let $x \mathrel{\mathcal{J}} y$ (resp. $x \mathrel{\mathcal{R}} y$, resp. $x \mathrel{\mathcal{L}} y$) if $x \leq_{\mathcal{J}} y$ and $y \leq_{\mathcal{J}} x$ (resp. $x \leq_{\mathcal{R}} y$ and $y \leq_{\mathcal{R}} x$, resp. $x \leq_{\mathcal{L}} y$ and $y \leq_{\mathcal{L}} x$). Therefore, $x \mathrel{\mathcal{J}} y$ (resp. $x \mathrel{\mathcal{R}} y$, resp. $x \mathrel{\mathcal{L}} y$) if and only if x and y generate the same ideal (resp. right ideal, resp. left ideal) in S. The relations $\leq_{\mathcal{R}}$, $\leq_{\mathcal{L}}$, and $\leq_{\mathcal{J}}$ form preorders on S; therefore $\mathcal{R}$, $\mathcal{L}$, and $\mathcal{J}$ are equivalence relations.

Let $\leq$ be a preorder on S. A set $P \subseteq S$ is a *$\leq$-order ideal* if $x \leq y$ and $y \in P$ implies $x \in P$. Note that every $\leq_{\mathcal{R}}$-order ideal (resp. $\leq_{\mathcal{L}}$-order ideal, resp. $\leq_{\mathcal{J}}$-order ideal) is a right ideal (resp. left ideal, resp. ideal) and vice versa. An *ordered semigroup* S is equipped with a compatible partial order $\leq$, *i.e.*, if $p \leq q$ and $s \leq t$, then $ps \leq qt$. Every semigroup is an ordered semigroup with equality as partial order. A language $L \subseteq A^+$ is *recognized* by an ordered semigroup S if there exists a homomorphism $h : A^+ \to S$ such that $L = h^{-1}(P)$ for some $\leq$-order ideal P. If the order of S is equality, then we obtain the usual notion of recognition. For a language $L \subseteq A^+$ the *syntactic preorder* $\leq_L$ on A^+ is given by $x \leq_L y$ if for all $u, v \in A^*$:

$$uyv \in L \text{ implies } uxv \in L.$$

The syntactic congruence $\equiv_L$ is defined by $x \equiv_L y$ if both $x \leq_L y$ and $y \leq_L x$. The equivalence classes $[x]_L = \{y \in A^+ \mid x \equiv_L y\}$ equipped with the canonical composition constitute the *syntactic semigroup* $\mathrm{Synt}(L)$ and the preorder $\leq_L$ on A^+ induces a compatible partial order on $\mathrm{Synt}(L)$. The *syntactic homomorphism* is $h_L : A^+ \to \mathrm{Synt}(L)$ with $h_L(x) = [x]_L$. The syntactic semigroup of L is finite if and only if L is regular. Moreover, every language is recognized by its syntactic semigroup.

By $[\![x^\omega y x^\omega \leq x^\omega]\!]$ we denote the class of finite ordered semigroups S such that $x^\omega y x^\omega \leq x^\omega$ for all elements $x, y \in S$. We let $\mathbf{B}_1$ be the class of finite semigroups S such that $(exfy)^\omega exf(tesf)^\omega = (exfy)^\omega esf(tesf)^\omega$ for all idempotents $e, f \in E(S)$ and all elements $s, t, x, y \in S$.

Lemma 2 *Let $(S, \leq)$ be an ordered semigroup such that $x^\omega y x^\omega \leq x^\omega$ for all $x, y \in S$. Then $S \in \mathbf{B}_1$.*

Proof: We have $f \geq fy(exfy)^{\omega-1}esf$ for all $s, x, y \in S$ and all idempotents $e, f \in E(S)$. Hence $(exfy)^\omega exf(tesf)^\omega \geq (exfy)^\omega ex(fy(exfy)^{\omega-1}esf)(tesf)^\omega = (exfy)^\omega esf(tesf)^\omega$. By symmetry $(exfy)^\omega exf(tesf)^\omega \leq (exfy)^\omega esf(tesf)^\omega$. □

Lemma 3 *Let $S \in \mathbf{B}_1$ and let $u, v \in S$ with $u = ue$ and $v = ve$ for some idempotent $e \in E(S)$. If $u \mathrel{\mathcal{R}} v$, then $u = v$.*

Proof: Let $x, y \in S$ such that $v = ux$ and $u = vy$. The $\mathbf{B}_1$-equation yields $v = u(exey)^\omega exe(eeee)^\omega = u(exey)^\omega eee(eeee)^\omega = u$. □

3 Dot-Depth 1/2

A language $L \subseteq A^+$ has dot-depth 1/2 if it is a positive Boolean combination of monomials $w_1 A^* w_2 \cdots A^* w_n$ with $w_i \in A^*$. By a result of Thomas [26], a language has dot-depth 1/2 if and only if it is definable in existential first-order logic $\Sigma_1[<, +1, \min, \max]$. Pin and Weil [16] have shown that L has dot-depth 1/2 if and only if $\mathrm{Synt}(L) \in [\![x^\omega y x^\omega \leq x^\omega]\!]$. In this section, we give a new proof of these equivalences. The key step in the proof is to show that if $L \subseteq A^+$ is recognized by some semigroup in $[\![x^\omega y x^\omega \leq x^\omega]\!]$, then L is a union of monomials $w_1 A^* w_2 \cdots A^* w_n$. The main advantage of the proof given here is that the degree $|w_1 \cdots w_n|$ is polynomially bounded (Proposition 1), whereas in the proof of Pin and Weil, the bound is exponential.

Theorem 1 (Pin/Weil [16], Thomas [26]) *Let $L \subseteq A^+$. The following assertions are equivalent:*

1. *L is definable in $\Sigma_1[<, +1, \min, \max]$.*
2. *L is a finite union of monomials $w_1 A^* w_2 \cdots A^* w_n$.*
3. *L is a positive Boolean combination of monomials $w_1 A^* w_2 \cdots A^* w_n$.*
4. *$\mathrm{Synt}(L) \in [\![x^\omega y x^\omega \leq x^\omega]\!]$.*

The remainder of this section is devoted to the proof of the above theorem.

Lemma 4 *Let $L \subseteq A^+$ be definable by a sentence in $\Sigma_1[<, +1, \min, \max]$ with m variables. Then L is a finite union of languages $w_1 A^+ w_2 \cdots A^+ w_n$ with $|w_1 \cdots w_n| \leq m$. In particular, L is a finite union of monomials of the form $w_1 A^* w_2 \cdots A^* w_n$ of degree at most $2m + 1$.*

