Languages of Dot-Depth $3/2$

Christian Glaßer[*] *Heinz Schmitz*[**]
Lehrstuhl für theoretische Informatik
Julius-Maximilians-Universität Würzburg
97074 Würzburg, Germany

Abstract

We prove an effective characterization of languages having dot-depth $3/2$. To our knowledge, only the classes of levels 0, $1/2$ and 1 of the dot-depth hierarchy were known to be decidable. Let $\mathcal{B}_{3/2}$ denote the class of languages having dot-depth $3/2$, i.e., languages that can be written as finite unions of languages of the form $u_0L_1u_1L_2u_2\cdots L_nu_n$, where $u_i \in A^*$ and L_i are languages of dot-depth one.

1. Let $k \geq 0$. We define and study in detail classes $\mathcal{B}_{3/2,k}$ for which $\mathcal{B}_{3/2} = \bigcup_{k\geq 0} \mathcal{B}_{3/2,k}$. Closure properties are investigated and a useful normalform representation for languages in $\mathcal{B}_{3/2,k}$ is proved.
2. We provide a combinatorial lemma which keeps control of the lengths of words leading to loops in a given deterministic finite automaton (dfa). It is a basic tool in our subsequent proofs and may be of independent interest.
3. Let $k \geq 0$ and F be a dfa accepting some language L. We show that L belongs to $\mathcal{B}_{3/2,k}$ if and only if F does not have Pattern $\mathbf{P}_{\boldsymbol{k}}$ in it's transition graph (cf. Figure 1). This yields an NL–algorithm for the decision problem for $\mathcal{B}_{3/2,k}$.
4. We identify a Pattern **P** such that L belongs to $\mathcal{B}_{3/2}$ if and only if F does not have Pattern **P** in it's transition graph (cf. Figure 2). This yields an NL–algorithm for the decision problem for $\mathcal{B}_{3/2}$. Moreover the proof is such that an algorithm can be derived to determine the exact level k of a given language inside $\mathcal{B}_{3/2}$.

Due to known relations between the dot-depth hierarchy and symbolic logic, the decidability of the class of languages definable by Σ_2 – formulas of the logic $\mathrm{FO}[<, \min, \max, S, P]$ follows. We also give an algebraic interpretation of our result.

This paper is self-contained. In particular (namely Item 3 above for the case $k = 0$), we give as a by-product a reproof of the known forbidden pattern characterization of level $3/2$ of the so-called Straubing-Thérien hierarchy.

[*] Supported by the Studienstiftung des Deutschen Volkes.
[**] Supported by the Deutsche Forschungsgemeinschaft (DFG), grant Wa 847/4-1.

Contents

1 Introduction

We contribute to the theory of finite automata and regular languages, with consequences in logic as well as in algebra. Particularly we deal with starfree regular languages. These are languages which are constructed from alphabet letters using only Boolean operations together with concatenation. Alternating these two kinds of operations in order to distinguish between combinatorial and sequential aspects leads to the definition of concatenation hierarchies that exhaust the class of starfree languages.

Prominent examples are the dot-depth hierarchy, first studied in [CB71], and the Straubing-Thérien hierarchy (cf. [Str81, Thé81, Str85]). Both are known to be strict and closely related to each other (cf. [BK78, Str85]). Most naturally arising questions concerning these hierarchies are of major interest in different research areas since surprisingly close connections to finite model theory, theory of finite semigroups, topology, Boolean circuits, complexity theory and others have been exposed. For an overview or as a good starting point to this rich field of research see e.g. the articles [Brz76, Pin96a, Pin96b].

Here we deal with the dot-depth hierarchy. Let A be some finite alphabet with $|A| \geq 2$. For a class $\mathcal{C}$ of languages over A^+ let $\mathrm{POL}(\mathcal{C})$ be its polynomial closure, i.e., the class of languages L that can be written as a finite union of languages $u_0L_1u_1L_2u_2 \cdots L_nu_n$, where $u_i \in A^*$, $L_i \in \mathcal{C}$ and $n \geq 0$. Denote by $\mathrm{BC}(\mathcal{C})$ its Boolean closure, i.e., the closure of $\mathcal{C}$ under finite union, finite intersection and complementation, where the latter is taken with respect to A^+ (which we denote by $\overline{L}^+$ for some language $L \subseteq A^+$). Then the dot-depth hierarchy can be defined as the following family of classes $\mathcal{B}_{n/2}$ (notations are adopted from [PW97]).

1. $\mathcal{B}_0 =_{\mathrm{def}} \{\emptyset, A^+\}$,
2. $\mathcal{B}_{n+1/2} =_{\mathrm{def}} \mathrm{POL}(\mathcal{B}_n)$ for $n \geq 0$, and
3. $\mathcal{B}_{n+1} =_{\mathrm{def}} \mathrm{BC}(\mathcal{B}_{n+1/2})$ for $n \geq 0$.

For a language $L \subseteq A^+$ and a minimal n with $L \in \mathcal{B}_{n/2}$ we say that L *has dot-depth* $n/2$. We will also consider the classes $\mathrm{co}\mathcal{B}_{n+1/2}$, where $\mathrm{co}\mathcal{C} =_{\mathrm{def}} \left\{ \overline{L}^+ \mid L \in \mathcal{C} \right\}$ for a class $\mathcal{C}$. One obtains the same hierarchy classes when setting $\mathcal{B}_{n+1/2} =_{\mathrm{def}} \mathrm{POL}(\mathrm{co}\mathcal{B}_{n-1/2})$ as proved in [Gla98]. By definition, all $\mathcal{B}_{n+1/2}$ are closed under union and it is known, that these classes are also closed under intersection (see [Arf87, Arf91]).

To our knowledge, only the classes of levels 0, $1/2$ and 1 of the dot-depth hierarchy were known to be decidable. Especially, the case of dot-depth $3/2$ was mentioned as an open question in the Handbook of Formal Languages [Pin96b] and in [PW97]. This can be seen in contrast to the Straubing-Thérien hierarchy, for which beside levels 0, $1/2$ and 1 also level $3/2$ is known to be decidable. The latter was first proved in [Arf87, Arf91] (see also [PW97]). Due to a fundamental result in [Str85] it is known that level n of the dot-depth hierarchy is decidable if and only if level n of the Straubing-Thérien hierarchy is decidable — for integers n. It was conjectured in [Pin96b] that this holds in fact for all levels $n/2$ and, in a sense, we support this opinion here. For a more detailed overview on the strong influence of these long-standing open decidability questions and the continuously ongoing research see again [Pin95a, Pin96b]. Recent results can be found in [PW97].

In this paper we prove an effective characterization of languages having dot-depth $3/2$. With an automata-theoretic approach we study the class $\mathcal{B}_{3/2}$ in detail. First fix some $k \geq 0$. We look at a word w as a word over A^{k+1} by packing together each $k+1$ consecutive letters and call this the k-decomposition of w (where k indicates the number of overlapping letters). In this way we obtain classes $\mathcal{B}_{3/2,k}$, where in case $k = 0$ we encounter level $3/2$ of the Straubing-Thérien hierarchy (up to the empty word) and for which it holds that $\mathcal{B}_{3/2} = \bigcup_{k \geq 0} \mathcal{B}_{3/2,k}$. The k-decomposition approach was used before in several contexts, among them in the study of locally testable events (see [BS73, McN74]), when relating the dot-depth hierarchy and the Straubing-Thérien hierarchy (as mentioned above, see [Str85]), for a levelwise analysis of dot-depth one languages (see [Ste85a], also [Sch99b]), when studying the notion of generalized deterministic languages (see [Sch99a]) and it was mentioned recently in [TW98].

We first look at the family of classes $\mathcal{B}_{3/2,k}$. With help of a number of closure properties and a series of technical lemmas we prove a useful normalform representation for languages in $\mathcal{B}_{3/2,k}$. This normalform was known before for the case $k = 0$ (cf. [Arf87, Arf91]).

Then we provide a combinatorial lemma which keeps control of the lengths of inputs w to deterministic finite automata such that $w = xuy$ and the state reached after input xu has a loop with label u. This can be compared to the algebraic notion of idempotents in finite semigroup theory. Here we keep track of the lengths one has to consider and bound it in the size of the automaton in question. Iterated applications of this lemma will be a basic tool in subsequent proofs.

Let F be a dfa accepting some language L. We show that L belongs to $\mathcal{B}_{3/2,k}$ if and only if F does not have Pattern $\mathbf{P}_{\boldsymbol{k}}$ in it's transition graph (cf. Figure 1). This yields an NL–algorithm for the decision problem for $\mathcal{B}_{3/2,k}$. As a by-product, we provide a self-contained reproof of the forbidden pattern characterization previously known for level $3/2$ of the Straubing-Thérien hierarchy (simply read our proof having in mind '$k = 0$'). The two proofs known before, namely given in [Arf87] and [PW97], use deep results from [Has83] and [Sim90], respectively.

Our generalization to arbitrary k enables us to identify a Pattern **P** such that L belongs to $\mathcal{B}_{3/2}$ if and only if F does not have Pattern **P** in it's transition graph (cf. Figure 2). This yields an NL–algorithm for the decision problem for $\mathcal{B}_{3/2}$. Moreover the proof is such that an algorithm can be derived to determine the exact level k of a given language inside $\mathcal{B}_{3/2}$.

We draw some consequences which are due to various relations of the classes of the dot-depth hierarchy to other fields of research. The connection between first-order logic and the class of starfree languages goes back to the work of McNaughton and Papert [MP71]. The dot-depth hierarchy is related to the first-order logic $\mathrm{FO}[<, \min, \max, S, P]$ having unary relations for the alphabet symbols from A, the binary relation $<$, the successor (predecessor) function S (P, resp.), and constants $\min$ and $\max$. Let Σ_n be the subclass of this logic which is defined by at most $n-1$ quantifier alternations, starting with an existential quantifier. It has been proved by W. Thomas in [Tho82] (see also [PP86, PW97]) that Σ_n formulas describe just the $\mathcal{B}_{n-1/2}$ languages and that the Boolean combinations of Σ_n formulas describe just the $\mathcal{B}_n$ languages. Due to this characterization we can conclude the decidability of the class of languages definable by Σ_2 – formulas of the logic $\mathrm{FO}[<, \min, \max, S, P]$.

We also give an algebraic interpretation of our result and characterize the languages of dot-depth $3/2$ by a condition on their ordered syntactic semigroups. Spoken in algebraic terms, this yields an effective characterization of the so-called variety of finite ordered semigroups corresponding to the positive variety of languages, as which $\mathcal{B}_{3/2}$ can be understood. An ineffective characterization of this class of finite ordered semigroups was given before in [PW97].

Our paper is self-contained and all proofs are given in full length, which we believe makes it broadly accessible. In particular, we think that it possibly provides a number of starting points helpful for investigating higher levels of the dot-depth hierarchy (see Section Conclusions).

2 Preliminaries

In this section we introduce notations and prepare the proofs of the following sections. In order to keep the paper self-contained, we will proof all properties needed, although they are partially known from literature. We mention the respective citations. Readers familiar with the dot-depth problem may skip one or the other proof in this section if the assertions sound reasonable.

2.1 Basic Definitions

We make all definitions for an arbitrary finite alphabet A with $|A| \geq 2$. The set of all words over A is denoted by A^*, the set of all non-empty words over A is denoted by A^+, and the set of all words from A^* of length k is denoted by A^k for $k \geq 0$. Let ϵ denote the empty word and let $|x|$ be the length of a word $x \in A^*$. Throughout the paper (if no other definition is made) we consider languages as subsets of A^+. In addition to the usual complement $\overline{L}$ with respect to A^* we denote by $\overline{L}^{+}$ the complement of some language $L \subseteq A^+$ with respect to A^+. If $L \subseteq A^+$ then it holds that $\overline{L}^{+} = \overline{L} \backslash \{\epsilon\}$. For $k \geq 0$ we denote by $A^{\leq k}$ ($A^{<k}$, $A^{\geq k}$, $\ldots$) the set of words from A^+ of length less or equal to k (less than k, greater or equal to k, $\ldots$ respectively). Note that none of these sets contains the empty word.

Let $x, y \in A^*$. We call y a *prefix* (*suffix*) of a word x if there exists a word z such that $x = yz$ ($x = zy$, resp.). For $k \geq 0$ the *k-prefix* and the *k-suffix* of $x = a_1 \cdots a_n$ are given by

$$p_k(x) =_{\text{def}} \begin{cases} a_1 a_2 \cdots a_k, & \text{if } n \geq k, \\ x, & \text{otherwise} \end{cases}$$

and

$$s_k(x) =_{\text{def}} \begin{cases} a_{n-k+1} a_{n-k+2} \cdots a_n, & \text{if } n \geq k, \\ x, & \text{otherwise.} \end{cases}$$

We define the set of letters of x as $\alpha(x) =_{\text{def}} \{a_1, a_2, \ldots, a_k\}$ and denote by x^R its reverse, i.e. $x^R =_{\text{def}} a_n a_{n-1} \cdots a_1$. For a language L let $L^R =_{\text{def}} \{ x^R \mid x \in L \}$. For some $L \subseteq A^*$ and $w \in A^*$ we define the right and left quotients of L with respect to w as $Lw^{-1} =_{\text{def}} \{ v \in A^* \mid vw \in L \}$ and $w^{-1}L =_{\text{def}} \{ v \in A^* \mid wv \in L \}$. Infinite sequences of words $\{x_i\}_{i=1}^{\infty}$ will be denoted for short as $\{x_i\}$.

Proposition 2.1. *For any $L, L' \subseteq A^*$ and $w \in A^*$ it holds that $w^{-1}(L \cup L') = w^{-1}L \cup w^{-1}L'$ and $w^{-1}\overline{L} = \overline{w^{-1}L}$. The same holds for right quotients.*

Proof. We have $v \in w^{-1}(L \cup L') \iff wv \in L \cup L' \iff wv \in L \vee wv \in L' \iff v \in w^{-1}L \vee v \in w^{-1}L' \iff v \in w^{-1}L \cup w^{-1}L'$. Also, $v \in w^{-1}\overline{L} \iff wv \in \overline{L} \iff wv \notin L \iff v \notin w^{-1}L \iff v \in \overline{w^{-1}L}$. By symmetry, the assertions for right quotients follow. ❑

The situation is a little different in case of complementation with respect to A^+ which is only defined for subsets of A^+. Here we have to care whether the empty word occurs when taking quotients.

Proposition 2.2. *For any $L \subseteq A^+$ and $w \in A^*$ it holds that $(w^{-1}\overline{L}^{+}) \backslash \{\epsilon\} = \overline{w^{-1}L}^{+}$. The same holds for right quotients.*

Proof. Let $L \subseteq A^+$ and $w \in A^*$. Then

$$\begin{aligned} v \in (w^{-1}\overline{L}^{+}) \backslash \{\epsilon\} &\iff wv \in \overline{L}^{+} \wedge v \neq \epsilon \\ &\iff wv \in \overline{L} \wedge v \neq \epsilon \\ &\iff v \in \overline{w^{-1}L} \wedge v \neq \epsilon \\ &\iff v \in \overline{w^{-1}L} \backslash \{\epsilon\} \\ &\iff v \in \overline{w^{-1}L}^{+} \end{aligned}$$

By symmetry, the proposition for right quotients follows. ❑

A deterministic finite automaton (dfa) F is given by $F = (A, S, \delta, s_0, S')$, where A is its input alphabet, S is its set of states, $\delta : A \times S \to S$ is its total transition function, $s_0 \in S$ is the starting state and $S' \subseteq S$ is the set of accepting states. For a dfa F we denote by $L(F)$ the language accepted by F. As usual we extend transition functions to input words, and we denote by $|F|$ the number of states of F. In this paper we will look only at dfa's where the starting state is not accepting, i.e., for any given dfa F we assume $L(F) \subseteq A^+$. We say that a state $s \in S$ has a loop $v \in A^+$ (has a v-loop, for short) if and only if $\delta(s, v) = s$. A total mapping $\delta' : S \to S$ leads to a v-loop if and only if $\delta'(s)$ has a v-loop for all $s \in S$. Every $w \in A^*$ induces a total mapping $\delta^w : S \to S$ with $\delta^w(s) =_{\text{def}} \delta(s, w)$.

For a class $\mathcal{C}$ of languages of A^+, let $\mathrm{BC}(\mathcal{C})$ be the boolean closure of $\mathcal{C}$, i.e., $\mathrm{BC}(\mathcal{C})$ is the smallest class containing $\mathcal{C}$ and being closed under union, intersection and complementation, where the latter is taken with respect to A^+. For a class $\mathcal{C}$ of languages of A^+, let $\mathrm{POL}(\mathcal{C})$ be its polynomial closure defined as follows: A language $L \subseteq A^+$ is in $\mathrm{POL}(\mathcal{C})$ if and only if L can be written as a finite union of languages $u_0 L_1 u_1 L_2 u_2 \cdots L_n u_n$, where $u_i \in A^*$, $L_i \in \mathcal{C}$ and $n \geq 0$ (if $n = 0$ then it follows that $u_0 \neq \epsilon$). The dot-depth hierarchy, first studied in [CB71], can be defined as the following family of classes of languages of A^+:

1. $\mathcal{B}_0 =_{\text{def}} \{\emptyset, A^+\}$,
2. $\mathcal{B}_{n+1/2} =_{\text{def}} \mathrm{POL}(\mathcal{B}_n)$ for $n \geq 0$, and
3. $\mathcal{B}_{n+1} =_{\text{def}} \mathrm{BC}(\mathcal{B}_{n+1/2})$ for $n \geq 0$.

For a language $L \subseteq A^+$ and a minimal n with $L \in \mathcal{B}_{n/2}$ we say that L *has dot-depth* $n/2$. To our knowledge, levels 0, $1/2$ and 1 were known to be decidable. By definition, all $\mathcal{B}_{n+1/2}$ are closed under union and it is known, that these classes are also closed under intersection [Arf87, Arf91].

It will be useful for us to look at a word $w \in A^+$ as a word over A^{k+1} by packing together each $k+1$ consecutive letters. We start with defining what the k-decomposition of a word is. In order to avoid confusion we will denote elements from A^{k+1} as $\alpha, \beta, \gamma, \ldots$ and subsets of A^{k+1} as $\Sigma, \Gamma, \ldots$. Let $k \geq 0$ and $w = a_1 a_2 \cdots a_{k+l} \in A^+$ for some $l \geq 1$. We call

$$\widehat{w} =_{\text{def}} (\alpha_1, \alpha_2, \ldots, \alpha_l)$$

the *k-decomposition* of w if $\alpha_i = a_i \cdots a_{i+k}$ for $1 \leq i \leq l$. The value of k will always be clear from the context when we use the notation $\widehat{w}$. Intuitively, k indicates by how many letters from A consecutive α_i overlap. We set $\alpha(\widehat{w}) =_{\text{def}} \{\alpha_1, \alpha_2, \ldots, \alpha_l\}$ as the set of elements from A^{k+1} occuring in the k-decomposition of w.

Next we want to define languages of words of A^+ that admit the same k-decomposition with respect to given elements and subsets of A^{k+1}.

Definition 2.3. Let $k \geq 0$. Let $\alpha_1, \ldots, \alpha_n \in A^{k+1}$ and $\Sigma_0, \Sigma_1, \ldots, \Sigma_n \subseteq A^{k+1}$ for some $n \geq 0$. For every $w \in A^+$ we say $w \in (\Sigma_0, \alpha_1, \Sigma_1, \ldots, \alpha_n, \Sigma_n)_k$ if and only if $|w| \geq k+1$, $\widehat{w} = (\beta_1, \ldots, \beta_l)$ and there exist $0 = j_0 < j_1 < j_2 < \ldots < j_n < j_{n+1} = l+1$ such that

(a) $\beta_{j_i} = \alpha_i$, for $1 \leq i \leq n$ and
(b) $\beta_j \in \Sigma_i$ for $0 \leq i \leq n$ and $j_i < j < j_{i+1}$.

If we write the expression '$(\Sigma_0, \alpha_1, \Sigma_1, \ldots, \alpha_n, \Sigma_n)_k$' we understand this as a syntactical object describing some language. Unless stated otherwise, we do not distinguish between such an object and the language it stands for, while being aware of the fact that there may be infinitely many expressions describing the same language (similarly to regular expression vs. regular languages).

So the language $(\Sigma_0, \alpha_1, \Sigma_1, \alpha_2, \Sigma_2, \ldots, \alpha_n, \Sigma_n)_k$ consists of those words $w \in A^{\geq k+1}$, whose k-decomposition starts with a number (possibly zero) of elements from Σ_0, then α_1, followed by a number (possibly zero) of elements from Σ_1, then α_2 and so on. Note that such languages only contain words of a length that admits a k-decomposition and that in case $k = 0$ we deal with the usual concatenation, e.g. $(A_0, a_1, A_1, a_2, A_2)_0 = A_0^* a_1 A_1^* a_2 A_2^*$ and $(A_0)_0 = A_0^+$.

Since the languages from the above definition will be important from Section 3 onwards, we capture for fixed $k \geq 0$ the classes of such languages with the following definition.

Definition 2.4. Let $\widetilde{\mathcal{B}}_k$ be the class of all languages that can be written as a finite union of languages L_i such that $L_i = D$ for $D \subseteq A^{\leq k}$ or $L_i = (\Sigma_0, \alpha_1, \Sigma_1, \dots, \alpha_n, \Sigma_n)_k$ where $n \geq 1$, $\alpha_1, \dots, \alpha_n \in A^{k+1}$ and $\Sigma_0, \Sigma_1, \dots, \Sigma_n \subseteq A^{k+1}$.

Note that we require here $n \geq 1$ which is no restriction since $(\Sigma_0)_k$ contains only words of length $\geq k+1$ and hence $(\Sigma_0)_k = \bigcup_{\beta \in \Sigma_0} (\Sigma_0, \beta, \Sigma_0)_k$. Furthermore, deterministic versions of these languages with requirements of the type '$\alpha_i \notin \Sigma_i$' were studied in [Sch99a]. These generalized deterministic languages were shown to be equal to languages definable by formulas of restricted temporal logic, which in turn are known to correspond to languages having locally $\mathcal{R}$–trivial syntactic semigroups and to languages having locally $\mathcal{L}$–trivial syntactic semigroups (cf. [CPP93]). Inclusion relations are easy to establish.

We introduce another convenient notation.

Definition 2.5. Let $w, v \in A^*$, $\alpha_1, \dots, \alpha_n \in A^{k+1}$ and $\Sigma_0, \dots, \Sigma_n \subseteq A^{k+1}$ for some $n \geq 0$. Then we write $(w|\Sigma_0, \alpha_1, \Sigma_1, \dots, \alpha_n, \Sigma_n|v)_k$ instead of $\big(wA^* \cap A^*v \cap (\Sigma_0, \alpha_1, \Sigma_1, \dots, \alpha_n, \Sigma_n)_k\big)$.

2.2 Characterizations of Languages of Dot-Depth One

We state characterizations of languages of dot-depth one and study the languages in $\mathcal{B}_1$ in terms of their representation over A^{k+1}.

Definition 2.6. Let $k \geq 0$ and $m \geq 1$. Then $\mathcal{B}_{1,(m,k)}$ is the class of languages $L \subseteq A^+$ that are in the Boolean algebra generated by languages L_i such that $L_i = D$ with $D \subseteq A^{<k+m}$ or

$$L_i = (w|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \dots, \alpha_m, A^{k+1}|v)_k$$

where $\alpha_1, \dots, \alpha_m \in A^{k+1}$ and $w, v \in A^k$.

The definition of these classes is motivated by a characterization of $\mathcal{B}_1$ in terms of a certain congruence $\sim_{m,k}$ with $k, m \geq 0$ introduced by I. Simon in [Sim72]. Roughly speaking, two words x and y are equivalent under this congruence if they have the same first k letters, the same last k letters and if the same m-tuples of words of length $k+1$ appear in x and y. We state and quickly reproof here known facts about this congruence, this time in terms of languages over A^{k+1}.

2.2.1 Inclusion Relations between the Classes $\mathcal{B}_{1,(m,k)}$

The next two propositions show that we obtain broader classes if we increase the parameters k and m. Let, for example, $L =_{\mathrm{def}} \big(w|A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_m, A^{k+1}|v\big)_k$. The idea of the proofs is the following: Words $w \in L$ of length $k+m$ form a subset of $A^{<k+m+1}$. Words $w \in L$ of length $> k+m$ have at least one position p in their k-decomposition which is not associated with an α_i. If we expand all α_i left (right) of p by one letter on the right side (left side, resp.), we obtain a representation with an increased k. If we explicitely add the element at position p to the description of L, we obtain a representation with an increased m. Finally, from these propositions we conclude that $\mathcal{B}_{1,(m,k)}$ is included in $\mathcal{B}_{1,(m+1,k)} \cap \mathcal{B}_{1,(m,k+1)}$.

Proposition 2.7. *Let $L =_{\mathrm{def}} \big(w|A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_m, A^{k+1}|v\big)_k$ with $k \geq 0$, $m \geq 1$, $w, v \in A^k$ and $\alpha_1, \dots, \alpha_m \in A^{k+1}$. Then*

$$L = \bigcup \Big(w'|A^{k+2}, \beta_1, A^{k+2}, \dots, \beta_m, A^{k+2}|v'\Big)_{k+1} \cup \ (L \cap A^{k+m})$$

where the union runs over all $n, w', v', \beta_1, \dots, \beta_m$ such that $w' \in wA$, $v' \in Av$, $0 \le n \le m$, $\beta_l \in \alpha_l A$ for $1 \le l \le n$ and $\beta_l \in A\alpha_l$ for $n < l \le m$.

Proof. '$\supseteq$': If u is an element of the right hand side, then either $|u| = k+m$ (in this case we are done) or $u \in \big(w'|A^{k+2}, \beta_1, A^{k+2}, \dots, \beta_m, A^{k+2}|v'\big)_{k+1}$ for suitable $n, w', v', \beta_1, \dots, \beta_m$ such that $w' \in wA$, $v' \in Av$, $0 \le n \le m$, $\beta_l \in \alpha_l A$ for $1 \le l \le n$ and $\beta_l \in A\alpha_l$ for $n < l \le m$. Choose suitable $c_1, \dots, c_{|u|} \in A$ and $\gamma_1, \dots, \gamma_{|u|-k-1} \in A^{k+2}$ such that $u = c_1 \cdots c_{|u|}$ and $\widehat{u} = \big(\gamma_1, \dots, \gamma_{|u|-k-1}\big)$. Note that the latter is a $(k+1)$-decomposition. Then there exist indices $1 \le j_1 < j_2 < \cdots < j_m \le |u| - k - 1$ such that $\beta_l = \gamma_{j_l}$ for $1 \le l \le m$. For $1 \le l \le m$ let

$$i_l =_{\text{def}} \begin{cases} j_l & : \text{ if } l \le n \\ j_l + 1 & : \text{ otherwise.} \end{cases}$$

It follows that $1 \le i_1 < i_2 < \cdots < i_m \le |u| - k$. Since $\beta_l \in \alpha_l A$ for $1 \le l \le n$ and $\beta_l \in A\alpha_l$ for $n < l \le m$, we obtain $\alpha_l = c_{i_l} \cdots c_{i_l+k}$ for $1 \le l \le m$. This shows $u \in \big(A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_m, A^{k+1}\big)_k$. Since $w' \in wA$, $v' \in Av$, we conclude $u \in L$.

'$\subseteq$': Now let $u \in L$ and choose suitable $c_1, \dots, c_{|u|} \in A$ and $\gamma_1, \dots, \gamma_{|u|-k} \in A^{k+1}$ such that $u = c_1 \cdots c_{|u|}$ and $\widehat{u} = \big(\gamma_1, \dots, \gamma_{|u|-k}\big)$. Here the latter is a k-decomposition. Thus we have indices $1 \le i_1 < i_2 < \cdots < i_m \le |u| - k$ such that $\alpha_l = \gamma_{i_l}$ for $1 \le l \le m$. If $|u| = k+m$ then $u \in L \cap A^{k+m}$ and we are done. Now let $|u| \ge k+1+m$, then there exists an index $1 \le r \le |u| - k$ such that $r \ne i_l$ for all $1 \le l \le m$. Therefore, we obtain $1 \le j_1 < j_2 < \cdots < j_m \le |u| - k - 1$ where

$$j_l =_{\text{def}} \begin{cases} i_l & : \text{ if } i_l < r \\ i_l - 1 & : \text{ otherwise (i.e., } i_l > r) \end{cases}$$

for $1 \le l \le m$. Let $n =_{\text{def}} \max\{1 \le l \le m : j_l < r\} \cup \{0\}$ and $\beta_l =_{\text{def}} c_{j_l} \cdots c_{j_l+k+1}$ for $1 \le l \le m$. Then we obtain $\beta_l \in \alpha_l A$ for $1 \le l \le n$ and $\beta_l \in A\alpha_l$ for $n < l \le m$. With $w' =_{\text{def}} p_{k+1}(u)$ and $v' =_{\text{def}} s_{k+1}(u)$ we conclude $u \in \big(w'|A^{k+2}, \beta_1, A^{k+2}, \dots, \beta_m, A^{k+2}|v'\big)_{k+1}$. ❑

Similarly, we can give a representation where m is increased.

Proposition 2.8. *Let $L =_{\text{def}} \big(w|A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_m, A^{k+1}|v\big)_k$ with $k \ge 0$, $m \ge 1$, $w, v \in A^k$ and $\alpha_1, \dots, \alpha_m \in A^{k+1}$. Then*

$$L = \bigcup \Big(w|A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_{n-1}, A^{k+1}, \beta, A^{k+1}, \alpha_n, A^{k+1}, \dots, \alpha_m, A^{k+1}|v\Big)_k \cup \; (L \cap A^{k+m})$$

where the union runs over all n, β with $1 \le n \le m+1$ and $\beta \in A^{k+1}$.

Proof. '$\subseteq$': Let $u \in L$ and $u = c_1 \cdots c_{|u|}$, $\widehat{u} = \big(\gamma_1, \dots, \gamma_{|u|-k}\big)$ for suitable $c_1, \dots, c_{|u|} \in A$, $\gamma_1, \dots, \gamma_{|u|-k} \in A^{k+1}$. Then there exist indices $1 \le i_1 < i_2 < \cdots < i_m \le |u| - k$ such that $\gamma_{i_l} = \alpha_l$ for $1 \le l \le m$. If $|u| = k+m$ then we are done. Otherwise there exists an index r with $r \ne i_l$ for all $1 \le l \le m$. Let $n =_{\text{def}} \min\{1 \le l \le m : r < i_l\} \cup \{m+1\}$ and $\beta =_{\text{def}} c_r c_{r+1} \cdots c_{r+k}$. It follows that $u \in \big(w|A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_{n-1}, A^{k+1}, \beta, A^{k+1}, \alpha_n, A^{k+1}, \dots, \alpha_m, A^{k+1}|v\big)_k$.

'$\supseteq$': Now let u be an element of the right hand side. If $|u| = k+m$ then we are done. Otherwise $u \in \big(w|A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_{n-1}, A^{k+1}, \beta, A^{k+1}, \alpha_n, A^{k+1}, \dots, \alpha_m, A^{k+1}|v\big)_k$ for suitable $\beta \in A^{k+1}$ and $1 \le n \le m+1$. It follows that $u \in \big(w|A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_m, A^{k+1}|v\big)_k$. ❑

Note that in the two previous propositions we did not need full Boolean operations, but only unions to express the increase in the parameters k and m (intersection is not needed, since $L \cap A^{k+m}$ is a language of the form $D \subseteq A^{<k+m+1}$).

