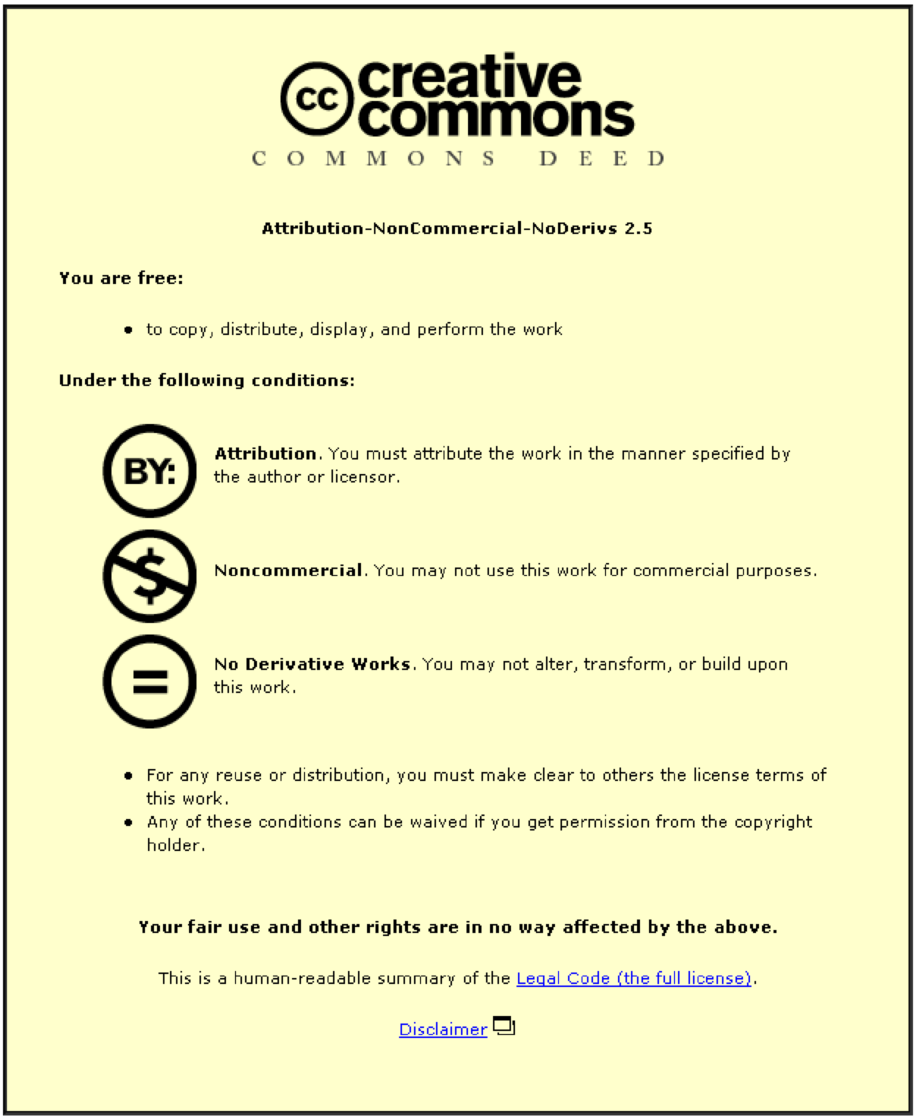

# Inferring Descriptive Generalisations of Formal Languages[☆]

Dominik D. Freydenberger[a,*], Daniel Reidenbach[b]

[a] Institut für Informatik, Johann Wolfgang Goethe-Universität,
Postfach 111932, 60054 Frankfurt am Main, Germany
[b] Department of Computer Science, Loughborough University,
Loughborough, Leicestershire, LE11 3TU, United Kingdom

**Abstract**

In the present paper, we introduce a variant of Gold-style learners that is not required to infer precise descriptions of the languages in a class, but that must find descriptive patterns, i. e., optimal generalisations within a class of pattern languages. Our first main result characterises those indexed families of recursive languages that can be inferred by such learners, and we demonstrate that this characterisation shows enlightening connections to Angluin's corresponding result for exact inference. Furthermore, this result reveals that our model can be interpreted as an instance of a natural extension of Gold's model of language identification in the limit. Using a notion of descriptiveness that is restricted to the natural subclass of terminal-free E-pattern languages, we introduce a generic inference strategy, and our second main result characterises those classes of languages that can be generalised by this strategy. This characterisation demonstrates that there are major classes of languages that can be generalised in our model, but not be inferred by a normal Gold-style learner. Our corresponding technical considerations lead to insights of intrinsic interest into combinatorial and algorithmic properties of pattern languages.

