Some Forbidden Patterns in Automata for Dot-Depth One Languages

Heinz Schmitz*
Theoretische Informatik
Universität Würzburg
Am Exerzierplatz 3
97072 Würzburg
schmitz@informatik.uni-wuerzburg.de

Abstract

We deal with the class $\mathcal{B}_1$ of dot-depth one languages which forms level one of the so-called dot-depth hierarchy. For some fixed alphabet A with $|A| \geq 2$, a language $L \subseteq A^+$ is in the class $\mathcal{B}_1$ if and only if it can be expressed as a Boolean combination of languages $u_0A^*u_1A^*u_2A^* \cdots A^*u_{n-1}A^*u_n$, where $u_i \in A^*$ and $n \geq 0$. This class is the highest level of the dot-depth hierarchy known to be decidable up to now and a number of different characterizations are known. We add here one more in terms of certain patterns that must not appear in the transition graph of deterministic finite automata accepting dot-depth one languages.

More precisely, the following levelwise characterization of languages having dot-depth one is known ($k \geq 0$): A language $L \subseteq A^+$ is in the class $\mathcal{B}_{1,k}$ if and only if any infinite sequence of words, where each word is a so-called k-extension of its predecessor, has a finite number of alternations with respect to L. It holds that $\mathcal{B}_{1,k} \subseteq \mathcal{B}_{1,k+1}$ and $\mathcal{B}_1 = \bigcup_{k\geq 0} \mathcal{B}_{1,k}$. We use this characterization to prove a uniform and levelwise "forbidden-pattern" characterization of dot-depth one languages having level $\leq k$ of the type: $L \in \mathcal{B}_{1,k}$ if and only if *none* of certain patterns P_k, P'_k and P''_k (cf. Figures 4, 5, 6) appears in the transition graph of a deterministic finite automaton accepting L. These patterns are different from other known characterizations of this type for these classes.

1 Introduction

We deal with dot-depth one languages which form level one of the dot-depth hierarchy, first studied in [CB71]. This hierarchy is known to be strict [BK78] and exhausts the class of regular starfree languages. For an overview or as a good starting point to the study of concatenation hierarchies and related problems see e.g. the articles [Brz76, Pin96a, Pin96b, Tho96].

We consider languages over a finite alphabet A with $|A| \geq 2$. All definitions will be made for arbitrary but fixed alphabet A. 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

*Supported by the Deutsche Forschungsgemeinschaft (DFG), grant Wa 847/1-2.

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 $u_0 L_1 u_1 L_2 u_2 \cdots u_{n-1} L_n u_n$, where $u_i \in A^*$, $L_i \in \mathcal{C}$ and $n \geq 0$ (if $n = 0$ then $u_0 \neq \epsilon$). Then the dot-depth hierarchy can be defined as the following family of classes of languages of A^+:

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* $n/2$. For example, for $A = \{a, b\}$ one verifies that $(ab)^* aaA^* = \big(A^* \cap \overline{(bA^* \cup A^*a \cup A^*aaA^* \cup A^*bbA^*)}\big) aaA^*$ is in $\mathcal{B}_{3/2}$. With a little more effort one can show, that this language is not in $\mathcal{B}_1$. Thus it has dot-depth $3/2$.

Unfortunately one main question about concatenation hierarchies, namely the question of the decidability of their classes, appears to be extremely difficult, although a lot of effort via different approaches has been invested. The decidability problem for some $n \geq 0$ can be stated as follows: given a regular language $L \subseteq A^+$ presented by a deterministic finite automaton, decide whether or not L has dot-depth $n/2$. To our knowledge, only levels $0, 1/2$ and 1 of the dot-depth hierarchy are known to be decidable (cf. [PW97]). We deal in this paper with level one, i.e. with the class $\mathcal{B}_1$.

The definitions given here are adopted from [Pin96b] and differ a little bit from the ones given in earlier literature, e.g. [CB71, BK78]. On one hand the levels $n + 1/2$ of the dot-depth hierarchy extent the corresponding classes defined in [CB71], but levels n coincide. On the other hand languages having dot-depth one have been studied in earlier literature as languages of A^*, i.e. possibly including the empty word. The restiction to A^+ now is due to the algebraic theory of finite automata. There it is convenient to exclude the empty word in the case of the dot-depth hierarchy to preserve a comprehensive theory (cf. [Pin96b]). In [Eil76] it was shown that the classes $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \ldots$ form a hierarchy of $+$-varieties of languages. We adopt this convention to be consistent with recent developments, whereas it is easy to verify that the characterizations we use remain valid.

In the introductory Section 2 we recall some known facts about languages having dot-depth one and start with the 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. It is known that L is in $\mathcal{B}_1$ if and only if L can be written as a union of equivalence classes of $\sim_{m,k}$ for some integers m, k. Then we take a closer look at the structure of $\mathcal{B}_1$ and fix the "level" k of a language within this class. For $k \geq 0$ the class $\mathcal{B}_{1,k}$ is the class of languages L that can be written as unions of equivalence classes of $\sim_{m,k}$ for some integer m. It holds that $\mathcal{B}_{1,k} \subseteq \mathcal{B}_{1,k+1}$ and $\mathcal{B}_1 = \bigcup_{k \geq 0} \mathcal{B}_{1,k}$. We conclude Section 2 with an interesting characterization of the classes $\mathcal{B}_{1,k}$ obtained by J. Stern in [Ste85a]. It involves infinite sequences of words, where each word is a so-called k-extension of its predecessor: A word y is a k-extension of a word x if again they have the same first k letters, the same last k letters and any letter of x appears in y and is followed by the same k letters in both words. A language L is in $\mathcal{B}_{1,k}$ if and only if any infinite sequence of k-extensions has finitely many alternations with respect to L. Additionally, in [Ste85a] a characterization of $\mathcal{B}_{1,k}$ in terms of the structure of the transition graph of deterministic finite automata was given: It looks at the strongly connected components and involves two forbidden patterns and a property called "k-stability". We recall this as well.

In order to make the characterization via k-extensions accessible to our techniques we introduce in Section 3 the notion of an *elementary* k-extension and show how to decompose a k-extension into elementary ones.

Section 4 is devoted to proof our main result (Theorem 4.6). We show a uniform characterization of languages having dot-depth one and level $\leq k$, namely: $L \in \mathcal{B}_{1,k}$ if and only if none of the patterns P_k, P'_k and P''_k (cf. Figures 4, 5, 6) appears in the transition graph of an arbitrary deterministic finite automaton accepting L.

Finally, in Section 5 we draw some consequences. For fixed k we see that our characterization provides a nondeterministic logspace decision algorithm for $\mathcal{B}_{1,k}$. It is known that the level k of a language L can be bound by $|F|^3$ where $|F|$ is the number of states of the minimal automaton F accepting L (cf. [Ste85a], Prop.4.1), so we can add to the known decision algorithms for $\mathcal{B}_1$ (cf. [Kna83, Ste85b, PW97]) another one.

Let ϵ denote the empty word and $|w|$ the length of a word $w \in A^*$. Let $x, y \in A^*$. We call y a *prefix* (*suffix*) of x if there exists a word z such that $x = yz$ ($x = zy$, resp.). Moreover y is a *factor* of x if there are word z_0, z_1 such that $x = z_0 y z_1$. We denote by $\preceq$ the *subword relation* on A^*, i.e. $y \preceq x$ if and only if there exist $z_0, z_1, \ldots, z_n \in A^*$ such that $y = a_1 a_2 \cdots a_n$ and $x = z_0 a_1 z_1 a_2 z_2 \cdots a_n z_n$. If trivial cases are excluded these relations are called *proper*. For a word $x = a_1 a_2 \cdots a_n$ we denote with x^R its *reverse*, i.e. $x^R =_{\mathrm{def}} a_n a_{n-1} \cdots a_1$, and for a language L let $L^R =_{\mathrm{def}} \left\{ x^R \mid x \in L \right\}$. We will denote infinite sequences of words $\{w_i\}_{i=1}^{\infty}$ (or numbers $\{\alpha_i\}_{i=1}^{\infty}$) for short as $\{w_i\}$ ($\{\alpha_i\}$, respectively).

As is standard, 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 transition function, $s_0 \in S$ is the starting state and $S' \subseteq S$ is the set of accepting states. As usual we extend transition functions to input words, and we denote by $|F|$ the number of states of F.

2 Known Characterizations of $\mathcal{B}_1$

We start with a formal definition of the basic congruence $\sim_{m,k}$.

Definition 2.1. Let $k, n \geq 0$ and $w = a_1 a_2 \cdots a_n \in A^*$ with $a_i \in A$. The *k-prefix* and the *k-suffix* of w are given by

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

and

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

Let $m, k \geq 0$ and $W_{m,k} =_{\mathrm{def}} (w_1, w_2, \ldots, w_m)$ with $w_i \in A^{k+1}$ be an m-tuple of words of length $k+1$. We say that $W_{m,k}$ *appears* in a word w, if there exist words $u_1, u_2, \ldots, u_m, v_1, v_2, \ldots, v_m$ with $|u_i| < |u_{i+1}|$ and w can be written as $w = u_i w_i v_i$. Let $\mu_{m,k}(w)$ be the set of all $W_{m,k}$ that appear in a given word w. By convention, $\mu_{0,k}(w) = \emptyset$ for all k.

Definition 2.2 [Sim72]. Let $m, k \geq 0$ and $x, y \in A^*$. We say that x and y are *(m,k)-equivalent*, in symbols $x \sim_{m,k} y$, if and only if

(a) $x = y$, if $|x| < m + k$, or
(b) $p_k(x) = p_k(y), s_k(x) = s_k(y)$ and $\mu_{m,k}(x) = \mu_{m,k}(y)$, otherwise.

It is easy to see that $\sim_{m,k}$ is a congruence of finite index on A^*. With $[w]_{m,k}$ we denote the equivalence class of a word w modulo $\sim_{m,k}$.

Proposition 2.3. *Let $m, k \geq 0$. Then $\sim_{m,k+1}$ and $\sim_{m+1,k}$ are refinements of $\sim_{m,k}$.*

Proof. Let $x, y \in A^*$. We show that $x \sim_{m,k+1} y$ implies $x \sim_{m,k} y$ and $x \sim_{m+1,k} y$ implies $x \sim_{m,k} y$.

(i) Let $x \sim_{m,k+1} y$. If $|x| < m + k + 1$ then $x = y$ and $x \sim_{m,k} y$. So we may assume that $|x| \geq m + k + 1$. Obviously $p_k(x) = p_k(y)$ and $s_k(x) = s_k(y)$ since $p_{k+1}(x) = p_{k+1}(y)$ and $s_{k+1}(x) = s_{k+1}(y)$ by assumption. If $m = 0$ then $\mu_{0,k}(x) = \emptyset = \mu_{0,k}(y)$ and $x \sim_{m,k} y$ follows. So let $m > 0$. Due to the length of x there are at least $m + 1$ proper factors of length $k + 1$ in x. Now let $W_{m,k} \in \mu_{m,k}(x)$. There must be a factor v of length $k + 1$ in x that was not used to build $W_{m,k}$. Let $1 \leq j \leq m$ such that $W_{m,k} = (w_1, w_2, \dots , w_j, w_{j+1}, \dots , w_m)$ and $w_1, w_2, \dots , w_j$ are factors in x left of v and $w_{j+1}, \dots , w_m$ are factors in x right of v. We then may choose letters $a_1, a_2, \dots , a_j, a_{j+1}, \dots , a_m$ from x such that $W_{m,k+1} =_{\text{def}} (w_1a_1, w_2a_2, \dots , w_ja_j, a_{j+1}w_{j+1}, \dots , a_mw_m)$ appears in x. Since $\mu_{m,k+1}(x) = \mu_{m,k+1}(y)$ we can conclude that $W_{m,k+1}$ also appears in y. Now we drop the letters a_i to obtain $W_{m,k}$ again and see that $W_{m,k}$ appears in y as well. This shows $\mu_{m,k}(x) \subseteq \mu_{m,k}(y)$ and the other inclusion follows by symmetry. All together we obtain $x \sim_{m,k} y$.

(ii) Let $x \sim_{m+1,k} y$ and again if $|x| < m + k + 1$ we are done. Otherwise we only have to show $\mu_{m,k}(x) = \mu_{m,k}(y)$ and if $m = 0$ we are also done. So let $m > 0$ and for a given $W_{m,k} \in \mu_{m,k}(x)$ we can determine a factor v of x as above and conclude that $W_{m+1,k} =_{\text{def}} (w_1, w_2, \dots , w_j, v, w_{j+1}, \dots , w_m)$ appears in x and also in y. We can drop v to see that $W_{m,k}$ appears in y as well and finally get $\mu_{m,k}(x) = \mu_{m,k}(y)$. ❑

Note that $[\epsilon]_{m,k} = \{\epsilon\}$ whenever $m+k > 0$. Especially we always find a refinement e.g. $\sim_{m',k}$ of $\sim_{m,k}$ with $[\epsilon]_{m',k} = \{\epsilon\}$ and we can state I. Simon's result for languages of A^+.

Theorem 2.4 [Sim72]. *Let $L \subseteq A^+$. Then $L \in \mathcal{B}_1$ if and only if L can be written as a union of equivalence classes modulo $\sim_{m,k}$ for some $m, k \geq 0$.*

Next we fix k and take a levelwise look at $\mathcal{B}_1$.

Definition 2.5. Let $k \geq 0$. Then $\mathcal{B}_{1,k}$ is the class of languages $L \subseteq A^+$ that can be written as a union of equivalence classes modulo $\sim_{m,k}$ for some $m \geq 0$.

It holds that $\mathcal{B}_{1,k} \subseteq \mathcal{B}_{1,k+1}$ by Proposition 2.3 and $\mathcal{B}_1 = \bigcup_{k\geq 0} \mathcal{B}_{1,k}$ by definition. For a language $L \subseteq A^+$ and a minimal k with $L \in \mathcal{B}_{1,k}$ we say that L *has level* k. In this terminology $\mathcal{B}_{1,k}$ is the class of languages having level $\leq k$. As a complementary result we show next that the classes $\mathcal{B}_{1,k}$ are closed under taking the reverse of a language.