Lemma 2.9. *Let $k \geq 0$ and $m \geq 1$. Then it holds that*

1. $\mathcal{B}_{1,(m,k)} \subseteq \mathcal{B}_{1,(m,k+1)}$ *and*
2. $\mathcal{B}_{1,(m,k)} \subseteq \mathcal{B}_{1,(m+1,k)}$.

Proof. We start with Statement 1. Since $\mathcal{B}_{1,(m,k+1)}$ is a Boolean algebra, it suffices to show that $\big(w|A^{k+1}, \alpha_1, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v\big)_k \in \mathcal{B}_{1,(m,k+1)}$ for every $w, v \in A^k$, $\alpha_1, \ldots, \alpha_m \in A^{k+1}$, and $D \in \mathcal{B}_{1,(m,k+1)}$ for $D \subseteq A^{<k+m}$. The latter is clear and to see the first part we apply Proposition 2.7. Note that $L \cap A^{k+m}$ is a subset $D \subseteq A^{<k+1+m}$. Statement 2 is proved completely analogously with help of Proposition 2.8. ❑

2.2.2 The Structure of $\mathcal{B}_1$

The next theorem establishes an important connection between $\mathcal{B}_1$ and $\mathcal{B}_{1,(m,k)}$, i.e., the fact that the classes $\mathcal{B}_{1,(m,k)}$ refine the class $\mathcal{B}_1$. Originally, this theorem was stated for languages of A^* and in terms of the above mentioned congruence $\sim_{m,k}$ in [Sim72]. For an overview on hierarchies which result from fixing one or the other parameter see [Brz76].

Theorem 2.10. *Let $L \subseteq A^+$. Then $L \in \mathcal{B}_1$ if and only if there exist $k \geq 0$ and $m \geq 1$ such that $L \in \mathcal{B}_{1,(m,k)}$.*

Proof. We need to swap between k-decompositions and the usual concatenation.
'if-part': Suppose some language $L \in \mathcal{B}_{1,(m,k)}$ for some $k \geq 0$ and $m \geq 1$ is given. Then L is a Boolean combination of languages $D \subseteq A^{<k+m}$ and languages L' with

$$L' = (w|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k$$

where $\alpha_1, \ldots, \alpha_m \in A^{k+1}$ and $w, v \in A^k$. Since $\mathcal{B}_1$ is a Boolean algebra which contains all finite sets of non-empty words, we only have to show that each L' is in $\mathcal{B}_1$. We have to consider that the elements $\alpha_i \in A^{k+1}$ in the description of L' may overlap and we have to express this with usual concatenations.

For a fixed $k \geq 0$ and for given $l \geq 1$ and $\beta_1, \ldots, \beta_l \in A^{k+1}$ we define the set $sh(\beta_1, \beta_2, \ldots, \beta_l)$ of *shuffle* words. A word u belongs to $sh(\beta_1, \beta_2, \ldots, \beta_l)$ if and only if $\widehat{u} = (\gamma_1, \ldots, \gamma_n)$ for some $n \geq 1$ and there exist $1 \leq j_1 < j_2 < \ldots < j_l \leq n$ with $\gamma_{j_i} = \beta_i$ for $1 \leq i \leq l$ and $j_{i+1} - j_i \leq k$ for $1 \leq i \leq l-1$. The latter condition ensures that all occurences of β_i in u overlap, e.g., we have $sh(abb, bbc) = A^*abbcA^* \cup A^*abbbcA^*$. Moreover we see that each $sh(\beta_1, \ldots, \beta_l)$ is a finite union of languages of the form A^*xA^* with $|x| \geq k+l$. To see this note that there is only a finite number of possibilities to arrange all β_i and each such possibility determines one factor in a word of length $\geq k+l$. Hence we can rewrite A^*xA^* as $A^+xA^+ \cup F$ with a suitable finite set $F \subseteq A^{|x|} \cup A^{|x|+1} \subseteq A^+$. Thus we have for given $l \geq 1$ and $\beta_1, \ldots, \beta_l \in A^{k+1}$ that $sh(\beta_1, \ldots, \beta_l) \in \mathcal{B}_{1/2}$.

Now we want to express L' in terms of concatenations of such sets of shuffle words and therefore rewrite L' as $(wA^+ \cap A^+v) \cap (L'' \cup sh(\alpha_1, \ldots, \alpha_m))$ where

$$L'' =_{\text{def}} \bigcup_{\substack{1 \leq n \leq m \\ 1 \leq i_1 < i_2 < \ldots < i_n \leq m}} sh(\alpha_1, \ldots, \alpha_{i_1}) sh(\alpha_{i_1+1}, \ldots, \alpha_{i_2}) \cdots sh(\alpha_{i_n+1}, \ldots, \alpha_m).$$

With the set L'' we guess a number n of positions where the occuring witnessing α_i from the description of L' do not overlap. We may rewrite terms of the form A^+A^+ as $\bigcup_{a \in A} aA^+$. Since $\mathcal{B}_{1/2}$ is closed under concatenation (as can be seen from the definition) we obtain $L'' \in \mathcal{B}_{1/2}$ and together $L' \in \mathcal{B}_1$.

'only-if-part': Now let $L \in \mathcal{B}_1$ be given, hence L is a Boolean combination of languages $L_j \subseteq A^+$ with $L_j = u_0A^+u_1A^+u_2 \cdots A^+u_n$ or $L_j = \emptyset$. If $L_j = \emptyset$ then $L_j \in \mathcal{B}_{1,(m,k)}$ for all $m \geq 1$, $k \geq 0$, because it is of the form $D \subseteq A^{<k+m}$. So from now on suppose L_j is of the form

$u_0A^+u_1A^+u_2\cdots A^+u_n$. We may assume that for $1 \le i \le n-1$ we have $u_i \ne \epsilon$ since otherwise we substitute $A^+\epsilon A^+$ in the concatenation by $\bigcup_{a\in A} aA^+$ and distribute the union over the concatenation. We also may assume $u_0 \ne \epsilon$ which is clear if $n = 0$. If $n > 0$ we rewrite L_j as $\bigcup_{a\in A} aA^+u_1A^+u_2\cdots A^+u_n \;\cup\; \bigcup_{a\in A} au_1A^+u_2\cdots A^+u_n$. The same argument shows that we may also assume $u_n \ne \epsilon$, so together we have for $0 \le i \le n$ that $u_i \ne \epsilon$ in each L_j.

We will show that we can determine for each set L_j some $m'_j \ge 1, k'_j \ge 0$ such that $L_j \in \mathcal{B}_{1,(m'_j,k'_j)}$. Then by Lemma 2.9 we have for all j that $L_j \in \mathcal{B}_{1,(m_{\max},k_{\max})}$ with $m_{\max} =_{\text{def}} \max_j m'_j$ and $k_{\max} =_{\text{def}} \max_j k'_j$. We then may perform the Boolean operations on the sets L_j in this Boolean algebra to see $L \in \mathcal{B}_{1,(m_{\max},k_{\max})}$. So let us fix a set L_j. We prove by induction on i that for $0 \le i \le n$ there are $m_i \ge 1, k_i \ge 0$ such that $u_0A^+u_1\cdots A^+u_i$ can be written as a finite union of languages $D \subseteq A^{<k_i+m_i}$ with languages

$$(w|A^{k_i+1}, \alpha_1, A^{k_i+1}, \alpha_2, A^{k_i+1}, \ldots, \alpha_{m_i}, A^{k_i+1}|v)_{k_i} \tag{1}$$

where $\alpha_1, \ldots, \alpha_{m_i} \in A^{k_i+1}$ and $w, v \in A^{k_i}$.

We prove the induction base for $i = 0, 1$. First assume $i = 0$. Then for $m_0 =_{\text{def}} 1$ and $k_0 =_{\text{def}} |u_0|$ we have that $\{u_0\}$ is a finite set D as above since $0 < |u_0| < k_0 + m_0$. Now let $i = 1$, $L' =_{\text{def}} u_0A^+u_1$ and $k_1 =_{\text{def}} |u_0| + |u_1|$. We see that $L' = \bigcup_{w,v}(w|A^{k_1+1}|v)_{k_1}$ where the union goes over all $w \in A^{k_1}$ with $p_{|u_0|}(w) = u_0$ and over all $v \in A^{k_1}$ with $s_{|u_1|}(v) = u_1$. Note that L' contains only words of length $\ge k_1 + 1$. So we may rewrite the above union as $\bigcup_{w,v,\alpha}(w|A^{k_1+1}, \alpha, A^{k_1+1}|v)_{k_1}$ where additionally the union goes over all $\alpha \in A^{k_1+1}$. We define $m_1 =_{\text{def}} 1$ to finish the induction base.

Let $2 \le i < n$. By hypothesis we can write $u_0A^+u_1\cdots A^+u_i$ as a finite union of languages as stated in (1) with languages $D \subseteq A^{<k_i+m_i}$. We will show how to determine m_{i+1} and k_{i+1} when concatenating each single member of this union with A^+u_{i+1}. We then obtain the desired values for $u_0A^+u_1\cdots A^+u_iA^+u_{i+1}$ as usual by maximizing over the occuring m_{i+1} and k_{i+1}, respectively.

So first look at $D \cdot A^+u_{i+1} = \bigcup_{w\in D} wA^+u_{i+1}$ where $D \subseteq A^{<k_i+m_i}$. How to choose in this case m_{i+1} and k_{i+1} has already been solved in the induction base or before when dealing with $\emptyset$. Now we have to deal with

$$L' =_{\text{def}} (w|A^{k_i+1}, \alpha_1, A^{k_i+1}, \ldots, \alpha_{m_i}, A^{k_i+1}|v)_{k_i} \cdot A^+u_{i+1} \tag{2}$$

where $\alpha_1, \ldots, \alpha_{m_i} \in A^{k_i+1}$ and $w, v \in A^{k_i}$.

Case 1. Assume $l =_{\text{def}} |u_{i+1}| \le k_i$. Then we claim that L' is equal to

$$\bigcup \; (w|A^{k_i+1}, \alpha_1, A^{k_i+1}, \ldots, \alpha_{m_i}, A^{k_i+1}, va, A^{k_i+1}, \beta_1, A^{k_i+1}, \ldots, \beta_l, A^{k_i+1}|z)_{k_i}$$

where the union goes over all $a \in A$, $z \in A^{k_i}$ with $s_l(z) = u_{i+1}$ and for $1 \le j \le l$ over all $\beta_j \in A^{k_i+1}$. To see this take some word $xvayu_{i+1} \in L'$ with $ay \in A^+$ and x is such that v is the suffix occurence from the above represention in (2). Then $|vayu_{i+1}| \ge k_i + 1 + l$ and we look at $v\widehat{ayu_{i+1}}$. This k_i-decomposition starts with va, followed by some $\beta_1, \ldots \beta_l \in A^{k_i+1}$, and ends with some $z \in A^{k_i}$ having l-suffix u_{i+1}. For the reverse inclusion it suffices to note that u_{i+1} does not begin before va ends, which is due to the l occurences of β_j right of va.

Case 2. Assume $l =_{\text{def}} |u_{i+1}| > k_i$. Then we provide by Proposition 2.7 a representation of $(w|A^{k_i+1}, \alpha_1, A^{k_i+1}, \ldots, \alpha_{m_i}, A^{k_i+1}|v)_{k_i}$ for a sufficiently large k'_i such that $l \le k'_i$, and apply Case 1. How to deal with the additionally occuring finite sets has also been shown before. ❑

Among others, a hierarchy in $\mathcal{B}_1$ obtained by fixing the "level" k of a language within this class has been studied. We do here the same thing in terms of our notations.

Definition 2.11. Let $k \ge 0$. Then $\mathcal{B}_{1,k} =_{\text{def}} \bigcup_{m\ge 1} \mathcal{B}_{1,(m,k)}$.

Proposition 2.12. *It holds that*

1. $\mathcal{B}_{1,k} \subseteq \mathcal{B}_{1,k+1}$ *for each* $k \geq 0$*, and*
2. $\mathcal{B}_1 = \bigcup_{k\geq 0} \mathcal{B}_{1,k}$.

Proof. To see Statement 1 let $L \in \mathcal{B}_{1,k}$. Then by definition $L \in \mathcal{B}_{1,(m,k)}$ for some fixed $m \geq 1$ and by Lemma 2.9 we have that $L \in \mathcal{B}_{1,(m,k+1)} \subseteq \mathcal{B}_{1,k+1}$. For Statement 2 note that by Theorem 2.10 we have $\mathcal{B}_1 = \bigcup_{k\geq 0} \bigcup_{m\geq 1} \mathcal{B}_{1,(m,k)}$. ❑

We may view $\mathcal{B}_{1,k}$ as the Boolean closure of languages, where m is not fixed.

Proposition 2.13. *Let* $k \geq 0$*. Then* $\mathcal{B}_{1,k}$ *is the Boolean algebra generated by languages* L_i *such that* $L_i = D$ *with* $D \subseteq A^{<k+m}$ *for some* $m \geq 1$*, or*

$$L_i = (w|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \dots, \alpha_m, A^{k+1}|v)_k$$

where $m \geq 1, \alpha_1, \dots, \alpha_m \in A^{k+1}$ *and* $w, v \in A^k$.

Proof. Let $L \in \mathcal{B}_{1,k}$, then by definition $L \in \mathcal{B}_{1,(m,k)}$ for some fixed $m \geq 1$. Thus it is a Boolean combination of languages of the desired form.

Conversely, let L be a Boolean combination of languages $L_1, \dots, L_n$ each having the form as stated in the proposition. Denote for each such language L_i the respective parameter m as m_i. Clearly each L_i is in $\mathcal{B}_{1,(m_i,k)}$. By Lemma 2.9 we know that for $1 \leq i \leq n$ we have $L_i \in \mathcal{B}_{1,(m_{\max},k)}$ with $m_{\max} =_{\text{def}} \max_{1\leq i\leq n} m_i$. Since $\mathcal{B}_{1,(m_{\max},k)}$ is a Boolean algebra we obtain $L \in \mathcal{B}_{1,(m_{\max},k)} \subseteq \mathcal{B}_{1,k}$. ❑

The hierarchy of classes $\{ \mathcal{B}_{1,k} \mid k \geq 0 \}$ is (up to the empty word) the so-called γ-hierarchy from [Brz76]. Various characterizations were given by J. Stern in [Ste85a] and also in [Sch99b]. Let us now define some relations on words and start with a generalization of the well-known subword relation, which we extend to k-decompositions of words.

Definition 2.14. Let $k \geq 0$. Let $u, v \in A^{>k}$ and let $\widehat{u} = (\alpha_1, \alpha_2, \dots, \alpha_m)$ and $\widehat{v} = (\beta_1, \beta_2, \dots, \beta_n)$. We define

$$u \lhd_k v \iff_{\text{def}} \text{there exist } 1 \leq j_1 < j_2 < \cdots < j_m \leq n \text{ such that } \beta_{j_i} = \alpha_i.$$

Moreover, if $|u| \leq k$ we write $u \lhd_k v$ if and only if $u = v$.

If we additionally require that $p_k(u) = p_k(v)$ and $s_k(u) = s_k(v)$ we denote this by $u \preceq_k v$. Note that these relations are the usual subword relations in case $k = 0$. Obviously, we have that $u \preceq_k v$ implies $u \lhd_k v$. A sequence of words $w_1, w_2, \dots$ is called a $\lhd_k$– chain, if $w_i \lhd_k w_{i+1}$ for all words in the sequence (similarly, we define $\preceq_k$– chains). We say that such a sequence has an alternation with respect to some given language L if $w_i \in L \iff w_{i+1} \notin L$ for some i. Both relations are reflexive and transitive.

In [Ste85a] a relation similar to $\preceq_k$ was used to characterize the classes $\mathcal{B}_{1,k}$ in terms of a finiteness condition on the number of alternations in chains with respect to this relation. There, additionally to $\preceq_k$, it was required that all words have the same prefix of length $2k$. For decision algorithms for $\mathcal{B}_1$ derived from the characterizations in [Ste85a] see [Ste85b, CH91, Sch99b].

Both relations defined so far (and the one defined in [Ste85a]) share an interesting property. We will concentrate here on $\preceq_k$ since it is more suitable for our purposes, namely, we want to establish in Subsection 2.4 a convenient representation for languages in $\mathcal{B}_{1,k}$.

2.3 Well Partial Orderings

Let $k \geq 0$. We show that A^+ together with $\lhd_k$ is a well partial ordered set. The same holds for $\preceq_k$ (and also Stern's relation) since restrictions of constant size in a word do not affect this notion.

For several equivalent properties, which may be used for defining well partial orderings, see [CK96, SS83]. We will show here that in A^+ there exists neither an infinite strictly descending $\lhd_k$– chain of words nor an infinite set of pairwise incomparable elements with respect to $\lhd_k$. By [CK96] this is equivalent to saying that for every nonempty subset of A^+ the set of minimal elements with respect to $\lhd_k$ in this subset is nonempty and finite.

First observe, by a typical length argument (namely that $u \lhd_k v$ with $u \neq v$ implies $|u| < |v|$), that we only have to show that any set of pairwise incomparable elements is finite. We will make use of the following arguments in the proof below.

Proposition 2.15. *Let $k \geq 0$, $w \in A^k$ and $a \in A$. Then it holds that*

1. *$u \lhd_k au$ for every $u \in A^{>k}$, and*
2. *$wx \lhd_k wy \implies awx \lhd_k awy$ for every $x, y \in A^*$.*

Proof. Statement 1 is obvious. To see 2 note that if $x = \epsilon$ then it must be that $y = \epsilon$. If $x \neq \epsilon$ then we only need to observe that aw as the first element of $\widehat{awx}$ appears also as the first element in $\widehat{awy}$. ❑

In case $k = 0$ (i.e., the subword relation, also called division ordering) the following theorem is a fundamental theorem from Higman [Hig52].

Theorem 2.16. *Let $k \geq 0$. Any subset of A^+ of pairwise incomparable words w.r.t. $\lhd_k$ is finite.*

Proof. We adapt the proof for the subword relation as stated in [SS83] for our purposes. So assume that there are infinite subsets of A^+ all consisting of pairwise incomparable words w.r.t. $\lhd_k$. In particular there exist infinite sequences $\{f_i\}$ of words such that from $i < j$ it follows that $f_i \not\lhd_k f_j$. We will show that this is not true. For this consider any such sequence $\{f_i\}$ and note that all words in such a sequence must be different since $\lhd_k$ is reflexive. We choose from these sequences an 'earliest' sequence $\{u_i\}$ as follows: Let u_1 be a shortest word beginning some sequence $\{f_i\}$, then let u_2 be a shortest second word of any sequence $u_1, f_2, f_3 \ldots$, then let u_3 be a shortest third word of any sequence $u_1, u_2, f_3 \ldots$, and so on. Clearly, also for $\{u_i\}$ it holds that from $i < j$ it follows $u_i \not\lhd_k u_j$. Since we have a finite alphabet there are words $u_{i_1} = awg_1, u_{i_2} = awg_2, \ldots$ with $i_1 < i_2 < \ldots$ for some $a \in A$ and some $w \in A^k$. Note that none of these g_j can be ϵ since $aw \lhd_k awv$ for arbitrary v.

Now we look at the sequence $u_1, u_2, \ldots, u_{i_1-1}, wg_1, wg_2, \ldots$ Denote this new sequence as $\{x_i\}$ which is 'earlier' than $\{u_i\}$ since $|wg_1| < |u_{i_1}|$. In order to obtain a contradiction to our construction we need to show that for all i, j with $i < j$ we can conclude $x_i \not\lhd_k x_j$. If $i, j \in \{1, \ldots, i_1 - 1\}$ then this is clear by the same property for $\{u_i\}$. Now suppose $i \in \{1, \ldots, i_1 - 1\}$ and $j \geq i_1$ and assume $x_i \lhd_k x_j$ where $x_i = u_i$ and $x_j = wg_l$ for some $l \geq 1$. Since $wg_l \in A^{>k}$ we have by Proposition 2.15.1 that $wg_l \lhd_k awg_l = u_{i_l}$ and together $u_i \lhd_k u_{i_l}$, a contradiction. Finally let $i, j \geq i_1$ with $i < j$. Assume $x_i \lhd_k x_j$ with $x_i = wg_l$ and $x_j = wg_m$ for some $l < m$. By Proposition 2.15.2 it follows that $awg_l \lhd_k awg_m$, so $u_{i_l} \lhd_k u_{i_m}$, again a contradiction. ❑

Theorem 2.17. *Let $k \geq 0$. Any subset of A^+ of pairwise incomparable words w.r.t. $\preceq_k$ is finite.*

Proof. Suppose there exist infinite subsets of A^+ of pairwise incomparable words w.r.t. $\preceq_k$. Then there is also an infinite subset L such that all words in L have the same k-prefix and k-suffix. So the words in L are pairwise incomparable w.r.t. $\lhd_k$ contradicting Theorem 2.16. ❑

2.4 The Structure of $\mathcal{B}_{1,k}$

If $u \preceq_k v$ we call v the k-extension from u and we define the set of k-extensions from a word u as $\langle u \rangle_k =_{\text{def}} \{ v \mid u \preceq_k v \}$. We can describe this set over A^{k+1}.

Proposition 2.18. *Let $k \geq 0$. If $u \in A^{\leq k}$ then $\langle u \rangle_k = \{u\}$. If $u \in A^{>k}$ and $\widehat{u} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ then*

$$\langle u \rangle_k = \Big(p_k(u)|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_m, A^{k+1}|s_k(u)\Big)_k$$

Proof. Clearly, any word from the right hand side is a k-extension of u. On the other hand, any k-extension of u must have u's k-prefix and k-suffix and all elements of u's k-decomposition must occur in this ordering. ❑

We write $\langle L \rangle_k =_{\text{def}} \bigcup_{u \in L} \langle u \rangle_k$ for the set of all words which are k-extensions of some word in a given language L. The following duality holds.

Proposition 2.19. *For $k \geq 0$ and $u, v \in A^+$ it holds that $u \preceq_k v \iff \langle u \rangle_k \supseteq \langle v \rangle_k$.*

The implication from left to right is due to the transitivity of $\preceq_k$. The other implication follows from its reflexivity. Next, we give another description of the languages in $\mathcal{B}_{1,k}$ over A^{k+1}. The difference to Proposition 2.13 is here that we can get rid of the parameter m occuring in the size of the finite set D. The main argument comes from the fact that $\preceq_k$ is a well partial ordering.

Theorem 2.20. *Let $k \geq 0$ and $L \subseteq A^+$. Then L is in $\mathcal{B}_{1,k}$ if and only if L is in the Boolean algebra generated by languages L_i such that $L_i = D$ with $D \subseteq A^{\leq k}$ or*

$$L_i = (w|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k$$

where $m \geq 1, \alpha_1, \ldots, \alpha_m \in A^{k+1}$ and $w, v \in A^k$.

Proof. We see that the languages mentioned in the theorem have a form which is a special case of the form of the languages from Proposition 2.13, so there is nothing to do to see the if-part.

To see the only-if part let $L \in \mathcal{B}_{1,k}$ which is by Proposition 2.13 a Boolean combination of languages L_i such that $L_i = D$ with $D \subseteq A^{\leq k+m_i}$ for some $m_i \geq 1$ or

$$L_i = (w_i|A^{k+1}, \alpha_1^i, A^{k+1}, \alpha_2^i, A^{k+1}, \ldots, \alpha_{m_i}^i, A^{k+1}|v_i)_k$$

with $m_i \geq 1, \alpha_1^i, \ldots, \alpha_{m_i}^i \in A^{k+1}$ and $w_i, v_i \in A^k$. We only have to care about the finite sets D_i, i.e., we have to show that we can express the words $u \in D_i$ with $|u| > k$ by Boolean combinations of languages of the form $(w|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_l, A^{k+1}|v)_k$ for some $l \geq 1, \alpha_1, \ldots, \alpha_l \in A^{k+1}$ and $w, v \in A^k$.

Observe that $u \in \langle u \rangle_k$ which has the desired form by Proposition 2.18. We want to show that we can subtract all other words in $\langle u \rangle_k$, i.e., describe $S =_{\text{def}} \langle u \rangle_k \backslash \{u\}$ by a Boolean combination of the requested type. Let $y \in S$ and $x \in \langle S \rangle_k$ with $y \preceq_k x$. Then $u \preceq_k y$ and $y \neq u$. It follows that $u \preceq_k y \preceq_k x$ and $x \neq u$ (by a length argument), hence $S = \langle S \rangle_k$. Now we can write

$$S = \langle S \rangle_k = \bigcup_{x \in S} \langle x \rangle_k$$

and Theorem 2.17 tells us that this union is in fact finite. ❑

So each class $\mathcal{B}_{1,k}$ can be understood as the Boolean closure of the class of all languages having the form from the previous theorem. We name the class of these languages.

Definition 2.21. Let $k \geq 0$. A language $L \subseteq A^+$ belongs to $\mathcal{B}_{1/2,k}$ if and only if it can be written as a finite union of languages L_i such that $L_i = D$ with $D \subseteq A^{\leq k}$ or

$$L_i = (w|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k$$

where $m \geq 1, \alpha_1, \ldots, \alpha_m \in A^{k+1}$ and $w, v \in A^k$.

Then $\mathcal{B}_{1,k} = \mathrm{BC}(\mathcal{B}_{1/2,k})$ for each $k \geq 0$ by Theorem 2.20. As a further application of Theorem 2.17 we obtain that the closure of an arbitrary language under k-extensions is in $\mathcal{B}_{1/2,k}$. In particular, we obtain a regular language. We state the following theorem since it may be interesting on its own. It will not be cited in the sequel.

Theorem 2.22. *Let $k \geq 0$ and $L \subseteq A^+$. Then $\langle L\rangle_k \in \mathcal{B}_{1/2,k}$.*

Proof. By definition, $\langle L\rangle_k = \bigcup_{u\in L}\langle u\rangle_k$ and by Proposition 2.18 we have $\langle u\rangle_k \in \mathcal{B}_{1/2,k}$. By Theorem 2.17 we conclude that only a finite number of different elements from L strictly contribute to the union. ❑

2.5 Closure Properties of $\mathcal{B}_{1/2,k}$ and $\mathcal{B}_{1,k}$

We proof closure properties that will be useful in forthcoming sections and we start with some general remarks. For some word $w \in A^*$ and a language $L \subseteq A^{\geq |w|+1}$ of a given class we will show that the left quotients of L w.r.t. w belongs to this class too. We understand this as the closure of this class w.r.t. left quotients. In the proof we can even restict ourselves to the case $|w| = 1$ since $(av)^{-1}L = v^{-1}(a^{-1}L)$, and if $L \subseteq A^{\geq k+1}$ for some $k \geq 1$ then $a^{-1}L \subseteq A^{\geq k}$. Moreover, from the closure of a class under reversion and under left quotients the closure under right quotients follows.

Proposition 2.23. *Let $k \geq 0$, $y \in A^*$ and $L, L' \subseteq \mathcal{B}_{1/2,k}$. Then*

1. *$L \cap L' \in \mathcal{B}_{1/2,k}$ and $L \cup L' \in \mathcal{B}_{1/2,k}$,*
2. *$L^R \in \mathcal{B}_{1/2,k}$ and*
3. *if $L \subseteq A^{\geq |y|+1}$ then $y^{-1}L, Ly^{-1} \in \mathcal{B}_{1/2,k}$.*

Proof. 1. The closure of $\mathcal{B}_{1/2,k}$ under union is by definition. So take two languages from $\mathcal{B}_{1/2,k}$ each of which being a finite union of languages L_i such that $L_i = D$ with $D \subseteq A^{\leq k}$ or

$$L_i = (w|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k$$

where $m \geq 1, \alpha_1, \ldots, \alpha_m \in A^{k+1}$, $w, v \in A^k$, and look at their intersection. First we may distribute the intersection over all unions. Clearly, the intersection of two finite sets D as above is again a subset of $A^{\leq k}$. If we have to consider mixed terms with a set D and an infinite set then the intersection is empty. Therefore we only have to care about intersections $L_1 \cap L_2$ where $L_i = (w_i|A^{k+1}, \alpha^i_1, A^{k+1}, \ldots, \alpha^i_{m_i}, A^{k+1}|v_i)_k$ for $i = 1, 2$. If $w_1 \neq w_2$ or $v_1 \neq v_2$ then again the intersection is empty. So we may assume $w_1 = w_2$ and $v_1 = v_2$ and we state that $L_1 \cap L_2 = L'$ with

$$L' =_{\mathrm{def}} \bigcup\ (w_1|A^{k+1}, \beta_1, A^{k+1}, \ldots, A^{k+1}, \beta_m, A^{k+1}|v_1)_k$$

where the union goes over all tuples $(\beta_1, \ldots, \beta_m)$ with $1 \leq m \leq m_1+m_2$ and $\beta_i \in A^{k+1}$ for $1 \leq i \leq m$ such that we find $(\alpha^1_1, \ldots, \alpha^1_{m_1})$ and $(\alpha^2_1, \ldots, \alpha^2_{m_2})$ in this tuple. Hence there exist $1 \leq i_1 < i_2 < \ldots <$

$i_{m_1} \leq m$ and $1 \leq j_1 < j_2 < \ldots < j_{m_2} \leq m$ such that $(\alpha_1^1, \alpha_2^1, \ldots, \alpha_{m_1}^1) = (\beta_{i_1}, \beta_{i_2}, \ldots, \beta_{i_{m_1}})$ and $(\alpha_1^2, \alpha_2^2, \ldots, \alpha_{m_2}^2) = (\beta_{j_1}, \beta_{j_2}, \ldots, \beta_{j_{m_2}})$.

Observe that $L' \in \mathcal{B}_{1/2,k}$ since m is bound by a constant. In the k-decomposition of each word $u \in L'$ we find by the choice of $(\beta_1, \ldots, \beta_m)$ the required elements of A^{k+1} to see that u is in L_1 and also in L_2, so $u \in L_1 \cap L_2$.

Now let $u \in L_1 \cap L_2$ and $\widehat{u} = (\gamma_1, \ldots, \gamma_l)$ for some $l \geq 1$. Thus there must be sequences of indices $1 \leq i_1 < i_2 < \ldots < i_{m_1} \leq l$ and $1 \leq j_1 < j_2 < \ldots < j_{m_2} \leq l$ such that $(\alpha_1^1, \ldots, \alpha_{m_1}^1) = (\gamma_{i_1}, \ldots, \gamma_{i_{m_1}})$ and $(\alpha_1^2, \ldots, \alpha_{m_2}^2) = (\gamma_{j_1}, \ldots, \gamma_{j_{m_2}})$. Then there is also a sequence of indices $1 \leq l_1 < l_2 < \ldots < l_m \leq l$ with $m \leq m_1 + m_2$ and $\{l_1, \ldots, l_m\} = \{i_1, \ldots, i_{m_1}, j_1, \ldots, j_{m_2}\}$. So $(\gamma_{l_1}, \gamma_{l_2}, \ldots, \gamma_{l_m})$ is one of the tuples $(\beta_1, \ldots, \beta_m)$ occuring in the definition of L' and we see that $u \in L'$.