*Keywords:* Inductive inference, Descriptive generalisation, Pattern languages, Descriptive patterns, Upper approximate identification from positive data

## 1. Introduction

In Gold's intensively studied learning paradigm of language identification in the limit from positive data (cf. Gold [12]), it is a requirement for the computational learner to infer, for any positive presentation of any language in some class, an *exact* description of that language. While this maximum accuracy of

[☆] A preliminary version of this work was presented at COLT 2010 [10].

[*] Corresponding author.

*Email addresses:* `freydenberger@em.uni-frankfurt.de` (Dominik D. Freydenberger), `D.Reidenbach@lboro.ac.uk` (Daniel Reidenbach)

the output of the inference procedure is clearly a natural goal, it has a number of downsides, the most obvious one being the fact that it can lead to significant limitations to the learning power of the model. From a more applied point of view, there is another important reason why one might wish to relax it and settle for receiving an approximation of the language from the learner: depending on the class of languages to be inferred, the corresponding grammars or acceptors might have undesirable properties, i. e., they might have computationally hard decision problems or be incomprehensible to a (human) user. Thus, in various settings it might be perfectly acceptable for an inference procedure to output a compact and reasonably precise approximation of the language instead of producing a precise yet arbitrarily complex grammar.

In the present paper, we introduce and study such a variant of Gold's model, where the requirement of exact language identification is dropped and replaced with that of inference of easily interpretable approximations. More precisely, we consider a learner that, for any language it reads, must converge to a *consistent pattern*, i. e., a finite string that consists of variables and of terminal symbols and that can be turned into any word of the language by substituting arbitrary strings of terminal symbols for the variables. In addition to being seen as mere descriptions of common features of words in a given language, such a pattern $\alpha$ can also be interpreted as a generator of a formal language $L(\alpha)$, the so-called *pattern language* (cf. Angluin [1]), which is simply the maximum set of words the pattern is consistent with. Hence, referring to this terminology, we can state that our learner has to output a pattern generating a language that is a superset of the input language, which means that our approach does not yield an arbitrary approximation of a language, but rather a *generalisation*. Even though many classes of pattern languages have a number of NP-complete or undecidable basic decision problems (see, e. g., Angluin [1], Jiang et al. [16], and Freydenberger and Reidenbach [8]), patterns (or related concepts, such as regular expressions and their various extensions implemented in today's programming languages and text editors, see Câmpeanu et al. [4]) are widely used when commonalities of words are to be specified or interpreted by a human user, which demonstrates that they are a worthwhile concept in the context of our paper.

When inferring consistent patterns instead of precise descriptions, it is of course vital to develop and employ a notion of high-quality patterns, so that the inference procedure does not lead to an overly imprecise result. Otherwise, the learner could always output the pattern $\alpha := x_1$ (where $x_1$ is a variable), which is consistent with every language, and this approach would obviously neither lead to a rich theory nor to practically relevant results. In our model, the inference procedure shall therefore be required to converge to a pattern $\delta$ that is *descriptive* of the language $L$ (with respect to a class $\mathrm{PAT}_\star$ of pattern languages). This means that $\delta$ must be consistent with $L$, $L(\delta)$ must be included in $\mathrm{PAT}_\star$, and there is no pattern $\delta'$ satisfying $L(\delta') \in \mathrm{PAT}_\star$ and $L \subseteq L(\delta') \subset L(\delta)$; in other words, a pattern is descriptive of a language if there is no other pattern providing a closer match for the language. Since descriptiveness captures a natural understanding of patterns providing a desirable generalisation of languages and, furthermore, descriptive patterns can be used to devise Gold-style learners precisely identifying classes of pattern languages from positive data, this concept has been thoroughly investigated (see, e. g., Angluin [1], Jiang et al. [16], and Freydenberger and Reidenbach [9]), and the same holds for optimal approximations of other types of languages (see, e. g., Arimura et al. [3]). While established definitions of descriptiveness often restrict their view to patterns covering finite languages and normally use the full class of E- or NE-pattern languages (to be formally introduced in Section 2) as the class $\mathrm{PAT}_\star$ of admissible pattern languages, we allow a descriptive pattern to cover a finite or an infinite language, and we have a class $\mathrm{PAT}_\star$ that can be arbitrarily chosen. Both of these extensions of the original definition are absolutely straightforward.