Proof: Let $L = L(\varphi)$ for some $\Sigma_1[<, +1, \min, \max]$-sentence φ. We can write

$$\varphi = \exists x_1 \cdots \exists x_m \colon \psi(x_1, \ldots, x_m)$$

such that $\psi(x_1, \ldots, x_m)$ is quantifier free. Suppose $u \models \varphi$, *i.e.*, there exist positions $j_1, \ldots, j_m$ of u such that $u \models \psi(j_1, \ldots, j_m)$. The latter notation means that ψ is true on u when every x_i is interpreted by j_i. We say that a position j of u is *marked* if $j = j_i$ for some i. Assume that the first and the last position of u are marked. Let $u = w_1 u_1 w_2 \cdots u_{n-1} w_n$ for $u_i \in A^+$ such that the factors w_i consist of the marked positions. Now, for $P_u = w_1 A^+ w_2 \cdots A^+ w_n$ we have $|w_1 \cdots w_n| \leq m$ and $u \in P_u$. Moreover, $P_u \subseteq L(\varphi)$ since the satisfying assignment of u can be adapted to all $v \in P_u$. Suppose now that the first position is marked but the last position is not marked and let $u = w_1 u_1 \cdots w_n u_n$ for $u_i \in A^+$ such that the factors w_i consist of the marked positions. For the language $P_u = w_1 A^+ \cdots w_n A^+$ we have $|w_1 \cdots w_n| \leq m$ and $u \in P_u \subseteq L(\varphi)$. In case the last position is marked but the first position is not, we take the language $P_u = A^+ w_1 \cdots A^+ w_n$ and if both positions are not marked we take $P_u = A^+ w_1 \cdots A^+ w_n A^+$. It follows $L(\varphi) = \bigcup_{u \models \varphi} P_u$ and this union is finite since there are only finitely many languages of the form $w_1 A^+ w_2 \cdots A^+ w_n$ with $|w_1 \cdots w_n| \leq m$.

Every monomial $w_1 A^+ \cdots w_{n-1} A^+ w_n$ is a union of monomials of the form

$$w_1 a_1 A^* \cdots w_{n-1} a_{n-1} A^* w_n$$

for $a_1, \ldots, a_{n-1} \in A$. The worst case here is $w_1 = 1 = w_n$ and $w_i \in A$ for $1 < i < n$. $\square$

Lemma 5 *Let $P = w_1 A^* w_2 \cdots A^* w_n$ for $w_1, \ldots, w_n \in A^+$ and let $m = |w_1 \cdots w_n|$.*

1. *P is definable by a $\Sigma_1[<, +1, \min, \max]$-sentence with m variables.*
2. *PA^* is definable by a $\Sigma_1[<, +1, \min]$-sentence with m variables.*
3. *$A^* P A^*$ is definable by a $\Sigma_1[<, +1]$-sentence with m variables.*

Proof: The proof is straightforward. For variable vectors $\underline{x} = (x_1, \ldots, x_k)$ and $\underline{y} = (y_1, \ldots, y_\ell)$ we use the shortcuts $\exists \underline{x}$ for $\exists x_1 \cdots \exists x_k$, and $\min(\underline{x})$ for $\min(x_1)$ and $\max(\underline{x})$ for $\max(x_k)$, and $\underline{x} < \underline{y}$ means $x_k < y_1$. Moreover, $\lambda(\underline{x}) = a_1 \cdots a_k$ is a shortcut for

$$\bigwedge_{1 \leq j \leq k} \lambda(x_j) = a_j \;\wedge\; \bigwedge_{1 \leq j < k} x_{j+1} = x_j + 1.$$

For every $i \in \{1, \ldots, n\}$ we introduce a variable vector $\underline{x}_i = (x_{i,1}, \ldots, x_{i,|w_i|})$. Now, the monomial $A^* P A^*$ is defined by the following sentence φ

$$\exists \underline{x}_1 \cdots \exists \underline{x}_n \colon \bigwedge_{1 \leq i \leq n} \lambda(\underline{x}_i) = w_i \wedge \bigwedge_{1 \leq i < n} \underline{x}_i < \underline{x}_{i+1}$$

where the first term of the conjunction ensures that each $\underline{x}_i$ corresponds to a factor w_i, and the second term ensures that the factors w_i occur in the correct order. The sentence for PA^* is $\varphi \wedge \min(\underline{x}_1)$ and the sentence for P is given by $\varphi \wedge \min(\underline{x}_1) \wedge \max(\underline{x}_n)$. $\square$

Lemma 6 *Let $L \subseteq A^+$ be a positive Boolean combination of monomials $w_1 A^* w_2 \cdots A^* w_n$. Then* $\mathrm{Synt}(L) \in [\![x^\omega y x^\omega \leq x^\omega]\!]$.

Proof: Consider $P = w_1 A^* w_2 \cdots A^* w_n$ and let $m \geq \max\{|w_1|, \ldots, |w_n|\}$. Let $x, y \in A^+$ and $u, v \in A^*$ be such that $u x^m v \in P$. Let i be maximal such that $u x^m \in w_1 A^* \cdots w_i A^* = Q_i$ and let j be minimal such that $x^m v \in A^* w_j \cdots A^* w_n = R_j$. By the choice of m we have $j \leq i + 1$. Therefore, $u x^m y x^m v \in Q_i R_j \subseteq P$.

Let L be a positive Boolean combination of monomials $w_1 A^* w_2 \cdots A^* w_n$. Choose $m \geq 0$ large enough such that each w_i has length at most m and such that all m-th powers are idempotent in $\mathrm{Synt}(L)$. The above observation yields that $u x^m v \in L$ implies $u x^m y x^m v \in L$. This shows $\mathrm{Synt}(L) \in [\![x^\omega y x^\omega \leq x^\omega]\!]$. $\square$

Lemma 7 *Let S be a finite semigroup. For every $w \in S^+$ there exists a factorization $w = x_1 w_1 y_1 \cdots x_m w_m y_m s$ with*

1. $0 \le m \le |E(S)|$ *and* $|x_1 y_1 \cdots x_m y_m s| < 2\,|S| \cdot |E(S)| + |S|$,
2. $w_i, s \in S^*$, $x_i, y_i \in S^+$, $|y_i| \le |S|$,
3. *there exist* $e_1, \ldots, e_m \in E(S)$ *such that* $x_i = x_i e_i$ *in* S *and* $y_i = y_i e_i$ *in* S.

Proof: For $w \in S^*$ let $E(w)$ be the set of all $e \in E(S)$ such that there exists a factor $x \in S^+$ of w with $|x| \le |S|$ and $xe = x$ in S. We prove the existence of the factorization by induction on $|E(w)|$ with the stronger assertions that $m \le |E(w)|$ and $|x_1 y_1 \cdots x_m y_m s| < 2\,|S| \cdot |E(w)| + |S|$ instead of condition "1". Suppose $|E(w)| = 0$. By Lemma 1 we have $|w| < |S|$. Hence, we can choose $m = 0$ and $s = w$.

If $|E(w)| \ge 1$, then Lemma 1 yields a nonempty prefix x of w with $|x| \le |S|$ such that $xe = x$ in S for some idempotent $e \in E(S)$. Write $w = xw'$. We have to distinguish two cases. The first case is $e \notin E(w')$. By induction, there exists a factorization $w' = x_1 w_1 y_1 \cdots x_m w_m y_m s$ with $m \le |E(w')| < |E(w)|$ and $|x_1 y_1 \cdots x_m y_m s| < 2\,|S| \cdot |E(w')| + |S|$ satisfying conditions "2" and "3". If $m \ge 1$, then $w = (x x_1) w_1 y_1 \cdots x_m w_m y_m s$ is a desired factorization of w. If $w' = s$, then the factorization is $w = xs$ with $m = 0$.

