Around Dot-depth One*

Manfred Kufleitner Alexander Lauser

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

March 7, 2011

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 fragments 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 existential first-order logic without min- or max-predicate. For all fragments, we obtain effective algebraic characterizations. Moreover, we give new combinatorial proofs for the decidability of the membership problem for dot-depth 1/2 and dot-depth 1.

1 Introduction

The dot-depth hierarchy $B_n$ for $n \in \mathbb{N} + \{1/2, 1\}$ has been introduced by Cohen and Brzozowski [2]. A very similar hierarchy is the Straubing-Thérien hierarchy $L_n$, see [16, 18]. Both hierarchies are strict [4] and they are exhausting 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 [9]. Thomas [20] 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 [13]. The latter property is effectively 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 $B_n$ is decidable if and only if membership in $L_n$ is decidable [17]. This result has been extended to the half-levels by Pin and Weil [12]. Simon has shown that the class of piecewise testable languages $L_1$ is decidable [14]. Later, Knast [6] gave an effective algebraic characterization of $B_1$. Decidability of $L_{1/2}$ was shown by Pin [10], and the levels $B_{1/2}$ and $L_{3/2}$ are decidable by

*Supported by the German Research Foundation (DFG), project DI 435/5-1.
a result of Pin and Weil \cite{11}. The most recent decidability result is for $B_{3/2}$ due to Glaßer and Schmitz \cite{4}. To date, no other levels are known to be decidable.

In this paper, we focus on subclasses of $B_{1/2}$ and $B_1$. For both $B_{1/2}$ and $B_1$ we give new proofs for their effective algebraic characterizations. The proof of Pin and Weil \cite{11} for $B_{1/2}$ is based on factorization forests \cite{15}, and the proof of Knast \cite{6} as well as the simplified version of Thérien \cite{19} for $B_1$ are based on a generalization of finite monoids, so-called finite categories \cite{21}. Our proofs are more combinatorial than algebraic. The proof for $B_1$ is a generalization of Klima’s proof \cite{5} for $L_1$. The main advantage of our proofs for $B_{1/2}$ and $B_1$ over previous ones is that the constants involved in finding language descriptions for given algebraic objects are more explicit (and therefore smaller).

Our main original contributions are effective algebraic characterizations of fragments of existential first-order logic over the signatures $[<,+1,\text{min}]$ without max-predicate, $[<,+1,\text{max}]$ without min, and $[<,+1]$ without min and max. These fragments also admit language characterizations in terms of subclasses of $B_{1/2}$ and $B_1$. The corresponding language classes are obtained by prohibiting the specification of prefixes or suffixes. In contrast to $B_{1/2}$ and $B_1$, the resulting subclasses do not form (positive) varieties of languages, but they still can be described using so-called lattice equations \cite{3}. Moreover, there is a tight connection with Cantor topologies over finite words \cite{7}. A more detailed overview of our results can be found in Section 7.

2 Preliminaries

Words and Languages Let $\Gamma$ be a finite non-empty alphabet. The set of finite words is $\Gamma^*$. By $\varepsilon$ we denote the empty word and $\Gamma^+ = \Gamma^* \setminus \{\varepsilon\}$ is the set of finite non-empty words. The length of a word $u \in \Gamma^*$ is $|u|$ and its alphabet is $\text{alph}(u) = \{a \in \Gamma \mid u \in \Gamma^* a \Gamma^*\}$. Similarly, $\text{alph}_k(u) = \{v \in \Gamma^k \mid u \in \Gamma^* v \Gamma^*\}$ is the set of all factors of $u$ of length $k$. A word $v \in \Gamma^*$ is a prefix (resp. suffix, resp. factor) of $u$ if $v \in \Gamma^* u$ (resp. $u \in \Gamma^* v$, resp. $u \in \Gamma^* v \Gamma^*$). We write $v \leq_p u$ if $v$ is a prefix of $u$ and $v <_p u$ if $v$ is a proper prefix of $u$. A quotient of $L \subseteq \Gamma^+$ is a language of the form $u^{-1} L = \{v \in \Gamma^+ \mid uv \in L\}$ or $Lu^{-1} = \{v \in \Gamma^+ \mid vu \in L\}$ for $u \in \Gamma^*$. A language $L$ is a monomial (of degree $m$) if $L = w_1 \Gamma^* w_2 \cdots \Gamma^* w_n$ or $L = w_1 \Gamma^* w_2 \cdots \Gamma^* w_n$ for some $n \geq 0$ and $w_1, \ldots, w_n \in \Gamma^*$ (with $|w_1 \cdots w_n| \leq m$). 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 $\text{FO} = \text{FO}[<,+1,\text{min},\text{max}]$ over finite words. We view words as sequences of labeled positions which are linearly ordered by $\prec$. 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 $\text{min}(x)$ (resp. $\text{max}(x)$) 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 $a \in \Gamma$. Formulas can be composed using Boolean operations, existential quantification, and universal quantification. The semantics is as usual. A sentence is a formula without free variables. For a sentence $\varphi$ of $\text{FO}$ we write $u \models \varphi$ if $u$ is a model of $\varphi$ and the language defined by $\varphi$ is $L(\varphi) = \{u \in \Gamma^+ \mid u \models \varphi\}$. 


2
The fragment $\Sigma_1$ consists of all FO-formulas in prenex normal form with only one block of quantifiers and these quantifiers are existential. Let $C \subseteq \{<,+1, \min, \max\}$. By $\Sigma_1[C]$ we denote the class of formulas in $\Sigma_1$ which only use predicates in $C$, equality, and the label predicate. The fragment $\exists \Sigma_1[C]$ comprises all Boolean combinations of formulas in $\Sigma_1[C]$.

**Finite Semigroups and Recognizable Languages**

Let $S$ be a semigroup. An element $x \in S$ is idempotent if $x^2 = x$. The set of idempotents is denoted by $E(S) = \{ e \in S \mid e^2 = e \}$. For every finite semigroup $S$ there exists a number $\omega \geq 1$ such that for every $x \in S$, the power $x^\omega$ 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 $u = \varepsilon = v$ or $u, v \in S^+$ evaluate to the same element of $S$.

**Lemma 1.** Let $S$ be a finite semigroup. For every word $u \in S^+$ with length $|u| = |S|$ there exists a non-empty prefix $p$ of $u$ and an idempotent $e \in E(S)$ such that $pe = p$ in $S$.

**Proof.** Let $a \in S$ be arbitrary and let $p_1 <_p \cdots <_p p_{|S|} <_p p_{|S|+1} = ua$ be the non-empty prefixes of $ua$. By the pigeonhole principle, there exist $1 \leq i < j \leq |S| + 1$ such that $p_i = p_j$ in $S$. In particular, $i \leq |S|$ and $p_i$ is a prefix of $u$. Let $p_iq = p_j$ for $q \in S^+$. We set $e = q^\omega$ to be the idempotent element generated by $q$. Now, $pe = p$ in $S$ for $p = p_i$. \hfill $\square$

Green’s relations are an important tool in the study of semigroups. They are defined as follows. Let $x \leq_R y$ (resp. $x \leq_L y$, resp. $x \leq_J y$) if there exist $s, t \in S \cup \{1\}$ such that $x = yt$ (resp. $x = sy$, resp. $x = syt$). Let $x \leq_R y$ (resp. $x \leq_L y$, resp. $x \leq_J y$) if $x \leq_R y$ and $y \leq_R x$ (resp. $x \leq_L y$ and $y \leq_L x$, resp $x \leq_J y$ and $y \leq_J x$). Here, $S \cup \{1\}$ is the monoid obtained by adding a new neutral element 1 to the semigroup $S$. The relations $\leq_R$, $\leq_L$, and $\leq_J$ are preorders on $S$; and $\leq_R$, $\leq_L$, and $\leq_J$ form equivalence relations.