2. Since reversion distributes in particular over unions it is enough to observe that for $D \subseteq A^{\leq k}$ we have $D^R \subseteq A^{\leq k}$ and if $L = (w|A^{k+1}, \alpha_1, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k$ where $m \geq 1, \alpha_1, \ldots, \alpha_m \in A^{k+1}$ and $w, v \in A^k$, then

$$L^R = (v^R|A^{k+1}, \alpha_m^R, A^{k+1}, \alpha_{m-1}^R, A^{k+1}, \ldots, \alpha_1^R, A^{k+1}|w^R)_k.$$

3. By Proposition 2.1 we may assume that $L = D$ with $D \subseteq A^{\leq k}$ or $L = L'$ with $L' = (w|A^{k+1}, \alpha_1, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k$ where $m \geq 1, \alpha_1, \ldots, \alpha_m \in A^{k+1}$ and $w, v \in A^k$. In light of the remarks at the beginning of this subsection we only need to show that if $L \subseteq A^{\geq 2}$ and $a \in A$ then $a^{-1}L \in \mathcal{B}_{1/2,k}$. Note that it must be that $L', D \subseteq A^{\geq 2}$.

Suppose $k = 0$. Then sets D do not occur and we have that $L = (A, a_1, A, \ldots, a_m, A)_0$ with $m \geq 2$ and $a_1, \ldots, a_m \in A$. If $a \neq a_1$ then $a^{-1}(A, a_1, A, \ldots, a_m, A)_0 = (A, a_1, A, \ldots, a_m, A)_0 \in \mathcal{B}_{1/2,k}$, otherwise $a^{-1}(A, a_1, A, \ldots, a_m, A)_0 = (A, a_2, A, \ldots, a_m, A)_0 \in \mathcal{B}_{1/2,k}$.

Now suppose that $k \geq 1$. Then $a^{-1}D$ is a subset of $A^{\leq k}$ as required. If the prefix w of L' does not start with a then $a^{-1}L' = \emptyset \in \mathcal{B}_{1/2,k}$. So let $w = au$, and let us for now assume that au is not a prefix of α_1, i.e., α_1 is not the first element in the k-decomposition of any word in L'. Then we have that

$$a^{-1}L' = \bigcup_{b \in A} (ub|A^{k+1}, \alpha_1, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k \in \mathcal{B}_{1/2,k}.$$

Now let au be a prefix of α_1. Then we additionally have to consider those words in L' starting with α_1. For $m \geq 2$ we claim $a^{-1}L' \in \mathcal{B}_{1/2,k}$, because

$$\begin{aligned} a^{-1}L' \quad &= \quad \bigcup_{b \in A} (ub|A^{k+1}, \alpha_1, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k \cup \\ & \qquad (s_k(\alpha_1)|A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k. \end{aligned} \tag{3}$$

To see the inclusion from left to right note that any word x with $ax \in L'$ and $\widehat{ax} = (\alpha_1, \ldots)$ is now included in the additional set. For the other inclusion we observe that $as_k(\alpha_1) = \alpha_1$ and that $s_k(\alpha_1)$ starts with u. Finally, if $m = 1$, we proceed as in the previous case, but we have to rewrite the set (3) as $\bigcup_{\beta \in A^{k+1}} (s_k(\alpha_1)|A^{k+1}, \beta, A^{k+1}|v)_k \cup (\{s_k(\alpha_1)\} \cap \{v\})$ to see that $a^{-1}L' \in \mathcal{B}_{1/2,k}$.

❑

By De Morgan's law, the classes $\mathrm{co}\mathcal{B}_{1/2,k}$ are closed under union and intersection as well. Next we show the closure of each $\mathcal{B}_{1,k}$ under reversion and quotients.

Proposition 2.24. *Let $k \geq 0$, $y \in A^*$ and $L, L' \in \mathcal{B}_{1,k}$. Then*

1. $\overline{L}^{+}, L \cap L', L \cup L', L \backslash L' \in \mathcal{B}_{1,k}$,
2. $L^R \in \mathcal{B}_{1,k}$ *and*

3. *if* $L \subseteq A^{\geq |y|+1}$ *then* $y^{-1}L, Ly^{-1} \in \mathcal{B}_{1,k}$.

Proof. 1. By Proposition 2.13, for each $k \geq 0$ the class $\mathcal{B}_{1,k}$ is a Boolean algebra.

2. Let $L \in \mathcal{B}_{1,k}$, so by Theorem 2.20 and Definition 2.21 we know that L is a Boolean combination of languages $L_i \in \mathcal{B}_{1/2,k}$. Since reversion distributes over Boolean operations we can apply Proposition 2.23 to each L_i.

3. Assume that L is a Boolean combination of languages $L'_i \in \mathcal{B}_{1/2,k}$ with $L'_i = D_i$ for $D_i \subseteq A^{\leq k}$ or $L'_i = L_i$ with $L_i = (w|A^{k+1}, \alpha_1, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k$ where $m \geq 1, \alpha_1, \ldots, \alpha_m \in A^{k+1}$ and $w, v \in A^k$. Again, we only need to show that if $L \subseteq A^{\geq 2}$ and $a \in A$ then $a^{-1}L \in \mathcal{B}_{1,k}$. Of course we would like to apply Proposition 2.23, but we only know that $L \subseteq A^{\geq 2}$ and we do not know whether each $D_i, L_i \subseteq A^{\geq 2}$.

We have to argue more detailed and rewrite L in disjunctive normalform as a finite union of languages K_j each of which is a finite intersection of languages $E_{j,l}$, which in turn are of the type L_i or D_i from above, or $\overline{L_i}^+$ or $\overline{D_i}^+$. Clearly, each $K_j \subseteq A^{\geq 2}$ and it suffices to show $a^{-1}K_j = \bigcap_l a^{-1}E_{j,l} \in \mathcal{B}_{1,k}$. Therefore we fix an arbitrary such set K_j and drop index j. Intuitively, since no word of length one can be in every set E_l we can take out these words from every E_l without changing K, while simultaneously taking left quotients.

Case $k = 0$. Here each E_l is of the form L_i or $\overline{L_i}^+$ with $L_i = (A, a_1, A, \ldots, a_m, A)_0$ and $m \geq 1$. If $E_l = L_i$ and in fact $m \geq 2$ we can apply Proposition 2.23 right away. If $m = 1$ we may rewrite $E_l = (A, a, A)_0$ as

$$\{a\} \cup \bigcup_{b \in A} \left((A, a, A, b, A)_0 \cup (A, b, A, a, A)_0\right).$$

In this term we can simply drop $\{a\}$ without changing L (since a is definitely not in K) and we are left with a subset of $A^{\geq 2}$ from $\mathcal{B}_{1/2,k}$. Hence we apply Proposition 2.23 again.

Now if $E_l = \overline{L_i}^+$ we also distinguish the cases $m \geq 2$ and $m = 1$. First assume $m \geq 2$, then $L_i \subseteq A^{\geq 2}$ and $\overline{L_i}^+$ definitely contains a. Thus, by Proposition 2.2 we have $a^{-1}\overline{L_i}^+ = \overline{a^{-1}L_i}^+ \cup \{\epsilon\}$. Since $a^{-1}K \subseteq A^{\geq 1}$ we can drop $\{\epsilon\}$ and apply Proposition 2.23 to $a^{-1}L_i$. If $m = 1$ then

$$E_l = \overline{(A, a, A)_0}^+ = \overline{\{a\}}^+ \cap \bigcap_{b \in A} \left(\overline{(A, a, A, b, A)_0}^+ \cap \overline{(A, b, A, a, A)_0}^+\right).$$

We only have to care about $\overline{\{a\}}^+$ because for the other members of the intersection we have $m \geq 2$ which is the case treated before. Since

$$\overline{\{a\}}^+ = \bigcup_{b \in A \setminus \{a\}} \{b\} \cup \bigcup_{b_1, b_2 \in A} (A, b_1, A, b_2, A)_0$$

we can again drop the words of length one and apply Proposition 2.23 to the remaining subsets of $A^{\geq 2}$.

Case $k \geq 1$. If $E_l = D_i$ we simply erase all words of length one from this set and obtain a finite subset of $A^{\geq 2} \cap A^{\leq k}$, hence we may apply Proposition 2.23. If $E_l = L_i$ we also have a subset of $A^{\geq 2}$ since $k \geq 1$ and we are done by Proposition 2.23. If $E_l = \overline{D_i}^+$ we note that $\overline{D_i}^+ = A^{\geq k+1} \cup D'$ for some set $D' \subseteq A^{\leq k}$ where both sets are of a form we already treated. If $E_l = \overline{L_i}^+$ we argue as above in case $k = 0$, $m \geq 2$ and have $a^{-1}\overline{L_i}^+ = \overline{a^{-1}L_i}^+ \cup \{\epsilon\}$, drop $\{\epsilon\}$ and apply Proposition 2.23 a last time. ❑

2.6 The Classes $\mathcal{B}_{3/2,k}$

We study in this paper for each $k \geq 0$ the polynomial closure of the classes $\mathcal{B}_{1,k}$.

Definition 2.25. Let $k \geq 0$. Then $\mathcal{B}_{3/2,k} =_{\text{def}} \text{POL}(\mathcal{B}_{1,k})$.

This is a levelwise view on level $3/2$ of the dot-depth hierarchy.

Proposition 2.26. *It holds that* $\mathcal{B}_{3/2} = \bigcup_{k\geq 0} \mathcal{B}_{3/2,k}$.

Proof. Suppose some language $L \in \mathcal{B}_{3/2,k} = \mathrm{POL}(\mathcal{B}_{1,k})$ for some $k \geq 0$ is given. Since POL is monotonic and $\mathcal{B}_{1,k} \subseteq \mathcal{B}_1$ (Proposition 2.12) we have $L \in \mathrm{POL}(\mathcal{B}_1) = \mathcal{B}_{3/2}$. Conversely, assume $L \in \mathrm{POL}(\mathcal{B}_1)$, i.e., L is a finite union of languages of the form $u_0L_1u_1L_2u_2\cdots L_nu_n$, where $u_i \in A^*$, $L_i \in \mathcal{B}_1$ and $n \geq 0$. Any occuring L_i is by Proposition 2.12 in some class $\mathcal{B}_{1,k_i}$. Now let k be the maximum of all these k_i. Then again by Proposition 2.12 we see that each L_i is in $\mathcal{B}_{1,k}$ and hence $L \in \mathrm{POL}(\mathcal{B}_{1,k}) = \mathcal{B}_{3/2,k}$. ❑

Proposition 2.27. *Let* $k \geq 0$, $w \in A^*$ *and* $L \in \mathcal{B}_{3/2,k}$. *Then*

1. $L^R \in \mathcal{B}_{3/2,k}$ *and*
2. *if* $L \subseteq A^{\geq |w|+1}$ *then* $w^{-1}L, Lw^{-1} \in \mathcal{B}_{3/2,k}$.

Proof. 1. Since by Proposition 2.24 each $\mathcal{B}_{1,k}$ is closed under reversion and due to the definition of the polynomial closure operation there is nothing to prove.

2. By Statement 1 it suffices to show the closure under left quotients. As before, we only need to show that if $L \subseteq A^{\geq 2}$ and $a \in A$ then $a^{-1}L \in \mathcal{B}_{3/2,k}$. Assume that L is a finite union of languages L' with $L' = u_0L_1u_1L_2u_2\cdots L_nu_n$, where $u_i \in A^*$, $L_i \in \mathcal{B}_{1,k}$ and $n \geq 0$. By assumption, it must be that $L' \subseteq A^{\geq 2}$ and we look at $a^{-1}L'$. If $n = 0$ then $|u_0| \geq 2$ and we are done. If $n \geq 1$ and $u_0 \neq \epsilon$ we are also done. So from now on assume $n \geq 1$ and $u_0 = \epsilon$.

We want to ensure that the first occuring set L_1 from $\mathcal{B}_{1,k}$ in the concatenation product has only words of length at least 2 in order to apply Proposition 2.24. We first show that for arbitrary $a \in A$ the set $\{a\}$ is in $\mathcal{B}_{1,k}$ for any $k \geq 0$. It suffices to show $\{a\} \in \mathcal{B}_{1,0}$. We have that

$$\{a\} = \ (A,a,A)_0 \ \setminus \ \Big(\bigcup_{b\in A\setminus\{a\}} (A,b,A)_0 \cup (A,a,A,a,A)_0 \Big).$$

Now we rewrite L_1 as $L_1 = L_1' \cup L_1''$ with $L_1' =_{\mathrm{def}} L_1 \cap A^{\geq 2}$ and $L_1'' =_{\mathrm{def}} \bigcup_{a\in A\cap L_1}\{a\}$. Note that $L_1', L_1'' \in \mathcal{B}_{1,k}$ and we may substitute L_1 in L' by $L_1' \cup L_1''$. It remains to show that $a^{-1}(L_1'u_1L_2u_2\cdots L_nu_n)$ and $a^{-1}(L_1''u_1L_2u_2\cdots L_nu_n)$ are in $\mathcal{B}_{3/2,k}$ which is clear in the latter case: $L_1'' = \emptyset$ or we rewrite L_1'' as the union of its elements. In the former case we know that $L_1' \subseteq A^{\geq 2}$ so we can apply Proposition 2.24 to see that $a^{-1}L_1' \in \mathcal{B}_{1,k}$ and we are done. ❑

We want to mention that one can define a dot-depth hierarchy for fixed k, as we did here for the lower levels, i.e., starting with $\mathcal{B}_{1/2,k}$, then $\mathcal{B}_{1,k} = \mathrm{BC}(\mathcal{B}_{1/2,k})$ and $\mathcal{B}_{3/2,k} = \mathrm{POL}(\mathcal{B}_{1,k})$, and so on. In this case the closure properties 'reversion' and 'quotients' hold on each level $(n/2, k)$. This can be seen by an induction where in the induction step one does basically the same as in our proofs for $\mathcal{B}_{1/2,k}, \mathcal{B}_{1,k}$ and $\mathcal{B}_{3/2,k}$.

3 A Normalform: $\mathcal{B}_{3/2,k} = \widetilde{\mathcal{B}}_k$

By definition, languages in $\mathcal{B}_{3/2,k}$ are finite unions of concatenations of words with Boolean combinations of languages from $\mathcal{B}_{1/2,k}$, which is a somewhat unstructured representation. In this section we show that each language in $\mathcal{B}_{3/2,k}$ can be written as a finite union of languages L_i such that $L_i = D$ for $D \subseteq A^{\leq k}$ or $L_i = (\Sigma_0, \alpha_1, \Sigma_1, \ldots, \alpha_n, \Sigma_n)_k$ where $n \geq 1$, $\alpha_1, \ldots, \alpha_n \in A^{k+1}$ and $\Sigma_0, \Sigma_1, \ldots, \Sigma_n \subseteq A^{k+1}$. Since the other inclusion also holds, we then have $\mathcal{B}_{3/2,k} = \widetilde{\mathcal{B}}_k$ (to see the definition of $\widetilde{\mathcal{B}}_k$ cf. Definition 2.4). This equation is already known for the case $k = 0$, i.e., the usual concatenation (cf. [Arf87]). We will even show a normalform for the classes $\widetilde{\mathcal{B}}_k$ itself, which is helpful to establish the above equation. First, we isolate in a series of proposition and lemmas the needed properties.

3.1 Basic Properties of $\widetilde{\mathcal{B}}_k$

With the next three propositions we show the closure of $\widetilde{\mathcal{B}}_k$ under quotients, under concatenation with words and under concatenation with languages from $\widetilde{\mathcal{B}}_k$. In particular, $\widetilde{\mathcal{B}}_k$ is closed under the polynomial closure operation. Note that the closure of $\widetilde{\mathcal{B}}_k$ under reversion can easily be seen from the definition.

Proposition 3.1. *Let $k \geq 0$, $w \in A^*$ and $L \in \widetilde{\mathcal{B}}_k$ with $L \subseteq A^{\geq |w|+1}$. Then $Lw^{-1}, w^{-1}L \in \widetilde{\mathcal{B}}_k$.*

Proof. It suffices to show $Lw^{-1} \in \widetilde{\mathcal{B}}_k$. By definition, languages of $\widetilde{\mathcal{B}}_k$ are finite unions of languages L_i such that $L_i = D$ for $D \subseteq A^{\leq k}$ or $L_i = (\Theta_0, \theta_1, \Theta_1, \ldots, \theta_{n'}, \Theta_{n'})_k$ where $n' \geq 1$, $\theta_1, \ldots, \theta_{n'} \in A^{k+1}$ and $\Theta_0, \ldots, \Theta_{n'} \subseteq A^{k+1}$. Notice that $Dw^{-1} \subseteq A^{\leq k}$ for every $D \subseteq A^{\leq k} \cap A^{\geq |w|+1}$. Therefore, by Proposition 2.1, it suffices to show the following claim.

Let $L = (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_m, \Gamma_m)_k$ with $k \geq 0, m \geq 1$, $\gamma_1, \ldots, \gamma_m \in A^{k+1}$, $\Gamma_0, \ldots, \Gamma_m \subseteq A^{k+1}$ and $L \subseteq A^{\geq 2}$. Then for $a \in A$ we have $La^{-1} \in \widetilde{\mathcal{B}}_k$. (4)

Furthermore, for all $\sigma_1 \in A^{k+1}, \Sigma_0, \Sigma_1 \subseteq A^{k+1}$ we have

$$(\Sigma_0, \sigma_1, \Sigma_1)_k = \bigcup_{\sigma \in \Sigma_0} (\Sigma_0, \sigma, \Sigma_0, \sigma_1, \Sigma_1)_k \cup \bigcup_{\sigma \in \Sigma_1} (\Sigma_0, \sigma_1, \Sigma_1, \sigma, \Sigma_1)_k \cup \{\sigma_1\}.$$

Observe that $\{\sigma_1\} a^{-1} \subseteq A^{\leq k}$ if $(\Sigma_0, \sigma_1, \Sigma_1)_k \subseteq A^{\geq 2}$. Therefore, we can assume without loss of generality that $m \geq 2$ in Claim (4) above.

Now take for $n \geq 2$ some $v \in (\Sigma_0, \sigma_1, \Sigma_1, \ldots, \sigma_n, \Sigma_n)_k$ with $\widehat{v} = (\beta_1, \ldots, \beta_l)$ for some $l \geq 1$. Then for β_l we have the following possibilities: Either $\beta_l \in \Sigma_n$ or, if $\beta_l \notin \Sigma_n$, then $\beta_l = \sigma_n$. Therefore, we can state

$$\begin{aligned}(\Sigma_0, \sigma_1, \Sigma_1, \ldots, \sigma_n, \Sigma_n)_k \;=\; & \bigcup_{\sigma \in \Sigma_n} (\Sigma_0, \sigma_1, \Sigma_1, \ldots, \sigma_{n-1}, \Sigma_{n-1}, \sigma_n, \Sigma_n, \sigma, \emptyset)_k \cup \\ & (\Sigma_0, \sigma_1, \Sigma_1, \ldots, \sigma_{n-1}, \Sigma_{n-1}, \sigma_n, \emptyset)_k .\end{aligned}$$

This shows that without loss of generality we may assume $\Gamma_m = \emptyset$ in Claim (4). The same argument, applied to the second last element in $\widehat{v}$, shows that we can also assume $\Gamma_{m-1} = \emptyset$. Thus it suffices to prove

$$(\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_m, \Gamma_m)_k \, a^{-1} \in \widetilde{\mathcal{B}}_k$$

for $m \geq 2$, $\gamma_1, \ldots, \gamma_m \in A^{k+1}$, $\Gamma_0, \ldots, \Gamma_m \subseteq A^{k+1}$, $\Gamma_{m-1} = \Gamma_m = \emptyset$ and $a \in A$. This is rather easy, because in this case we have

$$(\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_m, \Gamma_m)_k \, a^{-1} = \begin{cases} (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_{m-1}, \Gamma_{m-1})_k & : \text{ if } \gamma_m \text{ ends with letter } a \\ \emptyset & : \text{ otherwise.} \end{cases}$$

❑

Proposition 3.2. *Let $k \geq 0$. Then $\widetilde{\mathcal{B}}_k$ is closed under concatenation with elements of A^*.*

Proof. First we will show $Dw \in \widetilde{\mathcal{B}}_k$ and $wD \in \widetilde{\mathcal{B}}_k$ for every $k \geq 0$, $D \subseteq A^{\leq k}$, $w \in A^*$. With $D_1 =_{\text{def}} Dw \cap A^{\leq k}$ and $D_2 =_{\text{def}} Dw \cap A^{\geq k+1}$ we have $Dw = D_1 \cup D_2$. Every word $v \in D_2$ with $\widehat{v} = (\sigma_1, \ldots, \sigma_l)$ can be written as $\{v\} = (\emptyset, \sigma_1, \emptyset, \ldots, \sigma_l, \emptyset)_k$. Thus we obtain $Dw \in \widetilde{\mathcal{B}}_k$, and the same argument shows $wD \in \widetilde{\mathcal{B}}_k$.

Now it suffices to show that $L \cdot a \in \widetilde{\mathcal{B}}_k$, where $a \in A$ and $L =_{\text{def}} (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_n, \Gamma_n)_k$ with $k \geq 0, n \geq 1$, $\gamma_1, \ldots, \gamma_n \in A^{k+1}$, $\Gamma_0, \ldots, \Gamma_n \subseteq A^{k+1}$. This is easy because

$$L \cdot a = \bigcup_{\gamma \in A^k \cdot a} (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_n, \Gamma_n, \gamma, \emptyset)_k .$$

This shows $L \cdot a \in \widetilde{\mathcal{B}}_k$. Analogously, one can prove $a \cdot L \in \widetilde{\mathcal{B}}_k$. ❑

Proposition 3.3. *Let $k \geq 0$. Then $\widetilde{\mathcal{B}}_k$ is closed under concatenation.*

Proof. From Proposition 3.2 we obtain that $\widetilde{\mathcal{B}}_k$ is closed under concatenation with subsets of $A^{\leq k}$. Therefore, it suffices to show $L_1 \cdot L_2 \in \widetilde{\mathcal{B}}_k$ where $L_1 =_{\text{def}} (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_m, \Gamma_m)_k$ and $L_2 =_{\text{def}} (\Sigma_0, \sigma_1, \Sigma_1, \ldots, \sigma_n, \Sigma_n)_k$ for $m, n \geq 1$, $\gamma_1, \ldots, \gamma_m, \sigma_1, \ldots, \sigma_n \in A^{k+1}$ and $\Gamma_0, \ldots, \Gamma_m, \Sigma_0, \ldots, \Sigma_n \subseteq A^{k+1}$. We will prove that the following equation holds.

$$L_1 \cdot L_2 = \bigcup_{\theta_1, \ldots, \theta_k \in A^{k+1}} (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_m, \Gamma_m, \theta_1, \emptyset, \theta_2, \ldots, \emptyset, \theta_k, \Sigma_0, \sigma_1, \Sigma_1, \ldots, \sigma_n, \Sigma_n)_k \tag{5}$$

Denote the right hand side by L'. "$\subseteq$" Let $w = w_1 w_2$ with $w_1 \in L_1$, $w_2 \in L_2$, $\widehat{w_1} = (\alpha_1, \ldots, \alpha_g)$ and $\widehat{w_2} = (\beta_1, \ldots, \beta_h)$. Then we have $\alpha_g = a_1 \cdots a_{k+1}$ and $\beta_1 = b_1 \cdots b_{k+1}$ for suitable $a_1, \ldots, a_{k+1}, b_1, \ldots, b_{k+1} \in A$. With $\theta_1 =_{\text{def}} a_2 \cdots a_{k+1} b_1$, $\theta_2 =_{\text{def}} a_3 \cdots a_{k+1} b_1 b_2$, $\ldots$, $\theta_k =_{\text{def}} a_{k+1} b_1 \cdots b_k$, we obtain $w \in L'$.
"$\supseteq$" Let $w \in L'$. Then there exist $\alpha_1, \ldots, \alpha_g, \theta_1, \ldots, \theta_k, \beta_1, \ldots, \beta_h \in A^{k+1}$ and $w_1, w_2 \in A^{\geq k+1}$ with

- $\widehat{w} = (\alpha_1, \ldots, \alpha_g, \theta_1, \ldots, \theta_k, \beta_1, \ldots, \beta_h)$
- $\widehat{w_1} = (\alpha_1, \ldots, \alpha_g)$
- $\widehat{w_2} = (\beta_1, \ldots, \beta_h)$
- $w_1 \in (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_m, \Gamma_m)_k = L_1$
- $w_2 \in (\Sigma_0, \sigma_1, \Sigma_1, \ldots, \sigma_n, \Sigma_n)_k = L_2$

We denote the first letters of $\alpha_1, \ldots, \alpha_g, \theta_1, \ldots, \theta_k, \beta_1, \ldots, \beta_h$ by $a_1, \ldots, a_g, t_1, \ldots, t_k, b_1, \ldots, b_h$, respectively. We obtain $w_1 = a_1 a_2 \cdots a_{g-1} \alpha_g = a_1 a_2 \cdots a_g t_1 t_2 \cdots t_k$ and $w_2 = b_1 b_2 \cdots b_{h-1} \beta_h$. Since $w = a_1 \cdots a_g t_1 \cdots t_k b_1 \cdots b_{h-1} \beta_h$, we conclude $w = w_1 w_2 \in L_1 \cdot L_2$. ❑

Next we show a lemma which states that we can break up the representation of a language from $\widetilde{\mathcal{B}}_k$ and isolate any occuring α in this representation. The resulting factors are again languages from $\widetilde{\mathcal{B}}_k$, as can be seen by Proposition 3.1 and the fact that $(\Sigma_0)_k \in \widetilde{\mathcal{B}}_k$ for every $\Sigma_0 \subseteq A^{k+1}$.

Lemma 3.4. *Let $k \geq 0$ and $L = (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_m, \Gamma_m)_k$ for $m \geq 1$, $\gamma_1, \ldots, \gamma_m \in A^{k+1}$ and $\Gamma_0, \ldots, \Gamma_m \subseteq A^{k+1}$. Then for all $1 \leq h \leq m$ we have*

$$L = (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_{h-1}, \Gamma_{h-1})_k \, p_k(\gamma_h)^{-1} \;\cdot\; \gamma_h \;\cdot\; s_k(\gamma_h)^{-1} \, (\Gamma_h, \gamma_{h+1}, \Gamma_{h+1}, \ldots, \gamma_m, \Gamma_m)_k .$$

Proof. Let $L_1 =_{\text{def}} (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_{h-1}, \Gamma_{h-1})_k$ and $L_2 =_{\text{def}} (\Gamma_h, \gamma_{h+1}, \Gamma_{h+1}, \ldots, \gamma_m, \Gamma_m)_k$ for a fixed h with $1 \le h \le m$.

"$\subseteq$" Let $w \in L$ with $w = a_1 \cdots a_{l+k}$ and $\widehat{w} = (\alpha_1, \ldots, \alpha_l)$ for $l \ge 1$, $a_1, \ldots, a_{l+k} \in A$, $\alpha_1, \ldots, \alpha_l \in A^{k+1}$. By definition, there exist indices $0 = j_0 < j_1 < \cdots < j_m < j_{m+1} = l + 1$ such that

(a) $\alpha_{j_i} = \gamma_i$ for all $1 \le i \le m$, and

(b) $\alpha_j \in \Gamma_i$ for all $0 \le i \le m$ and $j_i < j < j_{i+1}$.

With $w_1 =_{\text{def}} a_1 a_2 \cdots a_{j_h - 1}$, $w_2 =_{\text{def}} a_{j_h+k+1} a_{j_h+k+2} \cdots a_{l+k}$ and $w_1' =_{\text{def}} w_1 \cdot p_k(\gamma_h)$, $w_2' =_{\text{def}} s_k(\gamma_h) \cdot w_2$ we obtain

$$\begin{aligned} w_1' &= a_1 a_2 \cdots a_{j_h-1} \cdot a_{j_h} a_{j_h+1} \cdots a_{j_h+k-1} \quad \text{and} \\ w_2' &= a_{j_h+1} a_{j_h+2} \cdots a_{j_h+k} \cdot a_{j_h+k+1} a_{j_h+k+2} \cdots a_{l+k}. \end{aligned}$$

Hence $w = w_1 \gamma_h w_2$, $\widehat{w_1'} = (\alpha_1, \ldots, \alpha_{j_h-1})$ and $\widehat{w_2'} = (\alpha_{j_h+1}, \ldots, \alpha_l)$. It follows that $w_1' \in L_1$ and $w_2' \in L_2$. Thus we obtain $w_1 \in L_1 p_k(\gamma_h)^{-1}$ and $w_2 \in s_k(\gamma_h)^{-1} L_2$. Therefore, we have $w \in L_1 p_k(\gamma_h)^{-1} \cdot \gamma_h \cdot s_k(\gamma_h)^{-1} L_2$.

"$\supseteq$" Now let $w \in L_1 p_k(\gamma_h)^{-1} \cdot \gamma_h \cdot s_k(\gamma_h)^{-1} L_2$, i.e., there exist words $w_1 \in L_1 p_k(\gamma_h)^{-1}$, $w_2 \in s_k(\gamma_h)^{-1} L_2$ such that $w = w_1 \gamma_h w_2$. For suitable $a_1, \ldots, a_{l+k} \in A, \alpha_1, \ldots, \alpha_l \in A^{k+1}$ we have $w = a_1 \cdots a_{l+k}$, $\widehat{w} = (\alpha_1, \ldots, \alpha_l)$ and $\alpha_{i_h} = \gamma_h$ with $i_h =_{\text{def}} |w_1| + 1$. Define $w_1', w_2' \in A^*$ such that $\widehat{w_1'} = (\alpha_1, \ldots, \alpha_{i_h-1})$ and $\widehat{w_2'} = (\alpha_{i_h+1}, \ldots, \alpha_l)$ (notice that this is well-defined). It follows that $w_1' = a_1 \cdots a_{i_h+k-1}$ and $w_2' = a_{i_h+1} \cdots a_{l+k}$, hence $w_1' = w_1 p_k(\gamma_h)$ and $w_2' = s_k(\gamma_h) w_2$. We obtain $w_1' \in L_1 = (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_{h-1}, \Gamma_{h-1})_k$ and $w_2' \in L_2 = (\Gamma_h, \gamma_{h+1}, \Gamma_{h+1}, \ldots, \gamma_m, \Gamma_m)_k$. Since $\widehat{w} = (\alpha_1, \ldots, \alpha_l)$, $\widehat{w_1'} = (\alpha_1, \ldots, \alpha_{i_h-1})$ and $\widehat{w_2'} = (\alpha_{i_h+1}, \ldots, \alpha_l)$, we conclude $w \in L = (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_m, \Gamma_m)_k$. ❑

In general, we must not distribute intersections over concatenation, i.e., for arbitrary languages G_1, G_2, S_1, S_2 we can not state $(G_1 \cap G_2) \cdot (S_1 \cap S_2) = (G_1 \cdot S_1) \cap (G_2 \cdot S_2)$. The situation changes if the concatenation is at a certain position, namely where some factor w occurs the first time.