The second case is $e \in E(w')$. Let $w' = w_0 y_0 w''$ such that $y_0 \in S^+$, $|y_0| \le |S|$, $y_0 e = y_0$ in S and $e \notin E(w'')$, *i.e.*, we take y_0 as the last short factor of w' which is stabilized by e. By induction, there exists a factorization $w'' = x_1 w_1 y_1 \cdots x_m w_m y_m s$. Now, $w = x_0 w_0 y_0 \cdots x_m w_m y_m s$ with $x_0 = x$ is a factorization of w of the desired form. □

Proposition 1 *Let $L \subseteq A^+$ be recognized by $S \in [\![x^\omega y x^\omega \le x^\omega]\!]$. Then L is a finite union of monomials $w_1 A^* w_2 \cdots A^* w_n$ with degree $|w_1 \cdots w_n| < 2\,|S|^3 + |S|^2$ and $n \le |S|^2$.*

Proof: Let $h : A^+ \to S$ be a homomorphism recognizing L. The order ideal of S generated by a subset $P \subseteq S$ is $\downarrow P = \{x \in S \mid x \le y \text{ for some } y \in P\}$. We define the *depth* of the word $u \in A^+$ as $d(u) = |\{s \in S \mid h(u) \le_{\mathcal{R}} s\}|$. For every $u \in A^+$ we are going to construct a language $P_u = w_1 A^* w_2 \cdots A^* w_n$ with $n \le d(u)\,|S|$ and $|w_1 \cdots w_n| < 2d(u)\,|S|^2 + d(u)\,|S|$ such that $u \in P_u \subseteq h^{-1}\big(\downarrow h(u)\big)$. With this claim, $L = \bigcup_{u \in L} P_u$ is a finite union since there are only finitely many monomials of degree less than $2\,|S|^3 + |S|^2$.

Write $u = vw$ with $v \in A^*$ and $w \in aA^*$ such that $h(va)\ \mathcal{R}\ h(u)$ and either $v = 1$ or $h(v) >_{\mathcal{R}} h(va)$. By Lemma 7 we find a factorization $w = x_1 w_1 y_1 \cdots x_m w_m y_m s$ such that $|x_1 y_1 \cdots x_m y_m s| < 2\,|S|^2 + |S|$ and for all $i \in \{1, \ldots, m\}$ there exists an idempotent e_i with $h(x_i) e_i = h(x_i)$ and $h(y_i) e_i = h(y_i)$. Using Lemma 2 and Lemma 3 we see that $h(u) = h(vw) = h(v x_1 \cdots x_m s)$.

If $v = 1$, then we set $P_u = x_1 A^* y_1 \cdots x_m A^* y_m s$. The degree of P_u is less than $2\,|S|^2 + |S|$. By construction, we have $u = w \in P_u$. Consider $w' \in P_u$ with $w' = x_1 w'_1 y_1 \cdots x_m w'_m y_m s$. Since $ese \le e$ for all $s \in S$ and all $e \in E(S)$ we see that $h(x_i) = h(x_i) e_i \ge h(x_i) e_i h(w'_i y_i) e_i = h(x_i w'_i y_i)$. Therefore, $h(u) = h(x_1 \cdots x_m s) \ge h(w')$. This shows $P_u \subseteq h^{-1}\big(\downarrow h(u)\big)$.

Let now $v \ne 1$. Then $d(v) < d(u)$ and thus, by induction, there exists a monomial P_v with $v \in P_v \subseteq h^{-1}(\downarrow h(v))$ of degree less than $2d(u)\,|S|^2 + d(u)\,|S| - 2\,|S|^2 - |S|$. We set $P_u = P_v\, x_1 A^* y_1 \cdots x_m A^* y_m s$. The degree of P_u is less than $2d(u)\,|S|^2 + d(u)\,|S|$. Note that $u \in P_u$. Suppose $v'w' \in P_u$ with $v' \in P_v$ and $w' = x_1 w'_1 y_1 \cdots x_m w'_m y_m s$. Then we have $h(v') \le h(v)$ and as before we see that $h(x_i) \ge h(x_i w'_i y_i)$. Therefore, $h(x_1 \cdots x_m s) \ge h(w')$ and $h(u) = h(v x_1 \cdots x_m s) \ge h(v'w')$. □

We are now ready to prove Theorem 1.

Proof (Proof of Theorem 1): "1 ⇒ 2": This is Lemma 4. "2 ⇒ 1" follows from property "1" of Lemma 5 and the fact that $\Sigma_1[<, +1, \min, \max]$ is closed under union. The implication "2 ⇒ 3" is trivial, and "3 ⇒ 4" is Lemma 6. Finally, "4 ⇒ 2" follows immediately from Proposition 1. □

4 Existential First-Order Logic without min or max

At higher levels of the quantifier alternation hierarchy, it is possible to specify the prefix and the suffix of a word by using successor $+1$ as the only numerical predicate. At the level Σ_1, the min-predicate is required to determine prefixes, and max is required for suffixes. We have the following inclusions:

$$\begin{array}{ccccccc} \Sigma_1[<] & \subsetneq & & \subsetneq & \Sigma_1[<, +1, \min] & \subsetneq & \\ & & \Sigma_1[<, +1] & & & & \Sigma_1[<, +1, \min, \max] \\ \Sigma_1[+1] & \subsetneq & & \subsetneq & \Sigma_1[<, +1, \max] & \subsetneq & \end{array}$$

Pin [14, 15] has given effective characterizations for the classes of languages definable in $\Sigma_1[<]$ and $\Sigma_1[+1]$. For $\Sigma_1[<, +1, \min, \max]$, decidability follows by a result of Pin and Weil [16] (or alternatively by Theorem 1). In this section, we characterize the languages definable in the remaining fragments and we show that definability within these fragments is decidable. The proofs easily follow from Theorem 1.

Theorem 2 *Let $L \subseteq A^+$. The following assertions are equivalent:*

1. *L is definable in $\Sigma_1[<, +1, \min]$.*
2. *L is a finite union of monomials $w_1 A^* \cdots w_n A^*$.*
3. $\mathrm{Synt}(L) \in [\![x^\omega y x^\omega \leq x^\omega]\!]$ *and $h_L(L)$ is a right ideal of* $\mathrm{Synt}(L)$.

Proof: "1 ⇒ 2": Let $L = L(\varphi)$ for $\varphi \in \Sigma_1[<, +1, \min]$. By Theorem 1, the language L is a finite union of monomials $w_1 A^* w_2 \cdots A^* w_n$. Let $u \models \varphi$. Then for every $v \in A^*$ the same assignment of the variables which makes φ true on u also satisfies φ on uv. Therefore, $LA^* \subseteq L$. Since $(P \cup Q)A^* = PA^* \cup QA^*$, it follows that L is a finite union of monomials $w_1 A^* w_2 \cdots A^* w_n A^*$. Assertion "2" in Lemma 5 yields "2 ⇒ 1".

"2 ⇒ 3": We have $\mathrm{Synt}(L) \in [\![x^\omega y x^\omega \leq x^\omega]\!]$ by Theorem 1. The language L is a right ideal of A^+. Since the image of a right ideal under a surjective homomorphism is again a right ideal, $h_L(L)$ is a right ideal of $\mathrm{Synt}(L)$.