Let $\leq$ be a preorder on $S$. A set $P \subseteq S$ is a $\leq$-order ideal if $x \leq y \in P$ implies $x \in P$. A subset $P \subseteq S$ is called $\leq$-covered if $x \in S$ and $x \leq y \in P$ for some $y \in P$. Any $\leq$-covered subset $P$ of $S$ 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 $L \subseteq \Gamma^+$ is recognized by an ordered semigroup $S$ if there exists a homomorphism $h : \Gamma^+ \rightarrow S$ 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 $L \subseteq \Gamma^+$ the syntactic preorder $\preceq_L$ over $\Gamma^+$ is given by $x \preceq_L y$ if $uv \in L$ implies $u \preceq_L v$. 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 of the syntactic congruence equipped with the canonical composition constitutes the syntactic semigroup $\text{Synt}(L)$ and the preorder $\preceq_L$ of $\Gamma^+$ becomes a compatible partial order for $\text{Synt}(L)$. The syntactic semigroup of $L$ is finite if and only if $L$ is regular and moreover, $L$ is recognized by its syntactic semigroup. By $[x^\omega y x^\omega \preceq x^\omega]$ we denote the class of finite ordered semigroups $S$ such that $x^\omega y x^\omega \preceq x^\omega$ for all elements $x, y \in S$. We let $B_1$ be the class of finite semigroups $S$ such that $(exf)^\omega esf(tesf)^\omega = (exf)^\omega esf(tesf)^\omega$ for all idempotents $e, f \in E(S)$ and all elements $x, y, t, s \in S$. Let $LR$ be the class of finite semigroups $S$ such that $(exye)^\omega exh = (exye)^\omega$ for all idempotents $e \in E(S)$ and all elements $x, y \in S$. We have the following inclusions among these classes of semigroups.

**Lemma 2.** We have $[x^\omega y x^\omega \preceq x^\omega] \subseteq B_1 \subseteq LR$. 

3
Proof. For a semigroup \( S \subseteq \Gamma^+ \) we have \( f \geq f(yfx)\omega^{-1}esf \) for all \( x, y, s \in S \) and all idempotents \( e, f \in S \). Hence \( (exy)f(ef(tesf))^\omega \geq (exy)e(xfyf(yfx)\omega^{-1}esf)(tesf)^\omega = (exy)e(ef(tesf))^\omega \). By symmetry \( (exy)e(ef(tesf))^\omega \leq (exy)e(ef(tesf))^\omega \) proving the first inclusion.

We have \( (exy)e = (exy)e(ex(eee))^\omega \) for all \( x \) and all idempotents \( e \) and for a semigroup in \( B_1 \) this is equal to \( (exy)e(eee)(eee)^\omega = (eeye)^\omega \). This shows the second inclusion. \( \square \)

3 Dot-depth 1/2

A language \( L \subseteq \Gamma^+ \) has dot-depth 1/2 if it is a positive Boolean combination of monomials \( w_1 \Gamma^*w_2 \cdots \Gamma^*w_n \) with \( w_i \in \Gamma^* \). By a result of Thomas [20], 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 [11] have shown that \( L \) has dot-depth 1/2 if and only if \( \text{Synt}(L) \in [x^\omega yx^\omega \leq x^\omega] \). In this section, we give a new proof of these equivalences. The main step in the proof is to show that if \( L \) is recognized by some homomorphism \( h : \Gamma^+ \rightarrow S \in [x^\omega yx^\omega \leq x^\omega] \), then \( L \) is a union of monomials \( w_1 \Gamma^*w_2 \cdots \Gamma^*w_n \). The main advantage of our proof is that the degree \( |w_1 \cdots w_n| \) is polynomially bounded (Proposition 9), whereas in the proof of Pin and Weil, the bound is exponential.

**Theorem 3** (Thomas [20], Pin/Weil [11]). Let \( L \subseteq \Gamma^+ \). The following assertions are equivalent:

1. \( L \) is definable in \( \Sigma_1[<,+1,\min,\max] \).
2. \( L \) is a finite union of monomials \( w_1 \Gamma^*w_2 \cdots \Gamma^*w_n \).
3. \( L \) is a positive Boolean combination of monomials \( w_1 \Gamma^*w_2 \cdots \Gamma^*w_n \).
4. \( \text{Synt}(L) \in [x^\omega yx^\omega \leq x^\omega] \).
5. There exists a homomorphism \( h : \Gamma^+ \rightarrow S \) with \( S \in [x^\omega yx^\omega \leq x^\omega] \) such that \( L = h^{-1}(P) \) for some \( \leq \)-order ideal \( P \).

In the remainder of this section we prove the above theorem.

**Lemma 4.** Let \( n \geq 0 \), and let \( w_1, \ldots, w_n \in \Gamma^* \).

1. The monomial \( w_1 \Gamma^*w_2 \cdots \Gamma^*w_n \) is definable in \( \Sigma_1[<,+1,\min,\max] \).
2. The monomial \( w_1 \Gamma^*w_2 \cdots \Gamma^*w_n \Gamma^* \) is definable in \( \Sigma_1[<,+1,\min] \).
3. The monomial \( \Gamma^*w_1 \Gamma^*w_2 \cdots \Gamma^*w_n \Gamma^* \) is definable in \( \Sigma_1[<,+1] \).

**Proof.** The proof is straightforward. For variable vectors \( x = (x_1, \ldots, x_\ell) \) and \( y = (y_1, \ldots, y_m) \) we use the shortcuts \( \exists x \) for \( \exists x_1 \cdots \exists x_\ell \), and \( \min(x) \) for \( \min(x_1) \) and \( \max(x) \) for \( \max(x_\ell) \), and \( x < y \) means \( x_\ell < y_1 \). Moreover, \( \lambda(x) = a_1 \cdots a_k \) is a shortcut for

\[
\bigwedge_{1 \leq j \leq k} \lambda(x_j) = a_j \land \bigwedge_{1 \leq j < k} x_{j+1} = x_j + 1.
\]

Let \( L = \Gamma^*w_1 \Gamma^*w_2 \cdots \Gamma^*w_n \Gamma^* \). We introduce variable vectors \( x_i = (x_{i,1}, \ldots, x_{i,w_i}) \) for every \( i \in \{1, \ldots, n\} \). Then, \( L \) is defined by the following sentence \( \varphi \):

\[
\exists x_1 \cdots \exists x_n : \bigwedge_{1 \leq i \leq n} \lambda(x_i) = w_i \land \bigwedge_{1 \leq i < n} x_i < x_{i+1}
\]
The first term of the conjunction ensures that each \( x_i \) corresponds to a factor \( w_i \), whereas the second term ensures that the factors \( w_i \) occur in the correct order. The sentence for \( w_1 \Gamma^+w_2 \cdots \Gamma^*w_n \Gamma^+ \) is \( \varphi \land \min(x_1) \) and the sentence for \( w_1 \Gamma^+w_2 \cdots \Gamma^*w_n \) is \( \varphi \land \min(x_1) \land \max(x_n) \).