To summarise our model of inference, we consider a learner that reads a positive presentation of a language and, after having seen a new input word, outputs a pattern, the so-called *hypothesis*. We then say that, for a class $\mathcal{L}$ of languages and a class $\mathrm{PAT}_\star$ of pattern languages, the learner $\mathrm{PAT}_\star$*-descriptively generalises* $\mathcal{L}$ if and only if, for every positive presentation of every language $L \in \mathcal{L}$, the sequence of hypotheses produced by the learner converges to a pattern $\delta$ that is descriptive of $L$ with respect to the class $\mathrm{PAT}_\star$. A more formal definition of our model is given in Section 4.1.

While the focus of research in inductive inference from positive data has been on exact identification, there are quite a few of studies of paradigms where – either directly or indirectly – approximations of languages are inferred; however, the motivation for this research often differs substantially from ours as described at the beginning of the present section. We now shall briefly summarise these approaches in order to highlight the differences to our model. Mukouchi [23] introduces the concept of *strong-minimal inference* and *minimal inference*, where the learner needs to converge to a minimum generalisation for *any* language (or, respectively, for any language where such a minimum generalisation exists among the admissible hypotheses). Hence, the model does not support an explicit restriction to a specific class $\mathcal{L}$ of languages that need to be generalised by the learner. The notion of *upper approximate identification* by Kobayashi and Yokomori [19] considers the inferrability of minimum generalisations of languages and features an explicit split between a class $\mathcal{L}$ of languages to be generalised and a class of admissible hypotheses. Hence – apart from the fact that our model, unlike upper approximate identification, is not restricted to indexed families $\mathcal{L}$, but rather restricts the nature of the hypotheses – this model is virtually identical to our approach. However, the focus of the paper [19] (and of subsequent studies; see, e. g., Kobayashi and Yokomori [20] and Fernau [6]) is on a different inference paradigm, namely *upper-best approximate identification*, where the topology of the class of hypotheses is restricted, as it needs to contain, for every language to be generalised, a semantically unique minimal generalisations. Such a property does not hold for the classes of pattern languages we shall use as a hypothesis space, and many of our technical results deal with exactly this aspect of *competing* descriptive generalisations, i. e., the existence of several descriptive patterns with incomparable languages for the language to be inferred. Finally, Jain and Kinber [14] introduce the model of *ResAllMWSubEx*, which does not directly draw its motivation from the wish of investigating the inference of approximations of languages, but is nevertheless similar to upper approximate identification studied by Kobayashi and Yokomori [19]. More precisely, this model again considers convergence to a minimal generalisation of the language that is presented to the learner, but it allows hardly any control over these languages, since they can be any sublanguage of any language in the hypothesis space.

In summary, the main difference between our notion of descriptive generalisation and all related approaches described above is that we have a distinct split between a class $\mathcal{L}$ of languages to be inferred and an arbitrary class $\mathrm{PAT}_\star$ of pattern languages determining the set of admissible hypotheses. This leads to a compact and powerful model that yields interesting insights into the question of to which extent the generalisability of $\mathcal{L}$ depends on properties of $\mathcal{L}$ or of $\mathrm{PAT}_\star$. We discuss this topic in Section 4.2, and we demonstrate in Section 4.3 that descriptive generalisation can be interpreted as a natural instance of a very general and simple inference model which, to the best of our knowledge, has not been considered so far.

In Section 5, we investigate our model for a fixed and rich class $\mathrm{PAT}_\star$, namely the class of *terminal-free E-pattern languages*, i. e., the class of all pattern languages generated by patterns not containing any terminal symbols, where the empty word may be substituted for the variables in the pattern. Our studies reveal that, for this choice of $\mathrm{PAT}_\star$, descriptive generalisation and inductive inference from positive data are incomparable, and they show that there are major and natural classes of formal languages that can be descriptively generalised according to our model, but not precisely inferred in Gold's model. Technically, our decision to focus on terminal-free E-pattern languages leads to a number of substantial combinatorial challenges for pattern languages, and we present various respective insights and tools of intrinsic interest (see Sections 3 and 5.1).