"3 ⇒ 2": By Theorem 1, the language L is a union of monomials of the form $P = w_1 A^* w_2 \cdots A^* w_n$. Suppose $P \subseteq L$. Since right ideals are closed under inverse homomorphisms, we see that $PA^* \subseteq L$. Hence, L is a union of monomials of the form $w_1 A^* w_2 \cdots A^* w_n A^*$. □

Of course, there also is a left-right dual of the above theorem: A language L is definable in $\Sigma_1[<, +1, \max]$ if and only if L is a union of monomials of the form $A^* w_1 \cdots A^* w_n$ if and only if $\mathrm{Synt}(L) \in [\![x^\omega y x^\omega \leq x^\omega]\!]$ and $h_L(L)$ is a left ideal of $\mathrm{Synt}(L)$. The following theorem is the analogue of Theorem 2 with neither min nor max predicates.

Theorem 3 *Let $L \subseteq A^+$. The following assertions are equivalent:*

1. *L is definable in $\Sigma_1[<, +1]$.*
2. *L is a finite union of monomials $A^* w_1 \cdots A^* w_n A^*$.*
3. $\mathrm{Synt}(L) \in [\![x^\omega y x^\omega \leq x^\omega]\!]$ *and $h_L(L)$ is an ideal of* $\mathrm{Synt}(L)$.

Proof: The proof is along the same lines as Theorem 2. □

The above characterizations yield the following decidability result.

Corollary 1 *Let $L \subseteq A^+$ be a regular language. It is decidable whether L is definable in* $\Sigma_1[<, +1]$ *(resp.* $\Sigma_1[<, +1, \min]$*, resp.* $\Sigma_1[<, +1, \max]$*).*

Proof: The syntactic homomorphism $h_L : A^+ \to \mathrm{Synt}(L)$ of L is effectively computable. Hence, one can verify whether property "3" in Theorem 3 (resp. "3" in Theorem 2, resp. the left-right dual of "3" in Theorem 2) holds. □

5 Dot-Depth One

A language $L \subseteq A^+$ has dot-depth one if it is a Boolean combination of monomials of the form $w_1 A^* w_2 \cdots A^* w_n$ with $w_i \in A^*$. Knast [9] has shown that a language L has dot-depth one if and only if $\mathrm{Synt}(L) \in \mathbf{B}_1$. Since the latter property is decidable, this yields decidability of dot-depth one. Later, Thérien [25] gave a simpler proof for Knast's result. Both proofs are based on an algebraic concept called *finite categories*, see [28]. In this section, we give a new (more combinatorial) proof of this theorem. The same techniques were used by the authors in order to obtain a characterization for languages of dot-depth one over infinite words [10]. As for dot-depth 1/2, the main advantage of the current proof is that the bounds involved are more explicit.

Theorem 4 (Knast [9], Thomas [26]) *Let $L \subseteq A^+$. The following are equivalent:*

1. *L is definable in* $\mathbb{B}\Sigma_1[<, +1, \min, \max]$.
2. *L is a Boolean combination of monomials $w_1 A^* w_2 \cdots A^* w_n$.*
3. $\mathrm{Synt}(L) \in \mathbf{B}_1$.

As for dot-depth 1/2, the equivalence of $\mathbb{B}\Sigma_1[<, +1, \min, \max]$ and dot-depth one is due to a result by Thomas [26]. The remainder of this section is devoted to the proof of the above theorem. The following lemma will serve as the link between the algebraic properties of $\mathbf{B}_1$ and the combinatorial properties in Lemma 9 below.

Lemma 8 *Let $S \in \mathbf{B}_1$, let $k \geq |S| + 1$, and let $a \in S$ and $u, v \in S^+$ with $|v| \geq k - 1$. If $u \mathrel{\mathcal{R}} uv >_{\mathcal{R}} uva$ in S, then* $\mathrm{alph}_k(v) \neq \mathrm{alph}_k(va)$.

Proof: Assume $u \mathrel{\mathcal{R}} uv >_{\mathcal{R}} uva$ and $\mathrm{alph}_k(v) = \mathrm{alph}_k(va)$. Let $va = v'wa$ with $|wa| = k$. Since $wa \in \mathrm{alph}_k(va) = \mathrm{alph}_k(v)$ we have $v = pwaq$ for some $p, q \in S^*$. Let $x = up$, $y = uv'$, and $wa = a_1 \cdots a_k$ for $a_i \in S$. By Lemma 1 there exist $i \in \{1, \ldots, |S|\}$ and $e \in E(S)$ such that $a_1 \cdots a_i = a_1 \cdots a_i e$ in S. In particular $i \leq k - 1$ and $x a_1 \cdots a_i \mathrel{\mathcal{R}} y a_1 \cdots a_i$. Lemma 3 yields $x a_1 \cdots a_i = y a_1 \cdots a_i$ in S. Thus $uva = ywa = xwa \mathrel{\mathcal{R}} u$ in S, a contradiction. □

Suppose $u = u_0 x u_1 = u'_0 y u'_1$ and $v = v_0 x v_1 = v'_0 y v'_1$ for words $x, y \in A^+$ and $u_i, u'_i, v_i, v'_i \in A^*$ for $i \in \{0, 1\}$. Let $\Delta_u = |u| - |u_0 u'_1|$ and $\Delta_v = |v| - |v_0 v'_1|$. We say that *the relative order of x and y is the same in u and v* if one of the following conditions applies:

- $\Delta_u > |xy|$ and $\Delta_v > |xy|$, *i.e.*, in each of the words u and v, all positions of x are on the left of all positions of y,
- $\Delta_u < 0$ and $\Delta_v < 0$, *i.e.*, in each of the words u and v, all positions of x are on the right of all positions of y,

- $\Delta_u = \Delta_v$, *i.e.*, if none of the previous conditions applies, then the occurrences of x and y have the same overlap in both words u and v.

The following lemma about relative orders is the main combinatorial ingredient for our proof of Knast's Theorem. It generalizes an idea of Klíma [8] to factors of words. The determinacy mechanism is similar to unambiguous interval logic with lookaround [11].

Lemma 9 *Let k, ℓ be positive integers, let $x_i, y_i, u_i, u'_i, v_i, v'_i \in A^+$ and $u_k, v_k, u'_1, v'_1 \in A^*$, and let*

$$u = x_1 u_1 \cdots x_k u_k = u'_1 y_1 \cdots u'_\ell y_\ell$$
$$v = x_1 v_1 \cdots x_k v_k = v'_1 y_1 \cdots v'_\ell y_\ell$$

such that $x_1 u_1 \cdots x_k$ (resp. $x_1 v_1 \cdots x_k$) is the shortest prefix of u (resp. v) in $x_1 A^+ x_2 \cdots A^+ x_k$, and $y_1 \cdots u'_\ell y_\ell$ (resp. $y_1 \cdots v'_\ell y_\ell$) is the shortest suffix of u (resp. v) in $y_1 A^+ y_2 \cdots A^+ y_\ell$. If u and v are contained in the same languages $w_1 A^+ w_2 \cdots A^+ w_n$ with $n \leq k + \ell$ and $|w_1 \cdots w_n| \leq |x_1 \cdots x_k \, y_1 \cdots y_\ell|$, then the relative orders of x_k and y_1 are the same in u and v.