**Lemma 5.** Let \( L \subseteq \Gamma^+ \) be definable by a sentence \( \varphi \in \Sigma[1]<+,1,\min,\max] \) with \( m \) variables. Then \( L \) is a finite union of monomials \( w_1 \Gamma^+w_2 \cdots \Gamma^*w_n \) of degree at most \( m \).

**Proof.** Let \( \varphi = \exists x_1 \cdots \cdot x_m : \psi \) for a propositional formula \( \psi \). Suppose \( (u,x_1,\ldots,x_m) \models \psi \) for positions \( x_i \) of \( u \). We say that a position \( j \) of \( u \) is marked if \( j = x_i \) for some \( i \). In order to avoid case distinctions we can introduce two new variables such that the first and the last position of \( u \) are marked. Let \( u = w_1u_1w_2 \cdots w_{n-1}w_n \) for \( u_i \in \Gamma^+ \) such that the factors \( w_i \) consist of the marked positions. Now, \( P_u = w_1 \Gamma^+w_2 \cdots \Gamma^*w_n \) is a monomial of degree \( |w_1 \cdots w_n| \leq m \) with \( u \in P_u \). Moreover, \( P_u \subseteq L(\varphi) \) since the satisfying assignment of \( u \) can be adapted to all \( v \in P_u \). It follows \( L(\varphi) = \bigcup_{u_i \models \varphi} P_u \) and this union is finite since there are only finitely many monomials of degree at most \( m \).

**Lemma 6.** Let \( L \subseteq \Gamma^+ \) be a finite union of monomials \( w_1 \Gamma^+w_2 \cdots \Gamma^*w_n \). Then \( \text{Synt}(L) \in [x^\prime x^\prime \leq x^\prime] \).

**Proof.** Let \( P = w_1 \Gamma^+w_2 \cdots \Gamma^*w_n \) and let \( u, x, y, v \in \Gamma^+ \) and choose \( m \) such that \( |x^m| > |w_1 \cdots w_n| \). Suppose \( ux^m v \in P \). Let \( i \) be maximal such that \( ux^m \in w_1 \Gamma^+w_2 \cdots \Gamma^*w_i \Gamma^+ = Q_i \) and let \( j \) be minimal such that \( x^m v \in \Gamma^*w_j \cdots \Gamma^*w_n = R_j \). By the choice of \( m \) we have \( j \leq i + 1 \). Hence, \( ux^m yx^m v \in Q_i R_j \subseteq P \).

**Lemma 7.** Let \( S \) be a finite semigroup. For every \( w \in S^+ \) there exists a factorization \( w = x_1w_1y_1 \cdots x_mw_my_m s \) with

1. \( w_i, s \in S^* \), \( x_i, y_i \in S^+ \), \( |y_i| \leq |S| \),
2. \( 0 \leq m \leq |S| \) and \( |x_1y_1 \cdots x_my_ms| < 2 |S|^2 + |S| \),
3. \( \forall i \in \{1, \ldots, m\} \exists e_i \in E(S): x_i = x_i e_i \text{ in } S \) and \( y_i = y_i e_i \text{ in } S \).

**Proof.** For \( w \in S^+ \), let \( E(w) \) be the set all \( e \in E(S) \) such that there exists a factor \( x \in S^+ \) of \( w \) with \( |x| \leq |S| \) and \( x = x \) in \( S \). We prove the existence of the factorization by induction on \( |E(w)| \) with the stronger assertions that \( m \leq |E(w)| \) and \( |x_1y_1 \cdots x_my_ms| < 2 |S| |E(w)| + |S| \) instead of condition 3. Suppose \( |E(w)| = 0 \). By Lemma 4, we have \( |w| < |S| \). Hence, we can choose \( m = 0 \) and \( s = w \).

If \( |E(w)| \geq 1 \), then Lemma 1 yields a non-empty prefix \( x \) of \( w \) with \( |x| \leq |S| \) such that \( x \neq x \) is in \( S \) for some idempotent \( e \in E(S) \). Write \( w = x w' \). We have to distinguish two cases. The first case is \( e \notin E(w') \). By induction, there exists a factorization \( w' = x_1w_1y_1 \cdots x_my_my_ms \) with \( m \leq |E(w')| < |E(w)| \) and \( |x_1y_1 \cdots x_my_ms| \leq 2 |S| |E(w')| + |S| \) satisfying conditions 1 and 3. If \( m \geq 1 \), then we set \( x_1 = x_1x_1 \). Now, \( w = x_1w_1y_1 \cdots x_my_my_ms \) is a desired factorization of \( w \). If \( m = 0 \), then the factorization for \( w \) is \( w = s \) with \( m = 0 \).

The second case is \( e \in E(w') \). Let \( w' = w_0y_0w'' \) such that \( y_0 \in S^+ \), \( |y_0| \leq |S| \), \( y_0e = y_0 \) in \( S \) and \( e \notin E(w'') \), i.e., we take \( y_0 \) as the last short factor of \( w' \) such that it is stabilized by \( e \). By induction, there exists a factorization \( w'' = x_1w_1y_1 \cdots x_my_my_ms \). Now, \( w = x_0w_0y_0 \cdots x_my_my_ms \) with \( x_0 = x \) is a factorization of \( w \) of the desired form.
Lemma 8. Let $S \subseteq \text{LR}$ be a finite semigroup. Let $u, x \in S^+$ and $e \in S$ be idempotent such that $u = ux$ and $x = xe$ in $S$. If $ux \not\subseteq u$, then $ux = u$ in $S$.

Proof. Let $y \in S^*$ such that $uxy = u$. In $S$ we have $u = u(exye)^e = u(exye)^eex = ux$ where the second equality follows because $S \subseteq \text{LR}$. 

Proposition 9. Let $L \subseteq \Gamma^+$ be recognized by $S \in \{x^\omega yx^\omega \leq x^\omega\}$. Then $L$ is a union of monomials $w_1 \Gamma^* w_2 \cdots \Gamma^* w_n$ of degree $|w_1w_2\cdots w_n| < 2|S|^3 + |S|^2$ and $n \leq |S|^2$.

Proof. Let $h : \Gamma^+ \rightarrow S$ be a homomorphism recognizing $L$. We define the depth of the word $u \in \Gamma^+$ as $d(u) = |\{s \in S \mid h(u) \not\subseteq _R s\}|$. For each $u \in \Gamma^+$ we construct a language $P_u = w_1 \Gamma^* w_2 \cdots \Gamma^* w_n$ with $|w_1w_2\cdots w_n| < 2d(u)|S|^2 + d(u)|S|$ such that $u \in P_u \subseteq h^{-1}(\{h(u)\})$.