Proposition 2.6. *Let $m, k \geq 0$ and $x, y \in A^*$. Then $x \sim_{m,k} y$ if and only if $x^R \sim_{m,k} y^R$.*

Proof. It is sufficient to show one implication and we assume $x \sim_{m,k} y$. If $|x| < m + k$ then $x = y$, $x^R = y^R$ and hence $x^R \sim_{m,k} y^R$. Now let $|x| \geq m + k$ and since $x \sim_{m,k} y$ we also have $|y| \geq m + k$. It holds that $p_k(x^R) = s_k(x)^R = s_k(y)^R = p_k(y^R)$ and $s_k(x^R) = p_k(x)^R = p_k(y)^R = s_k(y^R)$, so it remains to show $\mu_{m,k}(x^R) = \mu_{m,k}(y^R)$. We are done if $m = 0$. For the other case we claim that $W_{m,k} = (w_1, w_2, \dots , w_m)$ appears in x if and only if $(W_{m,k})^R =_{\text{def}} (w_m^R, w_{m-1}^R, \dots , w_1^R)$ appears in x^R. Since $((W_{m,k})^R)^R = W_{m,k}$ we restrict ourselves to the 'only if' part. Let $u_1, u_2, \dots , u_m, v_1, v_2, \dots , v_m$ witness that $W_{m,k}$ appears in

x. It holds that $x^R = (u_i w_i v_i)^R = v_i^R w_i^R u_i^R$ for $1 \leq i \leq m$ and if additionally $i < m$ we have $|v_{i+1}^R| < |v_i^R|$ since $|u_{i+1}^R| > |u_i^R|$. For $1 \leq i \leq m$ we define $u'_i =_{\text{def}} v_{m-i+1}^R$ and $v'_i =_{\text{def}} u_{m-i+1}^R$. Then we can conclude that $x^R = u'_i w_{m-i+1}^R v'_i$ for $1 \leq i \leq m$ and if additionally $i < m$ we have $|u'_i| = |v_{m-i+1}^R| < |v_{m-(i+1)+1}^R| = |u'_{i+1}|$. So the words u'_i and v'_i witness that $(W_{m,k})^R$ appears in x^R, which proves what we claimed. Now together with our assumption we can write $W_{m,k} \in \mu_{m,k}(x^R) \Leftrightarrow (W_{m,k})^R \in \mu_{m,k}(x) \Leftrightarrow (W_{m,k})^R \in \mu_{m,k}(y) \Leftrightarrow W_{m,k} \in \mu_{m,k}(y^R)$. ❑

Proposition 2.7. *For every language $L \subseteq A^+$ and every $k \geq 0$ it holds that $L \in \mathcal{B}_{1,k}$ if and only if $L^R \in \mathcal{B}_{1,k}$.*

Proof. We only prove the implication from left to right and suppose there is a finite set of words x_i such that L can be written as a union of their equivalence classes $[x_i]_{m,k}$ for some m. Then L^R is the union of all $[x_i^R]_{m,k}$ since

$$\begin{array}{rcll} w^R \in L^R & \Longleftrightarrow & w \in L & \\ & \Longleftrightarrow & w \in [x_i]_{m,k} \text{ for some i} & \\ & \Longleftrightarrow & w \sim_{m,k} x_i \text{ for some i} & \\ & \Longleftrightarrow & w^R \sim_{m,k} x_i^R \text{ for some i} & \text{(by Proposition 2.6)} \\ & \Longleftrightarrow & w^R \in [x_i^R]_{m,k} \text{ for some i} & \end{array}$$

Thus $L^R \in \mathcal{B}_{1,k}$ as well. ❑

To state the characterization of the classes $\mathcal{B}_{1,k}$ from [Ste85a] we first need to recall a relation on words which is a special kind of arbitrary word extensions induced by the subword relation.

Definition 2.8 [Ste85a]. Let $k \geq 0$, $u = u_1 u_2 \cdots u_m$ and $v = v_1 v_2 \cdots v_n \in A^*$ with $m, n \geq 0$ and letters u_i, v_i. A mapping $f : \{1, 2, \ldots, m\} \mapsto \{1, 2, \ldots, n\}$ is called a *k-extension from u to v* if

(a) $f(x+1) > f(x)$, for $1 \leq x < m$, and
(b) $f(x) = x$, for $1 \leq x \leq k$, and
(c) $f(m-x+1) = n-x+1$, for $1 \leq x \leq k$, and
(d) $u_{x+i} = v_{f(x)+i}$, for $1 \leq x \leq m$, $0 \leq i \leq k$, $x+i \leq m$.

Note that f is the identity function if and only if $u = v$, otherwise f is length-increasing and we call it a proper k-extension. We write $u \preceq_k v$ if there exists a k-extension from u to v and we say $u \preceq_k v$ *via* f if we want to emphasize on the witnessing function. For convenience, if $u \preceq_k v$ via some mapping f we may sometimes simply say that v is a k-extension of u. Note that a k-extension is a mapping between indices that preserves the k-prefix (b), the k-suffix (c), any letter of u appears in the same order in v (a) and is followed in both words by the same k letters (d) if the remaining suffix is long enough. Moreover $\preceq_0$ is equal to $\preceq$.

Proposition 2.9. *Let $k \geq 0$. Then $\preceq_k$ is a reflexive and transitive relation on A^*.*

Proof. The identity function is a k-extension from u to u for every $u \in A^*$. Let $u = u_1 u_2 \cdots u_m$, $w = w_1 w_2 \cdots w_l$, $v = v_1 v_2 \cdots v_n$, $u \preceq_k w$ via f and $w \preceq_k v$ via g. Then $f \circ g$ with $f \circ g(x) =_{\text{def}} g(f(x))$ is a mapping from $\{1, 2, \ldots, m\}$ to $\{1, 2, \ldots, n\}$ which obviously fulfills the requirements (a), (b) and (c) of the above definition. To see (d) we can conclude from the assumptions $u_{x+i} = w_{f(x)+i}$, for $1 \leq x \leq m$, $0 \leq i \leq k$, $x+i \leq m$ and $w_{x+i} = v_{g(x)+i}$, for $1 \leq x \leq l$, $0 \leq i \leq k$, $x+i \leq l$ that $u_{x+i} = w_{f(x)+i} = v_{g(f(x))+i}$ since $1 \leq f(x) \leq l$ and obtain $u_{x+i} = v_{f \circ g(x)+i}$, for $1 \leq x \leq m$, $0 \leq i \leq k$, $x+i \leq m$. ❑

A k-*extension chain* is an infinite sequence of words $\{w_i\}$ with $w_i \in A^*$ such that $w_i \preceq_k w_{i+1}$ for all i. Let L be a language. We say that $\{w_i\}$ has an *alternation with respect to* L if $w_i \in L \Leftrightarrow w_{i+1} \notin L$ for some i.

We recall conditions of a dfa also introduced in [Ste85a] that characterize its accepted languages as dot-depth one, level $\leq k$. A dfa $F = (A, S, \delta, s_0, S^{'})$ is called k-stable if for any two states s_1, s_2 and any word w of length k it holds that whenever $s_1, s_2, \delta(s_1, w), \delta(s_2, w)$ belong to the same strongly connected component of the transition graph of F, then $\delta(s_1, w) = \delta(s_2, w)$. To put it another way, a dfa F is not k-stable if and only if there are states $s_1, s_2, s_3, s_4, s, s' \in S$ and words $x, w, y, y', z \in A^*$ with $|w| = k$ and $s \in S^{'} \Leftrightarrow s' \notin S^{'}$ such that there is a diagram as given in Figure 1 in the transition graph of F. We say that F has pattern T_k in this case.

A dfa F has pattern T'_k if there are states $s_1, s_2, s_3, s_4, s_5, s_6, s, s' \in S$ and words $x, w, u, v, z \in A^*$ with $|w| = k$, $s \in S^{'} \Leftrightarrow s' \notin S^{'}$ and A and A' are distinct strongly connected components such that there is a diagram as given in Figure 2 in the transition graph of F.

A dfa F has pattern T''_k if there are states $s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s, s' \in S$ and words $x, w, w', y, y', u, v, z \in A^*$ with $|w| = |w'| = k$, $s \in S^{'} \Leftrightarrow s' \notin S^{'}$ and A and A' are distinct strongly connected components such that there is a diagram as given in Figure 3 in the transition graph of F.

Since the empty word changes the number of alternations with respect to a language in an infinite sequence at most by one (note that only the case of $k = 0$ is relevant) we can state J. Stern's result for languages of A^+.

Theorem 2.10 [Ste85a]. *Let $L \subseteq A^+$, $k \geq 0$ and let F be the minimal dfa accepting L. Then the following statements are equivalent:*

(1) $L \in \mathcal{B}_{1,k}$

(2) any k-extension chain has only a finite number of alternations with respect to L

(3) F does not have any of the patterns T_k, T'_k and T''_k

Although each class $\mathcal{B}_{1,k}$ is closed under taking the reverse of a language (Proposition 2.7) it is not the case that if $u \preceq_k v$ then also $u^R \preceq_k v^R$. Take as a counter-example $A = \{ab\}, k = 1, u =_{\mathrm{def}} aba$ and $v = abaa$. Then $u \preceq_1 v$ via f with $f(1) =_{\mathrm{def}} 1, f(2) =_{\mathrm{def}} 2$ and $f(3) =_{\mathrm{def}} 4$. But the first letter in u^R and the first letter in v^R are followed by different letters and since each 1-extension must preserve the 1-prefix we have $u^R \not\preceq_1 v^R$.

3 Elementary k-Extensions

We observe that if a word v is a k-extension of a word u there may be several positions in u where a new factor is inserted. We want to consider these insertions one after another and introduce in this section the notion of an elementary k-extension. First we look at the following easy example. Let $A = \{a, b, c\}, u = abacb, v = abacacbcb$ and thus $m = |u| = 5$ and $n = |v| = 9$. There is a 2-extension from u to v given by $f(1) =_{\mathrm{def}} 1$, $f(2) =_{\mathrm{def}} 2$, $f(3) =_{\mathrm{def}} 5$, $f(4) =_{\mathrm{def}} 8$ and $f(5) =_{\mathrm{def}} 9$. Note that every letter u_x mapped via f to $v_{f(x)}$ is followed in both words by the same two letters. We can read this mapping as inserting in u after position 2 the factor ac and after position 3 the factor cb to obtain v. Now we look at this insertions stepwise from left to right and define $w =_{\mathrm{def}} abacacb$. Then $u \preceq_2 w$ via g defined as $g(1) =_{\mathrm{def}} 1$, $g(2) =_{\mathrm{def}} 2$, $g(3) =_{\mathrm{def}} 5$,

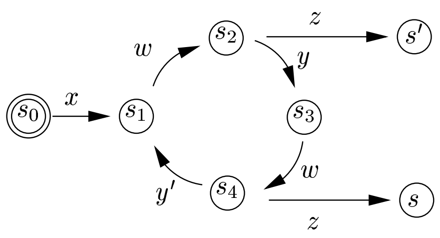

Figure 1: Pattern T_k with $|w| = k$ and $s \in S^{'} \Leftrightarrow s' \notin S^{'}$.

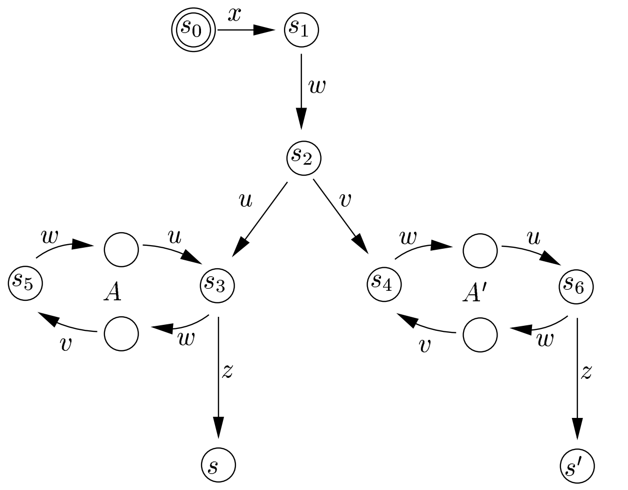

Figure 2: Pattern T'_k with $|w| = k$ and $s \in S^{'} \Leftrightarrow s' \notin S^{'}$.

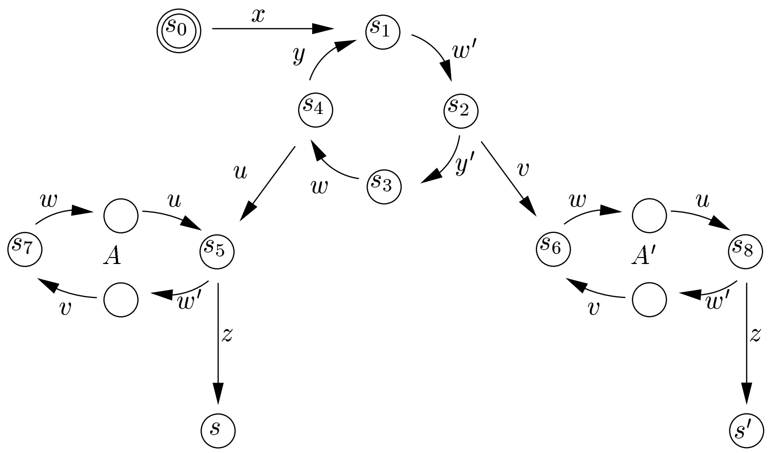

Figure 3: Pattern T''_k with $|w| = |w'| = k$ and $s \in S^{'} \Leftrightarrow s' \notin S^{'}$.

$g(4) =_{\text{def}} 6$, $g(5) =_{\text{def}} 7$ and $w \preceq_2 v$ via f' defined as $f'(x) =_{\text{def}} x$ for $x \in \{1, \ldots, 5\}$, $f'(6) =_{\text{def}} 8$ and $f'(7) =_{\text{def}} 9$. Here we have two single insertions and we get from u to v stepwise. First we make this notion precise and define such elementary k-extension.

Definition 3.1. Let $k \geq 0$ and $u \preceq_k v$ via f for $u, v \in A^*$ with $u = u_1 u_2 \cdots u_m$ and $v = v_1 v_2 \cdots v_n$. Then f is called an *elementary k-extension from u to v* if there is some j with $1 \leq j \leq m$ and

$$f(x) = \begin{cases} x, & \text{if } x < j, \\ f(j) + y, & \text{if } x = j + y \end{cases}$$

for all $1 \leq x \leq m$, $0 \leq y < m$.

Note that the identity function is an elementary k-extension and that it must be that $k < j \leq m - k + 1$ since k-extensions preserve the k-prefix and k-suffix. If there is an elementary k-extension from u to v we write $u \preceq_k^{\text{e}} v$ and since it is a special case of an arbitrary k-extension it is a reflexive and transitive relation as well. With the following lemma we generalize the above example and show that we can always decompose an arbitrary k-extension in an elementary one and a remaining k-extension.