Lemma 3.5. *Let $k \ge 0$, let $G_1, G_2, S_1, S_2 \subseteq A^*$ and $w \in A^{k+1}$ with $w = a_1 \cdots a_{k+1}$ for $a_1, \ldots, a_{k+1} \in A$. Furthermore, let $w_p =_{\text{def}} a_1 \cdots a_k$ and $w_s =_{\text{def}} a_2 \cdots a_{k+1}$. If $(G_1 \cup S_1) \subseteq \overline{A^* w A^*}$ then it holds that*

$$\left(G_1 w_p^{-1} \cdot w \cdot w_s^{-1} G_2\right) \cap \left(S_1 w_p^{-1} \cdot w \cdot w_s^{-1} S_2\right) = \left(G_1 \cap S_1\right) w_p^{-1} \cdot w \cdot w_s^{-1} \left(G_2 \cap S_2\right).$$

Proof. By Proposition 2.1, it suffices to show that

$$\left(G_1 w_p^{-1} \cdot w \cdot w_s^{-1} G_2\right) \cap \left(S_1 w_p^{-1} \cdot w \cdot w_s^{-1} S_2\right) = \left(G_1 w_p^{-1} \cap S_1 w_p^{-1}\right) \cdot w \cdot \left(w_s^{-1} G_2 \cap w_s^{-1} S_2\right).$$

"$\subseteq$" Let $v \in \left(G_1 w_p^{-1} \cdot w \cdot w_s^{-1} G_2\right) \cap \left(S_1 w_p^{-1} \cdot w \cdot w_s^{-1} S_2\right)$, then there exist words $g_1 \in G_1 w_p^{-1}, g_2 \in w_s^{-1} G_2, s_1 \in S_1 w_p^{-1}, s_2 \in w_s^{-1} S_2$ such that $v = g_1 w g_2 = s_1 w s_2$. Now we want to show $g_1 = s_1$. Assume $|g_1 w| < |s_1 w|$, thus $g_1 w$ is a proper prefix of $s_1 w$, i.e., there exist an $l > 0$ and letters $b_1, \ldots, b_l \in A$ such that $s_1 w = g_1 w b_1 \cdots b_l$. Since $s_1 \in S_1 w_p^{-1}$, we have $g_1 w b_1 \cdots b_{l-1} = s_1 w_p \in S_1$. This is a contradiction, because $S_1 \subseteq \overline{A^* w A^*}$, by assumption. We obtain $|g_1 w| \ge |s_1 w|$, and the same argument shows $|g_1 w| \le |s_1 w|$. Therefore, $g_1 = s_1$ and $g_2 = s_2$. It follows that $v \in \left(G_1 w_p^{-1} \cap S_1 w_p^{-1}\right) \cdot w \cdot \left(w_s^{-1} G_2 \cap w_s^{-1} S_2\right)$.

"$\supseteq$" Let $v \in \left(G_1 w_p^{-1} \cap S_1 w_p^{-1}\right) \cdot w \cdot \left(w_s^{-1} G_2 \cap w_s^{-1} S_2\right)$. Then there exist words $v_1 \in \left(G_1 w_p^{-1} \cap S_1 w_p^{-1}\right)$ and $v_2 \in \left(w_s^{-1} G_2 \cap w_s^{-1} S_2\right)$ with $v = v_1 w v_2$. It follows that $v_1 w v_2$ is an element of $\left(G_1 w_p^{-1} \cdot w \cdot w_s^{-1} G_2\right)$ and of $\left(S_1 w_p^{-1} \cdot w \cdot w_s^{-1} S_2\right)$. ❑

3.2 A Normalform for $\widetilde{\mathcal{B}}_k$

Suppose L is a finite union of languages of the form $(\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k$. We show in this subsection that we may assume that any word of length $k+1$ from some set Γ_i occurs in the representation of L the first time as some γ_j. In order to distinguish between languages and their formal representation we use now the term 'expression' and mean the syntactical object '$(\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k$' that describes a language. Let us formally describe what we mean with the notion of first occurence.

Definition 3.6. *For every $\alpha \in A^{k+1}$ and every expression $(\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k$ with $n \geq 0$, $\gamma_1, \dots, \gamma_n \in A^{k+1}$, $\Gamma_0, \dots, \Gamma_n \subseteq A^{k+1}$ we define the position of the first occurence of α as*

$$\begin{aligned} \alpha_{\min}(\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k &=_{\text{def}} \min(\{1 \leq i \leq n : \alpha = \gamma_i\} \cup \{n+1\}) \text{ and} \\ \alpha_{\text{MIN}}(\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k &=_{\text{def}} \min(\{0 \leq i \leq n : \alpha \in \Gamma_i\} \cup \{n+1\}). \end{aligned}$$

Here $\alpha_{\min}$ is the leftmost position of α as some γ_j in an expression, and α_{MIN} is the index of the leftmost set Γ_i in which α is contained. Next we look at two cardinalities. The first measures the size of an expression and the other one gives the number of words from A^{k+1} for which the property of first occurence is violated, i.e., it measures the number of different $\alpha \in A^{k+1}$ that occur in an expression at first in some set Γ_i. We call this the number of transpositions. These two cardinalities will be used in the proof of the following lemma by two nested inductions.

Definition 3.7. *We define the size of an expression and the number of its transpositions as*

$$\begin{aligned} \text{size}(\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k &=_{\text{def}} n + \sum_{0 \leq i \leq n} |\Gamma_i| \text{ and} \\ \text{Tr}(\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k &=_{\text{def}} \left| \left\{ \alpha \in A^{k+1} : \alpha_{\text{MIN}}(\Gamma_0, \dots, \Gamma_n)_k < \alpha_{\min}(\Gamma_0, \dots, \Gamma_n)_k \right\} \right|. \end{aligned}$$

We show that we can always find equivalent expressions, i.e., expressions that describe the same language, with zero transpositions.

Lemma 3.8. *Let $k \geq 0$. Every language given by an expression $(\Sigma_0, \sigma_1, \Sigma_1, \dots, \sigma_m, \Sigma_m)_k$ with $m \geq 0$ can be written as a finite union of languages, each of which represented by expressions of the form $(\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k$ such that $\text{Tr}(\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k = 0$ and $n \geq m$.*

Proof. We will show the lemma by induction on $\text{Tr}(\Sigma_0, \sigma_1, \Sigma_1, \dots, \sigma_m, \Sigma_m)_k$. The induction base is trivial. Now assume that we have shown the lemma for all languages given by expressions $(\Theta_0, \theta_1, \Theta_1, \dots, \theta_p, \Theta_p)_k$ with $\text{Tr}(\Theta_0, \theta_1, \Theta_1, \dots, \theta_p, \Theta_p)_k \leq l$, and we want to show the following:

> *Every language given by an expression $(\Sigma_0, \sigma_1, \Sigma_1, \dots, \sigma_m, \Sigma_m)_k$ with $m \geq 0$ and $\text{Tr}(\Sigma_0, \sigma_1, \Sigma_1, \dots, \sigma_m, \Sigma_m)_k = l+1$ can be written as a finite union of languages represented by expressions of the form $(\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k$ with $\text{Tr}(\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k = 0$ and $n \geq m$.* (6)

We will prove this claim by a second induction on $\text{size}(\Sigma_0, \sigma_1, \Sigma_1, \dots, \sigma_m, \Sigma_m)_k$. If $\text{size}(\Sigma_0, \sigma_1, \Sigma_1, \dots, \sigma_m, \Sigma_m)_k = 0$, then it must be that $m = 0$ and $\Sigma_0 = \emptyset$. Thus $\text{Tr}(\Sigma_0)_k = 0$, which proves our second induction base. Now we assume that Claim (6) has been shown for all languages given by expressions $(\Theta_0, \theta_1, \Theta_1, \dots, \theta_p, \Theta_p)_k$ with $\text{Tr}(\Theta_0, \theta_1, \Theta_1, \dots, \theta_p, \Theta_p)_k = l+1$ and $\text{size}(\Theta_0, \theta_1, \Theta_1, \dots, \theta_p, \Theta_p)_k \leq r$.

Let $E =_{\text{def}} (\Sigma_0, \sigma_1, \Sigma_1, \dots, \sigma_m, \Sigma_m)_k$ be an expression with $\text{Tr}(E) = l+1$ and $\text{size}(E) = r+1$. Thus there exist $\alpha \in A^{k+1}$ and indices $0 \leq i_1 < i_2 \leq m+1$ with $i_1 = \alpha_{\text{MIN}}(E)$ and $i_2 = \alpha_{\min}(E)$. We define the expressions

$$\begin{aligned} E_1 &=_{\text{def}} (\Sigma_0, \sigma_1, \Sigma_1, \dots, \sigma_{i_1}, \Sigma_{i_1} \setminus \{\alpha\}, \qquad \sigma_{i_1+1}, \Sigma_{i_1+1}, \dots, \sigma_m, \Sigma_m)_k \text{ and} \\ E_2 &=_{\text{def}} (\Sigma_0, \sigma_1, \Sigma_1, \dots, \sigma_{i_1}, \Sigma_{i_1} \setminus \{\alpha\}, \alpha, \Sigma_{i_1}, \sigma_{i_1+1}, \Sigma_{i_1+1}, \dots, \sigma_m, \Sigma_m)_k. \end{aligned}$$

Since $\alpha \in \Sigma_{i_1}$, it is easy to see that the language given by E is just the union of the languages given by E_1 and E_2. Moreover, the length (i.e., number of components) of E_1 and E_2 is greater or equal to the length of E and $\text{size}(E_1) = r$. E_1 was obtained from E by removing α from Σ_{i_1}, thus we have $\alpha'_{\text{MIN}}(E) \leq \alpha'_{\text{MIN}}(E_1)$ and $\alpha'_{\min}(E) = \alpha'_{\min}(E_1)$ for all $\alpha' \in A^{k+1}$. It follows that $\text{Tr}(E_1) \leq \text{Tr}(E) = l+1$. If $\text{Tr}(E_1) < l+1$ then the language given by E_1 can be written as a finite union of languages represented by expressions E' with $\text{Tr}(E') = 0$, by the first induction hypothesis. If $\text{Tr}(E_1) = l+1$ then the language given by E_1 can be written as a finite union of languages represented by expressions E' with $\text{Tr}(E') = 0$, by the second induction hypothesis (because $\text{size}(E_1) = r$).

Now we want to show that we reduced the number of transpositions with the contruction of E_2 and hence $\text{Tr}(E_2) < l+1$. Let $\alpha' \in A^{k+1} \setminus \{\alpha\}$. It follows from the definition of E_2 that $\alpha'_{\text{MIN}}(E_2) \neq i_1 + 1$ because from $\alpha' \in \Sigma_{i_1}$ also $\alpha' \in \Sigma_{i_1} \setminus \{\alpha\}$ would follow, and $\alpha'_{\min}(E_2) \neq i_1 + 1$. Furthermore, we have

$$\alpha'_{\text{MIN}}(E) = \begin{cases} \alpha'_{\text{MIN}}(E_2) & : \text{ if } \alpha'_{\text{MIN}}(E_2) \leq i_1 \\ \alpha'_{\text{MIN}}(E_2) - 1 & : \text{ otherwise, i.e., if } \alpha'_{\text{MIN}}(E_2) \geq i_1 + 2, \text{ and} \end{cases}$$

$$\alpha'_{\min}(E) = \begin{cases} \alpha'_{\min}(E_2) & : \text{ if } \alpha'_{\min}(E_2) \leq i_1 \\ \alpha'_{\min}(E_2) - 1 & : \text{ otherwise, i.e., if } \alpha'_{\min}(E_2) \geq i_1 + 2. \end{cases}$$

Therefore, for every $\alpha' \in A^{k+1} \setminus \{\alpha\}$ it holds that

$$\alpha'_{\text{MIN}}(E_2) < \alpha'_{\min}(E_2) \quad \Longrightarrow \quad \alpha'_{\text{MIN}}(E) < \alpha'_{\min}(E). \tag{7}$$

Since $i_1 = \alpha_{\text{MIN}}(E)$, we obtain

$$\alpha_{\text{MIN}}(E_2) \geq \alpha_{\min}(E_2). \tag{8}$$

From Equations (7), (8) and from $\alpha_{\text{MIN}}(E) < \alpha_{\min}(E)$ it follows that $\text{Tr}(E_2) < \text{Tr}(E) = l+1$. Hence E_2 can be represented as a finite union of expressions E' with $\text{Tr}(E') = 0$, by the first induction hypothesis. This completes our second induction and proves Claim (6). Thus the induction step of the first induction is completed, and the lemma is proved. ❑

Theorem 3.9. *Let $k \geq 0$. Every $L \in \widetilde{\mathcal{B}}_k$ can be written as a finite union of languages L_i such that $L_i = D$ for $D \subseteq A^{\leq k}$ or $L_i = (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_n, \Gamma_n)_k$ where $n \geq 1$, $\gamma_1, \ldots, \gamma_n \in A^{k+1}$, $\Gamma_0, \ldots, \Gamma_n \subseteq A^{k+1}$ such that for $0 \leq i \leq n$ it holds that*

$$\{\gamma_1, \ldots, \gamma_i\} \supseteq \Gamma_i.$$

Proof. Since $\text{Tr}\,(\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_n, \Gamma_n)_k = 0$ implies $\{\gamma_1, \ldots, \gamma_i\} \supseteq \Gamma_i$ for all $0 \leq i \leq n$, this theorem is an easy consequence of Lemma 3.8. Note that we can apply Lemma 3.8 in particular if $n \geq 1$. ❑

3.3 $\mathcal{B}_{3/2,k} = \widetilde{\mathcal{B}}_k$

Finally, we want to prove what we announced for this section. We proceed again stepwise and isolate our arguments. Our temporary goal will be Lemma 3.13 where we show that $\widetilde{\mathcal{B}}_k$ is closed under intersection with languages from $\text{co}\mathcal{B}_{1/2,k}$. These languages cause most of the difficulties when concatenating languages from $\mathcal{B}_{1,k}$. We show first that we can turn complements of languages from $\mathcal{B}_{1/2,k}$ into languages from $\widetilde{\mathcal{B}}_k$ of a certain form if we do not care about the k-prefix and the k-suffix.

Proposition 3.10. *Let* $k \geq 0$. *Let* $L = \overline{\left(A^{k+1}, \beta_1, A^{k+1}, \ldots, \beta_m, A^{k+1}\right)_k}^{+}$ *with* $m \geq 1$ *and* $\beta_1, \ldots, \beta_m \in A^{k+1}$. *Then it holds that*

$$L = \bigcup_{1 \leq i \leq m} \left(A^{k+1} \backslash \{\beta_1\}, \beta_1,\ A^{k+1} \backslash \{\beta_2\}, \beta_2,\ \ldots,\ A^{k+1} \backslash \{\beta_{i-1}\}, \beta_{i-1},\ A^{k+1} \backslash \{\beta_i\}\right)_k \cup A^{\leq k}$$

Proof. Let us first define for $1 \leq i \leq m$ the sets

$$L_i =_{\mathrm{def}} \left(A^{k+1} \backslash \{\beta_1\}, \beta_1,\ A^{k+1} \backslash \{\beta_2\}, \beta_2,\ \ldots,\ A^{k+1} \backslash \{\beta_{i-1}\}, \beta_{i-1},\ A^{k+1} \backslash \{\beta_i\}\right)_k.$$

"$\subseteq$" Let $w \in L$. If $1 \leq |w| \leq k$ then $w \in A^{\leq k}$ and we are done. Otherwise $w \in \left(A^{k+1}\right)_k$ and $\widehat{w} = (\alpha_1, \ldots, \alpha_l)$ for suitable $l > 0, \alpha_1, \ldots, \alpha_l \in A^{k+1}$. Define n as the minimal number ≥ 1, such that there do not exist indices $1 \leq i_1 < i_2 < \cdots < i_n \leq l$ with $\alpha_{i_j} = \beta_j$ for $1 \leq j \leq n$. Notice that $n \leq m$, otherwise we would obtain $w \in \left(A^{k+1}, \beta_1, A^{k+1}, \ldots, \beta_m, A^{k+1}\right)_k$, a contradiction. For $i_0 =_{\mathrm{def}} 0$ and $j = 1, \ldots, n-1$, let $i_j =_{\mathrm{def}} \min\{h > i_{j-1} : \alpha_h = \beta_j\}$ and observe that those minima exist. Thus for $1 \leq j < n$ we have $\alpha_{i_j} = \beta_j$ and $\alpha_h \in A^{k+1} \setminus \{\beta_j\}$ for $i_{j-1} < h < i_j$. From the definition of n it follows that $\alpha_h \in A^{k+1} \setminus \{\beta_n\}$ for all $i_{n-1} < h \leq l$. Therefore, $w \in L_n$.

"$\supseteq$" It is easy to see that $A^{\leq k} \subseteq L$. Observe $L_i \subseteq \overline{\left(A^{k+1}, \beta_1, A^{k+1}, \ldots, \beta_i, A^{k+1}\right)_k}^{+}$ for $1 \leq i \leq m$. Hence $L_i \subseteq L$ for $1 \leq i \leq m$. ❑

Proposition 3.11. *Let* $k \geq 0$. *Then* $\widetilde{\mathcal{B}}_k$ *is closed under intersection with languages of the form* $\overline{\left(A^{k+1}, \alpha_1, A^{k+1}, \ldots, \alpha_m, A^{k+1}\right)_k}^{+}$ *with* $m \geq 1$ *and* $\alpha_1, \ldots, \alpha_m \in A^{k+1}$.

Proof. Note that $\overline{\left(A^{k+1}, \alpha_1, A^{k+1}, \ldots, \alpha_m, A^{k+1}\right)_k}^{+} \cap D = D$ for every $D \subseteq A^{\leq k}$. Thus by Proposition 3.10 and distributive laws, it suffices to show the following claim:

Let $B = \left(A^{k+1} \backslash \{\beta_1\}, \beta_1, A^{k+1} \backslash \{\beta_2\}, \beta_2, \ldots, A^{k+1} \backslash \{\beta_{m-1}\}, \beta_{m-1},\ A^{k+1} \backslash \{\beta_m\}\right)_k$ (9) *with* $m \geq 1$ *and* $G = (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_n, \Gamma_n)_k$ *with* $n \geq 1$ *and* $\beta_1, \ldots, \beta_m, \gamma_1, \ldots, \gamma_n \in A^{k+1}$, $\Gamma_0, \ldots, \Gamma_n \subseteq A^{k+1}$. *Then* $B \cap G \in \widetilde{\mathcal{B}}_k$.

We will show this claim by induction on m. For $m = 1$ we have $B = \left(A^{k+1} \setminus \{\beta_1\}\right)_k$. By Theorem 3.9 and distributive laws, we may assume without loss of generality that $\{\gamma_1, \ldots, \gamma_i\} \supseteq \Gamma_i$ for $0 \leq i \leq n$. Hence if $\beta_1 \notin \{\gamma_1, \ldots, \gamma_n\}$ then $G \cap B = G \in \widetilde{\mathcal{B}}_k$, otherwise $G \cap B = \emptyset \in \widetilde{\mathcal{B}}_k$. This proves the induction base.

Now we assume that Claim (9) has been shown for $m \leq l$, and we want to prove it for $m = l+1$. Again by Theorem 3.9 and distributive laws, we may assume $\{\gamma_1, \ldots, \gamma_i\} \supseteq \Gamma_i$ for $0 \leq i \leq n$. If $\beta_1 \notin \{\gamma_1, \ldots, \gamma_n\}$, we have $G \subseteq \left(A^{k+1} \setminus \{\beta_1\}\right)_k$ and $G \cap B = \emptyset$. Otherwise, let $j =_{\mathrm{def}} \min\{i : \gamma_i = \beta_1\}$. By Lemma 3.4, for $\alpha =_{\mathrm{def}} \beta_1$, $\alpha_p =_{\mathrm{def}} p_k(\beta_1)$ and $\alpha_s =_{\mathrm{def}} s_k(\beta_1)$ we obtain

$$\begin{aligned} G &= \overbrace{(\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_{j-1}, \Gamma_{j-1})_k}^{G_1 =_{\mathrm{def}}} \alpha_p^{-1} \cdot \alpha \cdot \alpha_s^{-1} \overbrace{(\Gamma_j, \gamma_{j+1}, \Gamma_{j+1}, \ldots, \gamma_n, \Gamma_n)_k}^{G_2 =_{\mathrm{def}}} \quad \text{and} \\ B &= \underbrace{\left(A^{k+1} \backslash \{\beta_1\}\right)_k}_{B_1 =_{\mathrm{def}}} \alpha_p^{-1} \cdot \alpha \cdot \alpha_s^{-1} \underbrace{\left(A^{k+1} \backslash \{\beta_2\}, \beta_2, \ldots, A^{k+1} \backslash \{\beta_{m-1}\}, \beta_{m-1},\ A^{k+1} \backslash \{\beta_m\}\right)_k}_{B_2 =_{\mathrm{def}}}. \end{aligned}$$

Since $\{\gamma_1, \ldots, \gamma_i\} \supseteq \Gamma_i$ for $0 \leq i \leq n$, from the definition of j it follows that $G_1 \subseteq \overline{A^* \alpha A^*}$ and $B_1 \subseteq \overline{A^* \alpha A^*}$. Thus we can apply Lemma 3.5, we obtain $G \cap B = (G_1 \cap B_1)\, \alpha_p^{-1} \cdot \alpha \cdot \alpha_s^{-1} (G_2 \cap B_2)$. Observe that $G_1 \cap B_1 = G_1 \in \widetilde{\mathcal{B}}_k$ (if $j = 1$ then we have $G_1 = (\emptyset)_k = \emptyset \in \widetilde{\mathcal{B}}_k$). By induction hypothesis, $G_2 \cap B_2 \in \widetilde{\mathcal{B}}_k$ (if $j = n$ then $G_2 = \bigcup_{\gamma \in \Gamma_j} (\Gamma_j, \gamma, \Gamma_j)_k$).

Since $G_1, G_2, B_1, B_2 \subseteq A^{\geq k+1}$ and $|\alpha_p| = |\alpha_s| = k$, we have $G_1 \cap B_1 \subseteq A^{\geq |\alpha_p|+1}$ and $G_2 \cap B_2 \subseteq A^{\geq |\alpha_s|+1}$. By Proposition 3.1, $(G_1 \cap B_1)\, \alpha_p^{-1} \in \widetilde{\mathcal{B}}_k$ and $\alpha_s^{-1} (G_2 \cap B_2) \in \widetilde{\mathcal{B}}_k$. From Propositions 3.2 and 3.3 we conclude $G \cap B \in \widetilde{\mathcal{B}}_k$ which proves Claim (9). ❑

As one might expect, considering a particular k-prefix and k-suffix is no big difference in our case.

Proposition 3.12. *Let $k \geq 0$ and let $L = \overline{(w|A^{k+1}, \beta_1, A^{k+1}, \dots, \beta_m, A^{k+1}|v)_k}^{\top}$ with $m \geq 1$, $w, v \in A^k$ and $\beta_1, \dots, \beta_m \in A^{k+1}$. Then*

$$L = \overline{(A^{k+1}, \beta_1, A^{k+1}, \dots, \beta_m, A^{k+1})_k}^{\top} \cup \bigcup_{\alpha \in A^{k+1} \setminus wA} \left(\emptyset, \alpha, A^{k+1}\right)_k \cup \bigcup_{\beta \in A^{k+1} \setminus Av} \left(A^{k+1}, \beta, \emptyset\right)_k.$$

Proof. By definition,

$$\begin{aligned} L &= \overline{wA^* \cap A^*v \cap (A^{k+1}, \beta_1, A^{k+1}, \dots, \beta_m, A^{k+1})_k}^{\top} \\ &= \overline{(A^{k+1}, \beta_1, A^{k+1}, \dots, \beta_m, A^{k+1})_k}^{\top} \cup \overline{wA^*}^{\top} \cup \overline{A^*v}^{\top}. \end{aligned}$$

Since $A^{\leq k} \subseteq \overline{(A^{k+1}, \beta_1, A^{k+1}, \dots, \beta_m, A^{k+1})_k}^{\top}$, we obtain

$$L = \overline{(A^{k+1}, \beta_1, A^{k+1}, \dots, \beta_m, A^{k+1})_k}^{\top} \cup \left(\overline{wA^*}^{\top} \cap A^{\geq k+1}\right) \cup \left(\overline{A^*v}^{\top} \cap A^{\geq k+1}\right).$$

Therefore, it suffices to show

$$\begin{aligned} \overline{wA^*}^{\top} \cap A^{\geq k+1} &= \bigcup_{\alpha \in A^{k+1} \setminus wA} \left(\emptyset, \alpha, A^{k+1}\right)_k \quad \text{and} \\ \overline{A^*v}^{\top} \cap A^{\geq k+1} &= \bigcup_{\beta \in A^{k+1} \setminus Av} \left(A^{k+1}, \beta, \emptyset\right)_k. \end{aligned}$$

We will show the first equation, the second one follows by symmetry.

"$\subseteq$" Let $u \in \overline{wA^*}^{\top} \cap A^{\geq k+1}$ and $\widehat{u} = (\alpha_1, \dots, \alpha_l)$ for suitable $l \geq 1$, $\alpha_1, \dots, \alpha_l \in A^{k+1}$. It follows that $p_k(\alpha_1) = p_k(u) \neq w$. Thus $\alpha_1 \in A^{k+1} \setminus wA$ and $u \in \left(\emptyset, \alpha_1, A^{k+1}\right)_k$.

"$\supseteq$" Let $\alpha \in A^{k+1} \setminus wA$, $u \in \left(\emptyset, \alpha, A^{k+1}\right)_k$ and $\widehat{u} = (\alpha_1, \dots, \alpha_l)$ for suitable $l \geq 1$, $\alpha_1, \dots, \alpha_l \in A^{k+1}$. Then $\alpha = \alpha_1$ and $p_k(u) = p_k(\alpha_1) \neq w$. We obtain $u \in \overline{wA^*}^{\top} \cap A^{\geq k+1}$. ❑

Now we are ready to state the closure of $\widetilde{\mathcal{B}}_k$ under intersections with languages from $\mathrm{co}\mathcal{B}_{1/2,k}$.

Lemma 3.13. *Let $k \geq 0$. Then $\widetilde{\mathcal{B}}_k$ is closed under intersection with languages from $\mathrm{co}\mathcal{B}_{1/2,k}$.*

Proof. By definition, languages from $\mathrm{co}\mathcal{B}_{1/2,k}$ are finite intersections of languages L_i such that $L_i = \overline{D}^{\top}$ for $D \subseteq A^{\leq k}$ or $L_i = \overline{(w|A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_{n'}, A^{k+1}|v)_k}^{\top}$ where $n' \geq 1$. Since it is easy to see that $\widetilde{\mathcal{B}}_k$ is closed under intersection with $\overline{D}^{\top}$ for $D \subseteq A^{\leq k}$, it remains to show that $\widetilde{\mathcal{B}}_k$ is closed under intersection with languages of the form $\overline{(w|A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_{n'}, A^{k+1}|v)_k}^{\top}$. By Proposition 3.12, we may restrict ourselves to languages of the form $\left(\emptyset, \alpha, A^{k+1}\right)_k$, $\left(A^{k+1}, \alpha, \emptyset\right)_k$ or $\overline{(A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_{n'}, A^{k+1})_k}^{\top}$. Furthermore, by Proposition 3.11, it remains to show that $\widetilde{\mathcal{B}}_k$ is closed under intersection with languages being of the form $\left(\emptyset, \alpha, A^{k+1}\right)_k$ or $\left(A^{k+1}, \alpha, \emptyset\right)_k$.

Let us prove that $\widetilde{\mathcal{B}}_k$ is closed under intersection with languages of the form $\left(\emptyset, \alpha, A^{k+1}\right)_k$. It suffices to show $G \cap B \in \widetilde{\mathcal{B}}_k$ for $G =_{\mathrm{def}} (\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k$ and $B =_{\mathrm{def}} \left(\emptyset, \alpha, A^{k+1}\right)_k$ with $n \geq 1$, $\alpha, \gamma_1, \dots, \gamma_n \in A^{k+1}$ and $\Gamma_0, \dots, \Gamma_n \subseteq A^{k+1}$. We obtain

$$G \cap B = \begin{cases} (\emptyset, \alpha, \Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k \cup (\emptyset, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k & : \text{if } \alpha \in \Gamma_0 \wedge \gamma_1 = \alpha \\ (\emptyset, \alpha, \Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k & : \text{if } \alpha \in \Gamma_0 \wedge \gamma_1 \neq \alpha \\ (\emptyset, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k & : \text{if } \alpha \notin \Gamma_0 \wedge \gamma_1 = \alpha \\ \emptyset & : \text{if } \alpha \notin \Gamma_0 \wedge \gamma_1 \neq \alpha. \end{cases}$$

This shows $G \cap B \in \widetilde{\mathcal{B}}_k$ and we conclude that $\widetilde{\mathcal{B}}_k$ is closed under intersection with languages of the form $\left(\emptyset, \alpha, A^{k+1}\right)_k$. The same argument can be applied for languages being of the form $\left(A^{k+1}, \alpha, \emptyset\right)_k$. ❑

The more difficult inclusion is now a combination of our previous work in this section.

Lemma 3.14. *Let $k \geq 0$. Then $\mathcal{B}_{3/2,k} \subseteq \widetilde{\mathcal{B}}_k$.*

Proof. First we want to show $\mathcal{B}_{1/2,k} \subseteq \widetilde{\mathcal{B}}_k$. It suffices to show $L \in \widetilde{\mathcal{B}}_k$ for $L =_{\text{def}} \left(w|A^{k+1}, \beta_1, A^{k+1}, \dots, \beta_n, A^{k+1}|v\right)_k$ with $n \geq 1$, $w, v \in A^k$, $\beta_1, \dots, \beta_n \in A^{k+1}$. By definition of $\widetilde{\mathcal{B}}_k$, $L' =_{\text{def}} \left(A^{k+1}, \beta_1, A^{k+1}, \dots, \beta_n, A^{k+1}\right)_k \in \widetilde{\mathcal{B}}_k$. Let $L'_1 =_{\text{def}} w\left(w^{-1}L'\right)$ and observe that $L'_1 = L' \cap wA^*$. Since $|w| = k$ and $L' \subsetneq A^{\geq k+1}$ we can apply Proposition 3.1, and we obtain $w^{-1}L' \in \widetilde{\mathcal{B}}_k$. Hence, by Proposition 3.2 we have that $L'_1 \in \widetilde{\mathcal{B}}_k$ and $L'_1 \subsetneq A^{\geq k+1}$. Now let $L'_2 =_{\text{def}} \left(L'_1 v^{-1}\right) v$. The same argument shows $L'_2 = L'_1 \cap A^* v$ and $L'_2 \in \widetilde{\mathcal{B}}_k$. We conclude $L = wA^* \cap A^*v \cap L' = L'_2 \in \widetilde{\mathcal{B}}_k$ which shows $\mathcal{B}_{1/2,k} \subseteq \widetilde{\mathcal{B}}_k$.

By Lemma 3.13 we have in particular that $\mathrm{co}\mathcal{B}_{1/2,k} \subseteq \widetilde{\mathcal{B}}_k$ and $C \cap D \in \widetilde{\mathcal{B}}_k$ for all $C \in \mathcal{B}_{1/2,k}, D \in \mathrm{co}\mathcal{B}_{1/2,k}$. Since $\mathcal{B}_{1,k} = \mathrm{BC}(\mathcal{B}_{1/2,k})$ (see Theorem 2.20, Definition 2.21) and by Proposition 2.23.1 we have that languages of $\mathcal{B}_{1,k}$ can be written as finite unions of languages of the form C, D or $C \cap D$ with $C \in \mathcal{B}_{1/2,k}, D \in \mathrm{co}\mathcal{B}_{1/2,k}$. Therefore, $\mathcal{B}_{1,k} \subseteq \widetilde{\mathcal{B}}_k$. By Propositions 3.2 and 3.3 we have that $\widetilde{\mathcal{B}}_k = \mathrm{POL}(\widetilde{\mathcal{B}}_k)$. It follows that $\mathcal{B}_{3/2,k} = \mathrm{POL}(\mathcal{B}_{1,k}) \subseteq \widetilde{\mathcal{B}}_k$. ❑

It remains to establish the other inclusion.

