



This item was submitted to Loughborough's Institutional Repository by the author and is made available under the following Creative Commons Licence conditions.

A yellow rectangular box containing the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 license summary. At the top is the Creative Commons logo (CC) and the text 'creative commons' in a bold, lowercase sans-serif font, with 'COMMONS DEED' in a smaller, spaced-out font below it. The license title 'Attribution-NonCommercial-NoDerivs 2.5' is centered. Below this, the text 'You are free:' is followed by a bullet point: 'to copy, distribute, display, and perform the work'. Then, 'Under the following conditions:' is followed by three icons in circles: 'BY' (Attribution), a crossed-out dollar sign (Noncommercial), and an equals sign (No Derivative Works). Each icon is followed by a short explanation. At the bottom, there are two more bullet points, a statement about fair use, and a link to the legal code with a document icon.

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

A Discontinuity in Pattern Inference

Daniel Reidenbach*

Fachbereich Informatik, Technische Universität Kaiserslautern,
Postfach 3049, 67653 Kaiserslautern, Germany
reidenba@informatik.uni-kl.de

Abstract. This paper examines the learnability of a major subclass of E-pattern languages – also known as erasing or extended pattern languages – in Gold’s learning model: We show that the class of terminal-free E-pattern languages is inferrable from positive data if the corresponding terminal alphabet consists of three or more letters. Consequently, the recently presented negative result for binary alphabets is unique.

1 Introduction

Pattern languages have been introduced by Angluin (cf. [Ang79]), though some publications by Thue can already be interpreted as investigations of pattern structures (cf. e.g. [Thu06]). Following Angluin, a pattern is a finite string that consists of *variables* and of *terminal* symbols. A word of its language is generated by a uniform *substitution* of all variables with arbitrary strings of terminals. For instance, the language generated by the pattern $x_1 \mathbf{a} x_2 \mathbf{b} x_1$ (with x_1, x_2 as variables and \mathbf{a}, \mathbf{b} as terminals) contains, among others, the words $w_1 = \mathbf{a} \mathbf{a} \mathbf{b} \mathbf{a}$, $w_2 = \mathbf{a} \mathbf{a} \mathbf{b} \mathbf{b} \mathbf{a}$, $w_3 = \mathbf{a} \mathbf{b} \mathbf{b}$, whereas the following examples are not covered by α : $v_1 = \mathbf{a}$, $v_2 = \mathbf{b} \mathbf{a} \mathbf{b} \mathbf{b} \mathbf{a}$, $v_3 = \mathbf{b} \mathbf{b} \mathbf{b} \mathbf{b}$. Thus, numerous regular and nonregular languages can be described by patterns in a compact and “natural” way.

Pattern languages have been the subject of several analyses within the scope of formal language theory, e.g. in [JKS⁺94] and [JSSY95] (for a survey see [RS97]). These examinations reveal that a definition disallowing the substitution of variables with the empty word – as given by Angluin – leads to a language with particular features being quite different from the one allowing the empty substitution (that has been applied when generating w_3 in our example). Languages of the latter type have been introduced by Shinohara (cf. [Shi82]); they are referred to as *extended*, *erasing*, or *E-pattern languages*.

When dealing with pattern languages, manifold questions arise from the manifest problem of computing a pattern that is common to a given set of words. Therefore pattern languages have been a focus of interest of algorithmic learning theory from the very beginning. In the most elementary learning model of inductive inference – introduced by Gold (cf. [Gol67]) and known as *learning in the limit* or *Gold style learning* – a class of languages is said to be *inferrable from positive data* if and only if there exists a computable device (the so-called

* Supported by the Deutsche Forschungsgemeinschaft (DFG), Grant Wi 1638/1-2

learning strategy) that, for every of these (potentially infinite) languages and for every full enumeration of the words of the particular language to be learned, converges to a distinct and complete description of the language in finite time. According to [Gol67], this task is too challenging for many well-known classes of formal languages: All superfinite classes of languages – i.e. all classes that contain every finite and at least one infinite language – such as the regular, context-free and context-sensitive languages are not inferrable from positive data. Thus, the number of rich classes of languages that are known to be learnable is rather small.

The current state of knowledge concerning the learnability of pattern languages considerably differs when regarding standard or E-pattern languages, respectively: The learnability of the class of standard pattern languages was shown by Angluin when introducing its definition in 1979 (cf. [Ang79]). In the sequel there has been a variety of profound studies (e.g. in [LW91], [WZ94], [RZ98], and many more) on the complexity of learning algorithms, consequences of different input data, efficient strategies for subclasses, and so on. Regarding E-pattern languages, however, appropriate approaches presumably need to be more sophisticated and therefore progress has been rather scarce. Apart from positive results for the full class of E-pattern languages over the trivial unary and infinite terminal alphabets in [Mit98], the examinations in the past two decades restricted themselves to the learnability of subclasses (cf. [Shi82], [Mit98], [Rei], and – indirectly – [Wri89]). In spite of all effort, it took more than twenty years until at least for binary terminal alphabets the non-learnability of the subclass of *terminal-free* E-pattern languages (generated by patterns that only consist of variables) and, thus, of the full class could be proven (cf. [Rei02]).

In this paper we revert to the class of terminal-free E-pattern languages – that has been a subject of some language theoretical examinations (cf. [Fil88] and [JSSY95]) as well – with a rather surprising outcome: We show that it is inferrable from positive data *if and only if* the terminal alphabet does not consist of two letters. Thus, we present the first class of pattern languages to be known for which different non-trivial alphabets imply different answers to the question of learnability. Using several theorems in [Ang80] and [JSSY95], our respective reasoning is chiefly combinatorial; therefore it touches on some prominent topics within the research on word monoids and combinatorics of words.

2 Definitions and Preliminary Results

For standard mathematical notions and recursion-theoretic terms not defined explicitly in this paper we refer to [Rog92]; for unexplained aspects of formal language theory, [RS97] may be consulted.

For an arbitrary set A of symbols, A^+ denotes the set of all non-empty words over A and A^* the set of all (empty and non-empty) words over A . We designate the *empty* word as e . For the word that results from the n -fold concatenation of a letter \mathbf{a} we write \mathbf{a}^n . $|\cdot|$ denotes the size of a set or the length of a word, respectively, and $|w|_{\mathbf{a}}$ the frequency of a letter \mathbf{a} in a word w .