Lemma 3.2. *Let $k \geq 0$, $u, v \in A^*$, $u \neq v$ and $u \preceq_k v$ via f. Then there exists a word $w \in A^*$ with $w \neq u$, an elementary k-extension g and a k-extension f' such that $u \preceq_k^{\text{e}} w$ via g and $w \preceq_k v$ via f'.*

Proof. Let $u = u_1 u_2 \cdots u_m$ and $v = v_1 v_2 \cdots v_n$ with letters u_i, v_i and $u \preceq_k v$ via f. We define $j =_{\text{def}} \min\{\, x \mid 1 \leq x \leq m \text{ and } f(x) > x \,\}$, i.e. j is the leftmost insertion position in u. Since f is not the identity function a factor $v_j \cdots v_{f(j)-1}$ of length $p =_{\text{def}} f(j) - j \geq 1$ is inserted into u before position j. Therefore we define $w =_{\text{def}} v_1 v_2 \cdots v_{j-1} v_j \cdots v_{f(j)} u_{j+1} u_{j+2} \cdots u_m$ with $l =_{\text{def}} |w| = m + p$. Furthermore let

$$g(x) =_{\text{def}} \begin{cases} x, & \text{if } x < j, \\ f(j) + y, & \text{if } x = j + y \text{ for some } y \geq 0 \end{cases}$$

for $1 \leq x \leq m$, $0 \leq y < m$ and

$$f'(x) =_{\text{def}} \begin{cases} x, & \text{if } x < f(j), \\ f(x - p), & \text{otherwise} \end{cases}$$

for $1 \leq x \leq l$. Note that by the definition of g we have $g(j) = f(j)$ and if we can show that g is a k-extension from u to w at all, it even has the additional property required in Definition 3.1 and is thus elementary. The rest of the proof consists of two parts in which we show $u \preceq_k w$ via g and $w \preceq_k v$ via f'. To do so we verify each time requirements (a) to (d) of Definition 2.8.

(i) $u \preceq_k w$ via g.

To prove (a) for g we have to show that $g(x+1) > g(x)$ for all $1 \leq x < m$. By definition of g and j we can state the following inequalities. First suppose $1 \leq x < j - 1$. Then $g(x+1) = x + 1 > x = g(x)$. If $x = j - 1$ we have $g(x+1) = g(j) = f(j) > f(j-1) = j - 1 = x = g(x)$ since requirement (a) holds for f by assumption. And if $j \leq x < m$ we can write x as $j + y$ with a suitable choice of y and it holds that $g(x+1) = g(j+y+1) = f(j) + y + 1 > f(j) + y = g(j+y) = g(x)$.

Requirement (b) is fulfilled by g since it must be that $k < j$ by definition of j. Otherwise f violates (b) and is not a k-extension in contradiction to our assumption.

Since f fulfills also (c) we see as well that $j \leq m-k+1$, hence for every $1 \leq i \leq k$ we can write $m-k+i$ as $j+y$ with an appropriate choice of y. By definition of g we obtain $g(m-k+i) = g(j+y) = f(j)+y = m+f(j)-j-k+i = l-k+i$ for all $1 \leq i \leq k$ which shows (c) for g.

To verify (d) we have to show that $u_{x+i} = w_{g(x)+i}$ for $1 \leq x \leq m,\ 0 \leq i \leq k$ and $x+i \leq m$. We distinguish three cases.

Case 1. If $j \leq x \leq m$ we can write x as $j+y$ for some y. By definition of w and g it holds that $u_{x+i} = u_{j+y+i} = w_{f(j)+y+i} = w_{g(x)+i}$ for all $0 \leq i \leq k$ with $x+i \leq m$. Note that $w_{f(j)} = v_{f(j)} = u_j$.

Case 2. Now let $1 \leq x < j-k$. It follows that $x+i < j$ for all $0 \leq i \leq k$. By the choice of j we know that f is the identity function on all positions left of j. So by definition of w and g we have $u_{x+i} = v_{x+i} = w_{x+i} = w_{g(x)+i}$ for all $0 \leq i \leq k$.

Case 3. Finally let $j-k \leq x < j$. Since $x < j$ we have again $f(x) = x = g(x)$ and $u_x = v_x = w_x$. We must show that u_x and w_x are followed in u and w by the same k letters. But this time $w_x w_{x+1} \cdots w_{x+k}$ may have the inserted letters $v_j \cdots v_{f(j)-1}$ as a proper factor. It is not clear that we can maintain the k following letters if we simply concatenate the remaining letters of u to the leftmost inserted word as we did in the definition of w. As long as $x+k < f(j)$ we can surely argue with property (d) of f since we deal in w only with letters from v, i.e. $w_x w_{x+1} \cdots w_j \cdots w_{x+k} = v_x v_{x+1} \cdots v_j \cdots v_{x+k} = u_x u_{x+1} \cdots u_j \cdots u_{x+k}$ and we are done. If $x+k \geq f(j)$ the situation is a little more complicated. For $w_x w_{x+1} \cdots w_{f(j)-1}$ we have equality with $u_x u_{x+1} \cdots u_{f(j)-1}$ by the previous argument. It remains to show that $w_{f(j)} w_{f(j)+1} \cdots w_{x+k}$ is equal to $u_{f(j)} u_{f(j)+1} \cdots u_{x+k}$. This can be seen as follows: On one hand we can rewrite by the definitions of w and p the string $w_{f(j)} w_{f(j)+1} \cdots w_{x+k}$ as $u_j u_{j+1} \cdots u_{x+k-p}$ and get $u_j u_{j+1} \cdots u_{x+k-p} = v_{f(j)} v_{f(j)+1} \cdots v_{x+k}$ if we use property (d) for f at position j. On the other hand we can apply property (d) for f at position x and get $u_{f(j)} u_{f(j)+1} \cdots u_{x+k} = v_{f(j)} v_{f(j)+1} \cdots v_{x+k}$ as well. This completes the third case, shows (d) and finishes part *(i)*.

(ii) $w \preceq_k v$ via f'.

To see (a) for f' we have to show $f'(x+1) > f'(x)$ for all $1 \leq x < l$. By definition of f' and w we can state the following inequalities. Suppose $1 \leq x < f(j)-1$. Then $f'(x+1) = x+1 > x = f'(x)$. If $x = f(j)-1$ we have $f'(x+1) = f(j) > f(j)-1 = f'(x)$. And if $f(j) \leq x < l$ it holds that $f'(x+1) = f(x+1-p) > f(x-p) = f'(x)$ since requirement (a) holds for f by assumption.

Since $f(j) > j$ by definition of j and $j > k$ by a previous argument we see that $f'(x) = x$ for all $1 \leq x \leq k$ and requirement (b) is fulfilled by f'.

To prove (c) for f' we use again that $j \leq m-k+1$ which is contradicted by $f(j) > l-k+1 = m+p-k+1 = f(j)-j+m-k+1$. So we may assume that $f(j) \leq l-k+1$. Then we can state by definition of f' for all $1 \leq i \leq k$ that $f'(l-k+i) = f(l-k+i-p) = f(m-k+i) = n-k+i$ since (c) holds for f.

To verify (d) we have to show that $w_{x+i} = v_{f'(x)+i}$ for $1 \leq x \leq l,\ 0 \leq i \leq k$ and $x+i \leq l$. Again we distinguish three cases.

Case 1. Suppose $f(j) \leq x \leq l$. We can write x as $f(j)+y$ for some y and we have by definition of w that $w_{x+i} = w_{f(j)+y+i} = u_{j+y+i}$ for all $0 \leq i \leq k$ and $x+i \leq l$. Now observe that from $f(j)+y \leq l = m+f(j)-j$ follows that $j+y \leq m$. Thus we can continue by requirement (d) for f and by definition of f' with $u_{j+y+i} = v_{f(j+y)+i} = v_{f(j+y-f(j)+f(j))+i} = v_{f(x-p)+i} = v_{f'(x)+i}$.

Case 2. If $1 \leq x < f(j)-k$ it follows that $x+i < f(j)$ for all $0 \leq i \leq k$. By definition of f' we know that it is the identity function on all positions left of $f(j)$. So by definition of w we have $w_{x+i} = v_{x+i} = v_{f'(x)+i}$ for all $0 \leq i \leq k$.

Case 3. Finally let $f(j) - k \leq x < f(j)$. Since $x < f(j)$ we have $f'(x) = x$ and $w_x = v_x$. We must show that w_x and v_x are followed in w and v by the same k letters. First we can state with the same argument that $w_x w_{x+1} \cdots w_{f(j)-1} = v_x v_{x+1} \cdots v_{f(j)-1}$. On the other hand we can rewrite the remaining part $w_{f(j)} w_{f(j)+1} \cdots w_{x+k}$ as $u_j u_{j+1} \cdots u_{x+k-p}$ by the definition of w. Note that the last index $x + k - p = j + x + k - f(j)$ and that $x + k - f(j) < k$ by assumption. Now since (d) holds for f we have $u_{j+i} = v_{f(j)+i}$ for all $0 \leq i \leq k$ and we finally obtain $u_j u_{j+1} \cdots u_{x+k-p} = v_{f(j)} v_{f(j)+1} \cdots v_{x+k}$. This completes the third case, shows (d), finishes part *(ii)* and the proof. ❑

Repeated applications of the previous lemma lead to a complete decomposition of k-extensions into elementary ones.

Theorem 3.3. *Let $k \geq 0$, $u, v \in A^*$ and $u \preceq_k v$. Then there exist words $w_0, w_1, \ldots, w_r \in A^*$ with $w_0 = u, w_r = v$ and for all $1 \leq i < r$ it holds that $w_i \preceq_k^{\mathrm{e}} w_{i+1}$.*

Proof. Let $u, v \in A^*$ and $u \preceq_k v$ via f and define $w_0 =_{\text{def}} u$. If f is an elementary k-extension we are done with $w_1 =_{\text{def}} v$. If not take w, g and f' from Lemma 3.2, let $w_1 =_{\text{def}} w$ and hence $w_0 \preceq_k^{\mathrm{e}} w_1$ via g. Now start again with w_1, v and f' and if f' is an elementary k-extension we are done with $w_2 =_{\text{def}} v$, if not apply again Lemma 3.2 and so on. It is clear that this procedure comes to an end since proper k-extensions are length-increasing and the identity function is an elementary k-extension. ❑

4 A Pattern Characterization of $\mathcal{B}_{1,k}$

In this section we give a "forbidden-pattern" characterization of the classes $\mathcal{B}_{1,k}$ (for other characterizations of this class see [Sim75, Ste85a]). We use the characaterization of the classes $\mathcal{B}_{1,k}$ given by J. Stern in terms of k-extension chains as stated in Theorem 2.10.

Before we do so we introduce some technical prerequisites that deal with taking factors and concatenating factors in words standing in the k-extension relation. They are needed only in the proof of Theorem 4.6. There we are concerned with the following situation: Let $u \preceq_k v$ via f and let l_1, l_2 be two positions in u. If we cut out the factor v' from $f(l_1) + 1$ to $f(l_2)$ in v it need not be a k-extension of the factor u' from $l_1 + 1$ to l_2 in u. Although u' and v' have the same k-prefix since u_{l_1} and $v_{f(l_1)}$ are followed by the same k letters, f does not necessarily map $l_1 + i$ to $f(l_1) + i$ for all i with $0 \leq i \leq k$. The same problem arises if we look at the k-suffixes of u' and v'. But we can repair what is missing if we take into account the word of length k before u' in u and the word of length k following u' in u. Let $\hat{w}$ and w respectively be these words. Then we can state that $\hat{w}u'w \preceq_k \hat{w}v'w$ via f', where f' is the same as f on u' and additionally maps the k-prefixes $\hat{w}$ and the k-suffixes w to each other. To see this note that any letter of $\hat{w}$ is followed by the same k letters in u' and v' due to their common k-prefix and since w also follows v' in v any letter in v' is followed by the same k letters as before.

There is a similar situation if we take factors from l_1 to $l_2 - 1$ and $f(l_1)$ to $f(l_2) - 1$. We make this relation between u' and v' precise in the following Definition 4.1, study when such a situation occurs in Lemma 4.2 (for the case of taking factors from $l_1 + 1$ to l_2) and in Lemma 4.3 (for the case of taking factors from l_1 to $l_2 - 1$), and see how we can maintain this partial k-extension under concatenation in Lemma 4.4.

Definition 4.1. Let $k \geq 0$ and $u, u', v' \in A^*$ such that u' is a factor of u, i.e. there are $z_0, z_1 \in A^*$ with $u = z_0 u' z_1$. Set $\hat{w} =_{\text{def}} s_k(z_0)$ and $w =_{\text{def}} p_k(z_1)$. We say that v' is a *partial k-extension of u' with respect to* u, in symbols $u' \preceq_k^{\text{part}} v'$, if $\hat{w}u'w \preceq_k \hat{w}v'w$.

Note that if this relation holds and each word $\hat{w}, u'$ and v' has at least length k, then $p_k(u') = p_k(v')$ since the last letter of $\hat{w}$ is followed by the same k letters in u' and v'. Moreover it will be always clear from the context which u we mean if $u' \preceq_k^{\text{part}} v'$ with respect to u.

Lemma 4.2. *Let $k \geq 0$ and $u, v \in A^*$ with $u \preceq_k v$ via f. Let $1 \leq l_1 < l_2 \leq |u|$ with $l_1 + 1 > k, l_1 - l_2 \geq k$ and $l_2 \leq |u| - k$. Furthermore let $u' =_{\text{def}} u_{l_1+1} u_{l_i+2} \cdots u_{l_2}$ and $v' =_{\text{def}} v_{f(l_1)+1} v_{f(l_1)+2} \cdots v_{f(l_2)}$. Then $u' \preceq_k^{\text{part}} v'$.*

Proof. We have ensured by assumption that there are at least k letters in u before u' and following u'. Set $\hat{w} =_{\text{def}} u_{l_1-(k-1)} \cdots u_{l_1}$ and $w =_{\text{def}} u_{l_2+1} \cdots u_{l_2+k}$. Let $m' =_{\text{def}} |u'| = l_2 - l_1 \geq k$ and $n' =_{\text{def}} |v'| = f(l_2) - f(l_1)$. We define

$$f'(x) =_{\text{def}} \begin{cases} x, & \text{if } x \leq k, \\ f(x - k + l_1) + k - f(l_1), & \text{if } k < x \leq m' + k, \\ n' + k + i, & \text{if } m' + k < x \leq m' + 2k \text{ and } x = m' + k + i \\ & \text{ for some } 1 \leq i \leq k \end{cases}$$

for all $1 \leq x \leq m' + 2k$.