## 2. Definitions

This paper is largely self-contained. For language theoretic and recursion theoretic notations not explicitly defined, Rozenberg and Salomaa [30] and Rogers [29] can be consulted, respectively.

Let $\mathbb{N} := \{0, 1, 2, 3, \ldots\}$ and let $\infty$ denote infinity. The symbols $\subseteq$, $\subset$, $\supseteq$ and $\supset$ refer to subset, proper subset, superset and proper superset relation, respectively. The symbols $\mathcal{P}$ and $\setminus$ denote the power set and the set difference, respectively. For an arbitrary alphabet $A$, a *string* (over $A$) is a finite sequence of symbols from $A$, and $\lambda$ stands for the *empty string*. The symbol $A^+$ denotes the set of all nonempty strings over $A$, and $A^* := A^+ \cup \{\lambda\}$.

For any alphabet $A$, a *language* $L$ (over $A$) is a set of strings over $A$, i. e., $L \subseteq A^*$. A language $L$ is *empty* if $L = \emptyset$; otherwise, it is *nonempty*. A *class* $\mathcal{L}$ of languages (over $A$) is a set of languages over $A$, i. e., $\mathcal{L} \subseteq \mathcal{P}(A^*)$. Let $\mathrm{FIN}_A$ denote the class of all finite languages over $A$.

For the *concatenation* of two strings $w_1, w_2$ we write $w_1 \cdot w_2$ or simply $w_1 w_2$. We say that a string $v \in A^*$ is a *factor* of a string $w \in A^*$ if there are $u_1, u_2 \in A^*$

such that $w = u_1 v u_2$. The notation $|K|$ stands for the size of a set $K$ or the length of a string $K$; the term $|w|_a$ refers to the number of occurrences of the symbol $a$ in the string $w$. For any $w \in \Sigma^*$ and any $n \in \mathbb{N}$, $w^n$ denotes the *n-fold concatenation of* $w$, with $w^0 := \lambda$. Furthermore, we use $\cdot$ and the regular operations $*$ and $+$ on sets and strings in the usual way.

For any alphabets $A, B$, a *morphism* is a function $h : A^* \to B^*$ that satisfies $h(vw) = h(v)h(w)$ for all $v, w \in A^*$. Given morphisms $g : A^* \to B^*$ and $g : B^* \to C^*$ (for alphabets $A$, $B$, $C$), their *composition* $h \circ g$ is defined by $h \circ g(w) := h(g(w))$ for all $w \in A^*$. For every morphism $h : A^* \to B^*$ and every language $L \subseteq A^*$, we define $h(L) := \bigcup_{w \in L} \{h(w)\}$.

A morphism $h : A^* \to B^*$ is said to be *nonerasing* if $h(a) \neq \lambda$ for all $a \in A$. For any string $w \in C^*$, where $C \subseteq A$ and $|w|_a \geq 1$ for every $a \in C$, the morphism $h : A^* \to B^*$ is called a *renaming (of w)* if $h : C^* \to B^*$ is injective and $|h(a)| = 1$ for every $a \in C$.

Let $\Sigma$ be a (finite or infinite) alphabet of so-called *terminal symbols* (or: *letters*) and $X$ an infinite set of *variables* with $\Sigma \cap X = \emptyset$. We normally assume $\{\mathtt{a}, \mathtt{b}, \ldots\} \subseteq \Sigma$ and $\{x_1, x_2, x_3 \ldots\} \subseteq X$. A *pattern* is a string over $\Sigma \cup X$, a *terminal-free pattern* is a string over $X$ and a *word* is a string over $\Sigma$. The set of all patterns over $\Sigma \cup X$ is denoted by $\mathrm{Pat}_\Sigma$. For any pattern $\alpha$, we refer to the set of variables in $\alpha$ as $\mathrm{var}(\alpha)$, and to the set of terminal symbols in $\alpha$ as $\mathrm{symb}(\alpha)$.