Σ is an alphabet of *terminal* symbols and $X = \{x_1, x_2, x_3, \dots\}$ an infinite set of *variable* symbols, $\Sigma \cap X = \emptyset$. We designate Σ as *trivial* if and only if $|\Sigma| = 1$ or $|\Sigma| = \infty$. Henceforth, we use lower case letters from the beginning of the Latin alphabet as terminal symbols; words of terminal symbols are named as u, v , or w . For every $j, j \geq 1, y_j \in X$ is an unspecified variable, i.e. there may exist $j, j' \in \mathbb{N}$ such that $j \neq j'$ and $y_j = y_{j'}$. A *pattern* is a word over $\Sigma \cup X$, a *terminal-free pattern* is a word over X ; naming patterns we use lower case letters from the beginning of the Greek alphabet. $\text{var}(\alpha)$ denotes the set of all variables of a pattern α . We write Pat for the set of all patterns and Pat_{tf} for the set of all terminal-free patterns.

A *substitution* is a morphism $\sigma : (\Sigma \cup X)^* \rightarrow \Sigma^*$ such that $\sigma(\mathbf{a}) = \mathbf{a}$ for every $\mathbf{a} \in \Sigma$. The *E-pattern language* $L_\Sigma(\alpha)$ of a pattern α is defined as the set of all $w \in \Sigma^*$ such that $\sigma(\alpha) = w$ for any substitution σ . For any word $w = \sigma(\alpha)$ we say that σ *generates* w , and for any language $L = L_\Sigma(\alpha)$ we say that α *generates* L . If there is no need to give emphasis to the concrete shape of Σ we denote the E-pattern language of a pattern α simply as $L(\alpha)$. We use ePAT_{tf} as an abbreviation for the full class of terminal-free E-pattern languages. For any class ePAT^* of E-pattern languages we write ePAT_Σ^* if the corresponding alphabet is of interest.

According to [MS94] we call a word w *ambiguous* (in respect of a pattern α) if and only if there exist at least two substitutions σ and σ' such that $\sigma(\alpha) = w = \sigma'(\alpha)$, but $\sigma(x_j) \neq \sigma'(x_j)$ for an $x_j \in \text{var}(\alpha)$. The word $w = \mathbf{a} \mathbf{a}$, for instance, is ambiguous in respect of the pattern $\alpha = x_1 \mathbf{a} x_2$ since it can be generated by the substitutions $\sigma, \sigma(x_1) = \mathbf{a}, \sigma(x_2) = \mathbf{e}$, and $\sigma', \sigma'(x_1) = \mathbf{e}, \sigma'(x_2) = \mathbf{a}$.

Following [Mit98], we designate a pattern α as *succinct* if and only if $|\alpha| \leq |\beta|$ for all patterns β with $L(\beta) = L(\alpha)$, and we call a pattern *prolix* if and only if it is not succinct. The pattern $\alpha = x_1 x_1$, for instance, is succinct because there does not exist any shorter pattern that exactly describes its language, whereas $\beta = x_1 x_2 x_1 x_2$ is prolix since $L(\beta) = L(\alpha)$ and $|\alpha| < |\beta|$.

Let ePAT^* be any set of E-pattern languages. We say that the *inclusion problem* for ePAT^* is *decidable* if and only if there exists a computable function which, given two arbitrary patterns α, β with $L(\alpha), L(\beta) \in \text{ePAT}^*$, decides whether or not $L(\alpha) \subseteq L(\beta)$. In [JSSY95] it is shown that the inclusion problem for the full class of E-pattern languages is not decidable. Fortunately, this fact does not hold for terminal-free E-pattern languages. As this is of great importance for the following studies, we now cite two respective theorems of [JSSY95]:

Fact 1. *Let Σ be an alphabet, $|\Sigma| \geq 2$, and $\alpha, \beta \in X^*$ two arbitrarily given terminal-free patterns. Then $L_\Sigma(\beta) \subseteq L_\Sigma(\alpha)$ iff there exists a morphism $\phi : X^* \rightarrow X^*$ such that $\phi(\alpha) = \beta$.*

Fact 2. *The inclusion problem for ePAT_{tf} is decidable.*

We now introduce our notions on Gold's learning model (cf. [Gol67]): Each function $t : \mathbb{N} \rightarrow \Sigma^*$ satisfying $\{t(n) \mid n \geq 0\} = L(\alpha)$ is called a *text* for $L(\alpha)$. Let S be any total computable function reading initial segments of texts and returning patterns. Each such function is called a *strategy*. If α is a pattern and

t a text for $L(\alpha)$ we say that S identifies $L(\alpha)$ from t , if and only if the sequence of patterns returned by S , when reading t , converges to a pattern β , such that $L(\beta) = L(\alpha)$. Any class ePAT^* of E-pattern languages is *learnable (in the limit)* if and only if there is a strategy S identifying each language $L \in \text{ePAT}^*$ from any corresponding text. In this case we write $\text{ePAT}^* \in \text{LIM-TEXT}$ for short.

The analysis of the learnability of certain classes of languages is facilitated by some profound criteria given by Angluin (cf. [Ang80]). Because of Fact 2 and since Pat_{tf} is recursively enumerable, we can use the following:

Fact 3. *Let Pat^* be an arbitrary, recursively enumerable set of patterns and ePAT^* the corresponding class of E-pattern languages, such that the inclusion problem for ePAT^* is decidable. Then $\text{ePAT}^* \in \text{LIM-TEXT}$ iff for every pattern $\alpha \in \text{Pat}^*$ there exists a set T_α such that*

- $T_\alpha \subseteq L(\alpha)$,
- T_α is finite, and
- there does not exist a pattern $\beta \in \text{Pat}^*$ with $T_\alpha \subseteq L(\beta) \subset L(\alpha)$.

If T_α exists, then it is called a *telltale* (for $L(\alpha)$) (in respect of ePAT^*).

Roughly speaking, ePAT^* is, thus, inferrable from positive data if and only if every of its languages contains a finite subset that may be interpreted (by a strategy) as an exclusive signal to distinguish between that distinct language and all of its sub-languages in ePAT^* .

We conclude this section with the seminal learnability result on ePAT_{tf} that has been presented in [Rei02]:

Fact 4. *Let Σ be an alphabet, $|\Sigma| = 2$. Then $\text{ePAT}_{\text{tf},\Sigma} \notin \text{LIM-TEXT}$.*

In [Rei02] it is stated that the proof of this theorem cannot easily be extended on finite alphabets with more than two letters and it is conjectured that even the opposite of Fact 4 holds true for these alphabets. In the following section we discuss this fairly counter-intuitive assumption.