We argue that $\hat{w}u'w \preceq_k \hat{w}v'w$ via f' to show $u' \preceq_k^{\text{part}} v'$. First note that by definition of f' it fulfills (a), (b) and (c) of Definition 2.8. To see (d) it is sufficient to show on one hand that each letter of $\hat{w}$ is followed by the same k letters in $\hat{w}u'w$ and $\hat{w}v'w$: it holds since $p_k(u') = p_k(v')$ because $u_{l_1} = v_{f(l_1)}$ and these two letters are followed by the same k letters in u and in v. On the other hand we need to show that each letter of $s_k(u')$ and its mapping under f' are followed by the same k letters in $\hat{w}u'w$ and $\hat{w}v'w$. Note that $u_{l_2} = v_{f(l_2)}$ so w follows v' in v. That means that any letter of $s_k(u')$ mapped under f' to v' is now followed by the same k letters as before in v. So (d) holds because f' is equal to f on these letters due to the index transformation in the definition of f'. ❑

Lemma 4.3. *Let $k \geq 0$ and $u, v \in A^*$ with $u \preceq_k v$ via f. Let $1 \leq l_1 < l_2 \leq |u|$ with $l_1 > k, l_1 - l_2 \geq k$ and $l_2 - 1 \leq |u| - k$. Furthermore let $u' =_{\text{def}} u_{l_1} u_{l_i+1} \cdots u_{l_2-1}$ and $v' =_{\text{def}} v_{f(l_1)} v_{f(l_1)+1} \cdots v_{f(l_2)-1}$. Then $u' \preceq_k^{\text{part}} v'$.*

Proof. Again we have ensured by assumption that there are in u at least k letters before u' and following u'. Set $\hat{w} =_{\text{def}} u_{l_1-k} \cdots u_{l_1-1}$ and $w =_{\text{def}} u_{l_2} \cdots u_{l_2+k-1}$. The proof is completely analogously to the previous one since $p_k(u') = p_k(v')$ holds as well and w follows u' in u and v' in v. ❑

Lemma 4.4. *Let $k \geq 0$ and $u, v \in A^*$ and let $u_1, u_2, v_1, v_2 \in A^*$ such that $v_1 v_2$ is a factor of v and $u_1 u_2$ is a factor of u, i.e. there are $z_0, z_1 \in A^*$ with $u = z_0 u_1 u_2 z_1$. Furthermore let $|z_0| \geq k, |z_1| \geq k, |u_2| \geq k$ and $p_k(u_2) = p_k(v_2)$. If $u_1 \preceq_k^{\text{part}} v_1$ and $u_2 \preceq_k^{\text{part}} v_2$ then $u_1 u_2 \preceq_k^{\text{part}} v_1 v_2$.*

Proof. Define $\hat{w} =_{\text{def}} s_k(z_0)$ and $w =_{\text{def}} p_k(z_1)$. By assumption we have $u_1 \preceq_k^{\text{part}} v_1$ and $u_2 \preceq_k^{\text{part}} v_2$, so let $\hat{w}u_1 p_k(u_2) \preceq_k \hat{w}v_1 p_k(u_2)$ via f_1 and $s_k(u_1)u_2 w \preceq_k s_k(u_1)v_2 w$ via f_2. We have to show that $\hat{w}u_1 u_2 w \preceq_k \hat{w}v_1 v_2 w$ via some k-extension f. Therefore set $m_1 =_{\text{def}} |u_1|, m_2 =_{\text{def}} |u_2|, n_1 =_{\text{def}} |v_1|$ and $n_2 =_{\text{def}} |v_2|$ and define

$$f(x) =_{\text{def}} \begin{cases} x, & \text{if } x \leq k, \\ f_1(x), & \text{if } k < x \leq m_1 + k, \\ f_2(x - m_1) + n_1, & \text{if } m_1 + k < x \leq m_1 + m_2 + k, \\ n_1 + n_2 + k + i, & \text{if } m_1 + m_2 + k < x \leq m_1 + m_2 + 2k \\ & \text{ and } x = m_1 + m_2 + k + i \text{ for some } 1 \leq i \leq k \end{cases}$$

for all $1 \leq x \leq m_1 + m_2 + 2k$. Again we easily see by definition of f that it fulfills (a), (b) and (c) of Definition 2.8. To see (d) we can use arguments as before. Since $\hat{w}u_1 p_k(u_2) \preceq_k \hat{w}v_1 p_k(u_2)$ we have that f_1 ensures (d) for all letters of $\hat{w}u_1$. And for the letters of $u_2 w$ (d) is given because $s_k(u_1)u_2 w \preceq_k s_k(u_1)v_2 w$ via f_2. ❑

Moreover we will make use of the following lemma in the proof of Theorem 4.6.

Lemma 4.5 [Wag98]. *Let $\{\alpha_i\}_{i=1}^{\infty}$ be a sequence of real numbers such that $0 < \alpha_i < 1$ and $\alpha_i \neq \alpha_j$ for $i \neq j$. Then there exists an infinite monotonic subsequence of $\{\alpha_i\}$.*

Proof. Suppose there is no infinite strictly decreasing subsequence of $\{\alpha_i\}$. Set $\alpha_0 =_{\mathrm{def}} 0$. We inductively define a sequence of indices $\{i_s\}_{s\geq 0}$ such that $i_s < i_{s+1}$ for all s and $\alpha_j > \alpha_{i_s}$ for all $j > i_s$. Therefore let $i_0 =_{\mathrm{def}} 0$. Since there is no infinite strictly decreasing subsequence of $\{\alpha_i\}$ there exists for each s an index $i_{s+1} > i_s$ such that $\alpha_j > \alpha_{i_{s+1}}$ for all $j > i_{s+1}$. ❑

Next we define patterns that may occur in the transition graph of a deterministic finite automaton F. As we will see they lead to k-extension chains with an infinite number of alternations with respect to $L(F)$. We say that the dfa $F = (A, S, \delta, s_0, S')$ has the pattern P_k (cf. Figure 4) if there exist $w, \hat{w}, x, v, y, v_1, z' \in A^*$ with $|w| = |\hat{w}| = k, |v| \geq k, |y| \geq k, |z'| \geq k, s_k(v) = \hat{w}, p_k(y) = p_k(z') = w$ and if there exist states $s_1, s_2, s_3 \in S$ such that

- $\hat{w}yw \preceq_k^{\mathrm{e}} \hat{w}yv_1 w$,
- $\hat{w}yv_1 w \preceq_k \hat{w}vw$,
- $\delta(s_0, x) = \delta(s_1, v) = s_1$,
- $\delta(s_1, y) = s_2, \delta(s_2, v_1) = s_3$ and
- $\delta(s_2, z') \in S' \Leftrightarrow \delta(s_3, z') \notin S'$.

We say that the dfa F has the pattern P'_k (cf. Figure 5) if there exist $w, \hat{w}, x, u, u_1, z, z' \in A^*$ with $|w| = |\hat{w}| = k, |x| \geq k, |z| \geq k, |u| \geq k, s_k(x) = \hat{w}, p_k(z) = p_k(u) = w$ and if there exist states $s_1, s_2, s_3, s_4 \in S$ such that

- $\hat{w}zw \preceq_k^{\mathrm{e}} \hat{w}u_1 zw$,
- $\hat{w}u_1 zw \preceq_k \hat{w}uw$,
- $\delta(s_0, x) = s_1$, $\delta(s_1, u_1) = s_2$,
- $\delta(s_1, z) = \delta(s_3, u) = s_3$,
- $\delta(s_2, z) = \delta(s_4, u) = s_4$ and
- $\delta(s_3, z') \in S' \Leftrightarrow \delta(s_4, z') \notin S'$.

We say that the dfa F has the pattern P''_k (cf. Figure 6) if there exist $w, \hat{w}, x, v, v_1, y, u, u_1, z, z' \in A^*$ with $|w| = |\hat{w}| = k, |v| \geq k, |y| \geq k, |u| \geq k, |u_1 z| \geq k, s_k(v) = \hat{w}, p_k(y) = p_k(u) = p_k(u_1 z) = w$ and if there exist states $s_1, s_2, s_3, s_4, s_5 \in S$ such that

- $\hat{w}yzw \preceq_k^{\mathrm{e}} \hat{w}yv_1 u_1 zw$,

- $\hat{w}yv_1w \preceq_k \hat{w}vw$,
- $\hat{w}u_1zw \preceq_k \hat{w}uw$,
- $\delta(s_0, x) = \delta(s_1, v) = s_1$,
- $\delta(s_1, y) = s_2$, $\delta(s_2, v_1u_1) = s_3$,
- $\delta(s_2, z) = \delta(s_4, u) = s_4$,
- $\delta(s_3, z) = \delta(s_5, u) = s_5$ and
- $\delta(s_4, z') \in S' \Leftrightarrow \delta(s_5, z') \notin S'$.

The technically more involved part is to show conversely that a k-extension chain with an infinite number of alternations with respect to some language L implies the occurence of one of those patterns in any automaton accepting L. For this end we will continuously select suitable infinite subsequences of a given k-extension chain and emphasize on different positions in a word to oversee when an inserted factor leads to alternation. We contrast these infinite events with the finiteness of an automaton. Three major cases have to be distinguished and lead to the previously defined patterns. We will show both implications of the following theorem by contraposition.

Theorem 4.6. *Let F be a dfa and $k \geq 0$. Then the following statements are equivalent:*

(1) $L(F) \in \mathcal{B}_{1,k}$

(2) F does not have any of the patterns P_k, P'_k *and* P''_k

Proof. (2) $\Rightarrow$ (1): Assume that $L(F) \notin \mathcal{B}_{1,k}$ for some dfa $F = (A, S, \delta, s_0, S')$. Then by Theorem 2.10 there exists a k-extension chain $\{w_s\}$ with an infinite number of alternations with respect to $L(F)$. Moreover by Theorem 3.3 we can assume that $w_s \preceq_k^{\mathrm{e}} w_{s+1}$ via some elementary k-extension f_s for all $s \geq 1$. Throughout the proof we will frequently select via a strictly increasing mapping $\tau : \mathbb{N} \mapsto \mathbb{N}$ an infinite subsequence $\{w_{\tau(s)}\}$ of $\{w_s\}$ with certain properties. Any time we do so we redefine w.l.o.g. each w_s as $w_{\tau(s)}$ and adapt all notations. As the result of a first such selection we may assume that we deal with a k-extension chain $\{w_s\}$ such that for all $i \geq 1$

- $w_{2i-1} = w'_i w''_i \preceq_k^{\mathrm{e}} w'_i v_i w''_i = w_{2i} \preceq_k w_{2i+1} = w'_{i+1} w''_{i+1}$,
- $|v_i| \geq 1$,
- $\delta(s_0, w_{2i-1}) \notin S'$ and $\delta(s_0, w_{2i}) \in S'$.

We denote by j_i the index of the first letter of w''_i in w_{2i-1}. So for all $i \geq 1$ we know that $k < j_i \leq |w_{2i-1}| - k + 1$ and it holds that $f_{2i-1}(j_i - 1) = j_i - 1$ and $f_{2i-1}(j_i) = j_i + |v_i|$.

Observe that each letter in a word w_s can be traced via the composition of $f_s, f_{s+1}, \ldots$ to any word following w_s in the sequence. We use this to attach for all $i \geq 1$ markers m_i to the last letter of v_i, i.e. $w_{2i-1} = w'_i w''_i$ has markers $m_1, m_2, \ldots, m_{i-1}$ and $w_{2i} = w'_i v_i w''_i$ has markers $m_1, m_2, \ldots, m_i$. Each marker gives rise to a position (index) in the word in which it appears. Next we inductively attach values $\alpha_i \in \mathbb{R}$ to each marker m_i as follows: Let $\alpha_{i+1} =_{\mathrm{def}} (\beta_{i+1} + \gamma_{i+1})/2$ with $\beta_{i+1} =_{\mathrm{def}} \max\left(\left\{\, \alpha_j \;\middle|\; 1 \leq j \leq i \text{ and marker } m_j \text{ is left to } m_{i+1} \,\right\} \cup \{0\}\right)$ and $\gamma_{i+1} =_{\mathrm{def}} \min\left(\left\{\, \alpha_j \;\middle|\; 1 \leq j \leq i \text{ and marker } m_j \text{ is right to } m_{i+1} \,\right\} \cup \{1\}\right)$. We observe that in any word of the sequence m_i is left to $m_{i'}$ if and only if $\alpha_i < \alpha_{i'}$. Now Lemma 4.5 tells us that there is an infinite strictly monotonic subsequence of $\{\alpha_i\}$, which leads us to the first case.

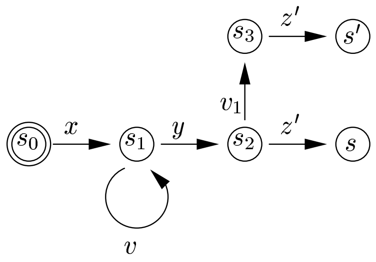

Figure 4: Pattern P_k with $\hat{w}yw \preceq_k^{\mathrm{e}} \hat{w}yv_1w \preceq_k \hat{w}vw$ where $\hat{w} = s_k(v), w = p_k(y) = p_k(z')$ and $s \in S^{'} \Leftrightarrow s' \notin S^{'}$.

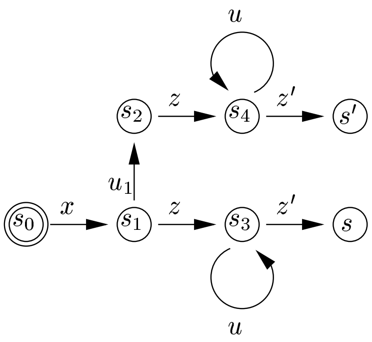

Figure 5: Pattern P'_k with $\hat{w}zw \preceq_k^{\mathrm{e}} \hat{w}u_1zw \preceq_k \hat{w}uw$ where $\hat{w} = s_k(x), w = p_k(z) = p_k(u)$ and $s \in S^{'} \Leftrightarrow s' \notin S^{'}$.

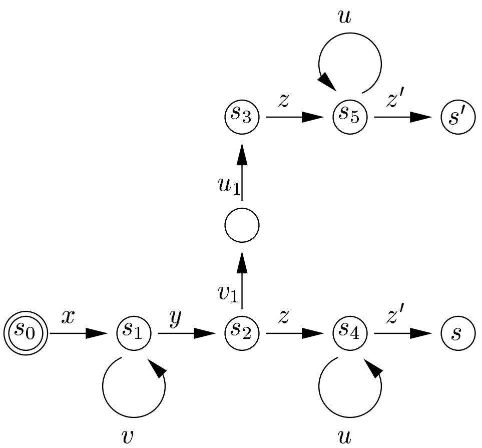

Figure 6: Pattern P''_k with $\hat{w}yzw \preceq_k^{\mathrm{e}} \hat{w}yv_1u_1zw$ and $\hat{w}yv_1w \preceq_k \hat{w}vw$ and $\hat{w}u_1zw \preceq_k \hat{w}uw$ where $\hat{w} = s_k(v), w = p_k(y) = p_k(u) = p_k(u_1z)$ and $s \in S^{'} \Leftrightarrow s' \notin S^{'}$.