In order to avoid unnecessary case distinctions, we set $P_e = \varepsilon$ and $h(\varepsilon) \supseteq_R h(u)$ for all $u \in \Gamma^+$. Let $u = vw$, $v \in \Gamma^*$, $w \in a\Gamma^*$ such that $h(v) \supseteq_R h(va) \supseteq_R h(u)$. Now, $d(v) < d(u)$ and hence by induction, there exists a monomial $P_v$ with $v \in P_v \subseteq h^{-1}(\{h(v)\})$ of degree less than $2d(u)|S|^2 + d(u)|S| - 2|S|^2 - |S|$. By Lemma 7 we find a factorization $w = x_1u_1y_1 \cdots x_mu_my_m$ such that $|x_1y_1 \cdots x_my_m| < 2|S|^2 + |S|$ 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 8 we see $h(u) = h(vw) = h(vx_1 \cdots x_m)$. Now, define the monomial $P_u = P_v x_1 \Gamma^* y_1 x_2 \Gamma^* \cdots y_m \Gamma^* y_m$ of degree less than $2d(u)|S|^2 + d(u)|S|$. By construction $u \in P_u$. Consider $v'w' \in P_u$ with $v' \in P_v$ and $w' = x_1w' y_1 x_2 w' \cdots y_m x_m w'_m$. We have $h(v') \leq h(v)$ and since $es e \leq e$ for all $s \in S$ and all $e \in E(S)$ we see that $h(x_i) = h(x_i)e_i \geq h(x_i)e_i h(w'_i y_i)e_i = h(x_i w'_i y_i)$. Therefore, $h(x_1 \cdots x_m) \geq h(w')$ and $u = h(vx_1 \cdots x_m) \geq h(v'w')$.

Now with the above properties, $L \subseteq \cup_{u \in L} P_u \subseteq L$ and this union is finite since there are only finite many monomials of degree less than $2|S|^3 + |S|^2$.

We are now ready to prove Theorem 3.

Proof (Theorem 3). $[1] \Rightarrow [2]:$ This is Lemma 5. $[2] \Rightarrow [3]$ follows from property $[1]$ of Lemma 4 and the fact that $\Sigma_1[<, +1, \min, \max]$ is closed under disjunction.


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 predicate (apart from labels $\lambda(x) = a$ for $a \in \Gamma$). At the level $\Sigma_1$, the min-predicate is required to determine prefixes, and max is required for suffixes. We have the following inclusions:

\[
\begin{align*}
\Sigma_1[<, +1, \min] & \subsetneq \Sigma_1[<, +1, \min, \max] \\
\Sigma_1[<] & \subsetneq \Sigma_1[<, +1] \subsetneq \Sigma_1[<, +1, \min, \max] \\
\Sigma_1[<, +1, \min, \max] & \subsetneq \Sigma_1[<, +1, \max]
\end{align*}
\]
By a result of Pin [10], it is decidable whether a given regular language is definable in \( \Sigma_1[<] \). For \( \Sigma_1[<,+1,\min,\max] \), decidability follows by a result of Pin and Weil [11] (or alternatively by Theorem 3). In this section, we characterize the languages definable in the other fragments and we show that definability within these fragments is decidable. As it turns out, these decidability results are a combination of effective algebraic and effective topological properties, cf. [7].

**Theorem 10.** Let \( L \subseteq \Gamma^+ \). The following assertions are equivalent:

1. \( L \) is definable in \( \Sigma_1[<,+1,\min] \).
2. \( L \) is a finite union of monomials \( w_1\Gamma^* \cdots w_n\Gamma^* \).
3. \( \text{Synt}(L) \in [x^\omega yx^\omega \leq x^\omega] \) and \( h_L(L) \) is a \( \leq_R \)-order ideal.

**Proof.**

1 \( \Rightarrow \) 2: Let \( L = \mathcal{L}(\varphi) \) for \( \varphi \in \Sigma_1[<,+1,\min] \). By Theorem 3 the language \( L \) is a finite union of monomials \( w_1\Gamma^*w_2 \cdots \Gamma^*w_n \). Let \( u \models \varphi \). Then for every \( v \in \Gamma^* \) the same assignment of the variables which makes \( \varphi \) true on \( u \) also satisfies \( \varphi \) on \( uv \). Therefore, \( L\Gamma^* \subseteq L \). Since \( (P \cup Q)\Gamma^* = P\Gamma^* \cup Q\Gamma^* \), it follows that \( L \) is a finite union of monomials \( w_1\Gamma^*w_2 \cdots \Gamma^*w_n\Gamma^* \).

2 \( \Rightarrow \) 3: This follows from \( '2' \) in Lemma 4.

3 \( \Rightarrow \) 1: \( \text{Synt}(L) \in [x^\omega yx^\omega \leq x^\omega] \) follows from Theorem 3. By [7, Theorem 1] we see that \( h_1(L) \) is a \( \leq_R \)-order ideal. The implication \( '3' \Rightarrow '2' \) follows from Proposition 9 and [7, Theorem 1].

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 \( \Gamma^*w_1 \cdots \Gamma^*w_n \Gamma^* \) if and only if \( \text{Synt}(L) \in [x^\omega yx^\omega \leq x^\omega] \) and \( h_L(L) \) is a \( \leq_L \)-order ideal. The following theorem is the analogue of Theorem 10 with neither min nor max predicates.

**Theorem 11.** Let \( L \subseteq \Gamma^+ \). The following assertions are equivalent:

1. \( L \) is definable in \( \Sigma_1[<,+1] \).
2. \( L \) is a finite union of monomials \( \Gamma^*w_1 \cdots \Gamma^*w_n\Gamma^* \).
3. \( \text{Synt}(L) \in [x^\omega yx^\omega \leq x^\omega] \) and \( h_L(L) \) is a \( \leq_J \)-order ideal.

**Proof.**

1 \( \Rightarrow \) 2: Let \( L \) be defined by \( \varphi \in \Sigma_1[<,+1] \). By Theorem 3 \( L \) is a finite union of monomials \( w_1\Gamma^*w_2 \cdots \Gamma^*w_n \). Let \( u,v,w \in \Gamma^* \). Every assignment satisfying \( \varphi \) on \( u \) also satisfies \( \varphi \) on \( uvw \). Hence, \( \Gamma^*L\Gamma^* \subseteq L \).

2 \( \Rightarrow \) 3: This follows from \( '3' \) in Lemma 4.

3 \( \Rightarrow \) 1: \( \text{Synt}(L) \in [x^\omega yx^\omega \leq x^\omega] \) follows from Theorem 3; the set \( h_L(L) \) is a \( \leq_J \)-order ideal by [7, Theorem 3]. The implication \( '3' \Rightarrow '2' \) follows from Proposition 9 and [7, Theorem 3].

The following decidability result is an immediate consequence of our characterizations.