Proof: Let $\Delta_u = |u| - |x_1 \cdots u_{k-1}| - |u'_2 \cdots y_\ell|$ and $\Delta_v = |v| - |x_1 \cdots v_{k-1}| - |v'_2 \cdots y_\ell|$. First suppose that $\Delta_u > |x_k y_1|$. Then $x_1 u_1 \cdots x_k$ is a proper prefix of u'_1. Thus $u \in x_1 A^+ \cdots x_k A^+ y_1 \cdots A^+ y_\ell$. This implies $v \in x_1 A^+ \cdots x_k A^+ y_1 \cdots A^+ y_\ell$ and $\Delta_v > |x_k y_1|$. By symmetry we conclude that $\Delta_u > |x_k y_1|$ if and only if $\Delta_v > |x_k y_1|$.

Let now $0 \leq \Delta_u \leq |x_k y_1|$. We can assume $\Delta_v \leq |x_k y_1|$. Now, u is contained in $P = x_1 A^+ \cdots x_i A^+ z A^+ y_j \cdots A^+ v_\ell$, where z is the factor of u comprising all $x_{i+1}, \ldots, x_k$ which are overlapping with (or adjacent to) y_1 and all $y_1, \ldots, y_{j-1}$ which are overlapping with (or adjacent to) x_k. Since $v \in P$, we conclude that x_k is not further to the right of y_1 in the word u as in the word v, *i.e.*, we have $\Delta_u \leq \Delta_v$. By symmetry, this shows that $0 \leq \Delta_u \leq |x_k y_1|$ if and only if $0 \leq \Delta_v \leq |x_k y_1|$. Moreover, if $0 \leq \Delta_u \leq |x_k y_1|$, then $\Delta_u = \Delta_v$.

By the above two cases, we also see that $\Delta_u < 0$ if and only if $\Delta_v < 0$. This shows that x_k and y_1 have the same relative order in u and v. $\square$

Lemma 10 *Let $S \in \mathbf{B}_1$ and let $u, v, x, s \in S$ and let $e, f \in E(S)$ be idempotent. If $u \mathrel{\mathcal{R}} uexf$ and $esfv \mathrel{\mathcal{L}} v$, then $uexfv = uesfv$.*

Proof: Since $u \mathrel{\mathcal{R}} uexf$ and $v \mathrel{\mathcal{L}} esfv$, there exist $y, t \in S^1$ with $u = uexfy$ and $v = tesfv$. In particular, $u = u(exfy)^\omega$ and $v = (tesf)^\omega v$. We conclude $uexfv = u(exfy)^\omega exf(tesf)^\omega v = u(exfy)^\omega esf(tesf)^\omega v = uesfv$, where the second equality uses $S \in \mathbf{B}_1$. $\square$

Proposition 2 *Let $L \subseteq A^+$ be recognized by a homomorphism $h : A^+ \to S$ with $S \in \mathbf{B}_1$ and let $u, v \in A^+$. If u and v are contained in the same languages $w_1 A^+ w_2 \cdots A^+ w_n$ with $n \leq 2\,|S|$ and $|w_1 \cdots w_n| \leq 4\,|S|^2 - 2\,|S|$, then $h(u) = h(v)$.*

Proof: This proof was inspired by Klíma's proof [8] of Simon's Theorem on piecewise-testable languages. The outline is as follows. We consider factorizations induced by the $\mathcal{R}$-factorization of u and the $\mathcal{L}$-factorization of v. Then we transfer the factorization of u to v and vice versa such that the respective orders of the factors in u and v are the same. Finally, we transform v into u by a sequence of h-invariant substitutions.

Consider the $\mathcal{R}$-factorization $u = a_1u_1 \cdots a_ku_k$ with $a_i \in A$ such that

$$h(a_1u_1 \cdots a_i) \mathrel{\mathcal{R}} h(a_1u_1 \cdots a_iu_i) >_{\mathcal{R}} h(a_1u_1 \cdots a_iu_ia_{i+1}) \qquad \text{for all } i.$$

We have $k \leq |S|$. Let j_i be the position of a_i in the above factorization. We color red all positions of u in all the intervals $[\, j_i - |S|\, ;\, j_i + |S| - 1\,]$. In particular, the a_i-positions j_i are red. Moreover in general, there is a neighborhood of size $2\,|S|$ around each a_i which contains only red positions. In the worst case, a_1 is the sole exception. Hence, there are at most $2\,|S|^2 - |S|$ red positions in u. Let R_i be the i-th consecutive factor consisting of red positions. Then $u = R_1u'_1 \cdots R_{k'}u'_{k'}$ for some $u'_i \in A^+$, $i < k'$, and $u'_{k'} \in A^*$. Note that $k' \leq k$ because some intervals could overlap. By Lemma 8, for each i the word $R_1u'_1 \cdots u'_{i-1}R_i$ is the shortest prefix of u contained in $R_1A^+ \cdots A^+R_i$.

Symmetrically, let $v = v_1b_1 \cdots v_\ell b_\ell$ with $b_i \in A$ be the $\mathcal{L}$-factorization such that

$$h(b_{i-1}v_ib_i \cdots v_\ell b_\ell) <_{\mathcal{L}} h(v_ib_i \cdots v_\ell b_\ell) \mathrel{\mathcal{L}} h(b_i \cdots v_\ell b_\ell) \qquad \text{for all } i.$$

Let j'_i be the position of b_i in the above factorization. We color blue all positions of v in all the intervals $[\, j'_i - |S| + 1;\, j'_i + |S|\,]$. As before, there are at most $2\,|S|^2 - |S|$ blue positions. Let B_i be the i-th consecutive factor of blue positions. Then $v = v'_1B_1 \cdots v'_{\ell'}B_{\ell'}$ for $\ell' \leq |S|$ and some $v'_i \in A^+$, $i > 1$, and $v'_1 \in A^*$. As before, $B_iv'_{i+1} \cdots v'_{\ell'}B_{\ell'}$ is the shortest suffix of v contained in $B_iA^+ \cdots A^+B_{\ell'}$.

Next, we transfer the red positions of u to v, and we transfer the blue positions of v to u. By assumption $v \in R_1A^+ \cdots R_{k'}A^+$ and therefore, there exists a factorization $v = R_1v''_1 \cdots R_{k'}v''_{k'}$ such that $R_1v''_1 \cdots v''_{i-1}R_i$ is the shortest prefix of v contained in $R_1A^+ \cdots A^+R_i$. We color the positions of the R_i's in v red. Similarly, there exists a factorization $u = u''_1B_1 \cdots u''_{\ell'}B_{\ell'}$ such that $B_iu''_{i+1} \cdots u''_{\ell'}B_{\ell'}$ is the shortest suffix of u contained in $B_iA^+ \cdots A^+B_{\ell'}$. We color the positions of the B_i's in u blue. Now, colored positions in u and v are either red or blue or both. By Lemma 9, the colored positions in u have the same order as the colored positions in v. Thus if w_i is the i-th consecutive factor of colored (red or blue) positions, then

$$\begin{aligned} u &= w_1x_1 \cdots w_{n-1}x_{n-1}w_n, \\ v &= w_1s_1 \cdots w_{n-1}s_{n-1}w_n. \end{aligned}$$