3 On the Learnability of ePAT_{tf}

Trivial alphabets, for which ePAT_{tf} is learnable (cf. [Mit98]), considerably ease the construction of telltales. Consequently, the recent negative result on binary alphabets (cf. Fact 4) – revealing that the assumed uniqueness of the approaches on trivial alphabets indeed might not be a matter of the methods, but of the subject – promotes the guess that ePAT_{tf} should not be learnable for every non-trivial alphabet. This surmise is supported by the fundamental algebraic theorem that for the free semigroup with two generators and for every $n \in \mathbb{N}$ there exists a free sub-semigroup with n generators and, thus, that the expressive power of words over three or more letters does not exceed that of words over two letters. Furthermore, there also exists a pattern specific hint backing this expectation since there seems to be no significant difference between terminal-free E-pattern languages over two and those over three letters (derived directly from Fact 1):

Theorem 1. *Let Σ_1, Σ_2 be finite alphabets, $|\Sigma_1| = 2$ and $|\Sigma_2| \geq 3$. Let α, β be terminal-free patterns. Then $L_{\Sigma_1}(\alpha) \neq L_{\Sigma_1}(\beta)$ iff $L_{\Sigma_2}(\alpha) \neq L_{\Sigma_2}(\beta)$.*

Thus, there is some evidence to suggest that Fact 4 might be extendable on all non-trivial terminal alphabets. In fact, our main result finds the opposite to be true:

Theorem 2. *Let Σ be a finite alphabet, $|\Sigma| \geq 3$. Then $\text{ePAT}_{\text{tf}, \Sigma} \in \text{LIM-TEXT}$.*

The proof of this theorem requires a broad combinatorial reasoning; it is accomplished in Section 3.1.

With Theorem 2 we can give a complete characterisation of the learnability of ePAT_{tf} , subject to alphabet size (for those cases not covered by Theorem 2, refer to [Mit98] or Fact 4, respectively):

Corollary 1. *Let Σ be an alphabet. Then $\text{ePAT}_{\text{tf}, \Sigma} \in \text{LIM-TEXT}$ iff $|\Sigma| \neq 2$.*

Consequently, we can state a discontinuity in the learnability of terminal-free E-pattern languages that – though it has been conjectured in [Rei02] – seems to be rather unexpected and that might explain the lack of comprehensive results on this subject in the past decades. The following section is dedicated to the proof of Theorem 2, but a precise and language theoretical explanation of the demonstrated singularity of terminal-free E-pattern languages over two letters is still open.

3.1 Proof of the Main Result

The proof of Theorem 2 consists of several steps: a characterisation of prolix patterns, a particular type of substitution, a learnability criterion for classes of terminal-free E-pattern languages, and some lemmata combining these elements.

To begin with, we give the characterisation of prolix patterns, that – although not implying a new decidability result (cf. Fact 2) – is a crucial instrument for our proof of the main theorem (see explanation after Theorem 4) as it gives a compact description of prolixness. Actually, in our reasoning we only use the *if* part of the following theorem, but we consider the characterisation of some interest since prolix terminal-free patterns may be seen as solution candidates for Post’s Correspondence Problem if the empty substitution is allowed (the other case has been analysed e.g. in [MS93] and [LP95]).

Theorem 3. *A terminal-free pattern α is prolix iff there exists a decomposition*

$$\alpha = \beta_0 \gamma_1 \beta_1 \gamma_2 \beta_2 \dots \beta_{n-1} \gamma_n \beta_n$$

for an $n \geq 1$, arbitrary $\beta_i \in X^*$ and $\gamma_i \in X^+$, $i \leq n$, such that

1. $\forall i : |\gamma_i| \geq 2$,
2. $\forall i, i' : \text{var}(\gamma_i) \cap \text{var}(\beta_{i'}) = \emptyset$,
3. $\forall i \exists y_i \in \text{var}(\gamma_i) : (|\gamma_i|_{y_i} = 1 \wedge \forall i' \leq n : (y_i \in \text{var}(\gamma_{i'}) \implies \gamma_i = \gamma_{i'}))$.

Proof. We first prove the *if* part of the theorem. Hence, let $\alpha \in \text{Pat}_{\text{tf}}$ be a pattern such that there exists a decomposition satisfying conditions 1, 2, and 3. We show that then there exist a pattern $\delta \in \text{Pat}_{\text{tf}}$ and two morphisms ϕ and ψ with $|\delta| < |\alpha|$, $\phi(\alpha) = \delta$, and $\psi(\delta) = \alpha$. Thus, we use Fact 1 as a criterion for the equivalence of E-pattern languages.

We define $\delta := \beta_0 y_1 \beta_1 y_2 \beta_2 \dots \beta_{n-1} y_n \beta_n$ with y_i derived from condition 3 for every $i \leq n$. Then condition 1 implies $|\delta| < |\alpha|$; the existence of ϕ and ψ (ϕ mapping y_i on γ_i and ψ mapping γ_i on y_i for every $i \leq n$, both of the morphisms leaving all other variables unchanged) results from conditions 2 and 3.

Due to space constraints and as it is not needed for our subsequent reasoning, the proof of the *only if* part is merely given as an extended sketch.

Assume that $\alpha \in \text{Pat}_{\text{tf}}$ is prolix. We show that then there exists at least one decomposition of α satisfying conditions 1, 2, and 3: Because of the assumption and Fact 1, there exist a succinct pattern $\delta \in \text{Pat}_{\text{tf}}$ and two morphisms ϕ and ψ with $|\delta| < |\alpha|$, $\phi(\alpha) = \delta$, and $\psi(\delta) = \alpha$. Since δ is succinct it is obvious that $|\psi(x_j)| \geq 1$ for every $x_j \in \text{var}(\delta)$. Moreover, we may conclude that for every $x_j \in \text{var}(\delta)$ there exists an $x_{j'} \in \text{var}(\psi(x_j))$ such that $|\delta|_{x_j} = |\alpha|_{x_{j'}}$, as otherwise δ would be prolix – according to the *if* part and because of $\phi(\alpha) = \delta$, leading to $x_j \in \text{var}(\phi(x_{j''}))$ for some $x_{j''} \in \text{var}(\alpha)$. Therefore the following fact (later referred to as (\star)) is evident: Without loss of generality, δ , ϕ , and ψ can be chosen such that $x_j \in \text{var}(\psi(x_j))$ for every $x_j \in \text{var}(\delta)$, $\phi(x_j) = x_j$ for every $x_j \in \text{var}(\alpha) \cap \text{var}(\delta)$, and $\phi(x_{j'}) = e$ for every $x_{j'} \in \text{var}(\alpha) \setminus \text{var}(\delta)$.

In order to provide a basic decomposition of α we now define some appropriate subsets of $\text{var}(\alpha)$: First, $Y_1 := \{x_{j_1} \in \text{var}(\alpha) \mid |\psi(\phi(x_{j_1}))| \geq 2\}$, second, $Y_2 := \{x_{j_2} \in \text{var}(\alpha) \mid \phi(x_{j_2}) = e\}$, and finally $Y_3 := \text{var}(\alpha) \setminus (Y_1 \cup Y_2)$. These definitions entail $Y_1 \cap Y_2 = \emptyset$, $Y_2 \neq \emptyset$ (because of $|\delta| < |\alpha|$), and $|\phi(x_{j_3})| = |\psi(\phi(x_{j_3}))| = 1$ for all $x_{j_3} \in Y_3$ (because of (\star)). Using these sets of variables we examine the following decomposition: $\alpha = \beta_0 \gamma_1 \beta_1 \gamma_2 \beta_2 \dots \beta_{m-1} \gamma_m \beta_m$ with $\beta_0, \beta_m \in Y_3^*$, $\beta_i \in Y_3^+$ for $0 < i < m$, and $\gamma_i \in (Y_1 \cup Y_2)^+$ for all $i \leq m$.