Proposition 3.15. *Let $k \geq 0$ and $L = (\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_n, \Gamma_n)_k$ for $n \geq 0$, $\gamma_1, \dots, \gamma_n \in A^{k+1}$ and $\Gamma_0, \dots, \Gamma_n \subseteq A^{k+1}$. Then $L \in \mathcal{B}_{3/2,k}$.*

Proof. We will show the lemma by induction on n. For $n = 0$ we have

$$(\Gamma_0)_k = \overline{\bigcup_{\substack{\alpha \in A^{k+1} \setminus \Gamma_0 \\ w,v \in A^k}} \left(w|A^{k+1}, \alpha, A^{k+1}|v\right)_k}^{+} \cap \underbrace{\left(\bigcup_{w,v \in A^k} \left(w|A^{k+1}|v\right)_k\right)}_{=\left(A^{k+1}\right)_k}.$$

Since $\left(w|A^{k+1}|v\right)_k \in \mathcal{B}_{1/2,k}$ and $\left(w|A^{k+1}, \alpha, A^{k+1}|v\right)_k \in \mathcal{B}_{1/2,k}$ for all $w, v \in A^k, \alpha \in A^{k+1}$, we obtain $(\Gamma_0)_k \in \mathcal{B}_{1,k}$. This proves the induction base, because $\mathcal{B}_{1,k} \subseteq \mathcal{B}_{3/2,k}$.

Assume that the lemma has been shown for $n = m \geq 0$, we want to prove it for $n = m + 1$. Let $L = (\Gamma_0, \gamma_1, \Gamma_1, \dots, \gamma_{m+1}, \Gamma_{m+1})_k$, then from Lemma 3.4 we obtain

$$L = \underbrace{(\Gamma_0)_k}_{L_1 =_{\text{def}}} p_k(\gamma_1)^{-1} \cdot \gamma_1 \cdot s_k(\gamma_1)^{-1} \underbrace{(\Gamma_1, \gamma_2, \Gamma_2, \dots, \gamma_{m+1}, \Gamma_{m+1})_k}_{L_2 =_{\text{def}}}.$$

By induction hypothesis, $L_1, L_2 \in \mathcal{B}_{3/2,k}$. Since $L_1, L_2 \subseteq A^{\geq k+1}$ and $|p_k(\gamma_1)| = |s_k(\gamma_1)| = k$, we can apply Proposition 2.27. It follows that $L_1 p_k(\gamma_1)^{-1}, s_k(\gamma_1)^{-1} L_2 \in \mathcal{B}_{3/2,k}$. We conclude $L \in \mathcal{B}_{3/2,k}$ which completes the induction. ❑

Theorem 3.16. *Let $k \geq 0$. Then $\mathcal{B}_{3/2,k} = \widetilde{\mathcal{B}}_k$.*

Proof. By definition, $\mathcal{B}_{3/2,k}$ is closed under finite union. Furthermore, $D \in \mathcal{B}_{1/2,k} \subseteq \mathcal{B}_{3/2,k}$ for every $D \subseteq A^{\leq k}$. Therefore, from Proposition 3.15 we obtain $\widetilde{\mathcal{B}}_k \subseteq \mathcal{B}_{3/2,k}$, since it applies in particular for languages from $\widetilde{\mathcal{B}}_k$. The theorem follows together with Lemma 3.14. ❑

4 Finding Automata Loops in Words

A useful tool for further proofs will be the fact that we can find factors in a word which are loops with respect to a given dfa. It is important to analyse the needed size to find such a factor. So let F be a dfa and let w be a word having a specified number (depending on $|F|$) of different markers attached to it. We show that we can find in w a factor u between two markers with $w = xuy$ such that F can not distinguish whether w or xu^iy for $i \geq 1$ was in the input.

More precisely, let $F = (A, S, \delta, s_0, S')$ be a finite automaton. Recall from the basic definitions in Subsection 2.1 that a mapping $\delta' : S \to S$ leads to an x-loop in F (for some $x \in A^+$) if and only if $\delta(\delta'(s), x) = \delta'(s)$ for all $s \in S$. We want to find a factor u in a word w such that δ^u leads to an u-loop with respect to F. As will be shown in Lemma 4.4, this is always possible if we look at a number of markers in w which is large enough, but bounded. This number will be $\lfloor \mathcal{K}(n)/2 \rfloor$ with $n = |F|$.

Definition 4.1. $\mathcal{K}(n) =_{\text{def}} (n+1)^{(n+1)^{(n+1)}}$

We first show that this number does not become too small if we repeatingly divide it by n^n. This will make the proof of Lemma 4.4 below better readable.

Proposition 4.2. *Let $n \geq 1$, $m_1 =_{\text{def}} \lfloor \mathcal{K}(n)/2 \rfloor$ and $m_{i+1} =_{\text{def}} \lfloor m_i/n^n \rfloor - 1$ for $i \geq 1$. Then it holds for $1 \leq i \leq n^n + 1$ that*

$$m_i \geq (2n^n)^{(n^n+3-i)}.$$

Proof. We will prove the lemma by induction on i with $1 \leq i \leq n^n + 1$. First we want to prove the induction base, i.e. the case $i = 1$. By definition of $\mathcal{K}(n)$, for $n = 1$ we have $m_1 = 8 = (2n^n)^{(n^n+3-1)}$. Now let $n > 1$. By the binomial theorem, $(n+1)^{n+1} \geq n^{n+1} + (n+1)\,n^n + (n+1)\,n + 1 \geq n^{n+1} + 2n + 2$ (for $n > 1$) and $n^n + n \cdot n^{n-1} \leq (n+1)^n$. Hence we obtain the following estimation which shows the induction base.

$$\begin{aligned}
(2n^n)^{(n^n+3-1)} = \left(n^n + n \cdot n^{n-1}\right)^{(n^n+3-1)} &\leq (n+1)^{n(n^n+2)} \\
&\leq (n+1)^{\left((n+1)^{(n+1)}-2\right)} \\
&\leq (n+1)^{\left((n+1)^{(n+1)}-1\right)} - 1 \\
&\leq \frac{(n+1)^{\left((n+1)^{(n+1)}\right)}}{2} - 1 \\
&\leq \lfloor \mathcal{K}(n)/2 \rfloor = m_1
\end{aligned}$$

Suppose now that we have already shown $m_k \geq (2n^n)^{(n^n+3-k)}$ with $1 \leq k < n^n+1$. By definition, $m_{k+1} = \lfloor m_k/n^n \rfloor - 1$. By induction hypothesis we obtain

$$m_{k+1} \geq \frac{(2n^n)^{(n^n+3-k)}}{n^n} - 2.$$

Since $n > 0$ and $k < n^n + 1$ we have $(2n^n)^{(n^n+3-k)}/n^n \geq 8$. It follows that

$$\frac{(2n^n)^{(n^n+3-k)}}{n^n} - 2 = 2 \cdot (2n^n)^{(n^n+3-(k+1))} - 2 \geq (2n^n)^{(n^n+3-(k+1))}.$$

❑

The key argument for the loop lemma below (Lemma 4.4) is the iterated use of the fact that there is only a finite number of mappings $\delta' : S \to S$ when S is given. We isolate the iteration step in the following lemma. Suppose some mappings δ' are given and also a word $w = u_0 u_1 \cdots u_m$ with 'large' m, where each δ' leads to u_i-loops for all $0 \le i \le m$. Using that in general, if $\delta^x = \delta^{xy}$ then δ^x leads to an y-loop, we can always find a mapping δ^x with $x = u_0 \cdots u_j$ that leads to y-loops where each y is a concatenation of factors u_i with $i > j$. Then in particular any δ' leads to an x-loop. In order to make this precise we define for given $u_0, u_1, \ldots, u_l \in A^+$ as $u[i,j] =_{\mathrm{def}} u_i u_{i+1} \cdots u_{j-1}$ for all $0 \le i < j \le l+1$ the concatenation of the respective words. We work with indices $i_0, \ldots, i_m$ in order to permit an iterated application.

Lemma 4.3. *Let $F = (A, S, \delta, s_0, S')$ be a finite automaton and $u_0, u_1, \ldots, u_l \in A^+$ be a list of words. Furthermore, let $0 \le i_0 < i_1 < \cdots < i_m \le l$ be a list of indices and let M be a set of total mappings $\delta' : S \to S$, such that every $\delta' \in M$ leads to an $u[i_j, i_{j+1}]$-loop for all $0 \le j < m$. Then there exist indices $i'_0 < i'_1 < \cdots < i'_n$ with $n =_{\mathrm{def}} \lfloor m/(|S|^{|S|}) \rfloor$, such that*

1. *$\{i'_0, i'_1, \ldots, i'_n\} \subseteq \{i_0, i_1, \ldots, i_m\}$,*
2. *every $\delta' \in M$ leads to an $u[i'_0, i'_1]$-loop (for $n > 0$) and*
3. *every $\delta' \in M \cup \left\{\delta^{u[i'_0, i'_1]}\right\}$ leads to an $u[i'_j, i'_{j+1}]$-loop for all $1 \le j < n$.*

Proof. First we set $i'_0 =_{\mathrm{def}} i_0$. This shows in particular the lemma for $n = 0$. If $n = 1$ we set $i'_1 =_{\mathrm{def}} i_1$ and we are done. Suppose now $n \ge 2$ and define $\delta_{i_j} =_{\mathrm{def}} \delta^{u[i_0, i_j]}$ for $1 \le j \le m$. Since there are at most $|S|^{|S|}$ total mappings $S \to S$, there exist mappings appearing several times on the list $\delta_{i_1}, \delta_{i_2}, \ldots, \delta_{i_m}$. Among these mappings we choose a mapping $\overline{\delta}$ which appears most frequently, say $\overline{\delta}$ appears n' times. Hence by the pigeon hole principle we have $n' \ge \lfloor m/(|S|^{|S|}) \rfloor = n$. So there exist indices $i'_1, i'_2, \ldots, i'_n \in \{i_1, i_2, \ldots, i_m\}$ such that $i'_1 < i'_2 < \cdots < i'_n$ and $\overline{\delta} = \delta_{i'_j}$ for all $1 \le j \le n$.

Clearly, $\{i'_1, i'_2, \ldots, i'_n\} \subseteq \{i_1, i_2, \ldots, i_m\}$, which shows Statement 1. By assumption, every $\delta' \in M$ leads to an $u[i_j, i_{j+1}]$-loop for all $0 \le j < m$. It follows that every $\delta' \in M$ leads also to an u-loop, where u is an arbitrary concatenation of words $u[i_j, i_{j+1}]$ with $0 \le j < m$. Particularly, every $\delta' \in M$ leads to an $u[i'_j, i'_{j+1}]$-loop for all $0 \le j < n$, thus Statement 2 follows and also Statement 3 for $\delta' \in M$.

It remains to show that $\delta^{u[i'_0, i'_1]}$ leads to an $u[i'_j, i'_{j+1}]$-loop for all $1 \le j < n$. By the choice of $\overline{\delta}$ we have that $\delta^{u[i'_0, i'_j]} = \overline{\delta} = \delta^{u[i'_0, i'_{j+1}]}$ for all $1 \le j < n$ and we see that $\overline{\delta}$ leads to an $u[i'_j, i'_{j+1}]$-loop for all $1 \le j < n$. Since $\delta^{u[i'_0, i'_1]} = \overline{\delta}$, Statement 3 follows. ❑

Note that the Statements 2 and 3 in the previous lemma for $\delta' \in M$ follow already from Statement 1. We explicitely give them here to make clear what is used in the following proof. Clearly, the mapping we add to M in Lemma 4.3 can not always be a new mapping. Hence, if the number of markers we start with is large enough to allow many applications of Lemma 4.3, then we find a mapping δ^u which has already been taken to M before, say δ'. But this means that δ' leads to an u-loop and hence $\delta^{uu} = \delta^u$.

Lemma 4.4. *For every finite automaton $F = (A, S, \delta, s_0, S')$ and for all $u_0, u_1, \ldots, u_l \in A^+$ with $l =_{\mathrm{def}} \lfloor \mathcal{K}(|F|)/2 \rfloor$ there exist indices $0 \le g < h \le l$ such that $\delta^{uu} = \delta^u$ with $u =_{\mathrm{def}} u_g u_{g+1} \cdots u_{h-1}$.*

Proof. Let $n =_{\mathrm{def}} |F|$. Initially, let $m^{(1)} =_{\mathrm{def}} l$, $M^{(1)} =_{\mathrm{def}} \emptyset$ and $i_r^{(1)} =_{\mathrm{def}} r$ for $0 \le r \le l$. Now we apply Lemma 4.3 for a first time and obtain for $n^{(1)} =_{\mathrm{def}} \lfloor m^{(1)}/n^n \rfloor$ indices ${i'}_r^{(1)}$ with $0 \le r \le n^{(1)}$ such that

1. $\left\{ {i'}_r^{(1)} : 0 \le r \le n^{(1)} \right\} \subseteq \left\{ i_r^{(1)} : 0 \le r \le m^{(1)} \right\}$ and
2. $\delta^{u[{i'}_0^{(1)}, {i'}_1^{(1)}]}$ leads to an $u[{i'}_r^{(1)}, {i'}_{r+1}^{(1)}]$-loop for all $1 \le r < n^{(1)}$.

Now we want to start over after position ${i'}_1^{(1)}$ and set $m^{(2)} =_{\text{def}} n^{(1)}-1$, $M^{(2)} =_{\text{def}} M^{(1)} \cup \left\{ \delta^{u[{i'}_0^{(1)},{i'}_1^{(1)}]} \right\}$ and $i_r^{(2)} =_{\text{def}} {i'}_{r+1}^{(1)}$ for $0 \leq r < n^{(1)}$. Then we can apply Lemma 4.3 again. In general, after the j-th application of Lemma 4.3, we obtain for $n^{(j)} =_{\text{def}} \lfloor m^{(j)}/n^n \rfloor$ the indices ${i'}_r^{(j)}$ with $0 \leq r \leq n^{(j)}$ such that

1. $\left\{ {i'}_r^{(j)} : 0 \leq r \leq n^{(j)} \right\} \subseteq \left\{ i_r^{(j)} : 0 \leq r \leq m^{(j)} \right\}$,
2. every $\delta' \in M^{(j)}$ leads to an $u[{i'}_0^{(j)}, {i'}_1^{(j)}]$-loop (for $n^{(j)} > 0$) and
3. every $\delta' \in M^{(j)} \cup \left\{ \delta^{u[{i'}_0^{(j)},{i'}_1^{(j)}]} \right\}$ leads to an $u[{i'}_r^{(j)}, {i'}_{r+1}^{(j)}]$-loop for all $1 \leq r < n^{(j)}$.

Moreover, with $m^{(j+1)} =_{\text{def}} n^{(j)} - 1$, $M^{(j+1)} =_{\text{def}} M^{(j)} \cup \left\{ \delta^{u[{i'}_0^{(j)},{i'}_1^{(j)}]} \right\}$ and $i_r^{(j+1)} =_{\text{def}} {i'}_{r+1}^{(j)}$ for $0 \leq r < n^{(j)}$ we can go to the $(j+1)$-st application of Lemma 4.3.

We chose l in the beginning large enough such that we can apply Lemma 4.3 sufficiently often to face the same mapping twice. By Proposition 4.2 we have that $m^{(j)} \geq (2n^n)^{(n^n+3-j)}$ for $1 \leq j \leq n^n + 1$. It follows that $n^{(j)} = \lfloor m^{(j)}/n^n \rfloor \geq (2n^n)^{(n^n+2-j)} - 1 \geq 1$ for $1 \leq j \leq n^n + 1$. Particularly, the indices ${i'}_0^{(j)}$ and ${i'}_1^{(j)}$ exist for $1 \leq j \leq n^n + 1$.

On one hand, at the end of each step j we take $\delta^{u[{i'}_0^{(j)},{i'}_1^{(j)}]}$ to $M^{(j)}$ and obtain $M^{(j+1)}$. On the other hand, there are at most n^n total mappings $S \to S$. Therefore, there exists a step t with $1 \leq t \leq n^n + 1$ such that $\overline{\delta} =_{\text{def}} \delta^{u[{i'}_0^{(t)},{i'}_1^{(t)}]}$ is already an element of $M^{(t)}$. From Statement 2 of Lemma 4.3 it follows that $\overline{\delta}$ leads to an $u[{i'}_0^{(t)}, {i'}_1^{(t)}]$-loop. With $g =_{\text{def}} {i'}_0^{(t)}$, $h =_{\text{def}} {i'}_1^{(t)}$ and $u =_{\text{def}} u_g u_{g+1} \cdots u_{h-1}$ we have $u = u[{i'}_0^{(t)}, {i'}_1^{(t)}]$. Thus $\overline{\delta} = \delta^u$ leads to an u-loop and hence $\delta^{uu} = \delta^u$. ❑

If we do not have the particular number l of words, but factors $u_0, u_1, \ldots, u_n$ for arbitrary n, we can partition them in a number of blocks u such that each block is 'short', i.e., only $\mathcal{K}(|F|)$ factors u_i are involved, and every second block u has in fact the property $\delta^{uu} = \delta^u$. Note that we can understand $u_0, u_1, \ldots, u_n$ as some word $w = u_0 u_1 \cdots u_n$ with n markers attached to it. Thus we obtain with the following lemma a factorization of w where each factor grasps a bounded number of markers, and additionally each second factor has the nice property $\delta^{uu} = \delta^u$.

Lemma 4.5. *For every finite automaton $F = (A, S, \delta, s_0, S')$ and for all $u_0, u_1, \ldots, u_n \in A^+$ there exist an $m \geq 0$ and indices $0 = i_0 < i_1 < \cdots < i_{2m+1} = n+1$ such that*

1. *$i_{j+1} - i_j \leq \mathcal{K}(|F|)$ for $0 \leq j \leq 2m$ and*
2. *$\delta^{uu} = \delta^u$ for all $u = u_{i_j} u_{i_j+1} u_{i_j+2} \cdots u_{i_{j+1}-1}$ with $1 \leq j < 2m$ and $j \equiv 1(2)$.*

Proof. Let $l =_{\text{def}} \lfloor \mathcal{K}(|F|)/2 \rfloor$. If $n < \mathcal{K}(|F|)$ then we set $m =_{\text{def}} 0$, $i_0 =_{\text{def}} 0$, $i_1 =_{\text{def}} n+1$ and we are done. Otherwise we divide the list $u_1, \ldots, u_n$ from left to right into blocks, such that every block contains $l+1$ words u_j. We obtain $m \geq 1$ blocks $B_1, \ldots, B_m$ and $r \leq l$ remaining words $u_{n-r+1}, \ldots, u_n$. For every block $B_t = (u_j, u_{j+1}, \ldots, u_{j+l})$ with $j = (t-1)(l+1)+1$ and $1 \leq t \leq m$ we can apply Lemma 4.4, and we obtain indices $j \leq g_t < h_t \leq j+l$ such that $\delta^{uu} = \delta^u$ with $u =_{\text{def}} u_{g_t} u_{g_t+1} \cdots u_{h_t-1}$. Now let $i_0 =_{\text{def}} 0$, $i_{2m+1} =_{\text{def}} n+1$ and $i_{2t-1} =_{\text{def}} g_t$, $i_{2t} =_{\text{def}} h_t$ for $1 \leq t \leq m$. Since $0 = i_0 < i_1 < \cdots < i_{2m+1} = n+1$ we already have Statement 2.

It remains to show Statement 1. For $1 \leq t \leq m$ we have $i_{2t} - i_{2t-1} = h_t - g_t \leq l < \mathcal{K}(|F|)$. For $1 \leq t < m$ we have $B_t = (u_j, u_{j+1}, \ldots, u_{j+l})$ and $B_{t+1} = (u_{j+l+1}, u_{j+l+2}, \ldots, u_{j+2l+1})$ with $j = (t-1)(l+1)+1$. Since $j \leq g_t < h_t \leq j+l < j+l+1 \leq g_{t+1} < h_{t+1} \leq j+2l+1$ it follows that $g_{t+1} - h_t \leq (j+2l) - (j+1) = 2l - 1 < \mathcal{K}(|F|)$. Moreover $j_1 - j_0 = g_1 \leq l < \mathcal{K}(|F|)$, so we have shown $i_{j+1} - i_j \leq \mathcal{K}(|F|)$ for $0 \leq j \leq 2m-1$.

We are left with $j_{2m+1} - j_{2m}$. Observe that $B_m = (u_{n-r-l}, u_{n-r-l+1}, \ldots, u_{n-r})$ and that $h_m > n-r-l$. So $j_{2m+1} - j_{2m} = n+1-j_{2m} = n+1-h_m < n+1-n+r+l = r+l+1 \leq 2l+1 \leq \mathcal{K}(|F|)+1$ and hence $j_{2m+1} - j_{2m} \leq \mathcal{K}(|F|)$. This proves the lemma. ❑

In particular we want to state here the case that all u_i are letters. For a given dfa F with transition function δ and an arbitrary word w, we can divide w into consecutive blocks of at most $\mathcal{K}(|F|)$ letters, such that every second section u has the property $\delta^{uu} = \delta^{u}$.

Corollary 4.6. *For every finite automaton $F = (A, S, \delta, s_0, S')$ and every $w \in A^+$ there exist words $w_0, \ldots, w_m, u_1, \ldots, u_m \in A^{\leq \mathcal{K}(|F|)}$ such that $w = w_0 u_1 w_1 \cdots u_m w_m$ and $\delta^{u_i} = \delta^{u_i u_i}$ for $1 \leq i \leq m$.*

Proof. This follows from Lemma 4.5, if $u_0, u_1, \ldots, u_n$ is the factorization of w into single letters. ❑

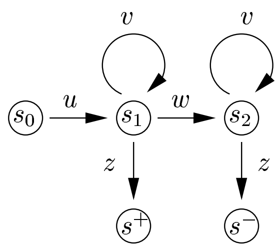

Figure 1: Pattern $\mathbf{P}_{\boldsymbol{k}}$ with $v \in A^{\geq k+1}$, $\alpha(\widehat{vwv}) \subseteq \alpha(\widehat{vv})$, $s^+ \in S'$ and $s^- \notin S'$.

5 Forbidden Pattern Characterization of $\mathcal{B}_{3/2,k}$

The goal of this section is to prove for fixed $k \geq 0$ a forbidden pattern characterization of the class $\mathcal{B}_{3/2,k}$, i.e., we want to show that a language L is in $\mathcal{B}_{3/2,k}$ if and only if the minimal automaton accepting L does not have a certain pattern in its transition graph. As it turns out, the emerging pattern (cf. Figure 1) has some requirements depending on k attached to the involved words. The case $k = 0$ is a result due to [PW97] who proved an effective characterization of level $3/2$ of the so-called Straubing-Thérien hierarchy. This was so far the highest level of a concatenation hierarchy for which a decidability result was known. In their paper the authors make heavy nontrivial use of the algebraic machinery and apply a fundamental result about factorization forests from I. Simon in [Sim90]. If one reads this section having in mind '$k = 0$' one can understand it as a self-contained reproof of their result, this time using an automata-theoretic approach. However, our generalization to arbitrary k will culminate in the following section in a decidability result for dot-depth $3/2$. As usual in forbidden pattern proofs, there is on one hand an easier to prove implication (see Subsection 5.2) and on the other hand a more difficult one (see Subsection 5.3).

5.1 The Pattern $\mathbf{P}_{\boldsymbol{k}}$

For every $k \geq 0$ we define the Pattern $\mathbf{P}_{\boldsymbol{k}}$, and we will show that the language accepted by some dfa F is in $\mathcal{B}_{3/2,k}$ if and only if F does not have this pattern.

Definition 5.1. Let $k \geq 0$. A dfa $F = (A, S, \delta, s_0, S')$ has Pattern $\mathbf{P}_{\boldsymbol{k}}$ if there are states $s_1, s_2, s^+, s^- \in S$ and words $u, v, w, z \in A^*$ with $|v| \geq k{+}1$ and $\alpha(\widehat{vwv}) \subseteq \alpha(\widehat{vv})$. Here the latter are k-decompositions of words. Moreover, it must be that $s^+ \in S'$, $s^- \notin S'$, $\delta(s_0, u) = s_1 = \delta(s_1, v)$, $\delta(s_1, w) = s_2 = \delta(s_2, v)$, $\delta(s_1, z) = s^+$ and $\delta(s_2, z) = s^-$ (cf. Figure 1).

A pattern equivalent to $\mathbf{P_0}$ is considered in [PW97]. If one reads the following subsections having in mind '$k = 0$' one obtains a reproof of the forbidden pattern characterization of the Straubing-Thérien hierarchy given in [PW97].

We want to give some intuition for $\mathbf{P}_{\boldsymbol{k}}$. Suppose we have a language L from $\mathcal{B}_{3/2,k} = \widetilde{\mathcal{B}_k}$ with $L = (\Sigma_0, \alpha_1, \Sigma_1, \dots, \alpha_n, \Sigma_n)_k$ with $\alpha_1, \dots, \alpha_n \in A^{k+1}$ and $\Sigma_0, \Sigma_1, \dots, \Sigma_n \subseteq A^{k+1}$. If there is $\mathbf{P}_{\boldsymbol{k}}$ in some dfa accepting L and we are after input u in state s_1 in $\mathbf{P}_{\boldsymbol{k}}$, we can pump up $uz \in L$ to arbitrary $uv^iz \in L$. Hence it must be that there is some j such that $\alpha(\widehat{vv}) \in \Sigma_j$. Now if $\alpha(\widehat{vwv}) \subseteq \alpha(\widehat{vv})$ we can insert w and still have a k-decomposition that shows that we remain in L, a contradiction to the occurence of Pattern $\mathbf{P}_{\boldsymbol{k}}$. This is more or less already the argumentation for the easy implication of the forbidden pattern characterization.

5.2 Easy Implication

We make precise what we just described. One can insert a word w into uv^nz if n is large enough compared to the language description of a given language from $\mathcal{B}_{3/2,k}$ and if w does not cause new elements in the respective k-decomposition.

Lemma 5.2. *Let $k \geq 0$, $n' \geq 1$, $n \geq 4n'+3$, $\gamma_1, \ldots, \gamma_{n'} \in A^{k+1}$ and $\Gamma_0, \ldots, \Gamma_{n'} \subseteq A^{k+1}$. Then for all $u, w, z \in A^*$, $v \in A^{\geq k+1}$ with $\alpha(\widehat{vwv}) \subseteq \alpha(\widehat{vv})$ there exists an $1 \leq m < n$ such that*

$$uv^n z \in (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_{n'}, \Gamma_{n'})_k \Longrightarrow uv^m wv^{n-m} z \in (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_{n'}, \Gamma_{n'})_k .$$

Proof. Assume that $uv^n z \in (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_{n'}, \Gamma_{n'})_k$. First we choose suitable $l \geq 1$, $a_1, \ldots, a_{l+k} \in A$, $\alpha_1, \ldots, \alpha_l \in A^{k+1}$ such that $uv^n z = a_1 \cdots a_{l+k}$ and $\widehat{uv^n z} = (\alpha_1, \ldots, \alpha_l)$. By assumption, there exist indices $0 = p_0 < p_1 < \cdots < p_{n'} < p_{n'+1} = l+1$ such that

(a) $\alpha_{p_i} = \gamma_i$ for all $1 \leq i \leq n'$ and

(b) $\alpha_j \in \Gamma_i$ for all $0 \leq i \leq n'$ and $p_i < j < p_{i+1}$.

We denote the position in $uv^n z$ where the i-th v starts by q_i, i.e. for $1 \leq i \leq n$ we set $q_i =_{\text{def}} 1 + |u| + (i-1)|v|$. Since $|v| \geq k+1$, we obtain n different positions such that $1 \leq q_1 < q_2 < \cdots < q_n \leq l$ and $q_i - q_{i-1} = |v|$ for $1 < i \leq n$.

By assumption we have that $n \geq 4n'+3$, so at least $3(n'+1)$ of the positions q_i are different from $p_1, \ldots, p_{n'}$. By the pigeon hole principle, there exist h, m with $0 \leq h \leq n'$ and $1 \leq m \leq n-2$ such that $p_h < q_m < q_{m+1} < q_{m+2} < p_{h+1}$. From now on let $\lambda =_{\text{def}} q_m$, $\mu =_{\text{def}} q_{m+1}$ and $\nu =_{\text{def}} q_{m+2}$. Since $a_\lambda a_{\lambda+1} \cdots a_{\mu-1} = a_\mu a_{\mu+1} \cdots a_{\nu-1} = v$ and $|v| \geq k+1$, we have

$$\alpha(\widehat{vv}) \subseteq \Gamma_h. \tag{10}$$

Now let $x =_{\text{def}} uv^m wv^{n-m} z$, and choose suitable $b_1, \ldots, b_{l+k+|w|} \in A$, $\beta_1, \ldots, \beta_{l+|w|} \in A^{k+1}$ such that $x = b_1 \cdots b_{l+k+|w|}$ and $\widehat{x} = \big(\beta_1, \ldots, \beta_{l+|w|}\big)$. Observe that

$$a_i = \begin{cases} b_i & : \text{ if } 1 \leq i \leq \mu - 1, \\ b_{i+|w|} & : \text{ if } \mu \leq i \leq l+k \end{cases} \tag{11}$$

and

$$\alpha_j = \begin{cases} \beta_j & : \text{ if } 1 \leq j \leq \mu - k - 1, \\ \beta_{j+|w|} & : \text{ if } \mu \leq j \leq l. \end{cases} \tag{12}$$

Let $r_i =_{\text{def}} p_i$ for $0 \leq i \leq h$ and $r_i =_{\text{def}} p_i + |w|$ for $h < i \leq n'+1$. Since $p_h \leq \lambda = \mu - |v| \leq \mu - k - 1$ and $p_{h+1} \geq \mu$, we have $r_h \leq \mu - k - 1$ and $r_{h+1} \geq \mu + |w|$. From Equation (12) it follows that $\beta_i = \alpha_i$ for $1 \leq i \leq r_h$ and $\beta_i = \alpha_{i-|w|}$ for $r_{h+1} \leq i \leq l + |w|$. Thus we have $\beta_{r_i} = \beta_{p_i} = \alpha_{p_i}$ for $1 \leq i \leq h$, and $\beta_{r_i} = \beta_{p_i+|w|} = \alpha_{p_i}$ for $h < i \leq n'$. Therefore, we obtain

$$\begin{aligned} &\beta_{r_i} = \gamma_i \text{ for all } i \text{ with } 1 \leq i \leq n', && (13) \\ &\beta_j = \alpha_j \in \Gamma_i \text{ for all } i, j \text{ with } 0 \leq i < h \text{ and } r_i = p_i < j < p_{i+1} = r_{i+1} \text{ and} \\ &\beta_j = \alpha_{j-|w|} \in \Gamma_i \text{ for all } i, j \text{ with } h < i \leq n' \text{ and } r_i = p_i + |w| < j < p_{i+1} + |w| = r_{i+1}. \end{aligned}$$

To see $x \in (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_{n'}, \Gamma_{n'})_k$, it remains to show $\beta_j \in \Gamma_h$ for all $r_h < j < r_{h+1}$. By Equation (12) we have that $\beta_j = \alpha_j \in \Gamma_h$ for all j with $r_h < j < \mu - k$, and also $\beta_j = \alpha_{j-|w|} \in \Gamma_h$ for all j with $\mu + |w| \leq j < r_{h+1}$. Therefore, it remains to show $\beta_j \in \Gamma_h$ for all $\mu - k \leq j < \mu + |w|$. Since $b_{\mu-k} b_{\mu-k+1} b_{\mu-k+2} \cdots b_{\mu+|w|+k-1} = s_k(v) w p_k(v)$, we have $\beta_j \in \alpha(\widehat{vwv})$ for all $\mu - k \leq j < \mu + |w|$. By assumption and Equation (10) it follows that $\alpha(\widehat{vwv}) \subseteq \Gamma_h$. So $\beta_j \in \Gamma_h$ for all $\mu - k \leq j < \mu + |w|$ and hence we have shown

$$\beta_j \in \Gamma_h \text{ for all } r_h < j < r_{h+1}. \tag{14}$$

From Equations (13) and (14) we conclude that

(a) $\beta_{r_i} = \gamma_i$ for all $1 \leq i \leq n'$ and
(b) $\beta_j \in \Gamma_i$ for all $0 \leq i \leq n'$ and $r_i < j < r_{i+1}$.