Case 1. Assume that there exists an infinite strictly increasing subsequence of $\{\alpha_i\}$. We select this subsequence $\{\alpha_{\tau(i)}\}$ and redefine $w_{2i-1} =_{\mathrm{def}} w_{2\tau(i)-1}$ and $w_{2i} =_{\mathrm{def}} w_{2\tau(i)}$ with all notations. Now in each w_{2i+1} it holds that m_i is strictly left to the new insertion position j_{i+1} (otherwise we would have $\alpha_{\tau(i)} > \alpha_{\tau(i+1)}$ in contradiction to the assumption of case 1). It inductively follows that each w'_{i+1} has from left to right markers $m_1, m_2, \ldots, m_i$.

We make two more assumptions that can be achieved by suitable selections of further subsequences. Since the sequence of markers is strictly increasing we can assume that there are at least k letters between marker m_i and position j_{i+1} in w_{2i+1}. From this it follows that there are also at least k letters between m_i and m_{i+1} in w_{2i+2}. Furthermore we know $|w''_i| \geq k$. So due to the infinite number of positions j_i where insertion leads to alternation we may select a subsequence such that for all i any w'_i (and hence any $w'_i v_i$) is followed by the same k letters (there are only $|A|^k$ different words of length k). Let $w =_{\mathrm{def}} p_k(w''_i)$. If we trace now an arbitrary marker m_i through all words of the sequence in which it appears it is always followed by w since any mapping under a k-extension preserves the k following letters. Note that the positions of any two markers m_i and $m_{i'}$ with $i < i'$ fulfill the properties needed to apply Lemma 4.2 and we will establish partial k-extension relations between such pairs of markers.

We summarize the properties of $\{w_s\}$ obtained here.

- $p_k(w''_i) = w$ for all $i \geq 1$
- each w'_{i+1} has from left to right markers $m_1, m_2, \ldots, m_i$ each followed by w and with a distance of at least k letters between each other

Next we emphasize on the state that F reaches when getting a word of the sequence up to a marker m_i as an input. Due to the finiteness of the set of states we can select an infinite subsequences such that this state is always the same for this marker. Moreover we can establish the partial k-extension relation from a factor between two markers in a word of the subsequence to the factor between the same two markers in a word later in the sequence. Finally the infinite number of alternations allows us to repeat the process of selection such subsequences infinitely often.

Claim 1. For every $l \geq 0$ there exists a state $s_l \in S$ and an infinite subsequence $\{w_{i,l}\}$ of $\{w'_i\}$ such that for all i there are words $v^1_{i,l}, v^2_{i,l}, \ldots, v^l_{i,l}, u_{i,l}, v_{i,l} \in A^*$ with

a. $w_{i,l} = v^1_{i,l} v^2_{i,l} \cdots v^l_{i,l} u_{i,l}$,

b. $v^j_{i,l} \preceq^{\mathrm{part}}_k v^j_{i+1,l}$ for $1 \leq j \leq l$,

c. $u_{1,l-1} v_{1,l-1} \preceq^{\mathrm{part}}_k v^l_{i,l}$ for $l \geq 1$,

d. there is a prefix $u'_{i+1,l}$ of $u_{i+1,l}$ such that $u_{i,l} v_{i,l} \preceq^{\mathrm{part}}_k u'_{i+1,l}$ and

e. $\delta(s_0, v^1_{i,l} v^2_{i,l} \cdots v^j_{i,l}) = s_j$ for $1 \leq j \leq l$.

Proof of Claim 1. We proceed by induction on l and prove the induction base for $l = 0$ and $l = 1$. So let $l = 0$ and set $u_{i,0} =_{\mathrm{def}} w'_i$ and $v_{i,0} =_{\mathrm{def}} v_i$. We only have to show d. Fix some i and let l_i be the index of marker m_i in $w'_{i+1} = u_{i+1,0}$. We set $u'_{i+1,0} =_{\mathrm{def}} p_{l_i}(u_{i+1,0})$. One can straightforward verify with Definition 4.1 that $u_{i,0} v_{i,0} \preceq^{\mathrm{part}}_k u'_{i+1,0}$.

Now let $l = 1$. Particularly we have $u_{1,0} v_{1,0} \preceq^{\mathrm{part}}_k u'_{2,0}$ and marker m_1 is attached to the last letter in both words. If we follow this marker for $i \geq 2$ in any $u_{i,0}$ we can identify words $v^1_{i,0}$ and $\hat{u}_{i,0}$ such that $u_{i,0} = v^1_{i,0} \hat{u}_{i,0}$ and m_1 is attached to any last letter of $v^1_{i,0}$. So we can verify $v^1_{i,0} \preceq^{\mathrm{part}}_k v^1_{i+1,0}$. Moreover we see that $u_{1,0} v_{1,0} \preceq^{\mathrm{part}}_k v^1_{i,0}$ since $v^1_{2,0} = u'_{2,0}$. For the prefix $\hat{u}'_{i+1,0}$

of $\hat{u}_{i+1,0}$ that begins after m_1 and includes all letters of $\hat{u}_{i+1,0}$ up to marker m_i we have by Lemma 4.2 that $\hat{u}_{i,0}v_{i,0} \preceq_k^{\text{part}} \hat{u}'_{i+1,0}$. Due to the finiteness of the set of states there is a state $s_1 \in S$ such that we can select a subsequence via some mapping $\tau : \mathbb{N} \mapsto \mathbb{N}$ with $\delta(s_0, v^1_{\tau(i),0}) = s_1$ for all $i \geq 1$. Now all assertions of the claim can be verified if we set $w_{i,1} =_{\text{def}} w_{\tau(i),0}$, $v^1_{i,1} =_{\text{def}} v^1_{\tau(i),0}$, $u_{i,1} =_{\text{def}} \hat{u}_{\tau(i),0}$, $v_{i,1} =_{\text{def}} v_{\tau(i),0}$ and $u'_{i+1,1} =_{\text{def}} \hat{u}'_{\tau(i+1),0}$. Note that the partial k-extension relations still hold after selecting this subsequence.

We repeat the argumentation of the case $l = 1$ to show the induction step and split successively the words $u_{i,l}$ into words $v^{l+1}_{i,l+1}u_{i,l+1}$ with the required properties. Assume that the claim holds for some $l \geq 1$. Particularly by hypothesis d. we have $u_{1,l}v_{1,l} \preceq_k^{\text{part}} u'_{2,l}$ and marker m_1 (the index '1' of this marker is due to the previous redefinition) is attached to the last letter of $u'_{2,l}$ by construction. If we follow this marker for $i \geq 2$ in any $u_{i,l}$ we can again identify words $v^{l+1}_{i,l}$ and $\hat{u}_{i,l}$ such that $u_{i,l} = v^{l+1}_{i,l}\hat{u}_{i,l}$ and $v^{l+1}_{i,l} \preceq_k^{\text{part}} v^{l+1}_{i+1,l}$ where m_1 is attached to each last letter. Note that by hypothesis b. we already know that $v^j_{i,l} \preceq_k^{\text{part}} v^j_{i+1,l}$ for $1 \leq j \leq l$ and we can write each $w_{i,l}$ as $w_{i,l} = v^1_{i,l}v^2_{i,l}\cdots v^l_{i,l}v^{l+1}_{i,l}\hat{u}_{i,l}$. Moreover we see that $u_{1,l}v_{1,l} \preceq_k^{\text{part}} v^{l+1}_{i,l}$ since $v^{l+1}_{2,l} = u'_{2,l}$. For the prefix $\hat{u}'_{i+1,l}$ of $\hat{u}_{i+1,l}$ up to marker m_i we have by Lemma 4.2 that $\hat{u}_{i,l}v_{i,l} \preceq_k^{\text{part}} \hat{u}'_{i+1,l}$.

Due to the finiteness of the set of states there is a state $s_{l+1} \in S$ such that we can select a subsequence via some mapping $\tau : \mathbb{N} \mapsto \mathbb{N}$ with $\delta(s_0, v^1_{\tau(i),l}v^2_{\tau(i),l}\cdots v^l_{\tau(i),l}v^{l+1}_{\tau(i),l}) = s_{l+1}$ for all $i \geq 1$. Again all assertions of the claim can be verified if we set $w_{i,l+1} =_{\text{def}} w_{\tau(i),l}$, $v_{i,l+1} =_{\text{def}} v_{\tau(i),l}$, $v^j_{i,l+1} =_{\text{def}} v^j_{\tau(i),l}$ for $1 \leq j \leq l+1$, $u_{i,l+1} =_{\text{def}} \hat{u}_{\tau(i),l}$ and $u'_{i+1,l+1} =_{\text{def}} \hat{u}'_{\tau(i+1),l}$. *(End proof of Claim 1)*

We keep the notations of the previous claim. Now, again due to the finiteness of S there exist p, q with $1 \leq p < q \leq |F| + 1$ and $s_p = s_q$. Hence we can define

- $x =_{\text{def}} v^1_{1,q-1}v^2_{1,q-1}\cdots v^p_{1,q-1}$,
- $y =_{\text{def}} v^{p+1}_{1,q-1}\cdots v^{q-1}_{1,q-1}u_{1,q-1}$,
- $v =_{\text{def}} v^{p+1}_{1,q}\cdots v^{q-1}_{1,q}v^q_{1,q}$,
- $v_1 =_{\text{def}} v_{1,q-1}$ and
- $\hat{w} =_{\text{def}} s_k(v)$.

Let r be an index such that $w_{1,q-1} = w'_r$. Then we set $z' =_{\text{def}} w''_r$. First note that $p_k(y) = p_k(z') = w$ since w follows any marker m_i and that $s_k(v) = \hat{w}$ by definition. Moreover we have $\delta(s_0, xyz') \notin S'$, $\delta(s_0, xyv_1z') \in S'$, and $\delta(s_0, x) = s_p = s_q = \delta(s_0, xv)$. Since $|y| \geq k$ we can state $yw \preceq_k^{\text{e}} yv_1w$ and $\hat{w}yw \preceq_k^{\text{e}} \hat{w}yv_1w$ follows immediately. On the other hand from assertions b. and c. of Claim 1 can be concluded that $v^j_{1,q-1} \preceq_k^{\text{part}} v^j_{\tau(1),q-1}$ for $p+1 \leq j \leq q-1$, where the latter word is equal to $v^j_{1,q}$ due to our redefinitions. In the same way we get $u_{1,q-1}v_{1,q-1} \preceq_k^{\text{part}} v^q_{1,q}$. Now with Lemma 4.4 we can put these pieces together and obtain $yv_1 \preceq_k^{\text{part}} v$ and thus by definition of the partial k-extension relation $\tilde{w}yv_1w \preceq_k \tilde{w}vw$ via some k-extension f and with $\tilde{w} =_{\text{def}} s_k(x)$. Finally note that we can exchange $\tilde{w}$ with $\hat{w}$ and obtain $\hat{w}yv_1w \preceq_k \hat{w}vw$ via f as well. Hence we found pattern P_k in F. *(End of Case 1)*

From now on we assume that there exists an infinite strictly decreasing subsequence of $\{\alpha_i\}$. We select this subsequence $\{\alpha_{\tau(i)}\}$ and redefine $w_{2i-1} =_{\text{def}} w_{2\tau(i)-1}$ and $w_{2i} =_{\text{def}} w_{2\tau(i)}$ with all notations as before. Now in each w_{2i+1} it holds that m_i is strictly left to m_{i-1} and j_{i+1} is less than or equal to m_i (otherwise we would have $\alpha_{\tau(i)} > \alpha_{\tau(i+1)}$ in contradiction to the actual assumption). It inductively follows that each w''_{i+1} has from right to left markers $m_1, m_2, \ldots, m_i$. But we can not assure that the new insertion position j_{i+1} is left to the mapping of the leftmost letter of v_i under f_{2i}.

In order to obtain a systematic relation between j_{i+1} and the letters of v_i in w_{2i+1} we repeat the procedure of setting markers, but this time we place them on the first letter of v_i. So we attach for all $i \geq 1$ markers n_i to the first letter of v_i, i.e. $w_{2i-1} = w'_i w''_i$ has in addition to the strictly decreasing positions of the markers $m_1, m_2, \ldots, m_{i-1}$ also the markers $n_1, n_2, \ldots, n_{i-1}$ and $w_{2i} = w'_i v_i w''_i$ has the additional markers $n_1, n_2, \ldots, n_i$. Note that for any i it holds that the position of n_i in any word is less or equal to the position of m_i. We did not exclude the case that n_i and m_i mark the same letter. Again we inductively attach values $\alpha'_i \in \mathbb{R}$ to each marker n_i as before: Let $\alpha'_{i+1} =_{\text{def}} (\beta_{i+1} + \gamma_{i+1})/2$ with $\beta_{i+1} =_{\text{def}} \max\left(\{\, \alpha'_j \mid 1 \leq j \leq i \text{ and marker } n_j \text{ is left to } n_{i+1} \,\} \cup \{0\}\right)$ and $\gamma_{i+1} =_{\text{def}} \min\left(\{\, \alpha'_j \mid 1 \leq j \leq i \text{ and marker } n_j \text{ is right to } n_{i+1} \,\} \cup \{1\}\right)$. Here we can also observe that in any word of the sequence n_i is left to $n_{i'}$ if and only if $\alpha'_i < \alpha'_{i'}$. We apply Lemma 4.5 again and proceed to the second case.

Case 2. Assume that there exists an infinite strictly decreasing subsequence of $\{\alpha'_i\}$. We select this subsequence $\{\alpha'_{\tau(i)}\}$ and redefine $w_{2i-1} =_{\text{def}} w_{2\tau(i)-1}$ and $w_{2i} =_{\text{def}} w_{2\tau(i)}$ with all notations. Now in each w_{2i+1} it holds that n_i is right or equal to the new insertion position j_{i+1} (otherwise we would have $\alpha'_{\tau(i)} < \alpha'_{\tau(i+1)}$ in contradiction to the assumption of case 2). It inductively follows that each w''_{i+1} has from right to left markers $n_1, n_2, \ldots, n_i$. Here in Case 2 we are not interested in the positions of the markers $m_1, m_2, \ldots, m_i$.

Again we make two more assumptions as in Case 1 that can be achieved by suitable selections of further subsequences. Since the sequence of markers is strictly decreasing we can assume that there are at least k letters between position j_{i+1} and marker n_i in w_{2i+1}. From this it follows that there are also at least k letters between n_{i+1} and n_i in w_{2i+2}. Furthermore we know $|w''_i| \geq k$. So again due to the infinite number of positions j_i we may select a subsequence such that for all i any w'_i (and hence any $w'_i v_i$) is followed by the same k letters. Let $w =_{\text{def}} p_k(w''_i)$. Here we will use that $p_k(v_i w''_i) = w$ since $f_{2i-1}(j_i - 1) = j_i - 1$ and the k following letters in w_{2i-1} and w_{2i} are w. It follows that n_i is placed on the first letter of w. Now if we trace an arbitrary marker n_i through all words of the sequence in which it appears it marks always the first letter of a factor w since any mapping under a k-extension preserves even the k following, and thus in particular the $k-1$ following letters. Note that this time the positions of any two markers n_i and $n_{i'}$ with $i > i'$ fulfill the properties needed to apply Lemma 4.3 and we will establish with the help of this lemma partial k-extension relations between such pairs of markers.