This decomposition is unique. Obviously, it satisfies condition 2, and because of (\star) we may state fact $(\star\star)$: $\gamma_i = \psi(\phi(\gamma_i))$ for every i , $1 \leq i \leq m$.

This leads to $\text{var}(\gamma_i) \cap Y_1 \neq \emptyset$ for all $i \leq m$, and therefore condition 1 is satisfied. Now we can identify the following two cases:

Case A: $\forall i, 1 \leq i \leq m : \sum_{x_{j_1} \in Y_1} |\gamma_i|_{x_{j_1}} = 1$

Consequently, if $\text{var}(\gamma_i) \cap \text{var}(\gamma_{i'}) \cap Y_1 \neq \emptyset$ for some $i, i', i \neq i'$, then $\phi(\gamma_i) = \phi(\gamma_{i'})$ and also $\psi(\phi(\gamma_i)) = \psi(\phi(\gamma_{i'}))$. Thus, with $(\star\star)$ we can state $\gamma_i = \gamma_{i'}$, and therefore condition 3 for the basic decomposition is satisfied.

Case B: $\exists \hat{i}, 1 \leq \hat{i} \leq m : \sum_{x_{j_1} \in Y_1} |\gamma_{\hat{i}}|_{x_{j_1}} = p \neq 1$ for a $p \in \mathbb{N}$

Because of condition 1 being satisfied, we can assume in this case that $p \geq 2$. Hence, we examine all modified decompositions of α that match the following principle for every \hat{i} meeting the requirement of Case B:

$$\alpha = \beta_0 \gamma_1 \beta_1 \gamma_2 \beta_2 \dots \beta_{\hat{i}-1} \overbrace{\gamma_{\hat{i}_1} \beta_{\hat{i}_1} \gamma_{\hat{i}_2} \beta_{\hat{i}_2} \dots \beta_{\hat{i}_{p-1}} \gamma_{\hat{i}_p}}^{\gamma_{\hat{i}}} \beta_{\hat{i}+1} \dots \beta_{m-1} \gamma_m \beta_m$$

such that β_i for all i and γ_i for all $i \neq \hat{i}$ derive from the previous definition, and, furthermore, such that $\beta_{i_k} = e$ for all k , $1 \leq k \leq p-1$, and $\gamma_{i_k} = \gamma_{i_k,l} x_{j_1, \hat{i}_k} \gamma_{i_k,r}$ for all k , $1 \leq k \leq p$, with $|\gamma_{i_k}| \geq 2$ and $\gamma_{i_k,l}, \gamma_{i_k,r} \in Y_2^*$ and $x_{j_1, \hat{i}_k} \in Y_1$. Then, for all of these newly created decompositions, conditions 1 and 2 still are satisfied. Because of $\psi(x_{j_1, \hat{i}_1} x_{j_1, \hat{i}_2} \dots x_{j_1, \hat{i}_p}) = \gamma_{\hat{i}}$, one of these decompositions meets the requirement of Case A. \square

As an illustration of Theorem 3 we now analyse some terminal-free patterns:

Example 1. $x_1x_2x_2x_1x_2x_2$ is prolix since $\gamma_1 = \gamma_2 = x_1x_2x_2$ and $\beta_0 = \beta_1 = \beta_2 = e$. $x_1x_2x_2x_1x_2x_2x_2$ and $x_1x_2x_1x_3x_4x_2x_4x_3$ are succinct since no variable for every of its occurrences has the same “environment” (i.e. a suitable γ) of length greater or equal 2 such that this environment does not share any of its variables with any potential β . $x_1x_2x_1x_2x_3x_3x_2x_4x_4x_5x_3x_2x_4x_4x_5$ is prolix since $\gamma_1 = \gamma_2 = x_1x_2$, $\gamma_3 = \gamma_4 = x_2x_4x_4x_5$, $\beta_0 = \beta_1 = \beta_4 = e$, $\beta_2 = x_3x_3$, $\beta_3 = x_3$.

As pointed out in [Rei02], certain words due to their ambiguity are unsuitable for being part of a telltale. In the following definition we introduce a particular type of substitution that – depending on the pattern it is applied to – may lead to ambiguous words as well; nevertheless, it can be used to generate telltale words as it imposes appropriate restrictions upon their ambiguity. This feature is relevant for the learnability criterion in Theorem 4.

Definition 1. Let α be a terminal-free pattern, $|\alpha| =: n$, and σ a substitution. For any $m \leq n$ let $\alpha \setminus_m = y_1 y_2 \dots y_m$ be the prefix of length m of α . Let r_1, r_2, \dots, r_{n-1} and l_2, l_3, \dots, l_n be the smallest natural numbers such that for every substitution σ' with $\sigma'(\alpha) = \sigma(\alpha)$ and for $m = 1, 2, \dots, n-1$:

$$|\sigma(\alpha \setminus_m)| - r_m \leq |\sigma'(\alpha \setminus_m)| \leq |\sigma(\alpha \setminus_m)| + l_{m+1}.$$

Furthermore, define $l_1 := 0 =: r_n$.

Then we call the substitution σ (λ, ρ) -significant (for α) iff there exist two mappings $\lambda, \rho : \mathbb{N} \rightarrow \mathbb{N}$ such that, for every $x_j \in \text{var}(\alpha)$, $\lambda(j) = \max\{l_m \mid y_m = x_j\}$, $\rho(j) = \max\{r_m \mid y_m = x_j\}$, and $|\sigma(x_j)| \geq \lambda(j) + \rho(j) + 1$. We designate a word w as significant (for α) iff for some λ, ρ there exists a (λ, ρ) -significant substitution σ such that $w = \sigma(\alpha)$.