**Corollary 12.** Let \( L \subseteq \Gamma^+ \) 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 \( h_L : \Gamma^+ \rightarrow \text{Synt}(L) \) of \( L \) is effectively computable. Hence, one can verify whether property \( '3' \) in Theorem 11 (resp. \( '3' \) in Theorem 10) resp. the left-right dual of \( '3' \) in Theorem 10 holds.
5 Dot-Depth One

A language \(L \subseteq \Gamma^+\) has dot-depth 1 if it is a Boolean combination of monomials of the form \(w_1\Gamma^*w_2 \cdots \Gamma^*w_n\) with \(w_i \in \Gamma^*\). Knast \cite{2} has shown that a language \(L\) has dot-depth 1 if and only if \(\text{Synt}(L) \in B_1\). Since the latter property is decidable, this gives decidability of dot-depth 1. Later, Thérien \cite{19} gave a simpler proof for Knast’s result. Both proofs are based on an algebraic concept called finite categories, see \cite{21}. In this section, we give a new (more combinatorial) proof of this theorem. As for dot-depth 1/2, the main advantage of our proof is that the bounds involved are more explicit.

**Theorem 13** (Thomas \cite{20}, Knast \cite{6}). Let \(L \subseteq \Gamma^+\). The following assertions are equivalent:

1. \(L\) is definable in \(B_1[<, +1, \min, \max]\).
2. \(L\) is a Boolean combination of monomials \(w_1\Gamma^*w_2 \cdots \Gamma^*w_n\).
3. \(\text{Synt}(L) \in B_1\).
4. \(L\) is recognized by some semigroup in \(B_1\).

As for dot-depth 1/2, the equivalence of \(B_1[<, +1, \min, \max]\) and dot-depth 1 is due to a result by Thomas \cite{20}. The remainder of this section is devoted to the proof of the above theorem.

**Lemma 14.** Let \(S\) be a finite semigroup and let \(u \in S^+\). Suppose there exists \(e \in S\) such that \(pe = p\) in \(S\) for some prefix \(p \leq u\). Choose \(|p|\) maximal with this property and let \(u = pv\). If \(xpey = x'pey'\) for some \(x \neq x'\), then there is at least one letter between the factors \(v\) in the two factorizations.

**Proof.** Let \(|x'| < |x|\) and assume that the claim is not true.

\[
\begin{array}{c|c|c|c|c}
 & x & p & v & y \\
\hline
x' & p & v & y' \\
\hline
p' & v'
\end{array}
\]

Then we find a factorization \(pv = p'v'\) such that \(p'e = p'\) and \(|p'| > |p|\) contradicting the maximality of \(|p|\). This also holds if the factors do not overlap but are adjacent, in which case \(p' = pv\). \hfill \Box

The following lemma will serve as the link between the algebraic properties of \(B_1\) and the combinatorial properties in Lemma \ref{lem16} below.

**Lemma 15.** Let \(S \in LR\) and let \(k \geq |S| + 1\). For every \(a \in S\) and for all \(u, x \in S^+\) with \(|x| \geq k\) we have: \(u \not\geq_{\mathcal{R}} uxax \rightarrow \alpha(x) = \alpha(xa)\).

**Proof.** Suppose \(u \not\geq_{\mathcal{R}} ux\) and \(\alpha(x) = \alpha(xa)\). Let \(w\) be the suffix of \(xa\) of length \(k\). By Lemma \ref{lem1} there exist \(p, v \in S^*\) such that \(w = pva\) in \(S^+\) and \(pe = p\) in \(S\) for some idempotent \(e \in S\). Let \(|p| \leq |S|\) be maximal with this property. Since \(w \in \alpha(xa) = \alpha(x)\) we can write

\[x = spvat\] for some \(s, t \in S^*\) such that \(p\) is a suffix of \(pvat\). Note that there is indeed at least one letter between the two occurrences of \(v\) by Lemma \ref{lem14}. For \(u' = usp\) and \(x' = vat\) we have \(u' = u'e, u'x' = u'x'e\), and \(u' \not\geq_{\mathcal{R}} u'x'\). Using Lemma \ref{lem8} we see that \(u' = u'x' = u'x'x'\). Hence, \(u \not\geq_{\mathcal{R}} u' = u'x'x' = uxat\) and therefore, \(u \not\geq_{\mathcal{R}} uxax\). \hfill \Box
The following lemma is the main combinatorial ingredient for our proof of Knast’s Theorem. It generalizes an idea of Klíma [5] to factors of words. The determinacy mechanism is similar to unary interval logic with lookahead [8].

**Lemma 16.** Let \( x_1, y_1, u, u_1', v_1, v_1' \in \Gamma^+ \) and \( u, v, u_1', v_1' \in \Gamma^* \), and let

\[
\begin{align*}
  u &= x_1 u_1 \cdots x_k u_k = u_1' y_1 \cdots u_\ell y_\ell \\
v &= x_1 v_1 \cdots x_k v_k = v_1' y_1 \cdots v_\ell' y_\ell \end{align*}
\]

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 \( x_1 \Gamma^+ x_2 \cdots \Gamma^+ x_k \), 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 \( y_1 \Gamma^+ y_2 \cdots \Gamma^+ y_\ell \).

If \( u \) and \( v \) are contained in the same monomials \( w_1 \Gamma^+ w_2 \cdots \Gamma^+ w_n \) with \( n \leq k + \ell \) and degree \( |w_1 \cdots w_n| \leq |x_1 \cdots x_k y_1 \cdots y_\ell| \), then the relative positions of \( x_k \) and \( y_1 \) are the same in \( u \) as in \( v \). More precisely,

1. \( x_1 u_1 \cdots x_k \) is a prefix of \( u_1' \) iff \( x_1 v_1 \cdots x_k \) is a prefix of \( v_1' \).
2. if \( x_k \) and \( y_1 \) overlap in \( u \) or in \( v \), then they have the same overlap in both words,
3. \( u_1' y_1 \) is a prefix of \( x_1 \cdots u_{k-1} \) iff \( v_1' y_1 \) is a prefix of \( x_1 \cdots v_{k-1} \).

**Proof.** [1]: Suppose that \( x_1 u_1 \cdots x_k \) is a prefix of \( u_1' \). Then \( u \) is contained in the language \( x_1 \Gamma^+ \cdots x_k \Gamma^+ v_1 \cdots v_\ell \) or in \( x_1 \Gamma^+ \cdots x_k v_1 \cdots \Gamma^+ v_\ell \). Hence \( v \) is contained in one of these two monomials, showing that \( x_1 v_1 \cdots x_k \) is a prefix of \( v_1' \).

[2]: We can assume that none of the conditions in [1] holds. We have to distinguish two cases. First, suppose that \( x_k \) and \( y_1 \) overlap in \( u \) such that \( x_1 u_1 \cdots x_k \) is a prefix of \( u_1' y_1 \) and let \( z \) be the word comprising all positions of \( x_k \) and \( y_1 \) in \( u \). Then \( u \in P = x_1 \Gamma^+ \cdots x_{k-1} \Gamma^+ z \Gamma^+ y_2 \cdots \Gamma^+ v_\ell \). Hence \( v \in P \), showing that \( x_k \) and \( y_1 \) in \( v \) have at most the same overlap as in \( u \).