We summarize the properties of $\{w_s\}$ obtained here.

- $p_k(w''_i) = w$ for all $i \geq 1$
- each w''_{i+1} has from right to left markers $n_1, n_2, \ldots, n_i$ each attached to the first letter of a factor w and with a distance of at least k letters between each other

In order to emphasize again on the state that F reaches when getting as an input a word of the sequence ending before a marker n_i, we do in the following claim essentially the same as in Claim 1. Only that we split words in factors from right to left with respect to the descending marker positions and that we have to remember two states since the insertion position is 'always left'.

Claim 2. For every $l \geq 0$ there exist states $r_l, t_l \in S$ and an infinite subsequence $\{w_{i,l}\}$ of $\{w''_i\}$ such that for all i there are words $x_{i,l}$, $v_{i,l}$, $u_{i,l}$, $v^l_{i,l}, \ldots, v^2_{i,l}, v^1_{i,l} \in A^*$ with

a. $w_{i,l} = u_{i,l} v_{i,l}^{l} \cdots v_{i,l}^{2} v_{i,l}^{1}$,

b. $v_{i,l}^{j} \preceq_k^{\mathrm{part}} v_{i+1,l}^{j}$ for $1 \leq j \leq l$,

c. $v_{1,l-1} u_{1,l-1} \preceq_k^{\mathrm{part}} v_{i,l}^{l}$ for $l \geq 1$,

d. there is a suffix $u'_{i+1,l}$ of $u_{i+1,l}$ such that $v_{i,l} u_{i,l} \preceq_k^{\mathrm{part}} u'_{i+1,l}$,

e. $\delta(s_0, x_{i,l} u_{i,l} v_{i,l}^{l} v_{i,l}^{l-1} \cdots v_{i,l}^{j+1}) = r_j$ for $1 \leq j \leq l-1$ and $\delta(s_0, x_{i,l} u_{i,l}) = r_l$, and

f. $\delta(s_0, x_{i,l} v_{i,l} u_{i,l} v_{i,l}^{l} v_{i,l}^{l-1} \cdots v_{i,l}^{j+1}) = t_j$ for $1 \leq j \leq l-1$ and $\delta(s_0, x_{i,l} v_{i,l} u_{i,l}) = t_l$.

Proof of Claim 2. We proceed by induction on l and prove the induction base for $l = 0$ and $l = 1$. So let $l = 0$ and set $u_{i,0} =_{\mathrm{def}} w''_i$, $v_{i,0} =_{\mathrm{def}} v_i$ and $x_{i,0} =_{\mathrm{def}} w'_i$. We only have to show d, e and f. Fix some i and let l_i be the index of marker n_i in $w''_{i+1} = u_{i+1,0}$. We set $u'_{i+1,0} =_{\mathrm{def}} s_{l'}(u_{i+1,0})$ with $l' =_{\mathrm{def}} |w''_{i+1}| - l_i + 1$. One can straightforward verify with Definition 4.1 that $u_{i,0} v_{i,0} \preceq_k^{\mathrm{part}} u'_{i+1,0}$. Due to the finiteness of the set of states there are states $r_0, t_0 \in S$ such that we can select a subsequence via some mapping $\tau : \mathbb{N} \mapsto \mathbb{N}$ with $\delta(s_0, x_{\tau(i),0} u_{\tau(i),0}) = r_0$ and $\delta(s_0, x_{\tau(i),0} v_{\tau(i),0} u_{\tau(i),0}) = t_0$ for all $i \geq 1$.

Now let $l = 1$. Particularly we have $v_{1,0} u_{1,0} \preceq_k^{\mathrm{part}} u'_{2,0}$ and marker n_1 is attached to the first letter in both words. If we follow this marker for $i \geq 2$ in any $u_{i,0}$ we can identify words $\hat{u}_{i,0}$ and $v_{i,0}^{1}$ such that $u_{i,0} = \hat{u}_{i,0} v_{i,0}^{1}$ and n_1 is attached to any first letter of $v_{i,0}^{1}$. So we can verify $v_{i,0}^{1} \preceq_k^{\mathrm{part}} v_{i+1,0}^{1}$. Moreover we see that $v_{1,0} u_{1,0} \preceq_k^{\mathrm{part}} v_{i,0}^{1}$ since $v_{2,0}^{1} = u'_{2,0}$. For the suffix $\hat{u}'_{i+1,0}$ of $\hat{u}_{i+1,0}$ that begins with n_i and includes all letters of $\hat{u}_{i+1,0}$ before marker n_1 we have by Lemma 4.3 that $v_{i,0} \hat{u}_{i,0} \preceq_k^{\mathrm{part}} \hat{u}'_{i+1,0}$.

Due to the finiteness of the set of states there are states $r_1, t_1 \in S$ such that we can select a subsequence via some mapping $\tau : \mathbb{N} \mapsto \mathbb{N}$ with $\delta(s_0, x_{\tau(i),0} \hat{u}_{\tau(i),0}) = r_1$ and $\delta(s_0, x_{\tau(i),0} v_{\tau(i),0} \hat{u}_{\tau(i),0}) = t_1$ for all $i \geq 1$. Now all assertions of the claim can be verified if we set $w_{i,1} =_{\mathrm{def}} w_{\tau(i),0}$, $v_{i,1}^{1} =_{\mathrm{def}} v_{\tau(i),0}^{1}$, $u_{i,1} =_{\mathrm{def}} \hat{u}_{\tau(i),0}$, $x_{i,1} =_{\mathrm{def}} x_{\tau(i),0}$, $v_{i,1} =_{\mathrm{def}} v_{\tau(i),0}$ and $u'_{i+1,1} =_{\mathrm{def}} \hat{u}'_{\tau(i+1),0}$.

We repeat the argumentation of the case $l = 1$ to show the induction step and split successively the words $u_{i,l}$ into words $u_{i,l+1} v_{i,l+1}^{l+1}$ with the required properties. Assume that the claim holds for some $l \geq 1$. Particularly by hypothesis d. we have $v_{1,l} u_{1,l} \preceq_k^{\mathrm{part}} u'_{2,l}$ and marker n_1 (the index '1' of this marker is due to the previous redefinition) is attached to the first letter of $u'_{2,l}$ by construction. If we follow this marker for $i \geq 2$ in any $u_{i,l}$ we can again identify words $\hat{u}_{i,l}$ and $v_{i,l}^{l+1}$ such that $u_{i,l} = \hat{u}_{i,l} v_{i,l}^{l+1}$ and $v_{i,l}^{l+1} \preceq_k^{\mathrm{part}} v_{i+1,l}^{l+1}$ where n_1 is attached to each first letter. Note that by hypothesis b. we already know that $v_{i,l}^{j} \preceq_k^{\mathrm{part}} v_{i+1,l}^{j}$ for $1 \leq j \leq l$ and we can write each $w_{i,l}$ as $w_{i,l} = \hat{u}_{i,l} v_{i,l}^{l+1} v_{i,l}^{l} \cdots v_{i,l}^{2} v_{i,l}^{1}$. Moreover we see that $v_{1,l} u_{1,l} \preceq_k^{\mathrm{part}} v_{i,l}^{l+1}$ since $v_{2,l}^{l+1} = u'_{2,l}$.

For the suffix $\hat{u}'_{i+1,l}$ of $\hat{u}_{i+1,l}$ that begins with n_i and includes all letters of $\hat{u}_{i+1,l}$ before marker n_1 we have by Lemma 4.3 that $v_{i,l} \hat{u}_{i,l} \preceq_k^{\mathrm{part}} \hat{u}'_{i+1,l}$.

Due to the finiteness of the set of states there are states $r_{l+1}, t_{l+1} \in S$ such that we can select a subsequence via some mapping $\tau : \mathbb{N} \mapsto \mathbb{N}$ with $\delta(s_0, x_{\tau(i),l} \hat{u}_{\tau(i),l}) = r_{l+1}$ and $\delta(s_0, x_{\tau(i),l} v_{\tau(i),l} \hat{u}_{\tau(i),l}) = t_{l+1}$ for all $i \geq 1$. Note that $\delta(s_0, x_{\tau(i),l} \hat{u}_{\tau(i),l} v_{\tau(i),l}^{l+1}) = r_l$ and $\delta(s_0, x_{\tau(i),l} v_{\tau(i),l} \hat{u}_{\tau(i),l} v_{\tau(i),l}^{l+1}) = t_l$ by hypothesis. Again all assertions of the claim can be verified if we set $w_{i,l+1} =_{\mathrm{def}} w_{\tau(i),l}$, $v_{i,l+1}^{l+1} =_{\mathrm{def}} v_{\tau(i),l}^{l+1}$, $u_{i,l+1} =_{\mathrm{def}} \hat{u}_{\tau(i),l}$, $x_{i,l+1} =_{\mathrm{def}} x_{\tau(i),l}$, $v_{i,l+1} =_{\mathrm{def}} v_{\tau(i),l}$ and $u'_{i+1,l+1} =_{\mathrm{def}} \hat{u}'_{\tau(i+1),l}$. *(End proof of Claim 2)*

We keep the notations of the previous claim. Now, again due to the finiteness of S there exist p, q with $1 \leq p < q \leq (|F| + 1)^2$ and $(r_p, t_p) = (r_q, t_q)$. We define

- $x =_{\text{def}} x_{1,q-1}$,
- $u_1 =_{\text{def}} v_{1,q-1}$,
- $z =_{\text{def}} u_{1,q-1} v_{1,q-1}^{q-1} \cdots v_{1,q-1}^{p+1}$,
- $z' =_{\text{def}} v_{1,q-1}^{p} \cdots v_{1,q-1}^{2} v_{1,q-1}^{1}$,
- $u =_{\text{def}} v_{1,q}^{q} v_{1,q}^{q-1} \cdots v_{1,q}^{p+1}$ and
- $\hat{w} =_{\text{def}} s_k(x)$.

Note that $p_k(z) = p_k(u) = w$ since w starts at each marker n_i and that $s_k(x) = \hat{w}$ with $|x| \geq k$. Moreover we have $\delta(s_0, xzz') \notin S'$, $\delta(s_0, xu_1zz') \in S'$, $\delta(s_0, xz) = r_p = r_q = \delta(s_0, xzu)$ and $\delta(s_0, xu_1z) = t_p = t_q = \delta(s_0, xu_1zu)$. By the choice of the sequence and since $|x|, |z'| \geq k$ we have $\hat{w}zw \preceq_k^{\text{e}} \hat{w}u_1zw$. On the other hand from assertions b. and c. of Claim 2 can be concluded that $v_{1,q-1}u_{1,q-1} \preceq_k^{\text{part}} v_{1,q}^{q}$ and that $v_{1,q-1}^{j} \preceq_k^{\text{part}} v_{1,q}^{j}$ for $q-1 \leq j \leq p+1$. With Lemma 4.4 we put these pieces together and obtain $u_1z \preceq_k^{\text{part}} u$ and thus by definition of the partial k-extension relation $\hat{w}u_1zw \preceq_k \hat{w}uw$. Hence we found pattern P'_k in F. *(End of Case 2)*

We are left over with a strictly decreasing subsequence of $\{\alpha_i\}$ which we already selected before Case 2 to obtain a strictly decreasing sequence of markers m_i placed on the last letter of each v_i. We then additionally marked each first letter of v_i with n_i and treated the case of a strictly decreasing subsequence of $\{\alpha'_i\}$. Now we have to cope with a strictly increasing subsequence of $\{\alpha'_i\}$. So we select this subsequence as before and have the situation of a strictly decreasing sequence of markers m_i together with a strictly increasing sequence of markers n_i. In the previous cases we learned how to deal with each of these sequences and combine now both techniques in the third case.

Case 3. We may assume that for all i we have $|v_i| \geq 2$. Otherwise for infinitely many i markers m_i and n_i are attached to the same letter and both possibilities of an increasing or decreasing subsequence are treated in Case 1 and 2, respectively. We observe that for each i the marker n_i is mapped via f_{2i} to a position less or equal to $j_{i+1} - 1$ in w_{2i+1}, which means that it is mapped to w'_{i+1}, and the marker m_i is mapped via f_{2i} to a position greater or equal to j_{i+1}in w_{2i+1}, which means that it is mapped to w''_{i+1} (otherwise we would have $\alpha'_{i+1} < \alpha'_i$ or $\alpha_{i+1} > \alpha_i$ in contradiction to our selected subsequences). Thus for each i there are two nonempty words v'_i and v''_i with $v_i = v'_i v''_i$ such that each letter of v'_i is mapped to w'_{i+1} and each letter of v''_i is mapped to w''_{i+1}.

We introduce new markers to v'_i and v''_i and trace their positions in our sequence. The last letter of v'_i is marked by n'_i and to the first letter of v''_i we attach a marker m'_i. So both are placed next to each other and are mapped to different parts of the following word w_{2i+1}, i.e. w'_{i+1} has from left to right markers $n_1, n'_1, n_2, n'_2, \ldots, n_i, n'_i$ and w''_{i+1} has from right to left markers $m_1, m'_1, m_2, m'_2, \ldots, m_i, m'_i$. Note that by the strictness of the positions of markers n_i and m_i we are able to select a subsequence such that there are in any word at least k letters between two markers n'_i and n'_{i+1} on one hand, and between m'_i and m'_{i+1} on the other hand. From now on we will no longer be concerned with n_i and m_i. We make another assumption and achieve it as well by a suitable selection of a subsequence. If we look at the infinite number of words $v'_i v''_i w''_i$ we may suppose that each word $v_i''w_i''$ starts with the same k letters. Set $w =_{\text{def}} p_k(v_i''w_i'')$. Now each marker n_i' is followed in any word by w and each m_i' is attached to the first letter of a factor w in each word of the sequence from w_{2i} onwards.

We summarize the properties of $\{w_s\}$ obtained here.

- $p_k(v_i''w_i'') = w$ for all $i \geq 1$
- each w_{i+1}' has from left to right markers $n_1', n_2', \ldots, n_i'$ each followed by a factor w and with a distance of at least k letters between each other
- each w_{i+1}'' has from right to left markers $m_1', m_2', \ldots, m_i'$ each attached to the first letter of a factor w and with a distance of at least k letters between each other

Here each insertion of a word v_i leads to a significant tripel of states that appears infinitely often in the remaining sequence. We split words in factors from right to left and also from left to right since the insertion position is 'always in the middle'.