A morphism $\sigma : (\Sigma \cup X)^* \to (\Sigma \cup X)^*$ is called *terminal-preserving* if $\sigma(a) = a$ for every $a \in \Sigma$. A terminal-preserving morphism $\sigma : (\Sigma \cup X)^* \to \Sigma^*$ is called a *substitution*.

The *NE-pattern language* $L_{\mathrm{NE},\Sigma}(\alpha)$ of a pattern $\alpha \in \mathrm{Pat}_\Sigma$ is given by

$$L_{\mathrm{NE},\Sigma}(\alpha) := \{\sigma(\alpha) \mid \sigma : (\Sigma \cup X)^* \to \Sigma^* \text{ is a nonerasing substitution}\},$$

and the *E-pattern language* $L_{\mathrm{E},\Sigma}(\alpha)$ of $\alpha$ is given by

$$L_{\mathrm{E},\Sigma}(\alpha) := \{\sigma(\alpha) \mid \sigma : (\Sigma \cup X)^* \to \Sigma^* \text{ is a substitution}\}.$$

If the correspondence is clear, we write $L(\alpha)$ instead of $L_{\mathrm{E},\Sigma}(\alpha)$ or $L_{\mathrm{NE},\Sigma}(\alpha)$. Let $\mathrm{ePAT}_\Sigma$ denote the class of all E-pattern languages over $\Sigma$, and $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$ the class of all terminal-free E-pattern languages over $\Sigma$.

Let $\mathrm{PAT}_{\star,\Sigma}$ be a class of NE-pattern languages or a class of E-pattern languages over $\Sigma$. We say that a pattern $\delta \in (\Sigma \cup X)^+$ is $\mathrm{PAT}_{\star,\Sigma}$-*descriptive* of a language $L \subseteq \Sigma^*$ if and only if $L(\delta) \in \mathrm{PAT}_{\star,\Sigma}$, $L(\delta) \supseteq L$, and there is no pattern $\alpha$ with $L(\alpha) \in \mathrm{PAT}_{\star,\Sigma}$ satisfying $L \subseteq L(\alpha) \subset L(\delta)$. Furthermore, $D_{\mathrm{PAT}_{\star,\Sigma}}(L)$ denotes the set of all patterns that are $\mathrm{PAT}_{\star,\Sigma}$-descriptive of $L$.

Let $\mathcal{L}$ be a class of languages over some alphabet $A$. Then $\mathcal{L}$ is said to be indexable provided that there exists an indexing $(L_i)_{i \in \mathbb{N}}$ of languages $L_i$ such that, first, $\mathcal{L} = \{L_i \mid i \in \mathbb{N}\}$ and, second, there exists a total computable function $\chi$ which uniformly decides the membership problem for $(L_i)_{i \in \mathbb{N}}$ – i. e., for every $w \in A^*$ and for every $i \in \mathbb{N}$, $\chi(w, i) = 1$ if and only if $w \in L_i$ . In this case, we call $\mathcal{L} = (L_i)_{i \in \mathbb{N}}$ an *indexed family (of recursive languages)*. Of course, in this notation for an indexed family (which conforms with the use in the literature) the equality symbol "=" does not refer to an equality in the usual sense, but is merely a symbol indicating that $\mathcal{L}$ contains all languages in $(L_i)_{i\in\mathbb{N}}$ and vice versa.

## 3. Basic Results on Pattern Languages

Obviously, the definition of a descriptive pattern is based on the inclusion of pattern languages, which is an undecidable problem for both the full class of NE-pattern languages and the full class of E-pattern languages (cf. Jiang et al. [17], Freydenberger and Reidenbach [8]). A significant part of our subsequent technical considerations, however, is restricted to terminal-free E-pattern languages, where the inclusion problem is known to be decidable. This directly results from the following characterisation:

**Theorem 1** (Jiang et al. [17])**.** *Let $|\Sigma| \geq 2$. For every $\alpha, \beta \in X^+$, $L_{\mathrm{E},\Sigma}(\alpha) \subseteq L_{\mathrm{E},\Sigma}(\beta)$ holds if and only if there is a morphism $\phi : X^* \to X^*$ with $\phi(\beta) = \alpha$.*

Unfortunately, this problem is NP-complete:

**Theorem 2** (Ehrenfeucht and Rozenberg [5])**.** *Let $\Sigma$ be an alphabet with $|\Sigma| \geq 2$. Then the inclusion problem for* $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$ *is* NP*-complete.*

On the other hand, in conjunction with Reidenbach and Schneider [27], a recent result by Holub [13] demonstrates that the equivalence problem can be decided in polynomial time:

**Theorem 3.** *There is a polynomial-time algorithm deciding, for any pair of terminal-free patterns $\alpha, \beta$ and for any alphabet $\Sigma$ with $|\Sigma| \geq 2$, on whether $L_{\mathrm{E},\Sigma}(\alpha) = L_{\mathrm{E},\Sigma}(\beta)$.*

*Proof.* As mentioned by Reidenbach and Schneider [27], $L_{\mathrm{E},\Sigma}(\alpha) = L_{\mathrm{E},\Sigma}(\beta)$ holds if and only if the so-called *morphic roots* of $\alpha$ and $\beta$ are identical up to a renaming.

Thus, to decide whether $L_{\mathrm{E},\Sigma}(\alpha) = L_{\mathrm{E},\Sigma}(\beta)$ holds for patterns $\alpha, \beta \in X^+$, one first computes the morphic roots $\alpha_\rho$ and $\beta_\rho$ of $\alpha$ and $\beta$, respectively. According to Holub [13], this step can be executed in time that is polynomial in the length of the input.

In the second step, one checks whether there is a renaming mapping $\alpha_\rho$ to $\beta_\rho$. By definition of morphic roots, $|\alpha_\rho| \leq |\alpha|$ and $|\beta_\rho| \leq |\beta|$ hold; hence, this step can also be executed in polynomial time. □

As shown by Freydenberger and Reidenbach [9], not every language has an $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$- or an $\mathrm{ePAT}_\Sigma$-descriptive pattern:

**Theorem 4.** *There is an infinite sequence $(\beta_n)_{n\geq 0}$ over $X^+$ such that, for every alphabet $\Sigma$ with $|\Sigma| \geq 2$, $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L_\Sigma) = D_{\mathrm{ePAT}_\Sigma}(L_\Sigma) = \emptyset$ holds for the language $L_\Sigma := \bigcup_{n\geq 0} L_{\mathrm{E},\Sigma}(\beta_n)$.*

*Proof.* This follows immediately from the proof of Theorem 12 in [9]. □

Note that $L_\Sigma$ is an infinite language (and, in fact, an infinite union of languages from $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$). In contrast to this, Jiang et al. [16] show that every finite language has an $\mathrm{ePAT}_\Sigma$-descriptive pattern. This is also true when considering $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive patterns:

**Proposition 5.** *For every $\Sigma$ with $|\Sigma| \geq 2$ and every finite nonempty $S \in \mathrm{FIN}_\Sigma$, $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(S) \neq \emptyset$, and a $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(S)$ can be effectively computed.*

*Proof.* This can be shown using the same reasoning as in the proof of Theorem 8.1 in [16]. Let $S = \{w_1, \ldots, w_n\} \subset \Sigma^*$ for some $n \geq 1$ and denote

$$c_S := \sum_{i=1}^{n} |w_i|.$$

Our claim is that for every $\alpha \in X^+$ with $L_{\mathrm{E},\Sigma}(\alpha) \supset S$, there is a $\beta \in X^+$ with $L_{\mathrm{E},\Sigma}(\alpha) \supseteq L_{\mathrm{E},\Sigma}(\beta) \supseteq S$ and $|\beta| \leq c_S$. If $L_{\mathrm{E},\Sigma}(\alpha) \supset S$, there are morphisms $\phi_1, \ldots, \phi_n : X^* \to \Sigma^*$ with $\phi_i(\alpha) = w_i$ for $i \in \{1, \ldots, n\}$. Let $R := \{x \in \mathrm{var}(\alpha) \mid \phi_i(x) = \lambda \text{ for all } i\}$, define the morphism $\rho : X^* \to X^*$ through

$$\rho(x) := \begin{cases} \lambda & \text{if } x \in R, \\ x & \text{if } x \notin R, \end{cases}$$

and let $\beta := \rho(\alpha)$. It is easily seen that $|\beta| \leq c_S$ and $\phi_i(\beta) = (\phi_i \circ \rho)(\alpha) = \phi_i(\alpha) = w_i$ for every $i$. Thus, $L_{\mathrm{E},\Sigma}(\alpha) \supseteq L_{\mathrm{E},\Sigma}(\beta) \supset S$ holds.