The second case is that \( x_k \) and \( y_1 \) overlap in \( u \) such that \( x_1 u_1 \cdots x_k \) is not a prefix of \( u_1' y_1 \). Moreover, we can assume that \( x_1 v_1 \cdots x_k \) is not a prefix of \( v_1' y_1 \) since otherwise we are in the first case with \( u \) and \( v \) interchanged. Now, \( u \) is contained in \( P = x_1 \Gamma^+ \cdots x_i \Gamma^+ z \Gamma^+ y_j \cdots \Gamma^+ v_\ell \), where \( z \) is the factor of \( u \) comprising all \( x_{i+1}, \ldots, x_k \) which are overlapping (or adjacent) with \( y_1 \) and all \( y_1, \ldots, y_{j-1} \) which are overlapping (or adjacent) with \( x_k \). Since \( v \in P \), we conclude that \( x_k \) and \( y_1 \) in \( v \) have at least the same overlap as in \( u \).

[3]: If none of the conditions in [1] and [2] holds, then in both words \( u \) and \( v \), the factor \( y_1 \) is on the left-hand side of \( x_k \).

**Lemma 17.** Let \( S \in B_1 \). For all \( u, v, x, s \in S \) and for all \( e, f \in E(S) \), the following implication holds: \( u R u e x f, e s f v L v \Rightarrow u e x f v = u e s f v \).

**Proof.** Since \( u R u e x f \) and \( v L e s f v \), there exist \( y, t \in M \) with \( u = u e x f y \) and \( v = t e s f v \). In particular, \( u = u (e x f y)^\omega \) and \( v = (t e s f)^\omega v \). We conclude

\[ u e x f v = u (e x f y)^\omega e x f (t e s f)^\omega v = u (e x f y)^\omega e s f (t e s f)^\omega v = u e s f v, \]

where the second equality uses \( S \in B_1 \).

**Proposition 18.** Let \( L \subseteq \Gamma^+ \) be recognized by \( h : \Gamma^+ \rightarrow S \) with \( S \in B_1 \). If words \( u \) and \( v \) are contained in the same monomials \( w_1 \Gamma^+ w_2 \cdots \Gamma^+ w_n \) with \( n \leq 2|S| \) and degree \( |w_1 \cdots w_n| \leq 4|S|^2 - 2|S| \), then \( h(u) = h(v) \).
Proof. This proof was inspired by Klíma’s proof [5] of Simon’s Theorem on piecewise-testable languages. The outline of our proof 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 factor in $u$ and $v$ are the same. Finally, we transform $u$ into $v$ by a sequence of $h$-invariant substitutions.

Consider the $\mathcal{R}$-factorization $u = a_1u_1 \cdots a_k u_k$ such that

$$h(a_1u_1 \cdots a_i) \mathcal{R} h(a_1u_1 \cdots a_i a_{i+1})$$

for all $i$. We have $k \leq |S|$. Let $j_i$ be the position of $a_i$ in the above factorization. We color all positions of $u$ in any of the intervals $[j_i - |S|; j_i + |S| - 1]$ red. In particular, the $a_i$-positions $j_i$ are red. And 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 only exception. Hence, there are at most $2|S|^2 - |S|$ red positions in $u$. Let $R_i$ be the $i$-th consecutive factor of red positions. Then $u = R_1u_1' \cdots R_k u_k'$ for some $u_i' \in \Gamma^+$, $i < k'$, and $u_k' \in \Gamma^*$. Note that $k' \leq k$ since some intervals could overlap. By Lemma 13, the word $R_1u_1' \cdots R_k$ is the shortest prefix of $u$ contained in $R_1 \Gamma^+ \cdots R_k$.

Symmetrically, we consider the $\mathcal{L}$-factorization $v = v_1 b_1 \cdots v_{i-1} b_{i-1}$ such that

$$h(b_{i+1} v_i b_i \cdots v_1 b_{1}) \mathcal{L} h(v_i b_i \cdots v_1 b_{1})$$

for all $i$. Let $j'_i$ be the position of $b_i$ in the above factorization. We color all positions of $v$ in any of the intervals $[j'_i - |S| + 1; j'_i + |S|]$ blue. 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 $v = v_1 B_1 \cdots v_{i'} B_{i'}$ for some $v_i' \in \Gamma^+$, $i \geq 1$ and $v_{i'}' \in \Gamma^*$. As before, $B_1 \cdots v_{i'} B_{i'}$ is the shortest suffix of $v$ contained in $B_1 \cdots \Gamma^+ B_{i'}$.

Next, we transfer the red positions of $u$ to $v$, and we transfer the blue positions of $v$ to $u$. By assumption, $v \in R_1 \Gamma^+ \cdots R_k \Gamma^+$. Therefore, there exists a factorization $v = R_1v_1'' \cdots R_k v_k''$ such that $R_1v_1'' \cdots R_k$ is the shortest prefix of $v$ contained in $R_1 \Gamma^+ \cdots R_k$. We color the positions of the $R_i$’s in $v$ red. Similarly, there exists a factorization $u = u_1'' B_1 \cdots u_{i'}'' B_{i'}$ such that $B_1 \cdots u_{i'}'' B_{i'}$ is the shortest suffix of $u$ contained in $B_1 \cdots \Gamma^+ B_{i'}$. 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 16, the colored positions in $u$ have the same order as the colored positions in $v$. Let $w_i$ be the $i$-th consecutive factor of colored (red or blue) positions, and write

$$u = w_1 x_1 \cdots w_{n-1} x_{n-1} w_n,$$
$$v = w_1 s_1 \cdots w_{n-1} s_{n-1} w_n.$$

By Lemma 1 and its left-right dual, there exist $e_1, \ldots, e_{n-1} \in E(S)$ and $f_2, \ldots, f_n \in E(S)$ such that each $w_i$ admits a factorization $w_i = p_i r_i q_i$ with $|p_i| \leq |S| - 1$ and $|q_i| \leq |S| - 1$ satisfying

$$h(r_i) = h(r_i) e_i \quad \text{for } 1 \leq i < n,$$
$$h(r_i) = f_i h(r_i) \quad \text{for } 1 \leq i \leq n.$$

In particular, we can assume $p_1 = \varepsilon = q_n$. Let $x_i' = q_i x_i p_i + 1$ and $s_i' = q_i s_i p_i + 1$ for $1 \leq i < n$. Then

$$u = r_1 x_1' r_2 \cdots x_{n-1}' r_n,$$
$$v = r_1 s_1' r_2 \cdots s_{n-1}' r_n.$$
and the $r_i$’s in $u$ cover the positions of the $R$-factorization of $u$, whereas the $r_i$’s in $v$ cover the positions of the $L$-factorization of $v$. Therefore, we have
\[
\begin{align*}
  h(r_1x_1'\cdots r_i) &\in R \cdot h(r_1x_1'\cdots r_i) \cdot e_i h(x_i') f_{i+1} & \text{for all } 1 \leq i < n, \\
  h(r_i\cdots s_i r_n) &\in L \cdot e_{i-1} h(s_i' r_{i-1}) f_i \cdot h(r_i\cdots s_i r_n) & \text{for all } 1 < i \leq n.
\end{align*}
\]
By an $(n-1)$-fold application of Lemma 17 we obtain
\[
\begin{align*}
  h(u) &= h(r_1x_1'\cdots r_n-2x'_{n-2}r_{n-1}s'_{n-1}r_n) \\
  &= h(r_1x_1'\cdots r_n-2x'_{n-2}r_{n-1}s'_{n-1}r_n) \\
  &= \cdots \\
  &= h(r_1s_1'\cdots r_{n-2}s'_{n-2}r_{n-1}s'_{n-1}r_n) = h(v).
\end{align*}
\]
Note that the substitution rules $x'_i \rightarrow s'_i$ are $h$-invariant in their respective contexts only when applied from right to left when converting $h(u)$ into $h(v)$. \hfill \square

**Corollary 19.** Let $L \subseteq \Gamma^+$ be recognized by a finite semigroup $S \in B_1$. If words $u$ and $v$ are contained in the same monomials $w_1 \Gamma^* w_2 \cdots \Gamma^* w_n$ with $n \leq 2|S|$ and degree $|w_1 \cdots w_n| < 4|S|^2$, then $h(u) = h(v)$.