Claim 3. For every $l \geq 0$ there exist states $s_l, r_l, t_l \in S$ and an infinite subsequence $\{w_{i,l}\}$ of $\{w_i'w_i''\}$ such that for all i there are words $v_{i,l}^1$, $v_{i,l}^2$, $\ldots$, $v_{i,l}^l$, $u_{i,l}$, $v_{i,l}'$, $v_{i,l}''$, $o_{i,l}$, $o_{i,l}^l$, $\ldots$, $o_{i,l}^2$, $o_{i,l}^1 \in A^*$ with

a. $w_{i,l} = v_{i,l}^1 v_{i,l}^2 \ldots v_{i,l}^l u_{i,l} o_{i,l} o_{i,l}^l \ldots o_{i,l}^2 o_{i,l}^1$,

b. $v_{i,l}^j \preceq_k^{\text{part}} v_{i+1,l}^j$ for $1 \leq j \leq l$ and $o_{i,l}^j \preceq_k^{\text{part}} o_{i+1,l}^j$ for $1 \leq j \leq l$,

c. $u_{1,l-1}v_{1,l-1}' \preceq_k^{\text{part}} v_{i,l}^l$ and $v_{1,l-1}''o_{1,l-1} \preceq_k^{\text{part}} o_{i,l}^l$ for $l \geq 1$,

d. there is a prefix $u_{i+1,l}'$ of $u_{i+1,l}$ such that $u_{i,l}v_{i,l}' \preceq_k^{\text{part}} u_{i+1,l}'$ and there is a suffix $o_{i+1,l}'$ of $o_{i+1,l}$ such that $v_{i,l}''o_{i,l} \preceq_k^{\text{part}} o_{i+1,l}'$,

e. $\delta(s_0, v_{i,l}^1 v_{i,l}^2 \cdots v_{i,l}^j) = s_j$ for $1 \leq j \leq l$,

f. $\delta(s_0, v_{i,l}^1 \ldots v_{i,l}^l u_{i,l} o_{i,l} o_{i,l}^l o_{i,l}^{l-1} \cdots o_{i,l}^{j+1}) = r_j$ for $1 \leq j \leq l-1$ and $\delta(s_0, v_{i,l}^1 \ldots v_{i,l}^l u_{i,l} o_{i,l}) = r_l$,

g. $\delta(s_0, v_{i,l}^1 \ldots v_{i,l}^l u_{i,l} v_{i,l}' v_{i,l}'' o_{i,l} o_{i,l}^l o_{i,l}^{l-1} \cdots o_{i,l}^{j+1}) = t_j$ for $1 \leq j \leq l-1$ and $\delta(s_0, v_{i,l}^1 \ldots v_{i,l}^l u_{i,l} v_{i,l}' v_{i,l}'' o_{i,l}) = t_l$.

Proof of Claim 3. We proceed by induction on l and prove the induction base for $l = 0$ and $l = 1$. So let $l = 0$ and set $u_{i,0} =_{\text{def}} w_i', v_{i,0}' =_{\text{def}} v_i', v_{i,0}'' =_{\text{def}} v_i''$ and $o_{i,0} =_{\text{def}} w_i''$. We only have to show d, f and g. Fix some i and let l_i^n (l_i^m) be the index of marker n_i' in $w_{i+1}' = u_{i+1,0}$ (index of marker m_i' in $w_{i+1}'' = o_{i+1,0}$, respectively). We set $u_{i+1,0}' =_{\text{def}} p_{l_i^n}(u_{i+1,0})$ and $o_{i+1,0}' =_{\text{def}} s_{l'}(o_{i+1,0})$ with $l' =_{\text{def}} |w_{i+1}'w_{i+1}''| - l_i^m + 1$. One can straightforward verify with Definition 4.1 that $u_{i,0}v_{i,0}' \preceq_k^{\text{part}} u_{i+1,0}'$ and $v_{i,0}''o_{i,0} \preceq_k^{\text{part}} o_{i+1,0}'$. Again due to the finiteness of the set of states we find states $s_0, r_0, t_0 \in S$ with the desired property by selecting a subsequence.

Now let $l = 1$. Particularly we have on one hand $u_{1,0}v_{1,0}' \preceq_k^{\text{part}} u_{2,0}'$ with marker n_1' attached to the last letter of both words and on the other hand $v_{1,0}''o_{1,0} \preceq_k^{\text{part}} o_{2,0}'$ with marker m_1' attached to the first letter of both words. If we follow marker n_1' for $i \geq 2$ in any $u_{i,0}$ we can identify words $v_{i,0}^1$ and $\hat{u}_{i,0}$ such that $u_{i,0} = v_{i,0}^1 \hat{u}_{i,0}$ and n_1' is attached to any last letter of $v_{i,0}^1$. So we can verify $v_{i,0}^1 \preceq_k^{\text{part}} v_{i+1,0}^1$. Moreover we see that $u_{1,0}v_{1,0}' \preceq_k^{\text{part}} v_{i,0}^1$ since $v_{2,0}^1 = u_{2,0}'$. If we follow marker m'_1 for $i \geq 2$ in any $o_{i,0}$ we can identify words $\hat{o}_{i,0}$ and $o^1_{i,0}$ such that $o_{i,0} = \hat{o}_{i,0} o^1_{i,0}$ and m'_1 is attached to any first letter of $o^1_{i,0}$. So we can verify $o^1_{i,0} \preceq_k^{\text{part}} o^1_{i+1,0}$. Moreover we see that $v''_{1,0} o_{1,0} \preceq_k^{\text{part}} o^1_{i,0}$ since $o^1_{2,0} = o'_{2,0}$.

For the prefix $\hat{u}'_{i+1,0}$ of $\hat{u}_{i+1,0}$ that begins after n'_1 and includes all letters of $\hat{u}_{i+1,0}$ including marker n'_i we have by Lemma 4.2 that $\hat{u}_{i,0} v'_{i,0} \preceq_k^{\text{part}} \hat{u}'_{i+1,0}$. For the suffix $\hat{o}'_{i+1,0}$ of $\hat{o}_{i+1,0}$ that begins with m'_i and includes all letters of $\hat{o}_{i+1,0}$ before marker m'_1 we have by Lemma 4.3 that $v''_{i,0} \hat{o}_{i,0} \preceq_k^{\text{part}} \hat{o}'_{i+1,0}$.

Due to the finiteness of the set of states there are states $s_1, r_1, t_1 \in S$ such that we can select a subsequence via some mapping $\tau : \mathbb{N} \mapsto \mathbb{N}$ with $\delta(s_0, v^1_{\tau(i),0}) = s_1$, $\delta(s_0, v^1_{\tau(i),0} \hat{u}_{\tau(i),0} \hat{o}_{\tau(i),0}) = r_1$ and $\delta(s_0, v^1_{\tau(i),0} \hat{u}_{\tau(i),0} v'_{\tau(i),0} v''_{\tau(i),0} \hat{o}_{\tau(i),0}) = t_1$ for all $i \geq 1$. Now all assertions of the claim can be verified if we set $w_{i,1} =_{\text{def}} w_{\tau(i),0}$, $v^1_{i,1} =_{\text{def}} v^1_{\tau(i),0}$, $o^1_{i,1} =_{\text{def}} o^1_{\tau(i),0}$, $u_{i,1} =_{\text{def}} \hat{u}_{\tau(i),0}$, $o_{i,1} =_{\text{def}} \hat{o}_{\tau(i),0}$, $v'_{i,1} =_{\text{def}} v'_{\tau(i),0}$, $v''_{i,1} =_{\text{def}} v''_{\tau(i),0}$, $u'_{i+1,1} =_{\text{def}} \hat{u}'_{\tau(i+1),0}$ and $o'_{i+1,1} =_{\text{def}} \hat{o}'_{\tau(i+1),0}$.

We repeat the argumentation of the case $l = 1$ to show the induction step and split successively the words $u_{i,l}$ into words $v^{l+1}_{i,l+1} u_{i,l+1}$ and the words $o_{i,l}$ into words $o_{i,l+1} o^{l+1}_{i,l+1}$ with the required properties. Assume that the claim holds for some $l \geq 1$.

Particularly by hypothesis d. we have $u_{1,l} v'_{1,l} \preceq_k^{\text{part}} u'_{2,l}$ and marker n'_1 (recall that the index '1' of this marker is due to the previous redefinition) is attached to the last letter of $u'_{2,l}$ by construction. If we follow this marker for $i \geq 2$ in any $u_{i,l}$ we can again identify words $v^{l+1}_{i,l}$ and $\hat{u}_{i,l}$ such that $u_{i,l} = v^{l+1}_{i,l} \hat{u}_{i,l}$ and $v^{l+1}_{i,l} \preceq_k^{\text{part}} v^{l+1}_{i+1,l}$ where n'_1 is attached to each last letter. Note that by hypothesis b. we already know that $v^j_{i,l} \preceq_k^{\text{part}} v^j_{i+1,l}$ for $1 \leq j \leq l$. Moreover we see that $u_{1,l} v'_{1,l} \preceq_k^{\text{part}} v^{l+1}_{i,l}$ since $v^{l+1}_{2,l} = u'_{2,l}$.

Also by hypothesis d. we have $v''_{1,l} o_{1,l} \preceq_k^{\text{part}} o'_{2,l}$ and marker m'_1 is attached to the first letter of $o'_{2,l}$ by construction. If we follow now this marker for $i \geq 2$ in any $o_{i,l}$ we can again identify words $\hat{o}_{i,l}$ and $o^{l+1}_{i,l}$ such that $o_{i,l} = \hat{o}_{i,l} o^{l+1}_{i,l}$ and $o^{l+1}_{i,l} \preceq_k^{\text{part}} o^{l+1}_{i+1,l}$ where m'_1 is attached to each first letter. By hypothesis b. we already know that $o^j_{i,l} \preceq_k^{\text{part}} o^j_{i+1,l}$ for $1 \leq j \leq l$ and we see that $v''_{1,l} o_{1,l} \preceq_k^{\text{part}} o^{l+1}_{i,l}$ since $o^{l+1}_{2,l} = o'_{2,l}$. We can write each $w_{i,l}$ as $w_{i,l} = v^1_{i,l} v^2_{i,l} \cdots v^l_{i,l} v^{l+1}_{i,l} \hat{u}_{i,l} \hat{o}_{i,l} o^{l+1}_{i,l} o^l_{i,l} \cdots o^2_{i,l} o^1_{i,l}$.

For the prefix $\hat{u}'_{i+1,l}$ of $\hat{u}_{i+1,l}$ that begins after n'_1 and includes all letters of $\hat{u}_{i+1,l}$ including marker n'_i we have by Lemma 4.2 that $\hat{u}_{i,l} v'_{i,l} \preceq_k^{\text{part}} \hat{u}'_{i+1,l}$. For the suffix $\hat{o}'_{i+1,l}$ of $\hat{o}_{i+1,l}$ that begins with m'_i and includes all letters of $\hat{o}_{i+1,l}$ before marker m'_1 we have by Lemma 4.3 that $v''_{i,l} \hat{o}_{i,l} \preceq_k^{\text{part}} \hat{o}'_{i+1,l}$.

Due to the finiteness of the set of states there are states $s_{l+1}, r_{l+1}, t_{l+1} \in S$ such that we can select a subsequence via some mapping $\tau : \mathbb{N} \mapsto \mathbb{N}$ with

- $\delta(s_0, v^1_{\tau(i),l} v^2_{\tau(i),l} \cdots v^l_{\tau(i),l} v^{l+1}_{\tau(i),l}) = s_{l+1}$,
- $\delta(s_0, v^1_{\tau(i),l} v^2_{\tau(i),l} \cdots v^l_{\tau(i),l} v^{l+1}_{\tau(i),l} \hat{u}_{\tau(i),l} \hat{o}_{\tau(i),l}) = r_{l+1}$ and
- $\delta(s_0, v^1_{\tau(i),l} v^2_{\tau(i),l} \cdots v^l_{\tau(i),l} v^{l+1}_{\tau(i),l} \hat{u}_{\tau(i),l} v'_{\tau(i),l} v''_{\tau(i),l} \hat{o}_{\tau(i),l}) = t_{l+1}$

for all $i \geq 1$. Note that by hypothesis we have

- $\delta(s_0, v^1_{\tau(i),l} v^2_{\tau(i),l} \cdots v^l_{\tau(i),l}) = s_l$,
- $\delta(s_0, v^1_{\tau(i),l} v^2_{\tau(i),l} \cdots v^l_{\tau(i),l} v^{l+1}_{\tau(i),l} \hat{u}_{\tau(i),l} \hat{o}_{\tau(i),l} o^{l+1}_{\tau(i),l}) = r_l$ and
- $\delta(s_0, v^1_{\tau(i),l} v^2_{\tau(i),l} \cdots v^l_{\tau(i),l} v^{l+1}_{\tau(i),l} \hat{u}_{\tau(i),l} v'_{\tau(i),l} v''_{\tau(i),l} \hat{o}_{\tau(i),l} o^{l+1}_{\tau(i),l}) = t_l$.

Again all assertions of the claim can be verified if we set

- $w_{i,l+1} =_{\text{def}} w_{\tau(i),l}$,
- $v^{l+1}_{i,l+1} =_{\text{def}} v^{l+1}_{\tau(i),l}$ and $o^{l+1}_{i,l+1} =_{\text{def}} o^{l+1}_{\tau(i),l}$,
- $u_{i,l+1} =_{\text{def}} \hat{u}_{\tau(i),l}$ and $o_{i,l+1} =_{\text{def}} \hat{o}_{\tau(i),l}$,
- $v'_{i,l+1} =_{\text{def}} v'_{\tau(i),l}$ and $v''_{i,l+1} =_{\text{def}} v''_{\tau(i),l}$,
- $u'_{i+1,l+1} =_{\text{def}} \hat{u}'_{\tau(i+1),l}$ and $o'_{i+1,l+1} =_{\text{def}} \hat{o}'_{\tau(i+1),l}$.

(End proof of Claim 3)

We keep the notations of the previous claim. Now, again due to the finiteness of S there exist p, q with $1 \leq p < q \leq (|F|+1)^3$ and $(s_p, r_p, t_p) = (s_q, r_q, t_q)$. We define

- $x =_{\text{def}} v^1_{1,q-1} v^2_{1,q-1} \cdots v^p_{1,q-1}$,
- $v =_{\text{def}} v^{p+1}_{1,q} v^{p+2}_{1,q} \cdots v^q_{1,q}$ and $\hat{w} =_{\text{def}} s_k(v)$,
- $y =_{\text{def}} v^{p+1}_{1,q-1} \cdots v^{q-1}_{1,q-1} u_{1,q-1}$,
- $v_1 =_{\text{def}} v'_{1,q-1}$ and $u_1 =_{\text{def}} v''_{1,q-1}$,
- $z =_{\text{def}} o_{1,q-1} o^{q-1}_{1,q-1} \cdots o^{p+1}_{1,q-1}$,
- $z' =_{\text{def}} o^p_{1,q-1} \cdots o^2_{1,q-1} o^1_{1,q-1}$ and
- $u =_{\text{def}} o^q_{1,q} o^{q-1}_{1,q} \cdots o^{p+1}_{1,q}$.