Therefore, we have $x = uv^m wv^{n-m} z \in (\Gamma_0, \gamma_1, \Gamma_1, \ldots, \gamma_{n'}, \Gamma_{n'})_k$. ❑

Now it is easy to show that if some dfa F accepts some language from $\mathcal{B}_{3/2,k}$ then F must not have Pattern $\mathbf{P}_{\boldsymbol{k}}$.

Theorem 5.3. *Let $k \geq 0$. Let $L \in \mathcal{B}_{3/2,k}$ and let F be a dfa with $L(F) = L$. Then F does not have Pattern $\mathbf{P}_{\boldsymbol{k}}$.*

Proof. Let $L \in \mathcal{B}_{3/2,k}$ and let $F = (A, S, \delta, s_0, S')$ be a dfa with $L(F) = L$. By Theorem 3.16 we can write L as

$$L = \bigcup_{i=1}^{n'} (\Gamma_{i,0}, \gamma_{i,1}, \Gamma_{i,1}, \ldots, \gamma_{i,n_i}, \Gamma_{i,n_i})_k \cup D$$

for suitable $n' \geq 0$, $n_i \geq 1$, $\gamma_{i,j} \in A^{k+1}$, $\Gamma_{i,j} \subseteq A^{k+1}$ and $D \subseteq A^{\leq k}$. Suppose that F has Pattern $\mathbf{P}_{\boldsymbol{k}}$, then $uv^i z \in L$ and $uv^i wv^j z \notin L$ for all $i, j \geq 0$. Particularly, $uv^n z \in L$ with $n =_{\text{def}} \max\{4n_i + 3 : 1 \leq i \leq n'\}$ (note that $n' \geq 1$, because L contains an infinite subset). Hence $|uv^n z| \geq k+1$ and there exists an l with $uv^n z \in (\Gamma_{l,0}, \gamma_{l,1}, \Gamma_{l,1}, \ldots, \gamma_{l,n_l}, \Gamma_{l,n_l})_k$. Since $n \geq 4n_l + 3$, we can apply Lemma 5.2 and we obtain $uv^m wv^{n-m} z \in (\Gamma_{l,0}, \gamma_{l,1}, \Gamma_{l,1}, \ldots, \gamma_{l,n_l}, \Gamma_{l,n_l})_k \subseteq L$ for a suitable m with $1 \leq m < n$. This is a contradiction to the fact that in Pattern P_k the state s^- is rejecting. We conclude that F does not have Pattern $\mathbf{P}_{\boldsymbol{k}}$. ❑

5.3 Automata without Pattern $\mathbf{P}_{\boldsymbol{k}}$ accept Languages which are in $\boldsymbol{\mathcal{B}_{3/2,k}}$

In the following we will consider expressions E of the form $w_0 \cdot (v_1|\Sigma_1|v'_1)_k \cdot w_1 \cdots (v_n|\Sigma_n|v'_n)_k \cdot w_n$, where $k, n \geq 0$, $w_i \in A^+$, $v_j, v'_j \in A^k$ and $\Sigma_j \subseteq A^{k+1}$ for all $0 \leq i \leq n$ and $1 \leq j \leq n$. Again, in order not to overload notations, we will identify the expression E with the language described by E (cf. Definitions 2.3 and 2.5), i.e., we will at the same time make statements like '$x \in E$' and also talk about the size of E as a syntactical object. For a fixed $k \geq 0$ we denote the set of all such expressions by $\mathcal{E}_k$.

In the following we will start with an expression $E \in \mathcal{E}_k$, and we will replace 'long' factors of the w_i by 'short' expressions $E' \in \mathcal{E}_k$. Observe that expressions generated with this procedure are again in $\mathcal{E}_k$ (even if we replace w_i completely with an expression $E' \in \mathcal{E}_k$).

Definition 5.4. *Let $k \geq 0$ and $E \in \mathcal{E}_k$ with $E = w_0 \cdot (v_1|\Sigma_1|v'_1)_k \cdot w_1 \cdots (v_n|\Sigma_n|v'_n)_k \cdot w_n$, then we define the size of expression E as*

$$\|E\| =_{\text{def}} |w_0 w_1 \cdots w_n|.$$

Since the w_i are non-empty words, for any fixed $k \geq 0$ there exist only a finite number of expressions in $\mathcal{E}_k$ having the same size. We define a function that will help us to analyze the size of expressions in the following lemma. The variables k, a, f, n will be associated with the level k of $\mathcal{B}_{3/2,k}$, the size of the alphabet A, the size of the automaton F and the cardinality of $\alpha(\widehat{w})$ for a given word w, respectively.

Definition 5.5.

$$\mathcal{L}(k, a, f, n) =_{\text{def}} \begin{cases} k & : \text{ if } n = 0 \\ 2f^f + k + 1 & : \text{ if } n = 1 \\ 3\mathcal{K}(f) \cdot (5f^f a^k + 1) \cdot \mathcal{L}(k, a, f, n-1) & : \text{ otherwise} \end{cases}$$

Recall that $\mathcal{K}(f) = (f+1)^{(f+1)^{(f+1)}}$ is the function being used to find automata loops in words. Our main lemma states, under the assumption that a dfa F does not have Pattern $\mathbf{P}_{\boldsymbol{k}}$, that for any word x we can find an expression E_x of bounded size such that $x \in E_x$ and if $x \in L(F)$ then $E_x \subseteq L(F)$. In the formulation of the lemma we consider prefixes and suffixes in order to allow an inductive proof. From this lemma we can then follow the desired implication.

Lemma 5.6. *Let $k \geq 0$ and let F be a dfa which does not have Pattern $\mathbf{P}_{\boldsymbol{k}}$. Then for every $x \in A^+$ there exists an expression $E_x \in \mathcal{E}_k$ with $x \in E_x$ and $\|E_x\| \leq \mathcal{L}(k, |A|, |F|, |\alpha(\widehat{x})|)$ such that for all $x', x'' \in A^*$ it holds that*

$$x'xx'' \in L(F) \Longrightarrow x'E_xx'' \subseteq L(F).$$

Proof. Let $k \geq 0$ and let $F = (A, S, \delta, s_0, S')$ be a dfa which does not have Pattern $\mathbf{P}_{\boldsymbol{k}}$. For this proof we extend the definition of $\alpha(\widehat{x})$ to words $x \in A^{\leq k}$, namely we set $\alpha(\widehat{x}) =_{\mathrm{def}} \emptyset$ for such x. We will prove the theorem by induction on $N =_{\mathrm{def}} |\alpha(\widehat{x})|$ with $0 \leq N \leq |A|^{k+1}$. For the induction base we consider the cases $N = 0$ and $N = 1$.

INDUCTION BASE. For $N = 0$ we set $E_x =_{\mathrm{def}} x$ and we are done. Let $x \in A^+$ with $N = |\alpha(\widehat{x})| = 1$. If $|x| \leq \mathcal{L}(k, |A|, |F|, 1)$, again we set $E_x =_{\mathrm{def}} x$ and we are done. Otherwise, we have $|x| > 2|F|^{|F|} + k + 1$. Since $|x| \geq k+2$ we obtain by comparing letters that $x = a^{|x|}$ for an alphabet letter $a \in A$. Let us consider the mappings induced by prefixes of x. Then there exist i, j with $1 \leq i < j \leq |F|^{|F|} + 1$ with $\delta^{a^i} = \delta^{a^j}$. Thus for all $m \geq 0$ it holds that

$$\delta^{a^i} = \delta^{a^{i+m(j-i)}}. \tag{15}$$

We can choose an l with $1 \leq l \leq j-i$ in such a way that $|x|-i-l \equiv 0 \mod (j-i)$. From Equation (15) we obtain

$$\delta^{a^{i+l}} = \delta^x. \tag{16}$$

Now define $E_x =_{\mathrm{def}} a^i \cdot \left(a^k | \left\{a^{k+1}\right\} |a^k\right)_k \cdot a^l$. Observe that $1 \leq i, l \leq |F|^{|F|}$ and $|x| - i - l \geq k+1$. It follows that $x \in E_x$ and $E_x \in \mathcal{E}_k$, because $\|E_x\| = i + l \leq 2|F|^{|F|} \leq \mathcal{L}(k, |A|, |F|, 1)$.

Let $x', x'' \in A^*$ such that $x'xx'' \in L(F)$. Let $s_1 =_{\mathrm{def}} \delta(s_0, x'a^i)$, then by Equation (15) we have $\delta(s_1, a^{m(j-i)}) = s_1$ for all $m \geq 0$. Assume that there is some $h \geq k+1$ such that $\delta(s_1, a^h a^l x'') \notin L(F)$. Let $s_2 =_{\mathrm{def}} \delta(s_1, a^h)$ and define $u =_{\mathrm{def}} x'a^i$, $v =_{\mathrm{def}} a^{(k+1)(j-i)}$, $w =_{\mathrm{def}} a^h$ and $z =_{\mathrm{def}} a^l x''$. Then $\alpha(\widehat{vwv}) \subseteq \alpha(\widehat{vv})$, since only a's occur and $|v| = (j-i)(k+1) \geq k+1$. Since s_1 has a v-loop and v consists only of a's, we do not leave this loop with the word a^h, and we obtain that s_2 has also a v-loop. Furthermore, we have $\delta(s_0, u) = s_1$, $\delta(s_1, z) \in S'$ (by Equation (16)) and $\delta(s_2, z) \notin S'$ (by assumption). This shows that we found pattern $\mathbf{P}_{\boldsymbol{k}}$ in F, a contradiction. So we have for all $h \geq k+1$ that $\delta(s_0, x'a^i a^h a^l x'') = \delta(s_1, a^h a^l x'') \in L(F)$. Particularly, $x'E_xx'' \subseteq L(F)$. This proves the induction base.

INDUCTION STEP. We state the induction hypothesis.

> *For all $x \in A^+$ with $0 \leq |\alpha(\widehat{x})| \leq N < |A|^{k+1}$ there exists $E_x \in \mathcal{E}_k$ with $x \in E_x$ and $\|E_x\| \leq \mathcal{L}(k, |A|, |F|, |\alpha(\widehat{x})|)$ such that for all $x', x'' \in A^*$ it holds that*
>
> $$x'xx'' \in L(F) \Longrightarrow x'E_xx'' \subseteq L(F).$$

Let $x \in A^+$ with $\alpha(\widehat{x}) = N+1$ and $N \geq 1$. First of all we decompose x into so-called 'sectors', i.e., into factors s_i of x with $|\alpha(\widehat{s})| \leq N$ (actual we will have $|\alpha(\widehat{s})| = N$ for all sectors except for the last one). This is done in the following way: We start with x and determine the longest prefix s_1 of x such that $|\alpha(\widehat{s_1})| \le N$. Now we start over with $s_1^{-1}x$, determine the longest prefix s_2 of $s_1^{-1}x$ such that $|\alpha(\widehat{s_2})| \le N$ and proceed with $(s_1s_2)^{-1}x$. If we continue this procedure, we obtain a factorization of x into sectors $s_1, s_2, \dots, s_l$ for some $l \ge 2$ such that

1. $x = s_1 s_2 \cdots s_l$ with $|s_i| \ge k+1$ for $1 \le i < l$ (since $|\alpha(\widehat{s_i})| = N \ge 1$ for $1 \le i < l$),
2. $\alpha(\widehat{s_i}) \subsetneq \alpha(\widehat{x})$ for $1 \le i \le l$, and
3. $\alpha(\widehat{s_i s_{i+1}}) = \alpha(\widehat{x})$ for $1 \le i < l$ (since we have chosen longest prefixes s_i).

We give an outline of the further argumentation. Note that sectors do not have to be short factors of x, i.e., we do not know how to bound the length of a sector and we can not state a particular bound for the number of sectors. The main task of the induction step is to replace the unbounded number of consecutive sectors by a bounded number of terms of the form $(w|\Sigma|v)_k$ in such a way that (i) we do not leave $L(F)$ if we started with $x \in L(F)$ and (ii) we obtain an expression where only a bounded number of sectors and of terms of the form $(w|\Sigma|v)_k$ is left. The induction hypothesis will provide us with expressions of bounded size for the remaining sectors. Note that the closure under concatenation of the classs in question is necessary for our approach. All in all we will obtain an expression of bounded size containing x and being a subset of $L(F)$. This will prove our lemma. We distinguish two cases.

Case 1: Assume that the number of sectors is small, i.e., assume $l \le 3\mathcal{K}(|F|) \cdot (5|F|^{|F|} \cdot |A|^k + 1)$. Here there is not much to do since l is small and $|\alpha(\widehat{s_i})| \le N$ for all $1 \le i \le l$. Thus, by induction hypothesis, for $1 \le i \le l$ we find an expression $E_{s_i} \in \mathcal{E}_k$ such that $s_i \in E_{s_i}$, $\|E_{s_i}\| \le \mathcal{L}(k, |A|, |F|, N)$ and

$$s_i' E_{s_i} s_i'' \subseteq L(F) \text{ for every } s_i', s_i'' \in A^* \text{ with } s_i' s_i s_i'' \in L(F). \tag{17}$$

We define $E_x =_{\text{def}} E_{s_1} \cdot E_{s_2} \cdots E_{s_l}$. It follows that $x \in E_x$ and $\|E_x\| \le \mathcal{L}(k, |A|, |F|, N+1)$.

Now let $x', x'' \in A^*$ such that $x'xx'' \in L(F)$. We need to show that $x'E_x x'' \subseteq L(F)$. Let $y \in E_x$. Then for $1 \le i \le l$ there exist $y_i \in E_{s_i}$ such that $y = y_1 y_2 \cdots y_l$. Starting with the word $x' s_1 s_2 \cdots s_l x''$ we will step by step replace the sectors s_i by y_i without leaving $L(F)$. For the first step we define $s_l' =_{\text{def}} x' s_1 \cdots s_{l-1}$ and $s_l'' =_{\text{def}} x''$, thus we have $x'xx'' = s_l' s_l s_l'' \in L(F)$. From Equation (17) it follows that $s_l' E_{s_l} s_l'' \subseteq L(F)$, so in particular $s_l' y_l s_l'' \in L(F)$, which shows $x' s_1 \cdots s_{l-1} y_l x'' \in L(F)$. For the second step let $s_{l-1}' =_{\text{def}} x' s_1 \cdots s_{l-2}$ and $s_{l-1}'' =_{\text{def}} y_l x''$. Thus we have $s_{l-1}' s_{l-1} s_{l-1}'' \in L(F)$ by the previous step, and again from Equation (17) it follows that $s_{l-1}' y_{l-1} s_{l-1}'' \in L(F)$. This shows $x' s_1 \cdots s_{l-2} y_{l-1} y_l x'' \in L(F)$. If we continue this procedure, we finally obtain $x' y_1 \cdots y_l x'' = x' y x'' \in L(F)$. Hence $x' E_x x'' \subseteq L(F)$.

Case 2: Suppose the number of sectors is large, i.e., suppose $l > 3\mathcal{K}(|F|) \cdot (5|F|^{|F|} \cdot |A|^k + 1)$. Define $u_i =_{\text{def}} s_{2i-1} s_{2i}$ for $1 \le i \le \lfloor l/2 \rfloor - 1$ and $u_{\lfloor l/2 \rfloor} =_{\text{def}} s_{2\lfloor l/2 \rfloor - 1} \cdots s_l$. Note that if l is odd, then $u_{\lfloor l/2 \rfloor}$ contains three sectors. So $x = u_1 u_2 \cdots u_{\lfloor l/2 \rfloor}$ and every u_i contains at least two and at most three sectors with the effect that $\alpha(\widehat{u_i}) = \alpha(\widehat{x})$. Now we apply Lemma 4.5 to the list of words $u_1, u_2, \dots, u_{\lfloor l/2 \rfloor}$. We obtain a new list of words $x_0', x_1, x_1', x_2, x_2', \dots, x_m, x_m'$ with $m \ge 1$ (since l is large) such that $x = x_0' x_1 x_1' x_2 x_2' \cdots x_m x_m'$ and the following fact holds.

Fact 1.
1. Every x_i (x_i', resp.) is a concatenation of at least one and at most $\mathcal{K}(|F|)$ words u_j. Thus it contains at least two and at most $3\mathcal{K}(|F|)$ sectors.
2. For every x_i it holds that $\delta^{x_i x_i} = \delta^{x_i}$.

Let us consider the following example before we continue. Here l is odd and $u_{\lfloor l/2 \rfloor}$ contains three sectors. The picture shows the word x and each letter of x is represented by a star. We have obtained a description of x which eventually looks like this:

$$\underbrace{\underbrace{\overbrace{\star\star\star\star\star}^{s_1}\overbrace{\star\star\star}^{s_2}}_{u_1}\underbrace{\overbrace{\star\star\star}^{s_3}\overbrace{\star\star\star\star\star}^{s_4}}_{u_2}}_{x'_0}\underbrace{\underbrace{\overbrace{\star\star\star\star\star}^{s_5}\overbrace{\star\star\star\star}^{s_6}}_{u_3}\underbrace{\overbrace{\star\star\star}^{s_7}\overbrace{\star\star\star\star}^{s_8}}_{u_4}\underbrace{\overbrace{\star\star\star}^{s_9}\overbrace{\star\star\star\star}^{s_{10}}}_{u_5}}_{x_1}\underbrace{\underbrace{\overbrace{\star\star\star}^{s_{11}}\overbrace{\star\star\star}^{s_{12}}}_{u_6}\underbrace{\overbrace{\star\star\star}^{s_{13}}\overbrace{\star\star\star}^{s_{14}}}_{u_7}}_{x'_1}\cdots\underbrace{\underbrace{\overbrace{\star\star\star}^{s_{l-4}}\overbrace{\star\star\star\star}^{s_{l-3}}}_{u_{\lfloor l/2\rfloor-1}}\underbrace{\overbrace{\star\star\star}^{s_{l-2}}\overbrace{\star\star\star}^{s_{l-1}}\overbrace{\star\star\star}^{s_l}}_{u_{\lfloor l/2\rfloor}}}_{x'_m}$$

Next we assign for $1 \leq i \leq m$ to each x_i a tag representing the mapping $\delta_i =_{\text{def}} \delta^{x'_0 x_1 x'_1 \cdots x_i}$ and also $s_k(x_i)$, the k-suffix of x_i. Notice that there are at most $|F|^{|F|} \cdot |A|^k$ different tags. Now the task is to find maximal factors between some x_i and x_j having the same tags. We will call such a maximal factor a 'region' and will argue below, how the number of such regions can be bounded. Let us start with an algorithm which describes how to determine such regions. First of all we choose some j_1, j'_1 with $1 \leq j_1 < j'_1 \leq m$ such that x_{j_1} and $x_{j'_1}$ have the same tag and $j'_1 - j_1$ is maximal. With (j_1, j'_1) we have found the first region and we mark all x_i for $j_1 \leq i \leq j'_1$ as already used, i.e., we mark all x_i within this region. We call x_{j_1} ($x_{j'_1}$) the left border (right border, resp.) and the word $x_{j_1+1}x_{j_1+2}\cdots x_{j'_1-1}$ the content of the region (j_1, j'_1). In a second step we choose j_2, j'_2 with $1 \leq j_2 < j'_2 \leq m$ such that none of the $x_{j_2}, x_{j_2+1}, \ldots, x_{j'_2}$ is marked, x_{j_2} and $x_{j'_2}$ have the same tag and $j'_2 - j_2$ is maximal. With (j_2, j'_2) we have found the second region (with left, right borders x_{j_2}, $x_{j'_2}$ and content $x_{j_2+1}x_{j_2+2}\cdots x_{j'_2-1}$) and again we mark all x_i for $j_2 \leq i \leq j'_2$ as already used. Analogously we proceed in the following steps, until no more pairs of indices can be found, which fulfill the selection condition. We obtain nonintersecting regions $(j_1, j'_1), (j_2, j'_2), \ldots, (j_n, j'_n)$.

If we denote the left (right) borders of a region by r_i (r'_i, resp.), the content of a region by b_i and the gaps between consecutive regions by b'_i, we can write x as

$$x = b'_0 \;\; r_1 b_1 r'_1 \;\; b'_1 \;\; r_2 b_2 r'_2 \;\; b'_2 \;\; \cdots \;\; r_n b_n r'_n \;\; b'_n.$$

We will treat the content b_i of a region below and show that we can give a short representation. Before we do so, we want to establish a bound on the number of sectors which are not factors of the content of some region (see Statement 5 of the following fact). In order to keep the argumentation transparent we also give some auxiliary statements.

Fact 2.

1. There do not exist two different regions (i, i') and (j, j') both having the same tags attached to their borders.

2. There do not exist $i < j$ such that x_i, x_j are not marked and have the same tags.

3. The number of x_i which are not marked and the number n of regions is bounded by $|F|^{|F|} \cdot |A|^k$.

4. The total number of sectors which lie in some gap b'_i is bounded by $3\mathcal{K}(|F|) \cdot (3|F|^{|F|} \cdot |A|^k + 1)$.

5. The total number of sectors which lie in some r_i, r'_i or b'_i is bounded by $3\mathcal{K}(|F|) \cdot (5|F|^{|F|} \cdot |A|^k + 1)$.

6. For every r_i it holds that $\delta^{r_i r_i} = \delta^{r_i}$.

First we want to show Statement 1 of Fact 2. Suppose there exist regions (i, i') and (j, j') such that $i < i' < j < j'$ and $x_i, x_{i'}, x_j, x_{j'}$ have the same tags. If there is no region between (i, i') and (j, j'), we should have chosen (i, j') instead of (i, i') and (j, j') in order to maximize the size of the region. If there exist a region $(\tilde{i}, \tilde{i}')$ between (i, i') and (j, j'), we should have chosen (i, j') instead of $(\tilde{i}, \tilde{i}')$, again in order to maximize the size of the region. Thus in both cases we obtain a contradiction. This shows Statement 1. Analogously one shows Statement 2, and from Statements 1 and 2 we obtain Statement 3.

Each b'_i has the form $x'_j x_{j+1} x'_{j+1} \cdots x_{j'} x'_{j'}$ for $j \leq j'$ (this follows from Lemma 4.5 and the construction of the regions). The number of all last factors $x'_{j'}$ can be bounded by the total number of gaps

b'_i which is by Statement 3 less then or equal to $|F|^{|F|} \cdot |A|^k + 1$. Also by Statement 3 there are at most $|F|^{|F|} \cdot |A|^k$ factors x_i in x which are not marked. Due to the given factorization of x this also bounds the number of factors x'_i before each x_{i+1}. It follows that the total number of all x_i and x'_i which are not marked (i.e., which lie in some gap) is bounded by

$$2 \cdot \left(|F|^{|F|} \cdot |A|^k\right) + \left(|F|^{|F|} \cdot |A|^k + 1\right) = 3 \cdot |F|^{|F|} \cdot |A|^k + 1.$$

By Statement 1 of Fact 1, every x_i (x'_i, resp.) contains at most $3\mathcal{K}(|F|)$ sectors, which in turn gives Statement 4.

By the definition of regions, every r_i (r'_i, resp.) consists of exactly one x_i. Thus from Statement 1 of Fact 1 and Statement 3 of Fact 2 we obtain that the number of sectors which lie in some r_i, r'_i is bounded by

$$3\mathcal{K}(|F|) \cdot 2 \cdot |F|^{|F|} \cdot |A|^k.$$

If we add this number to the quantity given in Statement 4 we obtain what is required for Statement 5. Finally, from Statement 2 of Fact 1 we obtain Statement 6. This completes the proof of Fact 2.

As an intermediate step we define an expression E'_x for words x satisfying the above fact, where we replace each b_i by $(p_k(b_i)|\alpha(\widehat{x})|s_k(b_i))_k$.

$$\begin{aligned} E'_x \quad =_{\text{def}} \quad b'_0 \;\; r_1 \cdot \;\; & \big(p_k(b_1)|\alpha(\widehat{x})|s_k(b_1)\big)_k \;\; \cdot r'_1 \;\; b'_1 \cdot \\ r_2 \cdot \;\; & \big(p_k(b_2)|\alpha(\widehat{x})|s_k(b_2)\big)_k \;\; \cdot r'_2 \;\; b'_2 \cdot \\ & \vdots \\ r_n \cdot \;\; & \big(p_k(b_n)|\alpha(\widehat{x})|s_k(b_n)\big)_k \;\; \cdot r'_n \;\; b'_n \end{aligned}$$

Although the number of remaining sectors in the b'_i, r_i and r'_i is bounded (cf. Statement 5 of Fact 2), we can not bound the size of E'_x. This is because the size of each remaining sector is yet unbounded. We will treat these sectors below, using the induction hypothesis, and then finally obtain expression E_x for x. Let us first prove that E'_x has the properties stated in the lemma, except the size requirement.

Since for $1 \leq i \leq n$ each b_i is a factor of x and consists of at least one x'_j which in turn consists of at least two sectors, we have $\alpha(\widehat{b_i}) = \alpha(\widehat{x})$ and hence $x \in E'_x$. Now take arbitrary words $x', x'' \in A^*$ and suppose $x'xx'' \in L(F)$. We will show $x'E'_xx'' \subseteq L(F)$ using the argument that F does not have pattern $\mathbf{P}_{\boldsymbol{k}}$. So let $y \in E'_x$. Then there are for $1 \leq i \leq n$ words $y_i \in (p_k(b_i)|\alpha(\widehat{x})|s_k(b_i))_k$ such that $y = b'_0r_1y_1r'_1b'_1r_2y_2r'_2b'_2 \cdots r_ny_nr'_nb'_n$. Similar to Case 1, we turn $x'xx''$ into $x'yx''$ by showing that we can replace each b_i in x by y_i for $i = n$ downto 1, and always obtain a word in $L(F)$. We demonstrate the first step of this procedure and argue, why it can be repeated.

Define $y'_n =_{\text{def}} b'_0r_1b_1r'_1b'_1r_2b_2r'_2b'_2 \cdots b'_{n-1}$ and $y''_n =_{\text{def}} b'_n$. Then $x'xx'' = x'y'_nr_nb_nr'_ny''_nx'' \in L(F)$. Assume $x'y'_nr_ny_nr'_ny''_nx'' \notin L(F)$ and define

$$\begin{aligned} \tilde{u} &=_{\text{def}} x'y'_nr_n, \\ \tilde{v} &=_{\text{def}} r'_n, \\ \tilde{w} &=_{\text{def}} y_nr'_n, \\ \tilde{z} &=_{\text{def}} y''_nx'', \\ s_1 &=_{\text{def}} \delta(s_0, \tilde{u}) \quad \text{and} \\ s_2 &=_{\text{def}} \delta(s_0, \tilde{u}\tilde{w}). \end{aligned}$$

We claim that F has pattern $\mathbf{P}_{\boldsymbol{k}}$ witnessed by the words $\tilde{u}, \tilde{v}, \tilde{w}, \tilde{z}$ and the states s_1 and s_2. To see this first note that by assumption $\delta(s_2, \tilde{z}) \notin S'$. By the choice of the regions, we know that r_n and r'_n have the same tags, hence y'_nr_n and $y'_nr_nb_nr'_n$ induce the same mappings on the states of F. It follows that $\delta^{\tilde{u}} = \delta^{\tilde{u}b_n r'_n}$. In particular, $\delta(s_0, \tilde{u}b_n r'_n) = s_1$ and since $x'xx'' = \tilde{u}b_n r'_n \tilde{z} \in L(F)$ we conclude $\delta(s_1, \tilde{z}) \in S'$. Moreover, Statement 6 of Fact 2 holds for r'_n and hence both states s_1 and s_2 have a $\tilde{v}$-loop (since $s_1 = \delta(s_0, \tilde{u}b_n r'_n)$). So we found a subgraph in the transition graph of F which is a possible candidate for $\mathbf{P}_{\boldsymbol{k}}$.

It remains to verify the alphabet condition for the respective k-decompositions, more precisely, we need to show $\alpha(\widehat{\tilde{v}\tilde{w}\tilde{v}}) \subseteq \alpha(\widehat{\tilde{v}\tilde{v}})$ which is the same as $\alpha(\widehat{r'_n y_n r'_n r'_n}) \subseteq \alpha(\widehat{r'_n r'_n})$. Since by the choice of the factors r'_i each such factor contains at least two successive sectors (cf. Fact 1, Statement 1) and by the choice of sectors we have that $\alpha(\widehat{x}) = \alpha(\widehat{r'_n}) \subseteq \alpha(\widehat{r'_n r'_n})$. We show $\alpha(\widehat{r'_n y_n r'_n}) \subseteq \alpha(\widehat{x})$. Since r'_n is a factor of x the inclusion is clear for every element in the k-decomposition of $r'_n y_n r'_n$ which lies completely inside r'_n. By the definition of y_n we immediately obtain $\alpha(\widehat{y_n}) \subseteq \alpha(\widehat{x})$ and we are left with the factors of length $k+1$ in $r'_n y_n r'_n$ which contain letters from r'_n and also from y_n. It holds that

$$\alpha(s_k(\widehat{r'_n) p_k(}y_n)) = \alpha(s_k(\widehat{r_n) p_k(}b_n)) \subseteq \alpha(\widehat{x})$$

since r'_n and r_n have the same k-suffix, $y_n \in (p_k(b_n)|\alpha(\widehat{x})|s_k(b_n))_k$ and $s_k(r_n)p_k(b_n)$ is a factor of x. We can also state

$$\alpha(s_k(\widehat{y_n) p_k(}r'_n)) = \alpha(s_k(\widehat{b_n) p_k(}r'_n)) \subseteq \alpha(\widehat{x})$$

with the same arguments. Thus we have shown $\alpha(\widehat{r'_n y_n r'_n}) \subseteq \alpha(\widehat{x}) \subseteq \alpha(\widehat{r'_n r'_n})$ and it follows that $\alpha(\widehat{\tilde{v}\tilde{w}\tilde{v}}) \subseteq \alpha(\widehat{\tilde{v}\tilde{v}})$. Hence we found pattern $\mathbf{P}_{\boldsymbol{k}}$, a contradiction. We conclude that $x'y'_n r_n y_n r'_n y''_n x'' \in L(F)$.

The whole argumentation was independent of the particular prefix $x'y'_n$ and the suffix $y''_n x''$, so we can repeat this step for y_{n-1}. It is crucial here that we proceed from right to left, so all tags left to the actual substitution position remain valid, i.e., the tags still stand for the mapping induced by the respective prefix. If we define $y'_{n-1} =_{\text{def}} b'_0 r_1 b_1 r'_1 b'_1 r_2 b_2 r'_2 b'_2 \cdots b'_{n-2}$ and $y''_{n-1} =_{\text{def}} b'_{n-1} r_n y_n r'_n b'_n$ then $x'xx'' = x'y'_{n-1} r_{n-1} b_{n-1} r'_{n-1} y''_{n-1} x'' \in L(F)$ by the previous step. Now we want to substitute b_{n-1} by y_{n-1} and observe, that we have the same starting position as in the previous step. In the same way we obtain $x'y'_{n-1} r_{n-1} y_{n-1} r'_{n-1} y''_{n-1} x'' \in L(F)$. If we repeat this procedure n-times, we finally obtain $x'yx'' \in L(F)$, which shows $x'E'_x x'' \subseteq L(F)$.