The following example illustrates Definition 1:

Example 2. Let $\alpha := x_1x_2x_3x_4x_1x_4x_3x_2$. Obviously, α is terminal-free and prolix (cf. Theorem 3). Let the substitution σ be given by $\sigma(x_1) := \mathbf{a}$, $\sigma(x_2) := \mathbf{ab}$, $\sigma(x_3) := \mathbf{b}$, and $\sigma(x_4) := \mathbf{aa}$. With little effort it can be seen that there exists only one different substitution σ' such that $\sigma'(\alpha) = \sigma(\alpha)$, namely $\sigma'(x_1) = \mathbf{aa}$, $\sigma'(x_2) = \mathbf{b}$, $\sigma'(x_3) = \mathbf{ba}$, and $\sigma'(x_4) = \mathbf{a}$. In terms of Definition 1 this implies the following:

$$\begin{array}{llll} \sigma(\alpha \setminus_1) = \mathbf{a} & \sigma(\alpha \setminus_2) = \mathbf{aab} & \sigma(\alpha \setminus_3) = \mathbf{aabb} & \sigma(\alpha \setminus_4) = \mathbf{aabbba} \dots \\ \sigma'(\alpha \setminus_1) = \mathbf{aa} & \sigma'(\alpha \setminus_2) = \mathbf{aab} & \sigma'(\alpha \setminus_3) = \mathbf{aabba} & \sigma'(\alpha \setminus_4) = \mathbf{aabbba} \dots \end{array}$$

Thus, $l_2 = l_4 = l_6 = l_8 = 1$, $l_1 = l_3 = l_5 = l_7 = 0$, and $r_k = 0$ for $1 \leq k \leq 8$. Then, with $\lambda(1) = \lambda(3) = 0$, $\lambda(2) = \lambda(4) = 1$, and $\rho(j) = 0$ for $1 \leq j \leq 4$, the substitution σ is (λ, ρ) -significant for α , since $|\sigma(x_j)| \geq \lambda(j) + \rho(j) + 1$ for every j , $1 \leq j \leq 4$. Consequently, there are certain subwords in $w := \sigma(\alpha) = \sigma'(\alpha)$ that are generated for every possible substitution by the same variable; therefore we may regard the following variables and subwords – that, in a different example, of course can consist of more than one letter each – as “associated”:

$$w = \underbrace{\mathbf{a}}_{x_1} \mathbf{a} \underbrace{\mathbf{b}}_{x_2} \underbrace{\mathbf{b}}_{x_3} \mathbf{a} \underbrace{\mathbf{a}}_{x_4} \underbrace{\mathbf{a}}_{x_1} \mathbf{a} \underbrace{\mathbf{a}}_{x_4} \underbrace{\mathbf{b}}_{x_3} \mathbf{a} \underbrace{\mathbf{b}}_{x_2}.$$

That is the particular property of significant words which serves our purposes.

A second (and “comprehensive”) example for a substitution generating significant words is given in Lemma 1.

Now we present the learnability criterion to be used, that is a generalisation of two criteria in [Rei02]. As mentioned in Example 2, this criterion utilizes the existence of certain subwords in significant words that may be mapped to a distinct variable. In these subwords we place a single letter as a *marker* for its variable such that the exact shape of the generating pattern can be extracted from a suitable set of these words. The need for this distinct marker is an oblique consequence of a method used in [Rei02] – the so-called *inverse substitution*.

Theorem 4. *Let Σ be an alphabet. Let Pat_{tf}^* be a recursively enumerable set of terminal-free patterns and $\text{ePAT}_{\text{tf}, \Sigma}^*$ the corresponding class of E-pattern languages. Then $\text{ePAT}_{\text{tf}, \Sigma}^* \in \text{LIM-TEXT}$ if for every $\alpha \in \text{Pat}_{\text{tf}}^*$ there exists a finite set $\text{SUB} := \{\sigma_i \mid 1 \leq i \leq n\}$ of substitutions and mappings λ_i and ρ_i such that*

1. every $\sigma_i \in \text{SUB}$ is (λ_i, ρ_i) -significant for α and
2. for every $x_j \in \text{var}(\alpha)$ there exists a substitution $\sigma_{j'} \in \text{SUB}$ with $\sigma_{j'}(x_j) = u_{j',j} \mathbf{a} v_{j',j}$ for a letter $\mathbf{a} \in \Sigma$ and some $u_{j',j}, v_{j',j} \in \Sigma^*$, $|u_{j',j}| \geq \lambda_{j'}(j)$ and $|v_{j',j}| \geq \rho_{j'}(j)$, such that $|\sigma_{j'}(\alpha)|_{\mathbf{a}} = |\alpha|_{x_j}$.

Proof. Given $\alpha \in \text{Pat}_{\text{tf}}^*$, we define a set T_α of words by $T_\alpha := \{w_i \mid \sigma_i(\alpha) = w_i \text{ for a } \sigma_i \in \text{SUB}\}$. We now show that T_α is a telltale for $L(\alpha)$ in respect of $\text{ePAT}_{\text{tf}, \Sigma}^*$. For that purpose assume $T_\alpha \subseteq L(\beta) \subseteq L(\alpha)$ for some $\beta \in \text{Pat}_{\text{tf}}^*$. Then (due to Fact 1) there exists a morphism $\phi : X^* \rightarrow X^*$ such that $\phi(\alpha) = \beta$.

Because every $\sigma_i \in \text{SUB}$ is (λ_i, ρ_i) -significant for α and because of condition 2 we may conclude that for every $x_j \in \text{var}(\alpha)$ and for every σ' with $\sigma'(\beta) = w_{j'} = \sigma_{j'}(\alpha)$ – that necessarily exists since $T_\alpha \subseteq L(\beta)$ – holds the following: $\sigma'(\phi(x_j)) = u'_{j',j} \mathbf{a} v'_{j',j}$ for a letter $\mathbf{a} \in \Sigma$, and two words $u'_{j',j}, v'_{j',j} \in \Sigma^*$; for these words, due to the significance of $w_{j'}$, it is evident that $u'_{j',j}$ is a suffix of $u_{j',j}$ (or vice versa) and $v'_{j',j}$ is a prefix of $v_{j',j}$ (or vice versa). In addition it is obvious that $|\sigma'(\beta)|_{\mathbf{a}} = |\alpha|_{x_j}$ can be stated for the examined σ' . Therefore, in order to allow appropriate substitutions to generate the single letter \mathbf{a} , ϕ for all $x_j \in \text{var}(\alpha)$ must have the shape $\phi(x_j) = \gamma_1 x_{j_{\mathbf{a}}} \gamma_2$ with $\gamma_1, \gamma_2 \in X^*$ and a single variable $x_{j_{\mathbf{a}}} \in \text{var}(\phi(x_j))$, i.e. $|\beta|_{x_{j_{\mathbf{a}}}} = |\alpha|_{x_j}$. Hence, the morphism $\psi : X^* \rightarrow X^*$, defined by $\psi(x_k) := x_j$ for $k = j_{\mathbf{a}}$ and $\psi(x_k) := e$ for $k \neq j_{\mathbf{a}}$,

leads to $\psi(\beta) = \alpha$ and – with the assumption $L(\beta) \subseteq L(\alpha)$ – to $L(\beta) = L(\alpha)$. Consequently, $L(\beta) \not\subseteq L(\alpha)$ for every pattern β with $T_\alpha \subseteq L(\beta)$ and, thus, T_α is a telltale for $L(\alpha)$ in respect of $\text{ePAT}_{\text{tf}}^*$ (because of Fact 3). \square

In prolix patterns, there exist variables that cannot be substituted in such a way that the resulting word is significant. For instance, in the pattern $\alpha = x_1 x_2 x_1 x_2$ every subword generated by x_1 can be generated by x_2 as well, and vice versa. Therefore, when applying Theorem 4 to ePAT_{tf} , Pat_{tf}^* necessarily has to consist of succinct patterns only.