Note that $p_k(y) = p_k(u) = p_k(u_1 z) = w$ since w starts at each marker m'_i, each marker n'_i is followed by a factor w and there is enough space between each two markers. It holds that $s_k(v) = \hat{w}$ by defintion and $|\hat{w}| \geq k$.

Moreover we have $\delta(s_0, xyzz') \notin S'$, $\delta(s_0, xyv_1u_1zz') \in S'$, $\delta(s_0, x) = s_p = s_q = \delta(s_0, xv)$, $\delta(s_0, xyz) = r_p = r_q = \delta(s_0, xyzu)$ and $\delta(s_0, xyv_1u_1z) = t_p = t_q = \delta(s_0, xyv_1u_1zu)$. By the choice of the sequence and since $|x|, |z'| \geq k$ we have $\tilde{w}yzw \preceq^{\text{e}}_k \tilde{w}yv_1u_1zw$ with $\tilde{w} =_{\text{def}} s_k(x)$.

From assertions b. and c. of Claim 2 can be concluded that $v^j_{1,q-1} \preceq^{\text{part}}_k v^j_{1,q}$ for $1 \leq j \leq q-1$, $u_{1,q-1}v'_{1,q-1} \preceq^{\text{part}}_k v^q_{1,q}$, $v''_{1,q-1}o_{1,q-1} \preceq^{\text{part}}_k o^q_{1,q}$ and that $o^j_{1,q-1} \preceq^{\text{part}}_k o^j_{1,q}$ for $1 \leq j \leq q-1$. With Lemma 4.4 we put these pieces together and obtain $yv_1 \preceq^{\text{part}}_k v$ and $u_1z \preceq^{\text{part}}_k u$. Thus by definition of the partial k-extension relation we have $\tilde{w}yv_1w \preceq_k \tilde{w}vw$ and $\dot{w}u_1zw \preceq_k \dot{w}uw$ with $\dot{w} =_{\text{def}} s_k(yv_1)$. We can exchange $\tilde{w}$ and $\dot{w}$ with $\hat{w}$ and obtain $\hat{w}yzw \preceq^{\text{e}}_k \hat{w}yv_1u_1zw$, $\hat{w}yv_1w \preceq_k \hat{w}vw$ and $\hat{w}u_1zw \preceq_k \hat{w}uw$. Hence we found pattern P''_k in F. *(End of Case 3)*

We remark that this proof of (2) $\Rightarrow$ (1) even shows that the automaton F does not have the *two* instances of each pattern with $s \in S'$ and $s' \notin S'$ on one hand, and $s' \in S'$ and $s \notin S'$ on the other hand. To see this note that we can start the whole investigation with selecting a sequence $\{w_s\}$ such that $\delta(s_0, w_{2i-1}) \in S'$ and $\delta(s_0, w_{2i}) \notin S'$.

(1) $\Rightarrow$ (2): Assume (2) is wrong, i.e. F has one of the patterns P_k, P'_k or P''_k. We show that any such pattern leads to a k-extension chain with an infinite number of alternations with respect

to $L(F)$, hence $L(F) \notin \mathcal{B}_{1,k}$ by Theorem 2.10.

Case 1. Suppose F has pattern P_k. Then we can write

$$\begin{array}{rcll} xvyz' & = & xv_0\hat{w}ywz_0' & (v = v_0\hat{w}, z' = wz_0') \\ & \preceq_k^{\mathrm{e}} & xv_0\hat{w}yv_1wz_0' & (\hat{w}yw \preceq_k^{\mathrm{e}} \hat{w}yv_1w) \\ & \preceq_k & xv_0\hat{w}vwz_0' & (\hat{w}yv_1w \preceq_k \hat{w}vw) \\ & = & xvv_0\hat{w}wz_0' & \\ & \preceq_k^{\mathrm{e}} & xvv_0\hat{w}wy_0wz_0' & \\ & = & xvv_0\hat{w}ywz_0' & (y = wy_0) \\ & \preceq_k^{\mathrm{e}} & xvv_0\hat{w}yv_1wz_0' & (\hat{w}yw \preceq_k^{\mathrm{e}} \hat{w}yv_1w) \\ & \preceq_k & xvv_0\hat{w}vwz_0' & (\hat{w}yv_1w \preceq_k \hat{w}vw) \\ & = & \ldots & \end{array}$$

and so on. In this way we obtain for all $i \geq 1$ with $xv^iyz' \preceq_k^{\mathrm{e}} xv^iyv_1z' \preceq_k xv^{i+1}z' \preceq_k^{\mathrm{e}} xv^{i+1}yz'$ a k-extension chain with an infinite number of alternations with respect to $L(F)$.

Case 2. Suppose F has pattern P_k'. Then we can write

$$\begin{array}{rcll} xzuz' & = & x_0\hat{w}zwu_0z' & (x = x_0\hat{w}, u = wu_0) \\ & \preceq_k^{\mathrm{e}} & x_0\hat{w}u_1zwu_0z' & (\hat{w}zw \preceq_k^{\mathrm{e}} \hat{w}u_1zw) \\ & \preceq_k & x_0\hat{w}uwu_0z' & (\hat{w}u_1zw \preceq_k \hat{w}uw) \\ & = & x_0\hat{w}wu_0uz' & \\ & \preceq_k^{\mathrm{e}} & x_0\hat{w}wz_0wu_0uz' & \\ & = & x_0\hat{w}zwu_0uz' & (z = wz_0) \\ & \preceq_k^{\mathrm{e}} & x_0\hat{w}u_1zwu_0uz' & (\hat{w}zw \preceq_k^{\mathrm{e}} \hat{w}u_1zw) \\ & \preceq_k & x_0\hat{w}uwu_0uz' & (\hat{w}u_1zw \preceq_k \hat{w}uw) \\ & = & \ldots & \end{array}$$

and so on. In this way we obtain for all $i \geq 1$ with $xzu^iz' \preceq_k^{\mathrm{e}} xu_1zu^iz' \preceq_k xu^{i+1}z' \preceq_k^{\mathrm{e}} xzu^{i+1}z'$ a k-extension chain with an infinite number of alternations with respect to $L(F)$.

Case 3. Suppose F has pattern P_k''. Then we can write

$$\begin{array}{rcll} xvyzuz' & = & xv_0\hat{w}yzwu_0z' & (v = v_0\hat{w}, u = wu_0) \\ & \preceq_k^{\mathrm{e}} & xv_0\hat{w}yv_1u_1zwu_0z' & (\hat{w}yzw \preceq_k^{\mathrm{e}} \hat{w}yv_1u_1zw) \\ & = & xv_0\hat{w}yv_1wz_0wu_0z' & (u_1z = wz_0) \\ & \preceq_k & xv_0\hat{w}vwz_0wu_0z' & (\hat{w}yv_1w \preceq_k \hat{w}vw) \\ & = & xvv_0\hat{w}u_1zwu_0z' & (v = v_0\hat{w}, u_1z = wz_0) \\ & \preceq_k & xvv_0\hat{w}uwu_0z' & (\hat{w}u_1zw \preceq_k \hat{w}uw) \\ & = & xvvwu_0uz' & (u = wu_0) \\ & \preceq_k^{\mathrm{e}} & xvvwy_0zwu_0uz' & (y = wy_0) \\ & = & xvv_0\hat{w}yzwu_0uz' & (v = v_0\hat{w}, y = wy_0) \\ & \preceq_k^{\mathrm{e}} & xvv_0\hat{w}yv_1u_1zwu_0uz' & (\hat{w}yzw \preceq_k^{\mathrm{e}} \hat{w}yv_1u_1zw) \\ & = & xvv_0\hat{w}yv_1wz_0wu_0uz' & (u_1z = wz_0) \\ & \preceq_k & xvv_0\hat{w}vwz_0wu_0uz' & (\hat{w}yv_1w \preceq_k \hat{w}vw) \\ & = & xvvv_0\hat{w}u_1zwu_0uz' & (v = v_0\hat{w}, u_1z = wz_0) \\ & \preceq_k & xvvv_0\hat{w}uwu_0uz' & (\hat{w}u_1zw \preceq_k \hat{w}uw) \\ & = & xvvvuuuz' & (v = v_0\hat{w}, u = wu_0) \\ & = & \ldots & \end{array}$$

and so on. In this way we obtain for all $i \geq 1$ with $xv^iyzu^iz' \preceq_k^{\mathrm{e}} xv^iyv_1u_1zu^iz' \preceq_k xv^{i+1}u^{i+1}z' \preceq_k^{\mathrm{e}} xv^{i+1}yzu^{i+1}z'$ a k-extension chain with an infinite number of alternations with respect to $L(F)$. ❑

Finally, we ask in this section how our patterns relate to the ones given by J. Stern in [Ste85a]. At least in one direction we can give a partial answer here.

Proposition 4.7. *Let F be a dfa.*

1. *If F has pattern T_k then F has pattern P_k.*
2. *If F has pattern T''_k then F has pattern P''_k.*

Proof. Let F be a dfa having pattern T_k witnessed by $\bar{x}, \bar{w}, \bar{z}, \bar{y}$ and $\bar{y}'$. If we define $x =_{\text{def}} \bar{x}$, $w =_{\text{def}} \bar{w}$, $v =_{\text{def}} \bar{w}\bar{y}\bar{w}\bar{y}'$, $y =_{\text{def}} \bar{w}\bar{y}$, $v_1 =_{\text{def}} \bar{w}\bar{y}'$, $z' =_{\text{def}} \bar{w}\bar{z}$ and $\hat{w} =_{\text{def}} s_k(v)$, then these words show that there is also pattern P_k in F. The required relations hold since $p_k(v_1) = w$ and $v = yv_1$.

Now let F have pattern T''_k witnessed by $\bar{x}, \bar{w}, \bar{w}'\bar{y}, \bar{y}', \bar{u}, \bar{v}$ and $\bar{z}$. Again, if we define $x =_{\text{def}} \bar{x}$, $w =_{\text{def}} \bar{w}'$, $v =_{\text{def}} \bar{w}'\bar{y}'\bar{w}\bar{y}$, $y =_{\text{def}} \bar{w}'\bar{y}'$, $v_1 =_{\text{def}} \bar{w}\bar{y}$, $u_1 =_{\text{def}} \bar{w}'\bar{v}$, $u =_{\text{def}} \bar{w}'\bar{v}\bar{w}\bar{u}$, $z =_{\text{def}} \bar{w}\bar{u}$, $z' =_{\text{def}} \bar{z}$ and $\hat{w} =_{\text{def}} s_k(v)$, then these words show that there is also pattern P''_k in F. The required relations hold since $p_k(z) = p_k(v_1) = \bar{w}$, $v = yv_1$ and $u = u_1 z$. ❑

The reverse implications and the case of pattern T'_k remain open. The problem with the latter is that we can not guarantee that $|x| \geq k$ which seems to be necessary for P'_k.

5 Consequences

Using Theorem 4.6 we obtain a co-NL(=NL)-algorithm for the decision problem for $\mathcal{B}_{1,k}$ with fixed k simply by testing the occurence of each of the patterns P_k, P'_k and P''_k in a given dfa. To see this note that we can guess the relevant states of a pattern and follow nondeterministically possible words between them. Moreover we can guess subwords of such a word and ensure that the k following letters are the same. This algorithm is different from the one which follows the characterization given by Theorem 2.10. We can directly derive a characterization of dot-depth one languages from our Theorem 4.6.

Theorem 5.1. *Let $L \in A^+$ and let F some dfa with $L(F) = L$. Then the following statements are equivalent:*

(1) $L(F) \notin \mathcal{B}_1$

(2) for all $k \geq 0$ the dfa F has one of the patterns P_k, P'_k or P''_k

(3) the dfa F has $\forall_{k\geq 0}$ P_k or $\forall_{k\geq 0}$ P'_k or $\forall_{k\geq 0}$ P''_k

Proof. To see (1) $\Leftrightarrow$ (2) note that by the defintion of the classes $\mathcal{B}_{1,k}$ and by Theorem 4.6 it holds that $L(F) \notin \mathcal{B}_1 \Leftrightarrow \forall_{k\geq 0} L(F) \notin \mathcal{B}_{1,k} \Leftrightarrow \forall_{k\geq 0} F$ has one of the patterns P_k, P'_k or P''_k. Moreover note that if (2) holds then one the patterns P_k, P'_k or P''_k has to appear for infinitely many k in F. Additionally any pattern with index $k \geq 1$ is also the same pattern with index $k-1$. ❑

It is known that the level k of a language L can be bound by $|F|^3$ where $|F|$ is the number of states of the minimal automaton F accepting L (cf. [Ste85a], Prop.4.1).

Proposition 5.2 [Ste85a]. *Let F be a minimal dfa and set $N =_{\text{def}} |F|^3$. If $L(F) \in \mathcal{B}_1$ then $L(F) \in \mathcal{B}_{1,N}$.*

Theorem 5.3. *Let $L \in A^*$ and let F be the minimal dfa with $L(F) = L$. Set $N =_{\mathrm{def}} |F|^3$. Then the following statements are equivalent:*

(1) $L(F) \notin \mathcal{B}_1$

(2) F *has one of the patterns* P_N, P'_N *or* P''_N

Proof. (1) $\Leftarrow$ (2) follows from Theorem 5.1. For the other direction suppose that F has one of the patterns P_N, P'_N or P''_N. Then by Theorem 4.6 we have $L(F) \notin \mathcal{B}_{1,N}$ and by Proposition 5.2 $L(F) \notin \mathcal{B}_1$. ❑

Using the above Theorem we obtain a decision algorithm for $\mathcal{B}_1$ simply by testing the occurence of each of the patterns P_N, P'_N and P''_N in a given minimal dfa.

Acknowledgement. Thanks are due to Christian Glaßer for carefully reading the proofs.

References

[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.

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

[Eil76] S. Eilenberg. *Automata, languages and machines*, volume B. Academic Press, New York, 1976.

[Kna83] R. Knast. A semigroup characterization of dot-depth one languages. *RAIRO Inform. Théor.*, 17:321–330, 1983.

[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.

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

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

[Sim75] I. Simon. Piecewise testable events. In *Proceedings 2nd GI Conference*, volume 33 of *Lecture Notes in Computer Science*, pages 214–222. Springer-Verlag, 1975.

[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.

[Tho96] W. Thomas. Languages, automata, and logic. Technical Report 9607, Institut für Informatik und praktische Mathematik, Universität Kiel, 1996.

[Wag98] K. W. Wagner, 1998. Personal communication.