The remaining task is the construction of expression E_x which is of bounded size. We start with expression E'_x and apply the induction hypothesis for each sector s_i in E'_x (notice that we have $|\alpha(\widehat{s_i})| \leq N$ by the construction of the sectors), i.e., we replace each s_i with E_{s_i} and obtain the expression E_x.

Since $x \in E'_x$ and $s_i \in E_{s_i}$ for every sector s_i in E'_x, we obtain $x \in E'_x \subseteq E_x$. Now let $x', x'' \in A^*$ with $x'xx'' \in L(F)$. We already know that $x'E'_x x'' \subseteq L(F)$. Exactly as in Case 1 we can show that we do not leave $L(F)$ if we replace in $x'E'_x x''$ each sector s_i by E_{s_i}. This shows $x'E_x x'' \subseteq L(F)$. Finally let us consider the size of E_x. By definition, E'_x is a concatenation of sectors s_i and terms of the form $(p_k(b_i)|\alpha(\widehat{x})|s_k(b_i))_k$. Since the latter terms do not influence the size of expressions, we obtain

$$\|E_x\| = \sum_{\substack{\text{sectors } s_i \\ \text{of } E'_x}} \|E_{s_i}\|.$$

By induction hypothesis we have $\|E_{s_i}\| \leq \mathcal{L}(k, |A|, |F|, N)$ for all sectors s_i of E'_x. By Statement 5 of Fact 2 we know that the number of sectors of E'_x is less than or equal to $3\mathcal{K}(|F|) \cdot (5|F|^{|F|} \cdot |A|^k + 1)$. It follows that

$$\|E_x\| \leq 3\mathcal{K}(|F|) \cdot (5|F|^{|F|} \cdot |A|^k + 1) \cdot \mathcal{L}(k, |A|, |F|, N) = \mathcal{L}(k, |A|, |F|, N+1).$$

This completes the induction. ❑

Theorem 5.7. *Let $k \geq 0$ and let F be a dfa. If F does not have Pattern $\mathbf{P}_{\boldsymbol{k}}$, then $L(F) \in \mathcal{B}_{3/2,k}$.*

Proof. First of all we will show that languages of expressions of $\mathcal{E}_k$ are in $\widetilde{\mathcal{B}}_k$. Languages of the form $\{a\}$ with $a \in A$ are in $\widetilde{\mathcal{B}}_k$, because they can be written either as $\{a\} \subseteq A^{\leq k}$ (if $k > 0$) or as $(\emptyset, a, \emptyset)_k$ (if $k = 0$). Thus by Proposition 3.3 we have $\{u\} \in \widetilde{\mathcal{B}}_k$ for every $u \in A^+$. Let us consider languages of the form $(w|\Gamma|v)_k$ with $w, v \in A^k$, $\Gamma \subseteq A^{k+1}$. For $k = 0$ they can be written as $(\Gamma)_k \in \widetilde{\mathcal{B}}_k$. For $k > 0$ we have $(w|\Gamma|v)_k = ((w(w^{-1}(\Gamma)_k))v^{-1})v$. Since $|w| = |v| = k$ and $(\Gamma)_k \subseteq A^{\geq k+1}$ we can apply Proposition 3.1. Together with Proposition 3.3 we obtain $(w|\Gamma|v)_k \in \widetilde{\mathcal{B}}_k$. Thus we have shown that languages of the forms $\{u\}$ or $(w|\Gamma|v)_k$ are in $\widetilde{\mathcal{B}}_k$. Now from Proposition 3.3 we conclude that that languages of expressions of $\mathcal{E}_k$ are in $\widetilde{\mathcal{B}}_k$.

For every $x \in L(F)$ we apply Lemma 5.6 and we obtain corresponding expressions E_x with $x \in E_x$. Since $x = \varepsilon x \varepsilon \in L(F)$, we have $E_x = \varepsilon E_x \varepsilon \subseteq L(F)$. Thus it holds that

$$L(F) = \bigcup_{x \in L(F)} E_x$$

By Lemma 5.6, the size of E_x is bounded by $\mathcal{L}(k, |A|, |F|, |\alpha(\widehat{x})|)$, so particularly $\mathcal{L}(k, |A|, |F|, |A|^{k+1})$ bounds the size of each E_x. As was mentioned before, there is only a finite number of expressions in $\mathcal{E}_k$ having the same size. Therefore, the above union is finite, i.e., $L(F)$ is a finite union of languages of $\widetilde{\mathcal{B}}_k$. Hence $L(F) \in \widetilde{\mathcal{B}}_k = \mathcal{B}_{3/2,k}$ by Theorem 3.16. ❑

5.4 The Forbidden Pattern Theorem for $\mathcal{B}_{3/2,k}$

Taking together Theorems 5.3 and 5.7 we immediately obtain the main result of this section.

Theorem 5.8. *Let $k \geq 0$ and let F be a dfa. Then $L(F) \in \mathcal{B}_{3/2,k}$ if and only if F does not have Pattern $\mathbf{P}_k$.*

We exploit our forbidden pattern characterization from the previous theorem to prove an effective characterization of $\mathcal{B}_{3/2,k}$ for fixed k.

5.4.1 An Algorithm to decide $\mathcal{B}_{3/2,k}$

We construct an algorithm solving the decision problems for the classes $\mathcal{B}_{3/2,k}$. For a fixed $k \geq 0$, this algorithm works in nondeterministic, logarithmic space (NL).

Lemma 5.9. *On input (F, k) where $k \geq 0$, F is a dfa and n is the size of the input (F, k), the question '$L(F) \in \mathcal{B}_{3/2,k}$' is decidable in nondeterministic space $2^{O(k)} + O(\log n)$.*

Proof. By Theorem 5.8 it suffices to construct an algorithm deciding whether Pattern $\mathbf{P}_k$ appears in the transition graph of a given dfa F (cf. Figure 1). Without loss of generality we may assume that the word w in Pattern $\mathbf{P}_k$ is an element of $A^{\geq k+1}$, otherwise we take vwv instead of w. The following nondeterministic algorithm solves this task on input (F, k) where $k \geq 0$ and $F = (A, S, \delta, s_0, S')$ is a dfa.

1. Guess states s_1, s_2, s^+, s^- of Pattern $\mathbf{P}_k$ and store them. Check $s^+ \in S'$ and $s^- \notin S'$, and reject if this fails.
2. Begin in the starting state s_0 and follow with help of δ, letter-by-letter and nondeterministically all possible words $u \in A^*$. Continue until state s_1 is reached.
3. Start simultaneously in the States s_1, s_2 and follow as above nondeterministically all possible words $z \in A^*$. Continue until states s^+ is reached from s_1 and s^- is reached form s_2 by the same word z.

4. Start simultaneously in the states s_1, s_2 and follow as above nondeterministically all possible words $v \in A^{\geq k+1}$. Store $p_k(v)$. While guessing v store also the actual value of $s_k(v)$ and $M' =_{\text{def}} \alpha(\widehat{v})$. Continue until state s_1 (s_2) is reached from s_1 (s_2, resp.). Determine $M =_{\text{def}} \alpha(\widehat{vv})$ with help of $p_k(v)$ and $s_k(v)$, and store the sets M', M.

5. Begin in state s_1 and follow as before nondeterministically all possible words $w \in A^{\geq k+1}$. Similar to the previous step, we store $p_k(w)$, the actual value of $s_k(w)$ and also of $N' =_{\text{def}} \alpha(\widehat{w})$. Continue until state s_2 is reached. Determine the set $N =_{\text{def}} \alpha(\widehat{vwv})$ from N' with help of $s_k(v), p_k(v), s_k(w), p_k(w)$ and the set M'. We store the set N.

6. Accept if $N \subseteq M$, otherwise reject.

It is easy to see that this algorithm accepts if and only if F has Pattern $\mathbf{P}_{\boldsymbol{k}}$, which in turn is equivalent to $L(F) \notin \mathcal{B}_{3/2,k}$ by Theorem 5.8.

Now let n be the size of the input (F, k). We want to see that space at most $2^{O(k)} + O(\log n)$ is needed. For Step 1 the algorithm uses only logarithmic space $O(\log n)$ to store the numbers of the states s_1, s_2, s^+, s^-. In Steps 2 and 3 additional space $O(\log n)$ is needed to store the current states we have reached. So let us consider Steps 4 and 5. To store the current states we need space $O(\log n)$ as before. To store the prefix and suffix words we need space $O(k)$. For the sets M', M, N', M the algorithm uses space at most $O(k) \cdot |A|^{k+1} \leq 2^{O(k)}$. This is because we have fixed the alphabet A and M', M, N', N are subsets of A^{k+1} with $|A^{k+1}| = |A|^{k+1}$, and for every $w \in A^{k+1}$ we need space $O(k)$. Finally, in Step 6 only space $O(k)$ is needed. All in all we keep a space bound of $2^{O(k)} + O(\log n)$.

Since we test for the existence of Pattern $\mathbf{P}_{\boldsymbol{k}}$, we have shown that the decision problem for $\mathrm{co}\mathcal{B}_{3/2,k}$ is in $\mathrm{NSPACE}(2^{O(k)} + O(\log n))$. This class is closed under complementation (cf. [Sze87, Imm88]), so there also exists an $\mathrm{NSPACE}(2^{O(k)} + O(\log n))$-algorithm for the decidability problem for $\mathcal{B}_{3/2,k}$. ❑

Corollary 5.10. *Fix some $k \geq 0$. Then the decision problem for $\mathcal{B}_{3/2,k}$ is in* NL.

5.4.2 Strictness of the $\mathcal{B}_{3/2,k}$-Hierarchy

We show with the construction of witnessing languages, that the family of classes $\{ \mathcal{B}_{3/2,k} \mid k \geq 0 \}$ is infinite. In fact we obtain for every $k \geq 0$ a different class. This is not surprising, since for $k = 0$ we obtain level 3/2 of the Straubing-Thérien-hierarchy (up to the empty word), which is known to be different from $\mathcal{B}_{3/2}$ (not only by the empty word).

Theorem 5.11. *Let $k \geq 0$. Then $\mathcal{B}_{3/2,k} \subsetneq \mathcal{B}_{3/2,k+1}$.*

Proof. By the definition of $\mathcal{B}_{3/2,k}$ (cf. Definition 2.25 and Proposition 2.12), we have $\mathcal{B}_{3/2,k} \subseteq \mathcal{B}_{3/2,k+1}$, because the operator POL is monotonic. Together with Theorem 3.16 this implies that it suffices to show $\widetilde{\mathcal{B}}_k \neq \widetilde{\mathcal{B}}_{k+1}$. Let $a, b \in A$ with $a \neq b$, let $w_i =_{\text{def}} \left(a^{k+1}b\right)^i$ for $i \geq 1$ and let

$$L =_{\text{def}} \left(a^{k+1}b, \left\{a^i ba^{k+1-i} : 0 \leq i \leq k+1\right\}, a^{k+1}b\right)_{k+1}.$$

It is easy to see that $L \in \widetilde{\mathcal{B}}_{k+1}$ and $L = \{w_i : i \geq 2\}$. Suppose now that $L \in \widetilde{\mathcal{B}}_k$. Since $L \subseteq A^{\geq k+2}$, we have

$$L = \bigcup_{i=1}^{m} (\gamma_{i,0}, \Gamma_{i,1}, \gamma_{i,1}, \ldots, \Gamma_{i,n_i}, \gamma_{i,n_i})_k$$

for suitable $m, n_i \geq 1$, $\gamma_{i,j} \in A^{k+1}$ and $\Gamma_{i,j} \subseteq A^{k+1}$. Let n be the maximum over all n_i, then there exists an l with $1 \leq l \leq m$ such that $w_{2n+2} \in (\gamma_{l,0}, \Gamma_{l,1}, \gamma_{l,1}, \ldots, \Gamma_{l,n_l}, \gamma_{l,n_l})_k$. The k-decomposition of w_{2n+2} consists of $(2n+2)(k+2) - k \geq (n+1) + n(k+2)$ elements of A^{k+1}. By the pigeon hole principle there exists an $1 \leq h \leq n_l$, such that an interval of at least $k+2$ consecutive elements of this k-decomposition is assigned to $\Gamma_{l,h}$. By definition of w_{2n+2} the element a^{k+1} must appear in this sequence of words of length $k+1$. Therefore, we can pump up the word w_{2n+2} with letters a without leaving the languages L. But this is a contradiction, because in every word of L the number of consecutive letters a is bounded by $k+1$. So $L \in \widetilde{\mathcal{B}}_{k+1} \backslash \widetilde{\mathcal{B}}_k$. ❑

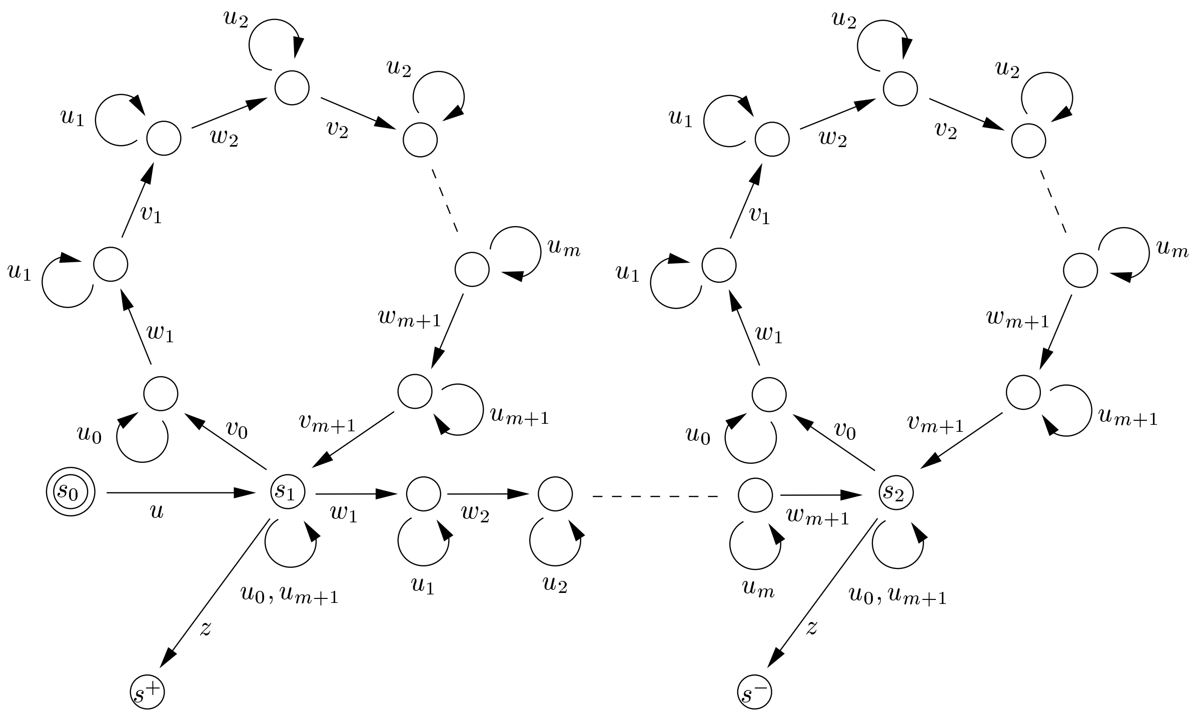

Figure 2: Pattern **P** with $m \geq 0$, $u_i, w_i \in A^+$, $v_i \in A^*$, $s^+ \in S'$ and $s^- \notin S'$.

6 Forbidden Pattern Characterization of $\mathcal{B}_{3/2}$

We identify a Pattern **P** which neither has a parameter k nor an alphabet condition attached to it, but characterizes $\mathcal{B}_{3/2}$. This yields an efficient algorithm for the decision problem for $\mathcal{B}_{3/2}$. Moreover, the proof is such that a bound on k in the size of the automaton can be established. So we can also find an algorithm which determines the minimal k such that $L(F) \in \mathcal{B}_{3/2,k}$ for a given dfa F.

6.1 The Pattern P

We make the following definition concerning the transition graph of a dfa.

Definition 6.1. *A dfa $F = (A, S, \delta, s_0, S')$ has Pattern* **P** *if there exist $m \geq 0$, states $s_1, s_2, s^+, s^- \in S$ and words $u_0, \ldots, u_{m+1}, w_1, \ldots, w_{m+1} \in A^+$, $v_0, \ldots, v_{m+1} \in A^*$ such that there is a subgraph in the transition graph of F as depicted in Figure 2. More precisely, for $0 \leq i \leq m+1$ and $1 \leq j \leq m+1$ it must be that the following holds:*

$$
\begin{aligned}
\delta(s_1, z) &= s^+ \in S' \\
\delta(s_2, z) &= s^- \notin S' \\
\delta(s_0, u) = s_1 &= \delta(s_1, v_0 w_1 v_1 \cdots w_{m+1} v_{m+1}) = \delta(s_1, u_0) \\
\delta(s_1, w_1 \cdots w_{m+1}) = s_2 &= \delta(s_2, v_0 w_1 v_1 \cdots w_{m+1} v_{m+1}) = \delta(s_2, u_0) \\
\delta(s_1, v_0 w_1 v_1 \cdots w_i v_i) &= \delta(s_1, v_0 w_1 v_1 \cdots w_i v_i u_i) \\
\delta(s_1, v_0 w_1 v_1 \cdots v_{j-1} w_j) &= \delta(s_1, v_0 w_1 v_1 \cdots v_{j-1} w_j u_j) \\
\delta(s_2, v_0 w_1 v_1 \cdots w_i v_i) &= \delta(s_2, v_0 w_1 v_1 \cdots w_i v_i u_i) \\
\delta(s_2, v_0 w_1 v_1 \cdots v_{j-1} w_j) &= \delta(s_2, v_0 w_1 v_1 \cdots v_{j-1} w_j u_j) \\
\delta(s_1, w_1 w_2 \cdots w_j) &= \delta(s_1, w_1 w_2 \cdots w_j u_j)
\end{aligned}
$$

6.2 The Forbidden Pattern Theorem for $\mathcal{B}_{3/2}$

Now we want to show that $\mathcal{B}_{3/2}$ is actually characterized by Pattern **P**. For this we need an auxiliary result which is stated in the following lemma. It says that if there is Pattern $\mathbf{P}_k$ for sufficiently large k in an automaton, then we can assume w.l.o.g. that the relevant witnessing words for this pattern lead to the same loop.

Lemma 6.2. *If a dfa F has Pattern $\mathbf{P}_k$ with $k > \mathcal{K}(|F|)$, then we can assume that the words v and w in the pattern are of the form $v = v'u_0$ and $w = w'u_0$ such that $1 \le |u_0| \le \mathcal{K}(|F|)$ and $\delta^{u_0u_0} = \delta^{u_0}$.*

Proof. Let F be a dfa, $k > \mathcal{K}(|F|)$ and suppose that F has Pattern $\mathbf{P}_k$ via words u, v, w, z. With the definition of words $\tilde{u}, \tilde{v}, \tilde{w}, \tilde{z}$ below we will find another instance of a pattern of type $\mathbf{P}_k$ in F which then has the properties stated in the lemma. By Corollary 4.6, we can write v as $v = v'_0u'_1v'_1 \cdots u'_mv'_m$ such that $u'_i, v'_i \in A^{\le \mathcal{K}(|F|)}$ and $\delta^{u'_iu'_i} = \delta^{u'_i}$. Since $|v| \ge k+1 > \mathcal{K}(|F|)$ we have $m \ge 1$. Therefore, with $v_0 =_{\text{def}} v'_0u'_1v'_1 \cdots u'_{m-1}v'_{m-1}$, $u_0 =_{\text{def}} u'_m$ and $v_1 =_{\text{def}} v'_m$ we can rewrite v as $v = v_0u_0v_1$ with $u_0, v_1 \in A^{\le \mathcal{K}(|F|)}$ and $\delta^{u_0u_0} = \delta^{u_0}$. Now define

$$
\begin{aligned}
\tilde{u} &=_{\text{def}} uv_0u_0,\\
\tilde{v} &=_{\text{def}} v_1v_0u_0,\\
\tilde{w} &=_{\text{def}} v_1wv_0u_0 \text{ and}\\
\tilde{z} &=_{\text{def}} v_1z
\end{aligned}
$$

Then we have $|\tilde{v}| = |v| \ge k+1$. Thus with $v' =_{\text{def}} v_1v_0$, $w' =_{\text{def}} v_1wv_0$ we can write $\tilde{v}, \tilde{w}$ as $\tilde{v} = v'u_0$ and $\tilde{w} = w'u_0$ such that $1 \le |u_0| \le \mathcal{K}(|F|)$ and $\delta^{u_0u_0} = \delta^{u_0}$. Furthermore, it is easy to see that $\tilde{u}, \tilde{v}, \tilde{w}, \tilde{z}$ define a subgraph of the form required for $\mathbf{P}_k$ in the automaton F. It remains to show the alphabet condition, i.e., $\alpha(\widehat{\tilde{v}\tilde{w}\tilde{v}}) \subseteq \alpha(\widehat{\tilde{v}\tilde{v}})$. Since $\tilde{v}\tilde{w}\tilde{v}$ is a factor of $vvwvv$ and $|v| = |\tilde{v}| \ge k+1$, we obtain $\alpha(\widehat{\tilde{v}\tilde{w}\tilde{v}}) \subseteq \alpha(\widehat{vvwvv}) \subseteq \alpha(\widehat{vv})$ and $\alpha(\widehat{vv}) = \alpha(\widehat{\tilde{v}\tilde{v}})$. ❑

Now we are able to prove some kind of equivalence of the Patterns **P** and $\mathbf{P}_k$, i.e., we show that **P** exists if and only if $\mathbf{P}_k$ exists for all $k \ge 0$. From this fact one can easily derive the characterization of $\mathcal{B}_{3/2}$ by means of the Pattern **P**.

Lemma 6.3. *For every dfa F it holds that F has Pattern **P** if and only if F has Pattern $\mathbf{P}_k$ for all $k \ge 0$.*

Proof. "$\Longrightarrow$" Suppose that F has Pattern **P** and let $k \ge 0$. Define

$$
\begin{aligned}
v &=_{\text{def}} u_0^kv_0u_0^kw_1u_1^kv_1u_1^k \cdots w_{m+1}u_{m+1}^kv_{m+1}u_{m+1}^k \text{ and}\\
w &=_{\text{def}} u_0^kw_1u_1^k \cdots w_{m+1}u_{m+1}^k.
\end{aligned}
$$

Thus with u, v, w, z we have found a subgraph in F which is of the form required for $\mathbf{P}_k$ (note that $|v| \ge k+1$). It remains to show $\alpha(\widehat{vwv}) \subseteq \alpha(\widehat{vv})$ w.r.t. k-decomposition.

Let $\gamma \in \alpha(\widehat{vwv})$, then we consider the following three cases.

Case 1: γ appears in v, then we have $\gamma \in \alpha(\widehat{vv})$ and we are done.

Case 2: γ appears in w. Since $|u_i| \ge 1$, we have $|u_i^k| \ge k$ for $0 \le i \le m+1$. Since $|\gamma| = k+1$ it follows that each occurence of γ in w overlaps at most one of the w_i. Particularly, γ is a factor of $u_i^kw_{i+1}u_{i+1}^k$ for some $0 \le i \le m$. By definition, each $u_i^kw_{i+1}u_{i+1}^k$ is a factor of v, so $\gamma \in \alpha(\widehat{vv})$ follows.

Case 3: γ appears neither in v nor in w, but it appears in vw or in wv (this case can not occur if $k = 0$). Thus γ is a factor of $s_k(v)p_k(w)$ or $s_k(w)p_k(v)$. Hence in both cases γ is a factor of $u_{m+1}^ku_0^k$. It follows that γ is also a factor of vv, so $\gamma \in \alpha(\widehat{vv})$.

"$\Longleftarrow$" Suppose now that F has Pattern $\mathbf{P}_k$ for all $k \ge 0$. We want to choose k sufficiently large such that the following holds: If we fix a decomposition of a word into blocks w'_i and u_i with $\delta^{u_iu_i} = \delta^{u_i}$

due to Corollary 4.6, then every factor $u_i w_i' u_{i+1}$ should appear in an element of the k-decomposition of this word, i.e., $|u_i w_i' u_{i+1}| \leq k+1$. Therefore, we consider a pattern $\mathbf{P}_{\boldsymbol{k}}$ with $k =_{\text{def}} 3\mathcal{K}(|F|)$, and we suppose that this pattern is defined via words u, v, w, z. By Lemma 6.2, we can assume that v and w are of the form $v = v'u_0$ and $w = w'u_0$ such that $1 \leq |u_0| \leq \mathcal{K}(|F|)$ and $\delta^{u_0u_0} = \delta^{u_0}$. By Corollary 4.6, for w' we find a decomposition $w' = w_0'u_1w_1' \cdots u_m w_m'$ such that $w_i', u_i \in A^{\leq \mathcal{K}(|F|)}$ and $\delta^{u_iu_i} = \delta^{u_i}$. Let $u_{m+1} =_{\text{def}} u_0$, then we have $u_0, w_0', u_1, \ldots, w_m', u_{m+1} \in A^{\leq \mathcal{K}(|F|)}$ and $\delta^{u_iu_i} = \delta^{u_i}$ for $0 \leq i \leq m+1$. It follows that $|u_i w_i' u_{i+1}| \leq 3\mathcal{K}(|F|) < k+1$ for $0 \leq i \leq m$. Since $u_0 w_0' u_1 \cdots w_m' u_{m+1}$ is a factor of vw, each $u_i w_i' u_{i+1}$ appears in some element of the k-decomposition of vwv. From the condition $\alpha(\widehat{vwv}) \subseteq \alpha(\widehat{vv})$ it follows that each $u_i w_i' u_{i+1}$ is a factor of vv, i.e., for $0 \leq i \leq m$ there exist words v_i', v_i'' such that $vv = v_i' u_i w_i' u_{i+1} v_i''$. Now define $v_0 =_{\text{def}} v_0'u_0$, $v_{m+1} =_{\text{def}} v_m''v$, $v_i =_{\text{def}} v_{i-1}''v_i'u_i$ for $1 \leq i \leq m$ and $w_j =_{\text{def}} w_{j-1}'u_j$ for $1 \leq j \leq m+1$.

Now we want to show that we have found Pattern **P**. By the definition of v_i and w_i we have

$$\begin{aligned}
v_0w_1v_1w_2v_2\cdots v_m w_{m+1}v_{m+1} &= \overbrace{v_0'u_0}^{v_0}\,\overbrace{w_0'u_1}^{w_1}\,\overbrace{v_0''\,v_1'u_1}^{v_1}\,\overbrace{w_1'u_2}^{w_2}\,\overbrace{v_1''\,v_2'u_2}^{v_2}\cdots\overbrace{v_{m-1}''\,v_m'u_m}^{v_m}\,\overbrace{w_m'u_{m+1}}^{w_{m+1}}\,\overbrace{v_m''\,v}^{v_{m+1}} \\
&= \underbrace{v_0'u_0\,w_0'u_1\,v_0''}_{vv}\,\underbrace{v_1'u_1\,w_1'u_2\,v_1''}_{vv}\,v_2'u_2\cdots v_{m-1}''\,\underbrace{v_m'u_m\,w_m'u_{m+1}\,v_m''}_{vv}\,v \\
&= v^{2m+3}.
\end{aligned}$$

It follows that s_1 and s_2 have a loop $v_0w_1v_1\cdots w_{m+1}v_{m+1}$. Since $w = w'u_{m+1} = w_0'u_1\cdots w_m'u_{m+1} = w_1\cdots w_{m+1}$, we have $\delta(s_1, w_1\ldots w_{m+1}) = s_2$. We already know that s_1 and s_2 have $u_0 = u_{m+1}$-loops. From the definition of the w_i and v_j it follows that every w_i leads to an u_i-loop and every v_j leads to an u_j-loop for $1 \leq i \leq m+1$ and $0 \leq j \leq m+1$. Moreover, all u_i and w_j are non-empty words. Finally $\delta(s_0, u) = s_1$, $\delta(s_1, z) \in S'$ and $\delta(s_2, z) \notin S'$. Thus we have found Pattern **P** in automaton F. ❑

We give the forbidden pattern theorem for $\mathcal{B}_{3/2}$.

Theorem 6.4. *Let F be a dfa accepting some language L. Then $L \in \mathcal{B}_{3/2}$ if and only if F does not have Pattern* **P**. *Furthermore, this equivalence remains true if we additionally require $u_0 = u_{m+1}$ in the definition of Pattern* **P**.

Proof. By Lemma 6.3, F does not have Pattern **P** if and only if there exists a $k \geq 0$ such that F does not have Pattern $\mathbf{P}_{\boldsymbol{k}}$. By Theorem 5.8 and Proposition 2.26, this is equivalent to $L(F) \in \mathcal{B}_{3/2}$.

From the proof of Lemma 6.3 we see that the existence of Pattern **P** is equivalent to the existence of a modified Pattern **P′** where we additionally require $u_0 = u_{m+1}$. ❑

6.3 An NL–Algorithm solving the Decision Problem for $\mathcal{B}_{3/2}$

As before in the case of the classes $\mathcal{B}_{3/2,k}$, we exploit now our forbidden pattern characterization of $\mathcal{B}_{3/2}$ to construct an algorithm for the decision problem of this class.

Lemma 6.5. *There exists an* NL*-algorithm which decides whether a given automaton has Pattern* **P**.