The next step is the construction of idempotent stabilizers near the beginning and near the end of each w_i. We do this from the inside to the outside by considering the first and the last $|S|$ letters in every word w_i: By Lemma 1 and its left-right dual, there exist idempotents $e_1, \ldots, e_{n-1} \in E(S)$ and $f_2, \ldots, f_n \in E(S)$ such that each w_i admits a factorization $w_i = p_ir'_ir''_iq_i$ with $|p_ir'_i| = |S|$, $|r'_i| \geq 1$ and $|q_i| \leq |S| - 1$ satisfying

$$\begin{aligned} h(r'_i) &= f_i\, h(r'_i) \qquad \text{for all } 1 < i \leq n, \\ h(r''_i) &= h(r''_i)\, e_i \qquad \text{for all } 1 \leq i < n. \end{aligned}$$

In particular, we can assume $p_1 = 1 = q_n$. Let $x'_i = q_ix_ip_{i+1}$ and $s'_i = q_is_ip_{i+1}$ for $1 \leq i < n$ and let $r_i = r'_ir''_i$ for $1 \leq i \leq n$. Then

$$\begin{aligned} u &= r_1x'_1r_2 \cdots x'_{n-1}r_n, \\ v &= r_1s'_1r_2 \cdots s'_{n-1}r_n. \end{aligned}$$

By construction, every position of the $\mathcal{R}$-factorization of u lies within some r_i''. We thus have $h(r_1 x_1' \cdots r_i) \; \mathcal{R} \; h(r_1 x_1' \cdots r_i x_i' r_{i+1}') = h(r_1 x_1' \cdots r_i) \cdot e_i h(x_i') f_{i+1} \cdot h(r_{i+1}')$ for all $1 \leq i < n$. Therefore, for all $1 \leq i < n$ we get

$$h(r_1 x_1' \cdots r_i) \;\; \mathcal{R} \;\; h(r_1 x_1' \cdots r_i) \cdot e_i h(x_i') f_{i+1}.$$

A symmetric argument shows

$$h(r_{i+1} \cdots s_n' r_n) \;\; \mathcal{L} \;\; e_i h(s_i') f_{i+1} \cdot h(r_{i+1} \cdots s_n' r_n).$$

By an $(n-1)$-fold application of Lemma 10 we obtain

$$\begin{aligned} h(v) &= h(r_1 s_1' r_2 s_2' r_3 \cdots s_{n-1}' r_n) \\ &= h(r_1 x_1' r_2 s_2' r_3 \cdots s_{n-1}' r_n) \\ &= h(r_1 x_1' r_2 x_2' r_3 \cdots s_{n-1}' r_n) \\ &\;\;\vdots \\ &= h(r_1 x_1' r_2 x_2' r_3 \cdots x_{n-1}' r_n) = h(u) \end{aligned}$$

Note that the substitution rules $s_i' \to x_i'$ are h-invariant in their respective contexts only when applied from left to right. $\square$

Corollary 2 *Let $L \subseteq A^+$ be recognized by a finite semigroup $S \in \mathbf{B}_1$ and let $u, v \in A^+$. If u and v are contained in the same monomials $w_1 A^* w_2 \cdots A^* w_n$ with $n \leq 2\,|S|$ and degree $|w_1 \cdots w_n| < 4\,|S|^2$, then $h(u) = h(v)$.*

Proof: Every monomial $w_1 A^+ \cdots w_{n-1} A^+ w_n$ is a finite union of monomials of the form $w_1 a_1 A^* \cdots w_{n-1} a_{n-1} A^* w_n$ for $a_1, \ldots, a_{n-1} \in A$. Therefore, the claim follows from Proposition 2. $\square$

Proof (Proof of Theorem 4): "1 $\Leftrightarrow$ 2": This follows from Theorem 1.

"2 $\Rightarrow$ 3": The syntactic semigroup of every monomial $w_1 A^* w_2 \cdots A^* w_n$ satisfies $x^\omega y x^\omega \leq x^\omega$ by Lemma 6, and by Lemma 2 it is in $\mathbf{B}_1$. Thus L is recognizable by a direct product $S \in \mathbf{B}_1$ of such semigroups. Since $\mathrm{Synt}(L)$ is a divisor of S, we see that $\mathrm{Synt}(L) \in \mathbf{B}_1$, *cf.* [13].

"3 $\Rightarrow$ 2": Let L be recognized by $h : A^+ \to S \in \mathbf{B}_1$. We write $u \equiv v$ if u and v are contained in the same monomials of the form $w_1 A^* w_2 \cdots A^* w_n$ of degree less than $4\,|S|^2$. We have $L = h^{-1}(P)$ for $P = h(L)$. Corollary 2 shows that every set $h^{-1}(p)$ is a union of $\equiv$-classes. Moreover, $\equiv$ has finite index since there are only finitely many monomials of bounded degree. Every $\equiv$-class is a finite Boolean combination of the required form by specifying which monomials hold and which do not hold. $\square$

6 Dot-Depth One without min or max

As for existential first-order logic, one cannot define min- or max-predicates in $\mathbb{B}\Sigma_1[<, +1]$. Therefore, the following inclusions hold:

$$
\begin{tikzcd}[column sep=small]
\mathbb{B}\Sigma_1[<] \arrow[dr, phantom, "\subsetneq"] & & \mathbb{B}\Sigma_1[<,+1,\min] \arrow[dr, phantom, "\subsetneq"] & \\
& \mathbb{B}\Sigma_1[<,+1] \arrow[ur, phantom, "\subsetneq"] \arrow[r, phantom, "\subsetneq"] & \mathbb{B}\Sigma_1[<,+1,\max] \arrow[r, phantom, "\subsetneq"] & \mathbb{B}\Sigma_1[<,+1,\min,\max] \\
\mathbb{B}\Sigma_1[+1] \arrow[ur, phantom, "\subsetneq"] \arrow[dr, phantom, "\subsetneq"] & & & \\
& \mathbb{B}\Sigma_1[+1,\min,\max] = \mathrm{FO}[+1] \arrow[uurr, phantom, "\subsetneq"] & &
\end{tikzcd}
$$

Simon's Theorem on piecewise testable languages [19] gives decidability of $\mathbb{B}\Sigma_1[<]$. An effective characterization of $\mathbb{B}\Sigma_1[+1]$ is due to Pin [15]. By a result of Thomas [26], we have $\mathbb{B}\Sigma_1[+1,\min,\max] = \mathrm{FO}[+1]$, *i.e.*, every language definable in full first-order logic with atomic predicates $x = y + 1$, $\lambda(x) = a$, and $x = y$ is already definable in the alternation-free fragment $\mathbb{B}\Sigma_1[+1,\min,\max]$. Moreover, this fragment coincides with the class of locally threshold testable languages, which is known to be decidable, see *e.g.* Theorem VI.3.1 in [23]. For the fragment $\mathbb{B}\Sigma_1[<,+1,\min,\max]$, decidability follows by Knast's Theorem [9], see Theorem 4. In this section, we give effective characterizations of the remaining fragments. Moreover, we obtain natural subclasses of dot-depth one for the languages definable by these fragments.