**Proof.** Every monominal $w_1 \Gamma^* w_2 \cdots w_{n-1} \Gamma^+ w_n$ is a union of monomials of the form
\[
w_1 a_1 \Gamma^* \cdots w_{n-1} a_{n-1} \Gamma^+ w_n
\]
for $a_1, \ldots, a_{n-1} \in \Gamma$. Therefore, the claim follows from Proposition 18. \hfill \square

We are now ready to prove Theorem 13.

**Proof (Theorem 13).** (1) $\Rightarrow$ (2): This follows from Theorem 3.

(2) $\Rightarrow$ (3): By Lemma 6 the syntactic semigroup of every monominal $w_1 \Gamma^* w_2 \cdots \Gamma^* w_n$ satisfies $x^\omega y x^\omega \leq x^\omega$ and by Lemma 2 it is in $B_1$. The claim follows since the class of languages recognizable in $B_1$ is closed under Boolean combinations. The implication (3) $\Rightarrow$ (4) is trivial.

(4) $\Rightarrow$ (5): Let $L$ be recognized by $h : \Gamma^+ \rightarrow S \subseteq B_1$. We write $u \equiv v$ if $u$ and $v$ are contained in the same monominals of the form $w_1 \Gamma^* w_2 \cdots \Gamma^* w_n$ of degree at most $4|S|^2$. We have $L = h^{-1}(P)$ for $P = h(L)$. Corollary 19 shows that every set $h^{-1}(p)$ is a union of $\equiv$-classes. Moreover, $\equiv$ has finite index since there are only finitely many monominals of bounded degree. Every $\equiv$-class is a finite Boolean combination of the required form by specifying which monominals hold and which do not. \hfill \square

### 6 Dot-depth One without min or max

As for $\Sigma_1$, one cannot define min- or max-predicates in $\mathbb{B}\Sigma_1[<,+1]$. Therefore, the following inclusions hold:
\[
\begin{align*}
  \mathbb{B}\Sigma_1[<,+1,\min] &\subseteq \mathbb{B}\Sigma_1[<,+1,\max] \\
  \mathbb{B}\Sigma_1[<] &\subseteq \mathbb{B}\Sigma_1[<,+1] \\
  \mathbb{B}\Sigma_1[<,+1,\min] &\subseteq \mathbb{B}\Sigma_1[<,+1,\max]
\end{align*}
\]
Simon’s Theorem on piecewise testable languages \[13\] gives decidability of \(\mathbb{B}\Sigma_1[<]\). For the fragment \(\mathbb{B}\Sigma_1[<,+1,\min,\max]\), decidability follows by Knast’s Theorem \[6\], see Theorem 13. In this section, we give effective characterizations of the remaining fragments. As for dot-depth \(1/2\), these characterizations are a combination of algebraic and topological properties, cf. \[2\]. Moreover, we obtain natural subclasses of dot-depth 1 for the languages definable by the above fragments.

**Theorem 22.** Let \(L \subseteq \Gamma^+\). The following assertions are equivalent:

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

**Proof.** The equivalence \([1] \Leftrightarrow [2]\) follows from Theorem 10.

\([2] \Rightarrow [3]\): We have \(\text{Synt}(L) \in \mathcal{B}_1\) by Theorem 13 and \(h_L(L)\) is a union of \(\mathcal{R}\)-classes by \([7]\) and Theorem 5.

\([3] \Rightarrow [2]\): By Lemma 21 there exists \(m \in \mathbb{N}\) such that \(h_L(u) \subseteq \mathcal{R}\) if \(u\) and \(v\) are contained in the same languages of the form \(w_1\Gamma^* \cdots w_n\Gamma^*\) with \(|w_1 \cdots w_n| \leq m\). Therefore, for each \(\mathcal{R}\)-class \(R\) of \(\text{Synt}(L)\), the language \(h_L^{-1}(R)\) is a Boolean combination of languages \(w_1\Gamma^* \cdots w_n\Gamma^*\) with \(|w_1 \cdots w_n| \leq m\). The claim follows, since \(L\) is a union of languages of the form \(h_L^{-1}(R)\). \(\Box\)

**Lemma 20.** Let \(P = w_1\Gamma^* w_2 \cdots \Gamma^* w_n\) and let \(uq \in P\). Then there exists a monomial \(P' = v_1\Gamma^* v_2 \cdots \Gamma^* v_n\) with \(|v_1 \cdots v_n| \leq |w_1 \cdots w_nq|\) such that \(uq \in P' \subseteq (Pq^{-1})q\).

**Proof.** Let \(j \leq n\) be minimal such that \(q \in \Gamma^* w_j \cdots \Gamma^* w_n\). If there exists a proper prefix \(y\) of \(w_{j-1}\) such that \(y\) is a suffix of \(u\) and \(yq \in w_{j-1} \cdots \Gamma^* w_n\), then we set \(q' = yq\), else we set \(q' = q\). We assume \(q'\) to be maximal with these properties. We can write \(uq = u''q'\). Moreover, by maximality of \(q'\), there exists an index \(j'\) such that \(q' \in \Gamma^* w_{j'} \cdots \Gamma^* w_n\) and \(q' \notin y'\Gamma^* w_{j'} \cdots \Gamma^* w_n\) for any non-empty suffix \(y'\) of \(w_{j'-1}\). We set \(P' = w_1\Gamma^* \cdots w_{j'-1}\Gamma^* q'\). Now, \(uq \in P'\) and for all \(w \in P'\) we have \(w \in P \cap \Gamma^* q = (Pq^{-1})q\). \(\square\)

**Lemma 21.** Let \(h : \Gamma^+ \to S \in \mathcal{B}_1\). If \(u, v \in \Gamma^+\) are contained in the same monomials \(w_1\Gamma^* \cdots w_n\Gamma^*\) of degree \(|w_1 \cdots w_n| < 8|S|^2\), then \(h(u) \cap h(v)\).

**Proof.** We write \(u \equiv_m v\), if \(u\) and \(v\) are contained in the same monomials \(w_1\Gamma^* \cdots \Gamma^* w_n\) of degree \(|w_1 \cdots w_n| \leq m\). Analogously, we write \(u \equiv_m v\) if \(u\) and \(v\) are contained in the same monomials \(w_1\Gamma^* \cdots w_n\Gamma^*\) of degree \(|w_1 \cdots w_n| \leq m\). If \(u \equiv_m v\) for \(m = 4|S|^2 - 1\), then by Corollary 19 we have \(h(u) = h(v)\).