Proof. First we define two reachability relation on tupels of states to simplify notations. Let $s_1, \ldots, s_n, s_1', \ldots, s_n' \in S$ for $n \geq 1$ then define:

$$\begin{aligned}
(s_1, \ldots, s_n) \longrightarrow (s_1', \ldots, s_n') \quad &\overset{df}{\Longleftrightarrow} \quad \text{there exists } w \in A^* \text{ such that } \delta(s_i, w) = s_i' \text{ for } 1 \leq i \leq n \\
(s_1, \ldots, s_n) \longrightarrow^+ (s_1', \ldots, s_n') \quad &\overset{df}{\Longleftrightarrow} \quad \text{there exists } w \in A^+ \text{ such that } \delta(s_i, w) = s_i' \text{ for } 1 \leq i \leq n
\end{aligned}$$

We note that for a fixed $n \geq 1$ these relations can be tested in NL. To see this we can proceed as follows: Our program uses the variable $\tilde{s}_i$ to store the number of the current state of the possible path from s_i to

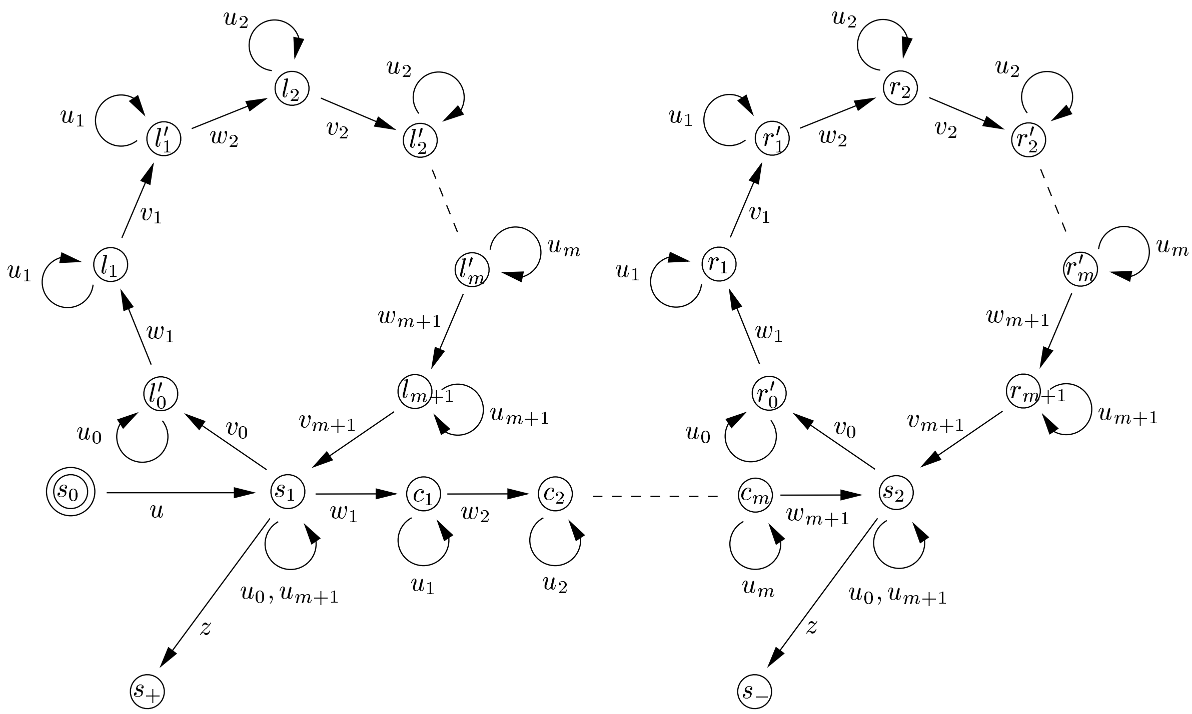

Figure 3: The labels of Pattern **P** used in the proof of Lemma 6.5 where $s_1 = c_0 = l_0 = l'_{m+1}$ and $s_2 = c_{m+1} = r_0 = r'_{m+1}$.

s'_i for $1 \leq i \leq n$. For a fixed $n \geq 1$ these variables can be stored in logarithmic space. We initialize the variables $\tilde{s_1}, \ldots, \tilde{s_n}$ with the values $s_1, \ldots, s_n$, respectively. Now we perform a loop where in each instance we guess the next letter of w as $a \in A \cup \{\varepsilon\}$. If the guess was $a = \varepsilon$, then we break the loop (i.e., we guessed the end of w) and test $\tilde{s_i} = s'_i$ for $1 \leq i \leq n$. Otherwise we apply a to the states $\tilde{s_1}, \ldots, \tilde{s_n}$, i.e., we set $\tilde{s_i} := \delta(\tilde{s_i}, a)$ for $1 \leq i \leq n$. Note that this can be done also in logarithmic space. Thus the first relation can be tested in NL for a fixed $n \geq 1$, analogously one shows this for the second one.

In the Table 1 we describe an algorithm which works in nondeterministic logarithmic space and which decides whether a given automaton F has Pattern **P**. Note that the starting state s_0 can be determined from the input in logarithmic space. We use the notations for Pattern **P** given by Figure 3. Moreover, to distinguish between the states of Pattern **P** and program variables, we mark the latter one with a bar.

It is easy to see that this algorithm decides the existence of Pattern **P** in a given automaton. Furthermore, all variables were only used to store ordinal numbers of states, thus logarithmic space suffices to store them. Since we have used the relations $(\cdot) \longrightarrow (\cdot)$ and $(\cdot) \longrightarrow^+ (\cdot)$ at most for 5-tupels, we can check them in logarithmic space (as mentioned at the beginning of this proof). All in all this shows that the given algorithm is in NL. ❑

Theorem 6.6. *The decidability problem for $\mathcal{B}_{3/2}$ is in* NL.

Proof. This follows from Theorem 6.4, Lemma 6.5 and the fact that $\mathrm{NL} = \mathrm{coNL}$ proved in [Sze87, Imm88]. ❑

As can be seen from the proof of Lemma 6.3, we can actually bound the level k of a language in $\mathcal{B}_{3/2}$ by $3\mathcal{K}(|F|)$. So we can also determine the minimal $k \geq 0$ such that $L(F) \in \mathcal{B}_{3/2,k}$ for a given dfa F, if we already know (by the previous theorem) that it is in $\mathcal{B}_{3/2}$ at all.

Label	**Command**	**Remark**
	Guess states $\overline{s_1}, \overline{s_2}, \overline{s_+}, \overline{s_-}$ and test: • $\overline{s_+} \in S'$ and $\overline{s_-} \notin S'$ • $(\overline{s_0}) \longrightarrow (\overline{s_1})$ • $(\overline{s_1}, \overline{s_2}) \longrightarrow (\overline{s_+}, \overline{s_-})$	Here we guess the states s_1, s_2, s_+, s_- of Pattern **P** in Figure 3 and check that s_+ is accepting, s_- is rejecting, s_1 is reachable from the stating state, and s_+, s_- are reachable from s_1, s_2 (resp.) with the same word, i.e., with the word z in Figure 3.
	Guess states $\overline{l_1}, \overline{l_4}, \overline{r_1}, \overline{r_4}$ and test: • $(\overline{s_1}, \overline{s_2}) \longrightarrow (\overline{l_1}, \overline{r_1})$ • $(\overline{l_4}, \overline{r_4}) \longrightarrow (\overline{s_1}, \overline{s_2})$ • $(\overline{s_1}, \overline{s_2}, \overline{l_1}, \overline{r_1}) \longrightarrow^+ (\overline{s_1}, \overline{s_2}, \overline{l_1}, \overline{r_1})$ • $(\overline{s_1}, \overline{s_2}, \overline{l_4}, \overline{r_4}) \longrightarrow^+ (\overline{s_1}, \overline{s_2}, \overline{l_4}, \overline{r_4})$	Here the program variables $\overline{l_1}, \overline{r_1}, \overline{l_4}, \overline{r_4}$ correspond to the states $l'_0, r'_0, l_{m+1}, r_{m+1}$ in Figure 3, respectively. Hence we check whether there are words v_0, v_{m+1} such that l'_0, r'_0 can be reached from s_1, s_2 (resp.) via v_0, and s_1, s_2 can be reached from l_{m+1}, r_{m+1} (resp.) via v_{m+1}. Furthermore, we test the existence of an u_0-loop (u_{m+1}-loop) at the states s_1, s_2, l'_0, r'_0 ($s_1, s_2, l_{m+1}, r_{m+1}$, resp.).
	Let $\overline{s'} := \overline{s_1}$	In the following loop we will simultaneously walk through the large loop at s_1 (using the program variables $\overline{l_1}, \ldots, \overline{l_4}$), through the large loop at s_2 (using $\overline{r_1}, \ldots, \overline{r_4}$) and through the path going from s_1 to s_2 (using $\overline{s'}, \overline{s''}$).
	Jump nondeterministically to `loop` or to `exit`.	Here we guess whether we have reached the states l'_m, c_m, r'_m with the program variables $\overline{l_1}, \overline{s'}, \overline{r_1}$, respectively (if we have reached them, we jump to `exit`).
`loop:`	Guess states $\overline{s''}, \overline{l_2}, \overline{l_3}, \overline{r_2}, \overline{r_3}$ and test: • $(\overline{l_1}, \overline{s'}, \overline{r_1}) \longrightarrow^+ (\overline{l_2}, \overline{s''}, \overline{r_2})$ • $(\overline{l_2}, \overline{r_2}) \longrightarrow (\overline{l_3}, \overline{r_3})$ • $(\overline{l_2}, \overline{l_3}, \overline{s''}, \overline{r_2}, \overline{r_3}) \longrightarrow^+ (\overline{l_2}, \overline{l_3}, \overline{s''}, \overline{r_2}, \overline{r_3})$	At the beginning of the i-th instance of the loop, the variables $\overline{l_1}, \overline{r_1}, \overline{l_4}, \overline{r_4}, \overline{s'}$ corresponds to the states $l'_{i-1}, r'_{i-1}, l_{m+1}, r_{m+1}, c_{i-1}$ in Figure 3, respectively. With $\overline{s''}, \overline{l_2}, \overline{r_2}, \overline{l_3}, \overline{r_3}$ we guess the states $c_i, l_{i-1}, r_{i-1}, l'_i, r'_i$, respectively. Now the first test ensures the existence of w_i, the second one the existence of v_i, and the third one checks the states for an u_i-loop.
	Let $\overline{l_1} := \overline{l_3}$, $\overline{r_1} := \overline{r_3}$, $\overline{s'} := \overline{s''}$ and jump nondeterministically to `loop` or to `exit`.	Now the variables $\overline{l_1}, \overline{r_1}, \overline{s'}$ corresponds to the states l'_i, r'_i, c_i, and we guess whether we have already reached the correct value of m, i.e., whether $m = i$ (if so, jump to `exit`).
`exit:`	If $(\overline{s'}, \overline{l_1}, \overline{r_1}) \longrightarrow^+ (\overline{s_2}, \overline{l_4}, \overline{r_4})$ then accept, otherwise reject.	Finally we check the existence of w_{m+1} and make the corresponding output.

Table 1: An NL-Algorithm which decides the existence of Pattern **P** in a given automaton

Proposition 6.7. *There is a recursive function g which assigns to a given dfa F with $L(F) \in \mathcal{B}_{3/2}$ the minimal $k \geq 0$ such that $L(F) \in \mathcal{B}_{3/2,k}$.*

Proof. By assumption we know that F does not have Pattern $\mathbf{P}_{\boldsymbol{k}}$ for $k = 3\mathcal{K}(|F|)$, since otherwise it also has Pattern $\mathbf{P}$ as carried out in the proof of Lemma 6.3. So $L(F)$ is in some class $\mathcal{B}_{3/2,k}$ for $0 \leq k \leq 3\mathcal{K}(|F|)$. To determine the minimal k we may apply the $\mathrm{NSPACE}(2^{O(k)} + O(\log n))$–algorithm given by Lemma 5.9 for $k = 0, 1, 2, \ldots$ until we find the smallest k such that F does not have Pattern $\mathbf{P}_{\boldsymbol{k}}$ (such a k exists and is at most $3\mathcal{K}(|F|)$). Note that this can be done by a binary search, since if an automaton has Pattern $\mathbf{P}_{\boldsymbol{k}}$ for some $k \geq 1$ then it also has Pattern $\mathbf{P}_{\boldsymbol{k-1}}$. ❑

6.4 Further Consequences

Due to the various characterizations of the classes of the dot-depth hierarchy our result has immediate consequences in other fields of research. Note that the usual representations of regular languages, e.g., finite automata, regular expressions, logical formulas and syntactical semigroups, can be effectively transformed from one to another.

6.4.1 Logical Interpretation

The dot-depth hierarchy is related to the first-order logic $\mathrm{FO}[<, \min, \max, S, P]$ having unary relations for the alphabet symbols from A, the binary relation $<$, the successor (predecessor) function S (P, resp.), and constants $\min$ and $\max$. Let Σ_n be the subclass of this logic which is defined by at most $n-1$ quantifier alternations, starting with an existential quantifier. It has been proved by W. Thomas in [Tho82] (see also [PP86, PW97]) that Σ_n formulas describe just the $\mathcal{B}_{n-1/2}$ languages and that the Boolean combinations of Σ_n formulas describe just the $\mathcal{B}_n$ languages. Due to this characterization we have the following corollary.

Corollary 6.8. *Given a regular language L, it is decidable whether L is definable by a Σ_2 –formula of the logic* $\mathrm{FO}[<, \min, \max, S, P]$.

6.4.2 Algebraic Interpretation

For an introduction to the algebraic theory of finite automata see [Pin96b]. We translate our result to the notations used in this theory and we recall some basic algebraic notions. By id_S we denote the *identity mapping* of a set S. An *order relation* on a set S is a binary relation $\leq$ which is reflexive ($x \leq x$), transitive ($x \leq y \wedge y \leq z \Longrightarrow x \leq z$) and antisymmetric ($x \leq y \wedge y \leq x \Longrightarrow x = y$). A *semigroup* is a set S with a law of composition which is associative. For every element x of a semigroup S we define the ω-power of x as $x^\omega =_{\mathrm{def}} x^i$ where $i = \inf\left\{j \geq 1 : x^j = x^{2j}\right\}$. Note that for a finite semigroup S x^ω is always an element of S. An *ordered semigroup* $(S, \leq)$ is a semigroup S equipped with some order relation $\leq$ on S which respects the law of composition of S, i.e., for every $u, v, x \in S$ with $u \leq v$ we have $ux \leq vx$ and $xu \leq xv$.

Let L be a regular language of A^+ and let $F_L = (A, S, \delta, s_0, S')$ be its unique minimal dfa. Then we define the *syntactic semigroup* of L as $S_L =_{\mathrm{def}} \{\delta^w : w \in A^+\}$ where the composition is defined as $\delta^u \cdot \delta^v =_{\mathrm{def}} \delta^{uv}$ for every $u, v \in A^+$. Such a syntactic semigroup S_L can be considered as an *ordered syntactic semigroup* $S_L^{\leq} =_{\mathrm{def}} (S_L, \leq)$ by setting $\mu \leq \eta$ if and only if $\alpha\eta\beta(s_0) \in S'$ implies $\alpha\mu\beta(s_0) \in S'$ for all $\alpha, \beta \in S_L \cup \{\mathrm{id}_S\}$. It is easy to see that $\leq$ is reflexive and transitive, and we want to show that it is also antisymmetric. Let $\mu, \eta \in S_L$ with $\mu \leq \eta$ and $\eta \leq \mu$. Thus we have $\alpha\mu\beta(s_0) \in S' \Longleftrightarrow \alpha\eta\beta(s_0) \in S'$ for all $\alpha, \beta \in S_L \cup \{\mathrm{id}_S\}$. Suppose $\mu(s') \neq \eta(s')$ for some $s' \in S$ and choose a suitable $u \in A^*$ with $\delta^u(s_0) = s'$ (such an u exists, because F_L is minimal). Since $\delta^u\mu\beta(s_0) \in S' \Longleftrightarrow \delta^u\eta\beta(s_0) \in S'$ for all $\beta \in S_L \cup \{\mathrm{id}_S\}$, with $s_1 =_{\mathrm{def}} \mu(s')$ and $s_2 =_{\mathrm{def}} \eta(s')$ it follows $\beta(s_1) \in S' \Longleftrightarrow \beta(s_2) \in S'$ for all $\beta \in S_L \cup \{\mathrm{id}_S\}$. Therefore, $\delta(s_1, w) \in S' \Longleftrightarrow \delta(s_2, w) \in$

S' for every $w \in A^*$, which is a contradiction to the fact that F_L is a minimal dfa. It follows that $\mu = \eta$ which shows that $\leq$ is antisymmetric. Thus $\leq$ is an order relation. Now we want to show $\mu\gamma \leq \eta\gamma$ and $\gamma\mu \leq \gamma\eta$ for every $\mu, \eta, \gamma \in S_L$ with $\mu \leq \eta$. Let $\mu, \eta, \gamma \in S_L$ with $\mu \leq \eta$, then by definition we have $\alpha\eta\gamma\beta(s_0) \in S' \implies \alpha\mu\gamma\beta(s_0) \in S'$ and $\alpha\gamma\eta\beta(s_0) \in S' \implies \alpha\gamma\mu\beta(s_0) \in S'$ for all $\alpha, \beta \in S_L \cup \{\mathrm{id}_S\}$. It follows that $\mu\gamma \leq \eta\gamma$ and $\gamma\mu \leq \gamma\eta$. This shows that $S_L^{\leq}$ is an ordered semigroup.

Theorem 6.9. *Let A be a fixed alphabet, let $L \subseteq A^+$ be a regular language and $S_L^{\leq}$ be its ordered syntactic semigroup. Then $L \in \mathcal{B}_{3/2}$ if and only if $S_L^{\leq}$ satisfies all equations $\{E_m : m \geq 0\}$ for any choice of τ_i, β_i and γ_i from $S_L^{\leq}$ where*

$$\begin{aligned} E_m &=_{\mathrm{def}} \beta^\omega\gamma\beta^\omega \leq \beta^\omega \text{ with} \\ \gamma &=_{\mathrm{def}} \tau_0^\omega\gamma_1\tau_1^\omega\gamma_2\tau_2^\omega\cdots\gamma_{m+1}\tau_{m+1}^\omega \text{ and} \\ \beta &=_{\mathrm{def}} \tau_0^\omega\beta_0\tau_0^\omega\gamma_1\tau_1^\omega\beta_1\tau_1^\omega\gamma_2\tau_2^\omega\beta_2\tau_2^\omega\cdots\gamma_{m+1}\tau_{m+1}^\omega\beta_{m+1}\tau_{m+1}^\omega. \end{aligned}$$

Proof. Let $F_L = (A, S, \delta, s_0, S')$ be the minimal dfa of L.

"$\implies$" Suppose $S_L^{\leq}$ violates Equation E_m for some $m \geq 0$. We will show that F_L has Pattern $\mathbf{P}_{\boldsymbol{k}}$ for all $k \geq 0$ which implies $L \notin \mathcal{B}_{3/2}$ by Proposition 2.26 and Theorem 5.8. So let $k \geq 0$. For $0 \leq i \leq m+1$ and $1 \leq j \leq m+1$ we can choose suitable $\tau_i, \beta_i, \gamma_j \in S_L^{\leq}$ such that $\beta^\omega\gamma\beta^\omega \not\leq \beta^\omega$. Moreover, for $0 \leq i \leq m+1$ and $1 \leq j \leq m+1$ there exist words $u_i, v_i, w_j \in A^+$ such that $\tau_i^\omega = \delta^{u_i}$ (note that $\delta^{u_i} = \delta^{u_i u_i}$), $\beta_i = \delta^{v_i}$ and $\gamma_j = \delta^{w_j}$. Hence with $v' =_{\mathrm{def}} u_0v_0u_0w_1u_1v_1u_1\cdots w_{m+1}u_{m+1}v_{m+1}u_{m+1}$ and $w' =_{\mathrm{def}} u_0w_1u_1\cdots w_{m+1}u_{m+1}$ we have $\beta = \delta^{v'}$ and $\gamma = \delta^{w'}$. Now choose a minimal r such that $\beta^\omega = \delta^{v'^r}$, it follows that $\delta^{v'^r} = \delta^{v'^r v'^r}$. Since $\beta^\omega\gamma\beta^\omega \not\leq \beta^\omega$, we must have $\delta^{u'}\delta^{v'^r}\delta^z(s_0) \in S'$ and $\delta^{u'}\delta^{v'^r}\delta^{w'}\delta^{v'^r}\delta^z(s_0) \notin S'$ for some $u', z \in A^*$ (note that $\delta^\varepsilon = \mathrm{id}_S$). Let $u =_{\mathrm{def}} u'v'^r$, $v =_{\mathrm{def}} (u_0^kv_0u_0^kw_1u_1^kv_1u_1^k\cdots w_{m+1}u_{m+1}^kv_{m+1}u_{m+1}^k)^r$ and $w =_{\mathrm{def}} u_0^kw_1u_1^k\cdots w_{m+1}u_{m+1}^kv$. Since $\delta^{u_i} = \delta^{u_iu_i}$ for $0 \leq i \leq m+1$, we have $\delta^{v'^r} = \delta^v$ (particularly $\delta^v = \delta^{vv}$) and $\delta^{w'v'^r} = \delta^w$. It follows that the states $s_1 =_{\mathrm{def}} \delta(s_0, u)$ and $s_2 =_{\mathrm{def}} \delta(s_1, w)$ have a v-loop. Furthermore, we have $\delta(s_1, z) = \delta(s_0, u'v'^rz) = \delta^{u'}\delta^{v'^r}\delta^z(s_0) \in S'$ and $\delta(s_2, z) = \delta(s_0, u'v'^rw'v'^rz) = \delta^{u'}\delta^{v'^r}\delta^{w'}\delta^{v'^r}\delta^z(s_0) \notin S'$ by assumption. Thus in F_L we have found a subgraph of the form $\mathbf{P}_{\boldsymbol{k}}$. Now we proceed analogously to the proof of the direction "$\implies$" of Lemma 6.3, and we show $\alpha(\widehat{vwv}) \subseteq \alpha(\widehat{vv})$ w.r.t. k-decompositions. Hence we have found Pattern $\mathbf{P}_{\boldsymbol{k}}$ in F_L.

"$\impliedby$" Suppose $L \notin \mathcal{B}_{3/2}$. Thus F_L has Pattern $\mathbf{P}$ by Theorem 6.4. Then there exist $m \geq 0$, $s_1, s_2 \in S$, $u, z \in A^*$, $u_0, \ldots u_{m+1}, w_1, \ldots, w_{m+1} \in A^+$ and $v_0, \ldots, v_{m+1} \in A^*$ such that we find the pattern defined in Figure 2 in automaton F_L. Note that without loss of generality we may assume that $v_i \in A^+$ (otherwise we could use the words v_iu_i instead of v_i). For $0 \leq i \leq m+1$ and $1 \leq j \leq m+1$ we define $\tau_i =_{\mathrm{def}} \delta^{u_i}$, $\beta_i =_{\mathrm{def}} \delta^{v_i}$, $\gamma_j =_{\mathrm{def}} \delta^{w_j}$, $\gamma =_{\mathrm{def}} \tau_0^\omega\gamma_1\tau_1^\omega\gamma_2\tau_2^\omega\cdots\gamma_{m+1}\tau_{m+1}^\omega$ and $\beta =_{\mathrm{def}} \tau_0^\omega\beta_0\tau_0^\omega\gamma_1\tau_1^\omega\beta_1\tau_1^\omega\gamma_2\tau_2^\omega\beta_2\tau_2^\omega\cdots\gamma_{m+1}\tau_{m+1}^\omega\beta_{m+1}\tau_{m+1}^\omega$. We want to show that $\beta^\omega\gamma\beta^\omega \not\leq \beta^\omega$. Since $\beta(s) = \delta(s, u_0^{i_0}v_0u_0^{i_0}w_1u_1^{i_1}v_1u_1^{i_1}\cdots w_{m+1}u_{m+1}^{i_{m+1}}v_{m+1}u_{m+1}^{i_{m+1}})$ for suitable $i_0, \ldots, i_{m+1} \geq 1$, we conclude $\beta(s_1) = s_1$ and $\beta(s_2) = s_2$. It follows that $\beta^\omega(s_1) = s_1$ and $\beta^\omega(s_2) = s_2$. Analogously we obtain $\gamma(s_1) = s_2$. Thus with $\mu =_{\mathrm{def}} \delta^u$ and $\eta =_{\mathrm{def}} \delta^z$ (note that $\delta^\varepsilon = \mathrm{id}_S$) we obtain $\mu\beta^\omega\gamma\beta^\omega\eta(s_0) = \gamma\beta^\omega\eta(s_1) = \eta(s_2) = \delta(s_2, z) \notin S'$ and $\mu\beta^\omega\eta(s_0) = \eta(s_1) = \delta(s_1, z) \in S'$. This shows $\beta^\omega\gamma\beta^\omega \not\leq \beta^\omega$. Therefore, $S_L^{\leq}$ does not satisfy Equation E_m. ❑

Due to [Arf87] the class $\mathcal{B}_{3/2}$ can be understood as a positive $+$–variety of languages. For a reproof and the definition of the notion of positive varieties see [PW97]. An Eilenberg-like theorem was given for the case of positive varieties in [Pin95b] which states that positive $+$–varieties of languages and varieties of finite ordered semigroups are in one–one correspondence via the operation of taking ordered syntactic semigroups. So the equations $\{E_m : m \geq 0\}$ from Theorem 6.9 characterize the variety of finite ordered semigroups corresponding to $\mathcal{B}_{3/2}$. It is known that this variety of finite ordered semigroups is equal to the Mal'cev product of the variety of finite semigroups corresponding to dot-depth one languages with a certain other variety of finite ordered semigroups, as stated in Theorem 5.8 in [PW97]. Contrary to their result, our characterization is effective (and even efficient) as can be followed from Lemma 6.5.

7 Conclusions

Our detailed levelwise analysis of languages having dot-depth $3/2$ was inspired by normalform and decidability results known for languages of level $3/2$ of the Straubing-Thérien hierarchy, and by previous work concerning k-decompositions. This is reflected by the following observation. If we omit the loops u_i in our Pattern **P** we obtain an equivalent version of the forbidden pattern for level $3/2$ of the Straubing-Thérien hierarchy from [PW97]. We also believe that it is very likely that the decidability questions of both hierarchies are related on all levels $n/2$, and at least for $n \leq 3$ this is true.

We see the contribution of our paper not only as a stand-alone result providing the decidability of $\mathcal{B}_{3/2}$, but in what we can learn for the general case. First we note that our proof is such that it bounds in a computable way (in terms of the automaton size and the alphabet size) the descriptional complexity of a language L, i.e., it bounds the length of an expression that witnesses that L is of dot-depth $3/2$. We believe that this kind of relation also holds for all levels $n + 1/2$ of the hierarchy. Let us further describe on a very informal level our ongoing work.

1. If we compare the forbidden pattern characterization of $\mathcal{B}_{1/2}$ (via some Pattern $\mathbf{P'}$, proved in [PW97]) and the one of $\mathcal{B}_{3/2}$ (via our Pattern **P**), we observe that subgraphs of type $\mathbf{P'}$ also appear in Pattern **P**. Generalizing this, we define for $n \geq 1$ a Pattern $\mathbf{R_n}$ whose occurence in a transition graph can be effectively checked and for which $\mathbf{R_1}$=$\mathbf{P'}$ and $\mathbf{R_2}$=**P** holds.

2. We look at $\mathbf{R_n}$ as a candidate to characterize the classes $\mathcal{B}_{n-1/2}$ in a forbidden pattern manner. Denote by $\mathcal{C}_n$ the class of languages that can be accepted exactly by those dfa's which do not have $\mathbf{R_n}$ in their transition graph.

3. We can prove for all $n \geq 1$ that $\mathcal{C}_n$ is a class of starfree languages.

4. We can prove for all $n \geq 1$ that $\mathcal{B}_{n-1/2} \subseteq \mathcal{C}_n$.

5. We can prove for all $n \geq 1$ that $\mathcal{C}_n \subsetneq \mathcal{C}_{n+1}$ using the languages L_n from [BK78] witnessing the strictness of the dot-depth hierarchy.

Acknowledgements. We thank Bernd Borchert, José Ruiz, Victor Selivanov, Klaus W. Wagner and Thomas Wilke for discussions related to our subject.

References

[Arf87] M. Arfi. Polynomial operations on rational languages. In *Proceedings 4th Symposium on Theoretical Aspects of Computer Science*, volume 247 of *Lecture Notes in Computer Science*, pages 198–206. Springer-Verlag, 1987.

[Arf91] M. Arfi. Opérations polynomiales et hiérarchies de concaténation. *Theoretical Computer Science*, 91:71–84, 1991.

[BK78] J. A. Brzozowski and R. Knast. The dot-depth hierarchy of star-free languages is infinite. *Journal of Computer and System Sciences*, 16:37–55, 1978.

[Brz76] J. A. Brzozowski. Hierarchies of aperiodic languages. *RAIRO Inform. Theor.*, 10:33–49, 1976.

[BS73] J. A. Brzozowski and I. Simon. Characterizations of locally testable languages. *Discrete Math.*, 4:243–271, 1973.

[CB71] R. S. Cohen and J. A. Brzozowski. Dot-depth of star-free events. *Journal of Computer and System Sciences*, 5:1–16, 1971.

[CH91] S. Cho and D. T. Huynh. Finite-automaton aperiodicity is PSPACE-complete. *Theoretical Computer Science*, 88:99–116, 1991.

[CK96] C. Choffrut and J. Karhumäki. Combinatorics of words. In G.Rozenberg and A.Salomaa, editors, *Handbook of formal languages*, volume I, pages 329–438. Springer, 1996.

[CPP93] J. Cohen, D. Perrin, and J. E. Pin. On the expressive power of temporal logic. *Journal of Computer and System Sciences*, 46:271–294, 1993.

[Gla98] C. Glaßer. A normalform for classes of concatenation hierarchies. Technical Report 216, Inst. für Informatik, Univ. Würzburg, 1998.

[Has83] K. Hashiguchi. Representation theorems on regular languages. *Journal of Computer and System Sciences*, 27:101–115, 1983.

[Hig52] G. Higman. Ordering by divisibility in abstract algebras. In *Proc. London Math. Soc.*, volume 3, pages 326–336, 1952.

[Imm88] N. Immerman. Nondeterministic space is closed under complementation. *SIAM Journal on Computing*, 17:935–938, 1988.

[McN74] R. McNaughton. Algebraic decision procedures for local testablility. *Math. Syst. Theor.*, 8:60–76, 1974.

[MP71] R. McNaughton and S. Papert. *Counterfree Automata*. MIT Press, Cambridge, 1971.

[Pin95a] J. E. Pin. Finite semigroups and recognizable languages: an introduction. In J. Fountain, editor, *NATO Advanced Study Institute: Semigroups, Formal Languages and Groups*, pages 1–32. Kluwer Academic Publishers, 1995.

[Pin95b] J. E. Pin. A variety theorem without complementation. *Russian Math.*, 39:74–83, 1995.

[Pin96a] J. E. Pin. Logic, semigroups and automata on words. *Annals of Mathematics and Artificial Intelligence*, 16:343–384, 1996.

[Pin96b] J. E. Pin. Syntactic semigroups. In G.Rozenberg and A.Salomaa, editors, *Handbook of formal languages*, volume I, pages 679–746. Springer, 1996.

[PP86] D. Perrin and J. E. Pin. First-order logic and star-free sets. *Journal of Computer and System Sciences*, 32:393–406, 1986.

[PW97] J. E. Pin and P. Weil. Polynomial closure and unambiguous product. *Theory of computing systems*, 30:383–422, 1997.

[Sch99a] H. Schmitz. Generalized deterministic languages and their automata: A characterization of restricted temporal logic. Technical Report 226, Inst. für Informatik, Univ. Würzburg, 1999. An extended abstract appeared in the Preproceedings of DCAGRS'99, Magdeburg.

[Sch99b] H. Schmitz. Some forbidden patterns in automata for dot-depth one languages. Technical Report 220, Inst. für Informatik, Univ. Würzburg, 1999.

[Sim72] I. Simon. *Hierarchies of events with dot-depth one*. PhD thesis, University of Waterloo, 1972.

[Sim90] I. Simon. Factorization forests of finite height. *Theoretical Computer Science*, 72:65–94, 1990.

[SS83] J. Sakarovitch and I. Simon. Subwords. In M. Lothaire, editor, *Combinatorics on Words*, Encyclopedia of mathematics and its applications, pages 105–142. Addison-Wesley, 1983.

[Ste85a] J. Stern. Characterizations of some classes of regular events. *Theoretical Computer Science*, 35:17–42, 1985.

[Ste85b] J. Stern. Complexity of some problems from the theory of automata. *Information and Control*, 66:163–176, 1985.

[Str81] H. Straubing. A generalization of the Schützenberger product of finite monoids. *Theoretical Computer Science*, 13:137–150, 1981.

[Str85] H. Straubing. Finite semigroups varieties of the form V * D. *J.Pure Appl.Algebra*, 36:53–94, 1985.

[Sze87] R. Szelepcsényi. The method of forcing for nondeterministic automata. *Bull.of the EATCS*, 33:96–100, 1987.

[Thé81] D. Thérien. Classification of finite monoids: the language approach. *Theoretical Computer Science*, 14:195–208, 1981.

[Tho82] W. Thomas. Classifying regular events in symbolic logic. *Journal of Computer and System Sciences*, 25:360–376, 1982.

[TW98] D. Thérien and Th. Wilke. Over words, two variables are as powerful as one quantifier alternation: $\mathrm{FO}^2 = \Sigma_2 \cap \Pi_2$. In *Proceedings of the 30th annual ACM symposium on Theory of computing*, pages 234–240, Dallas, Texas, 1998. ACM Press.