The following two lemmata prove that for every succinct terminal-free pattern (and, thereby, for every language in ePAT_{tf}) there exists a set of substitutions satisfying the conditions of Theorem 4.

Lemma 1. *Let α be a succinct pattern, $\alpha \in \text{Pat}_{\text{tf}}$, and Σ an alphabet such that $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\} \subseteq \Sigma$. Let for every $x_j \in \text{var}(\alpha)$ and for every $i \in \{j \mid x_j \in \text{var}(\alpha)\}$ the substitution σ_i^{tf} be given by*

$$\sigma_i^{\text{tf}}(x_j) := \begin{cases} \mathbf{a} \mathbf{b}^{3j-2} \mathbf{a} \mathbf{a} \mathbf{b}^{3j-1} \mathbf{a} \mathbf{a} \mathbf{b}^{3j} \mathbf{a} & , \quad i \neq j, \\ \mathbf{a} \mathbf{b}^{3j-2} \mathbf{a} \mathbf{c} \mathbf{a} \mathbf{b}^{3j-1} \mathbf{a} \mathbf{a} \mathbf{b}^{3j} \mathbf{a} & , \quad i = j, \end{cases}$$

Then for every σ'_i with $\sigma'_i(\alpha) = \sigma_i^{\text{tf}}(\alpha)$, for every $x_j \in \text{var}(\alpha)$, and for some $u_j, v_j \in \Sigma^*$:

$$\sigma'_i(x_j) = \begin{cases} u_j \mathbf{a} \mathbf{a} \mathbf{b}^{3j-1} \mathbf{a} \mathbf{a} v_j & , \quad i \neq j, \\ u_j \mathbf{a} \mathbf{c} \mathbf{a} \mathbf{b}^{3j-1} \mathbf{a} \mathbf{a} v_j & , \quad i = j, \end{cases}$$

Proof. To begin with we explain the following terms that are used frequently: A *segment* of $\sigma_i^{\text{tf}}(x_j)$ is a subword $\mathbf{a} \mathbf{b}^{3j-q} \mathbf{a}$, $0 \leq q \leq 2$. As the natural extension thereof, the term *segment* of $\sigma_i^{\text{tf}}(\delta)$ for every $\delta \in X^*$ designates any segment of $\sigma_i^{\text{tf}}(x_j)$ with $x_j \in \text{var}(\delta)$. An *outer segment* of $\sigma_i^{\text{tf}}(x_j)$ is the subword $\mathbf{a} \mathbf{b}^{3j-2} \mathbf{a}$ or the subword $\mathbf{a} \mathbf{b}^{3j} \mathbf{a}$. The *inner segment* of $\sigma_i^{\text{tf}}(x_j)$ is the subword $\mathbf{a} \mathbf{b}^{3j-1} \mathbf{a}$. $\sigma'_i(x_j)$ *contains segments* of $\sigma_i^{\text{tf}}(x_j)$ means that the segments of $\sigma_i^{\text{tf}}(x_j)$ occur in natural order (i.e. in that order specified by σ_i^{tf}), consecutively (apart from the potential necessity of inserting the letter \mathbf{c}), and non-overlapping.

Let σ'_i be an arbitrary substitution with $\sigma'_i(\alpha) = \sigma_i^{\text{tf}}(\alpha)$ and $\sigma'_i(x_j) \neq \sigma_i^{\text{tf}}(x_j)$ for an $x_j \in \text{var}(\alpha)$. Then we define the following subsets of $\text{var}(\alpha)$: Let Y_1 be the set of all $x_{j_1} \in \text{var}(\alpha)$ such that $\sigma'_i(x_{j_1})$ contains the inner segment of $\sigma_i^{\text{tf}}(x_{j_1})$, of every outer segment at least one letter, and at least one segment of the substitution σ_i^{tf} of a neighbouring variable. Consequently, $\sigma'_i(x_{j_1})$ contains at least two segments of $\sigma_i^{\text{tf}}(x_{j_1})$, and for all $x_{j_1} \in Y_1$: $\alpha \neq \delta_1 x_{j_1} x_{j_1} \delta_2$ with $\delta_1, \delta_2 \in X^*$. Let Y_2 be the set of all $x_{j_2} \in \text{var}(\alpha)$ such that $\sigma'_i(x_{j_2})$ contains of at least one segment of $\sigma_i^{\text{tf}}(x_{j_2})$ no letter. Then $Y_1 \cap Y_2 = \emptyset$. Finally, let Y_3 be given by $Y_3 := \text{var}(\alpha) \setminus (Y_1 \cup Y_2)$. Then $\sigma'_i(x_{j_3})$ for all $x_{j_3} \in Y_3$ contains the inner segment of $\sigma_i^{\text{tf}}(x_{j_3})$ and of both outer segments at least one letter, but no complete segment of a neighbouring variable.

Now assume to the contrary $Y_2 \neq \emptyset$, that implies $Y_1 \neq \emptyset$ as for every variable there are three unique corresponding segments (for two segments, depending on α and σ'_i , we might face the situation that $Y_2, Y_3 \neq \emptyset$, but $Y_1 = \emptyset$). We show that

this assumption entails α being prolix. The subsequent argumentation on this utilizes Theorem 3 and an evident fact (referred to as (\star)): For every $\delta \in Y_2^+$, $\sigma'_i(\delta)$ contains of at least one segment of $\sigma_i^{\text{tf}}(\delta)$ no letter.

As the starting point of our reasoning we use the following decomposition: $\alpha = \beta_0 \gamma_1 \beta_1 \gamma_2 \beta_2 \dots \beta_{n-1} \gamma_n \beta_n$ with $n \geq 1$, $\beta_0, \beta_n \in Y_3^*$, $\beta_k \in Y_3^+$ for $0 < k < n$, and $\gamma_k \in (Y_1 \cup Y_2)^+$ for $1 \leq k \leq n$. This decomposition is unique, and it obviously satisfies condition 2 and – due to (\star) – condition 1 of Theorem 3. However, concerning condition 3 it possibly deserves some modifications. To this end, the following procedure reconstructs the above decomposition such that in every γ_k there is exactly one occurrence of a variable in Y_1 (using (\star) again):

PROCEDURE:

Define $k := 1$.