Let \(u \sim_2 v\). We want to show \(h(u) \cap h(v)\). We can assume \(|u|, |v| \geq 2m\), 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 \(u \in \Gamma^q\Gamma^* \ni v\). We claim \(u \equiv_k v'q\) and therefore, \(h(v) \subseteq h(v'q) = h(u)\). Symmetry then yields \(h(u) \cap h(v)\).

We now prove the claim. First, let \(v'q \in P = w_1\Gamma^* \cdots \Gamma^* w_n\) with \(w_1 \cdots w_n \leq m\). Then \(v \in \Gamma^* q\) and \(u \in \Gamma^* q\). Since \(w_n\) is a suffix of \(q\), we conclude \(u \in P\).

Next, suppose \(u \in P = w_1\Gamma^* \cdots \Gamma^* w_n\) with \(|w_1 \cdots w_n| \leq m\). By Lemma 20 there exists a monomial \(P' = v_1\Gamma^* v_2 \cdots \Gamma^* v_n\) with \(|v_1 \cdots v_n| \leq |w_1 \cdots w_nq| \leq 2m\) and \(uq \in P' \subseteq (Pq^{-1})q\). Since \(u \in \Gamma^* q\), we obtain \(v \in \Gamma^* q\). By choice of \(x\), we have \(v'q \in P' \subseteq \Gamma^* q\). Since \(w_n\) is a suffix of \(q\), we conclude \(v'q \in w_1\Gamma^* \cdots \Gamma^* w_n\). \(\Box\)
There also is 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 \(\Gamma^*w_1 \cdots \Gamma^*w_n\) if and only if \(\text{Synt}(L) \in \mathcal{B}_1\) 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 23.** Let \(h : \Gamma^+ \to S \in \mathcal{B}_1\). If \(u, v \in \Gamma^+\) are contained in the same monomials \(\Gamma^*w_1 \Gamma^* \cdots \Gamma^*w_n\) of degree \(|w_1 \cdots w_n| < 12 |S|^2\), then \(h(u) = h(v)\).

**Proof.** This proof is only a slight variation of the proof of Lemma 21. We write \(u \equiv_m v\), if \(u\) and \(v\) are contained in the same monomials \(\Gamma^*w_1 \Gamma^* \cdots \Gamma^*w_n\) of degree \(|w_1 \cdots w_n| \leq m\). Analogously, we write \(u \sim_m v\), if \(u\) and \(v\) are contained in the same monomials \(\Gamma^*w_1 \Gamma^* \cdots \Gamma^*w_n\) of degree \(|w_1 \cdots w_n| \leq m\). If \(u \equiv_m v\) for \(m = 4|S|^2 - 1\), then by Corollary 19 we have \(h(u) = h(v)\).

Let \(u \sim_3 v\). We want to show \(h(u) \equiv h(v)\). We can assume \(|u|, |v| \geq 3m\), because otherwise \(u = v\). Let \(u = pu'q\) with \(|p| = |q| = m\). Consider the factorization \(v = spv'q\) 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 \(q\). This factorization exists, since \(u \in \Gamma^*p^*q^* \Gamma^* \ni v\). We claim \(u \equiv_m pv'q\) and therefore, \(h(v) \equiv h(pv'q) = h(u)\). Symmetry then yields \(h(u) \equiv h(v)\).

We now prove the claim. First, let \(pv'q \in P\) for \(P = w_1 \Gamma^* \cdots \Gamma^*w_n\) with \(w_1 \cdots w_n \leq m\). Then \(v \in \Gamma^*P\Gamma^*\) and \(u \in \Gamma^*P\Gamma^*\). Since \(w_1\) is a prefix of \(p\) and \(w_n\) is a suffix of \(q\), we conclude \(v \in P\).

Next, suppose \(u \in P\) with \(|w_1 \cdots w_n| \leq m\). By Lemma 20 and its left-right dual, there exists a monomial \(P' = v_1 \Gamma^*v_2 \cdots \Gamma^*v_n\) with \(|v_1 \cdots v_n| \leq |pv_1 \cdots v_nq| \leq 3m\) and \(u = pu'q \in P' \subseteq p(p^{-1}Pq^{-1})q\). Since \(u \in \Gamma^*P\Gamma^*\), we obtain \(v \in \Gamma^*P\Gamma^*\). By choice of \(s\) and \(x\), we have \(pv'q \in \Gamma^*P\Gamma^* \subseteq \Gamma^*P\Gamma^*\). Since \(w_1\) is a prefix of \(p\) and \(w_n\) is a suffix of \(q\), we conclude \(pv'q \in w_1 \Gamma^* \cdots \Gamma^*w_n\).

**Theorem 24.** Let \(L \subseteq \Gamma^+\). The following assertions are equivalent:

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

**Proof.** The equivalence \(1 \iff 2\) follows from Theorem 11.

\(2 \Rightarrow 3\): We have \(\text{Synt}(L) \in \mathcal{B}_1\) by Theorem 13 and \(h_L(L)\) is a union of \(\mathcal{R}\)-classes by Theorem 7.

\(3 \Rightarrow 2\): By Lemma 23 there exists \(m \in \mathbb{N}\) such that \(h_L(u) \equiv h_L(v)\), if \(u\) and \(v\) are contained in the same languages of the form \(\Gamma^*w_1 \Gamma^* \cdots \Gamma^*w_n\) with \(|w_1 \cdots w_n| \leq m\). Since \(h_L(L)\) is a union of \(\mathcal{J}\)-classes, the language \(L\) is a Boolean combination of languages of the form \(\Gamma^*w_1 \Gamma^* \cdots \Gamma^*w_n\) of degree \(|w_1 \cdots w_n| \leq m\).

The following decidability result is an immediate consequence of our characterizations.

**Corollary 25.** Let \(L \subseteq \Gamma^+\) be a regular language. It is decidable whether \(L\) is definable in \(\mathbb{B} \Sigma_1[<, +1, \min]\) (resp. \(\mathbb{B} \Sigma_1[<, +1, \max]\), resp. \(\mathbb{B} \Sigma_1[<, +1, \max]\)).

**Proof.** The syntactic homomorphism \(h_L : \Gamma^+ \to \text{Synt}(L)\) of \(L\) is effectively computable. Hence, one can verify whether property \(3\) in Theorem 24 (resp. \(3\) in Theorem 22) resp. the left-right dual of \(3\) in Theorem 22 holds. \(\square\)
7 Summary

We considered subclasses of languages with dot-depth $1/2$ and of languages with dot-depth $1$. These subclasses admit counterparts in terms of fragments of existential first-order logic $\Sigma_1$ and its Boolean closure $B\Sigma_1$. For all fragments, we give effective algebraic characterizations. At closer look, the characterizations are a conjunction of an algebraic and a topological property, cf. [7]. We summarize our main results in Table 1. To shorten notation, we write $B_{1/2}$ instead of $J_{x^\omega}y^x$ $\leq x^\omega$.

In addition, we give new proofs for Pin and Weil’s Theorem on dot-depth $1/2$ and for Knast’s Theorem on dot-depth $1$. The proofs are combinatorial and they improve the bounds involved in computing a language description for a given recognizing semigroup.

References