Lemma 11 *Let $P = w_1 A^* w_2 \cdots A^* w_n$ and let $uq \in P$. There exists a monomial $Q = v_1 A^* v_2 \cdots A^* v_\ell$ with $|v_1 \cdots v_\ell| \le |w_1 \cdots w_n|$ and $\ell \le n$ such that $u \in Q \subseteq Pq^{-1}$.*

Proof: Let $uq = w_1 s_1 w_2 \cdots s_{n-1} w_n$. First consider the case $u = w_1 s_1 \cdots w_{i-1} s_{i-1} v$, $w_i = vv'$ and $q = v' s_i w_{i+1} \cdots s_{n-1} w_n$ for some i. Setting $Q = w_1 A^* \cdots w_{i-1} A^* v$ yields the claim. In the other case we have $u = w_1 s_1 \cdots s_{i-1} w_i t$, $s_i = tt'$ and $q = t' w_{i+1} \cdots s_{n-1} w_n$ for some i. In this case, we set $Q = w_1 A^* w_2 \cdots w_i A^*$. $\square$

Apart from Theorem 4, the following lemma is the main ingredient in the proof of Theorem 5 below.

Lemma 12 *Let $h : A^+ \to S \in \mathbf{B}_1$ and let $u, v \in A^+$. If u and v are contained in the same monomials $w_1 A^* \cdots w_n A^*$ with $|w_1 \cdots w_n| < 8\,|S|^2$, then $h(u)\ \mathcal{R}\ h(v)$.*

Proof: We write $u \equiv_m v$ if u and v are contained in the same monomials $w_1 A^* w_2 \cdots A^* w_n$ of degree $|w_1 \cdots w_n| \le m$. Analogously, we write $u \sim_m v$ if u and v are contained in the same monomials $w_1 A^* \cdots w_n A^*$ of degree $|w_1 \cdots w_n| \le m$. If $u \equiv_m v$ for $m = 4\,|S|^2 - 1$, then by Corollary 2 we have $h(u) = h(v)$.

Let $u \sim_{2m} v$. We want to show $h(u)\ \mathcal{R}\ h(v)$. We can assume $|u|, |v| \ge 2m$ because otherwise $u = v$. Let $u = u'q$ with $|q| = m$. Consider the factorization $v = v'qx$ such that qx is the shortest suffix of v admitting q as a factor, *i.e.*, v is factorized at the last occurrence of q. This factorization exists since both u and v belong to $A^* q A^*$. We claim that $u \equiv_m v'q$ and therefore, $h(v) \le_{\mathcal{R}} h(v'q) = h(u)$. Symmetry then yields $h(u)\ \mathcal{R}\ h(v)$.

We now prove the claim. Let $P = w_1 A^* w_2 \cdots A^* w_n$ with $|w_1 \cdots w_n| \le m$. First suppose that $v'q \in P$. Then $v \in PA^*$ and $u \in PA^*$. Since w_n is a suffix of q, we conclude $u \in P$. Next, suppose $u \in P$. By Lemma 11, there exists a monomial $Q = v_1 A^* v_2 \cdots A^* v_\ell$ with $|v_1 \cdots v_\ell| \le |w_1 \cdots w_n|$ and $u' \in Q \subseteq Pq^{-1}$. Since $u'q \in QqA^*$ and the degree of the monomial QqA^* is at most $2m$, we obtain $v \in QqA^*$. By choice of x we have $v'q \in QqA^* \subseteq PA^*$. Since w_n is a suffix of q, we conclude $v'q \in w_1 A^* w_2 \cdots A^* w_n$. $\square$

Theorem 5 *Let $L \subseteq A^+$. The following assertions are equivalent:*

1. *L is definable in $\mathbb{B}\Sigma_1[<, +1, \min]$.*
2. *L is a Boolean combination of monomials $w_1 A^* \cdots w_n A^*$.*
3. *$\mathrm{Synt}(L) \in \mathbf{B}_1$ and the syntactic homomorphism $h_L : A^+ \to \mathrm{Synt}(L)$ has the property that $h_L(L)$ is a union of $\mathcal{R}$-classes.*

Proof: The equivalence “1 $\Leftrightarrow$ 2” follows from Theorem 2.

“2 $\Rightarrow$ 3”: We have $\mathrm{Synt}(L) \in \mathbf{B}_1$ by Theorem 4. The set $h_L(w_1 A^* \cdots w_n A^*)$ is a right ideal. Hence, $h_L(L)$ is a Boolean combination of right ideals. The claim follows since every Boolean combination of right ideals is a union of $\mathcal{R}$-classes.

“3 $\Rightarrow$ 2”: By Lemma 12, there exists $m \in \mathbb{N}$ such that $h_L(u) \mathrel{\mathcal{R}} h_L(v)$ if u and v are contained in the same languages of the form $w_1 A^* \cdots w_n A^*$ with $|w_1 \cdots w_n| \leq m$. Therefore, for each $\mathcal{R}$-class R of $\mathrm{Synt}_L(L)$, the language $h_L^{-1}(R)$ is a Boolean combination of languages $w_1 A^* \cdots w_n A^*$ with $|w_1 \cdots w_n| \leq m$. The claim follows, since L is a union of languages of the form $h_L^{-1}(R)$. ☐

There is also a left-right dual of the above theorem: A language L is definable in $\mathbb{B}\Sigma_1[<, +1, \max]$ if and only if L is a Boolean combination of monomials $A^* w_1 \cdots A^* w_n$ if and only if $\mathrm{Synt}(L) \in \mathbf{B}_1$ and $h_L(L)$ is a union of $\mathcal{L}$-classes. Next, we consider the fragment $\mathbb{B}\Sigma_1[<, +1]$ with neither min nor max.

Lemma 13 *Let $h : A^+ \to S \in \mathbf{B}_1$ and let $u, v \in A^+$. If u and v are contained in the same monomials $A^* w_1 A^* \cdots w_n A^*$ with $|w_1 \cdots w_n| < 12\,|S|^2$, then $h(u) \mathrel{\mathcal{J}} h(v)$.*

Proof: The proof is along the same lines as Lemma 12. The major difference is that we need to consider the factorization $u = pu'q$ with $|p| = |q| = m$ as well as the factorization $v = spv'qx$ such that sp is the shortest prefix of v admitting p as a factor and qx is the shortest suffix of v admitting q as a factor, *i.e.*, v is factorized at the first occurrence of p and the last occurrence of q. ☐

Theorem 6 *Let $L \subseteq A^+$. The following assertions are equivalent:*

1. *L is definable in $\mathbb{B}\Sigma_1[<, +1]$.*
2. *L is a Boolean combination of monomials $A^* w_1 \cdots A^* w_n A^*$.*
3. *$\mathrm{Synt}(L) \in \mathbf{B}_1$ and the syntactic homomorphism $h_L : A^+ \to \mathrm{Synt}(L)$ has the property that $h_L(L)$ is a union of $\mathcal{J}$-classes.*

Proof: The proof is similar to Theorem 5 with “3 $\Rightarrow$ 2” relying on Lemma 13. ☐

The condition of $h_L(L)$ being a union of $\mathcal{J}$-classes in Theorem 6 has also been used by Beauquier and Pin for an effective characterization of strongly locally testable languages [1].

The above characterizations immediately give the following decidability result.