STEP 1: Let y_1, y_2 be the leftmost variables in γ_k with $y_1, y_2 \in Y_1$ and $\gamma_k = \delta_1 y_1 y_2 \delta_2$ for $\delta_1, \delta_2 \in X^*$. IF these variables exist, THEN define $\gamma_{n+1} := \gamma_n, \dots, \gamma_{k+2} := \gamma_{k+1}, \beta_{n+1} := \beta_n, \dots, \beta_{k+1} := \beta_k, \gamma_k := \delta_1 y_1, \beta_k = e$, and $\gamma_{k+1} := y_2 \delta_2$; finally, define $n := n + 1$. END IF. IF $k < n$, THEN define $k := k + 1$ and go to STEP 1. ELSE rename all pattern fragments as follows: $\alpha := \beta_{0_1} \gamma_{1_1} \beta_{1_1} \gamma_{2_1} \beta_{2_1} \dots \beta_{(n-1)_1} \gamma_{n_1} \beta_{n_1}$. Finally, define $k_1 := 1_1$ and go to STEP 2. END IF.

STEP 2: Let $y_1 \in Y_2$ and $y_2 \in Y_1$ be the leftmost variables in γ_{k_1} with $\gamma_{k_1} = \delta_1 y_1 y_2 \delta_2$ for $\delta_1, \delta_2 \in X^*$, such that at least one segment of $\sigma_i^{\text{tf}}(y_1)$ is generated by variables in δ_1 . IF these variables exist, THEN IF $\sigma'_i(y_2)$ contains a segment of $\sigma_i^{\text{tf}}(\delta_2)$ (thus, necessarily the leftmost segment), THEN extend the decomposition as described in STEP 1. Finally, define $k_1 := (k+1)_1, n_1 := (n+1)_1$, and go to STEP 2. ELSE define $Y_1 := Y_1 \setminus \{y_2\}$ and $Y_3 := Y_3 \cup \{y_2\}$, reconstruct all pattern fragments β_{k_1} and γ_{k_1} accordingly and go to STEP 2. END IF. ELSE IF $k_1 < n_1$, THEN define $k_1 := (k+1)_1$ and go to STEP 2. ELSE rename all pattern fragments as follows: $\alpha := \beta_{0_2} \gamma_{1_2} \beta_{1_2} \gamma_{2_2} \beta_{2_2} \dots \beta_{(n-1)_2} \gamma_{n_2} \beta_{n_2}$. Finally, define $k_2 := 1_2$ and go to STEP 3. END IF. END IF.

STEP 3: Let $y_1 \in Y_1$ and $y_2 \in Y_2$ be the leftmost variables in γ_{k_2} with $\gamma_{k_2} = \delta_1 y_1 y_2 \delta_2$ for $\delta_1, \delta_2 \in X^*$, such that at least one segment of $\sigma_i^{\text{tf}}(y_2)$ is generated by variables in δ_2 . Modify γ_{k_2} analogously to Step 2. When this has been done for every k_2 , then rename all pattern fragments as follows: $\alpha := \beta_{0_3} \gamma_{1_3} \beta_{1_3} \gamma_{2_3} \beta_{2_3} \dots \beta_{(n-1)_3} \gamma_{n_3} \beta_{n_3}$. Define $k_3 := 1_3$ and go to STEP 4.

STEP 4: Let y_1, y_2 be the leftmost variables in γ_{k_3} with $y_1, y_2 \in Y_2$ and $\gamma_{k_3} = \delta_1 y_1 y_2 \delta_2$ for $\delta_1, \delta_2 \in X^*$, such that at least one segment of $\sigma_i^{\text{tf}}(y_1)$ is contained in $\sigma'_i(\delta_1)$ and at least one segment of $\sigma_i^{\text{tf}}(y_2)$ is contained in $\sigma'_i(\delta_2)$. IF these variables exist, THEN extend the decomposition as described in STEP 1. Finally, define $k_3 := (k+1)_3, n_3 := (n+1)_3$, and go to STEP 4. END IF. IF $k_3 < n_3$, THEN define $k_3 := (k+1)_3$ and go to STEP 4. ELSE rename all pattern fragments as follows: $\alpha := \beta_{0_4} \gamma_{1_4} \beta_{1_4} \gamma_{2_4} \beta_{2_4} \dots \beta_{(n-1)_4} \gamma_{n_4} \beta_{n_4}$. END IF.

END OF PROCEDURE

Obviously, $\forall k_4, k'_4 : \text{var}(\gamma_{k_4}) \cap \text{var}(\beta_{k'_4}) = \emptyset$. So the following is left to be shown:

- a) $\forall k_4 : |\gamma_{k_4}| \geq 2$
b) $\forall k_4 : \sum_{x_{j_1} \in Y_1} |\gamma_{k_4}|_{x_{j_1}} = 1$
c) $\forall k_4, k'_4 : (\text{var}(\gamma_{k_4}) \cap \text{var}(\gamma_{k'_4}) \cap Y_1 \neq \emptyset \implies \gamma_{k_4} = \gamma_{k'_4})$

ad a) As a consequence of (\star) , the first decomposition of α is modified by Steps 1 - 4 if and only if there exists a γ_k that contains at least two variables $y_1, y_2 \in Y_1$. The procedure splits this γ_k in such a way that all new $\gamma_{k_1}, \gamma_{k_2}, \gamma_{k_3}$, and γ_{k_4} contain at least one $x_{j_2} \in Y_2$ and a sequence of variables that generates a segment of $\sigma_i^{\text{tf}}(x_{j_2})$ by σ'_i . Thus, $|\gamma_{k_4}| \geq 2$ for all k_4 .

ad b) As mentioned in a), every γ_k contains at least one variable $x_{j_2} \in Y_2$. Moreover, there must also be a variable $x_{j_1} \in Y_1$ in every γ_k (again due to (\star)). The procedure splits only those γ_k that contain at least two variables of Y_1 , and obviously – if possible – in such a way that every γ_{k_4} contains exactly one $x_{j_1} \in Y_1$. If this due to a) is not possible (e.g. for $\gamma_k = y_1 y_2 y_3$ with $y_1, y_3 \in Y_1$, $y_2 \in Y_2$, and necessarily $y_1 \neq y_3$), then in Step 2 or Step 3 either y_1 or y_3 is removed from Y_1 and therefore it is removed from γ_{k_1} or γ_{k_2} , respectively.

ad c) Because of a) and b) it is obvious that every γ_{k_4} begins or ends with a variable from Y_2 . We consider that case where γ_{k_4} begins with $x_{j_2, k_4} \in Y_2$; the second case is symmetrical and the argumentation for the case that both aspects hold true derives from the combination of both approaches.

Hence, without loss of generality $\gamma_{k_4} = x_{j_2, k_4} \delta_1$ for $\delta_1 \in X^*$. Due to (\star) and the construction of γ_{k_4} , to the right of x_{j_2, k_4} there is a pattern fragment $\gamma_{k_4, 1}$, $|\gamma_{k_4, 1}| \geq 1$, such that $\sigma'_i(\gamma_{k_4, 1})$ contains at least one segment of $\sigma_i^{\text{tf}}(x_{j_2, k_4})$. If $\gamma_{k_4, 1} \in Y_2^+$, then to the right of this pattern fragment there is a second pattern fragment $\gamma_{k_4, 2}$, $|\gamma_{k_4, 2}| \geq 1$, that – with σ'_i – again generates the segments “missing” in $\sigma'_i(\gamma_{k_4, 1})$ and so on. As every γ_{k_4} has finite length, there must exist a $x_{j_1, k_4} \in Y_1$ in γ_{k_4} concluding this argumentation.

Consequently, $\gamma_{k_4} = x_{j_2, k_4} \gamma_{k_4, 1} \gamma_{k_4, 2} \dots \gamma_{k_4, p-1} \gamma_{k_4, p} x_{j_1, k_4}$ for a $p \in \mathbb{N}$. However, since all variables in $\gamma_{k_4, l}$, $1 \leq l \leq p$ of at least one of their segments do not generate any letter, x_{j_1, k_4} by σ'_i exactly determines the shape of $\gamma_{k_4, p}$, $\gamma_{k_4, p}$ that of $\gamma_{k_4, p-1}$ etc. up to $\gamma_{k_4, 1}$ determining x_{j_2, k_4} .

This holds true for every occurrence of x_{j_1, k_4} and therefore $\forall k_4, k'_4 \forall x_{j_1} \in Y_1 : ((x_{j_1} \in \text{var}(\gamma_{k_4}) \wedge x_{j_1} \in \text{var}(\gamma_{k'_4})) \implies \gamma_{k_4} = \gamma_{k'_4})$. This proves c).

Thus, with a), b), c), and Theorem 3, α is prolix. This is a contradiction. Consequently, we may conclude $Y_1 = Y_2 = \emptyset$; this proves the lemma. \square

With Lemma 1, the major part of our reasoning is accomplished. Now the following lemma can be concluded without effort:

Lemma 2. *Let α be a succinct pattern, $\alpha \in \text{Pat}_{\text{tf}}$. Then for every i , $i \in \{j \mid x_j \in \text{var}(\alpha)\}$, there exist mappings $\lambda_i, \rho_i : \mathbb{N} \rightarrow \mathbb{N}$ such that σ_i^{tf} (cf. Lemma 1) is (λ_i, ρ_i) -significant for α .*

Proof. Directly from Lemma 1, since for every $x_j \in \text{var}(\alpha)$ we can state $\lambda_i(j) \leq 3j - 1$, $\rho_i(j) \leq 3j + 1$, and $|\sigma_i^{\text{tf}}(x_j)| \geq 9j + 3$. \square

Consequently, for every succinct terminal-free pattern there exists a set of significant words. However, no word generated by any σ_i^{tf} needs to consist of three different letters in order to be significant – **a** and **b** would be sufficient. Indeed, due to the marker **c**, the given set of all σ_i^{tf} satisfies the second condition of Theorem 4. Thus, Theorem 4 is applicable for ePAT_{tf} and therefore the main result of this paper, given in Theorem 2, is proven.

References

- [Ang79] D. Angluin. Finding patterns common to a set of strings. In *Proc. STOC 1979*, pages 130–141, 1979.
- [Ang80] D. Angluin. Inductive inference of formal languages from positive data. *Inf. Control*, 45:117–135, 1980.
- [Fil88] G. Filè. The relation of two patterns with comparable language. In *Proc. STACS 1988*, volume 294 of *LNCS*, pages 184–192, 1988.
- [Gol67] E.M. Gold. Language identification in the limit. *Inf. Control*, 10:447–474, 1967.
- [JKS⁺94] T. Jiang, E. Kinber, A. Salomaa, K. Salomaa, and S. Yu. Pattern languages with and without erasing. *Int. J. Comput. Math.*, 50:147–163, 1994.
- [JSSY95] T. Jiang, A. Salomaa, K. Salomaa, and S. Yu. Decision problems for patterns. *J. Comput. Syst. Sci.*, 50:53–63, 1995.
- [LP95] M. Lipponen and G. Păun. Strongly prime PCP words. *Discrete Appl. Math.*, 63:193–197, 1995.
- [LW91] S. Lange and R. Wiehagen. Polynomial-time inference of arbitrary pattern languages. *New Generat. Comput.*, 8:361–370, 1991.
- [Mit98] A.R. Mitchell. Learnability of a subclass of extended pattern languages. In *Proc. COLT 1998*, pages 64–71, 1998.
- [MS93] A. Mateescu and A. Salomaa. PCP-prime words and primality types. *RAIRO Inform. théor.*, 27(1):57–70, 1993.
- [MS94] A. Mateescu and A. Salomaa. Finite degrees of ambiguity in pattern languages. *RAIRO Inform. théor.*, 28(3–4):233–253, 1994.
- [Rei] D. Reidenbach. A non-learnable class of E-pattern languages. *Theor. Comp. Sci.*, to appear.
- [Rei02] D. Reidenbach. A negative result on inductive inference of extended pattern languages. In *Proc. ALT 2002*, volume 2533 of *LNAI*, pages 308–320, 2002.
- [Rog92] H. Rogers. *Theory of Recursive Functions and Effective Computability*. MIT Press, Cambridge, Mass., 1992. 3rd print.
- [RS97] G. Rozenberg and A. Salomaa. *Handbook of Formal Languages*, volume 1. Springer, Berlin, 1997.
- [RZ98] R. Reischuk and T. Zeugmann. Learning one-variable pattern languages in linear average time. In *Proc. COLT 1998*, pages 198–208, 1998.
- [Shi82] T. Shinohara. Polynomial time inference of extended regular pattern languages. In *Proc. RIMS Symposia on Software Science and Engineering, Kyoto*, volume 147 of *LNCS*, pages 115–127, 1982.
- [Thu06] A. Thue. Über unendliche Zeichenreihen. *Kra. Vidensk. Selsk. Skrifter. I Mat. Nat. Kl.*, 7, 1906.
- [Wri89] K. Wright. Identification of unions of languages drawn from an identifiable class. In *Proc. COLT 1989*, pages 328–333, 1989.
- [WZ94] R. Wiehagen and T. Zeugmann. Ignoring data may be the only way to learn efficiently. *J. Exp. Theor. Artif. Intell.*, 6:131–144, 1994.