Corollary 3 *Let $L \subseteq A^+$ be a regular language. It is decidable whether L is definable in $\mathbb{B}\Sigma_1[<, +1]$ (resp. $\mathbb{B}\Sigma_1[<, +1, \min]$, resp. $\mathbb{B}\Sigma_1[<, +1, \max]$).*

Proof: The syntactic homomorphism $h_L : A^+ \to \mathrm{Synt}(L)$ of L is effectively computable. Hence, one can verify whether property “3” in Theorem 6 (resp. “3” in Theorem 5, resp. the left-right dual of “3” in Theorem 5) holds. ☐

Table 1: Languages around dot-depth one.

Languages	**Logics**	**Algebra**	**Reference**
$\bigcup w_1 A^* w_2 \cdots A^* w_n$	$\Sigma_1[<, +1, \min, \max]$	$\mathbf{B}_{1/2}$	[16], Thm. 1
$\bigcup w_1 A^* \cdots w_n A^*$	$\Sigma_1[<, +1, \min]$	right ideals in $\mathbf{B}_{1/2}$	Thm. 2
$\bigcup A^* w_1 \cdots A^* w_n$	$\Sigma_1[<, +1, \max]$	left ideals in $\mathbf{B}_{1/2}$	*cf.* Thm. 2
$\bigcup A^* w_1 \cdots A^* w_n A^*$	$\Sigma_1[<, +1]$	ideals in $\mathbf{B}_{1/2}$	Thm. 3
$\mathbb{B}(w_1 A^* w_2 \cdots A^* w_n)$	$\mathbb{B}\Sigma_1[<, +1, \min, \max]$	$\mathbf{B}_1$	[9], Thm. 4
$\mathbb{B}(w_1 A^* \cdots w_n A^*)$	$\mathbb{B}\Sigma_1[<, +1, \min]$	$\mathcal{R}$-classes in $\mathbf{B}_1$	Thm. 5
$\mathbb{B}(A^* w_1 \cdots A^* w_n)$	$\mathbb{B}\Sigma_1[<, +1, \max]$	$\mathcal{L}$-classes in $\mathbf{B}_1$	*cf.* Thm. 5
$\mathbb{B}(A^* w_1 \cdots A^* w_n A^*)$	$\mathbb{B}\Sigma_1[<, +1]$	$\mathcal{J}$-classes in $\mathbf{B}_1$	Thm. 6

7 Summary

We considered subclasses of languages with dot-depth 1/2 and of languages with dot-depth one. These subclasses admit counterparts in terms of fragments of existential first-order logic Σ_1 and its Boolean closure $\mathbb{B}\Sigma_1$. For all fragments, we gave effective algebraic characterizations. We summarize the main results of this paper in Table 1. To shorten notation, we write $\mathbf{B}_{1/2}$ instead of $[\![x^\omega y x^\omega \leq x^\omega]\!]$.

In addition, we gave new proofs for Pin and Weil's Theorem on dot-depth 1/2 and for Knast's Theorem on dot-depth one. Our proofs improve the bounds involved in computing a language description for a given recognizing semigroup.

Acknowledgments

We thank the anonymous referees for several suggestions which helped to improve the presentation of the paper, and we gratefully acknowledge the support by the German Research Foundation (DFG) under grant DI 435/5-1.

References

[1] D. Beauquier and J.-É. Pin. Languages and scanners. *Theor. Comput. Sci.*, 84(1):3–21, 1991.

[2] 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.

[3] R. S. Cohen and J. A. Brzozowski. Dot-depth of star-free events. *J. Comput. Syst. Sci.*, 5(1):1–16, 1971.

[4] Z. Ésik and K. G. Larsen. Regular languages definable by Lindström quantifiers. *RAIRO, Inform. Théor. Appl.*, 37(3):179–241, 2003.

[5] Ch. Glaßer and H. Schmitz. Languages of dot-depth 3/2. In *STACS'00, Proceedings*, volume 1770 of *LNCS*, pages 555–566. Springer, 2000.

[6] Ch. Glaßer and H. Schmitz. Languages of dot-depth 3/2. *Theory Comput. Syst.*, 42(2):256–286, 2008.

[7] J. Kallas, M. Kufleitner, and A. Lauser. First-order fragments with successor over infinite words. In *STACS'11, Proceedings*, volume 9 of *LIPIcs*, pages 356–367. Dagstuhl Publishing, 2011.

[8] O. Klíma. Piecewise testable languages via combinatorics on words. *Discrete Mathematics*, 311(20):2124–2127, 2011.

[9] R. Knast. A semigroup characterization of dot-depth one languages. *RAIRO, Inf. Théor.*, 17(4):321–330, 1983.

[10] M. Kufleitner and A. Lauser. Languages of dot-depth one over infinite words. In *LICS'11, Proceedings*, pages 23–32. IEEE Computer Society, 2011.

[11] K. Lodaya, P. K. Pandya, and S. S. Shah. Around dot depth two. In *DLT'10, Proceedings*, volume 6224 of *LNCS*, pages 305–316. Springer, 2010.

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

[13] J.-É. Pin. *Varieties of Formal Languages*. North Oxford Academic, London, 1986.

[14] J.-É. Pin. A variety theorem without complementation. In *Russian Mathematics (Iz. VUZ)*, volume 39, pages 80–90, 1995.

[15] J.-É. Pin. Expressive power of existential first-order sentences of Büchi's sequential calculus. *Discrete Math.*, 291(1-3):155–174, 2005.

[16] J.-É. Pin and P. Weil. Polynomial closure and unambiguous product. *Theory Comput. Syst.*, 30(4):383–422, 1997.

[17] J.-É. Pin and P. Weil. The wreath product principle for ordered semigroups. *Commun. Algebra*, 30(12):5677–5713, 2002.

[18] M. P. Schützenberger. On finite monoids having only trivial subgroups. *Inf. Control*, 8:190–194, 1965.

[19] I. Simon. Piecewise testable events. In *Autom. Theor. Form. Lang., 2nd GI Conf.*, volume 33 of *LNCS*, pages 214–222. Springer, 1975.

[20] I. Simon. Factorization forests of finite height. *Theor. Comput. Sci.*, 72(1):65–94, 1990.

[21] H. Straubing. A generalization of the Schützenberger product of finite monoids. *Theor. Comput. Sci.*, 13:137–150, 1981.

[22] H. Straubing. Finite semigroup varieties of the form $\mathbf{V} * \mathbf{D}$. *J. Pure Appl. Algebra*, 36(1):53–94, 1985.

[23] H. Straubing. *Finite Automata, Formal Logic, and Circuit Complexity*. Birkhäuser, 1994.

[24] D. Thérien. Classification of finite monoids: The language approach. *Theor. Comput. Sci.*, 14(2):195–208, 1981.

[25] D. Thérien. Categories et langages de dot-depth un. *RAIRO, Inf. Théor.*, 22(4):437–445, 1988.

[26] W. Thomas. Classifying regular events in symbolic logic. *J. Comput. Syst. Sci.*, 25:360–376, 1982.

[27] W. Thomas. Languages, automata and logic. In A. Salomaa and G. Rozenberg, editors, *Handbook of Formal Languages*, volume 3, Beyond Words. Springer, Berlin, 1997.

[28] B. Tilson. Categories as algebras: An essential ingredient in the theory of monoids. *J. Pure Appl. Algebra*, 48:83–198, 1987.