As there are only finitely many terminal-free patterns (modulo renaming) of length at most $c_S$, there must be a pattern $\delta \in X^+$ with $|\delta| \leq c_S$, $L_{\mathrm{E},\Sigma}(\alpha) \supseteq L_{\mathrm{E},\Sigma}(\delta)$ and $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(S)$.

This also leads to the desired effective procedure that returns a pattern $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(S)$: As an initial value, choose $\alpha := x_1$. In every step, try to find a pattern $\beta \in X^+$ with $L_{\mathrm{E},\Sigma}(\alpha) \supset L_{\mathrm{E},\Sigma}(\beta) \supseteq S$ and $|\beta| \leq c_S$. If such a pattern is found, let $\alpha := \beta$, and search again. If no such pattern is found, $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(S)$ must hold. As the search space is finite, and all involved inclusions problems are decidable (cf. Theorem 1), the whole procedure is effective. □

While Proposition 5 yields the existence of an $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive pattern for every finite nonempty set, its proof makes use of a procedure that computes a descriptive pattern in a costly manner, since it solves the NP-complete inclusion problem (cf. Theorem 2) for an exponential number of patterns. Indeed, there is probably no algorithm that solves this problem in polynomial time:

**Theorem 6.** *If $\mathrm{P} \neq \mathrm{NP}$, then there is no polynomial-time algorithm that, for every alphabet $\Sigma$ and every finite set $S \subset \Sigma^*$ of words, computes a pattern that is $\mathrm{ePAT}_\Sigma$-descriptive of $S$.*

*Proof.* We prove the contraposition of the theorem. Thus, we assume that there is an algorithm $\chi_\delta$ computing, for every alphabet $\Sigma$ and every finite set $S \subseteq \Sigma^*$ of words, a pattern that is $\mathrm{ePAT}_\Sigma$-descriptive of $S$, and we shall use $\chi_\delta$ to decide the inclusion problem for terminal-free E-pattern languages in polynomial time. Since this problem is NP-complete (see Theorem 2), we can conclude P = NP.

Let $\alpha, \beta \in X^*$ be any patterns. W. l. o. g., we assume that $\mathrm{var}(\alpha) \cap \mathrm{var}(\beta) = \emptyset$. Let $\Sigma$ be an alphabet with $|\Sigma| = |\mathrm{var}(\alpha)| + |\mathrm{var}(\beta)|$, and let the morphism $r : (\mathrm{var}(\alpha) \cup \mathrm{var}(\beta))^* \to \Sigma^*$ be a renaming. This implies that there is an inverse morphism $r^{-1}$ with $r^{-1}(r(\alpha)) = \alpha$ and $r^{-1}(r(\beta)) = \beta$, and $\mathrm{symb}(r(\alpha)) \cap \mathrm{symb}(r(\beta)) = \emptyset$. We define $S := \{r(\alpha), r(\beta)\}$, and we use $\chi_\delta$ to compute a pattern $\delta$ that is $\mathrm{ePAT}_\Sigma$-descriptive of $S$. Since $\mathrm{symb}(r(\alpha)) \cap \mathrm{symb}(r(\beta)) = \emptyset$, $\delta$ is terminal-free. Furthermore, we can use the reasoning given by Jiang et al. [16] for their Theorem 8.1 to verify that $|\delta| \leq |r(\alpha)| + |r(\beta)|$. We can therefore use a polynomial-time algorithm $\chi_=$ (the existence of which is ensured by Theorem 3) to decide on whether $L_{\mathrm{E},\Pi}(\beta) = L_{\mathrm{E},\Pi}(\delta)$ for any alphabet $\Pi$ with $|\Pi| \geq 2$. Since the question of whether $L_{\mathrm{E},\Pi}(\beta)$ equals $L_{\mathrm{E},\Pi}(\delta)$ does not depend on the actual size of $\Pi$ and since $|\Sigma| \geq 2$, we may, w. l. o. g., define $\Pi := \Sigma$.

We now show that $L_{\mathrm{E},\Sigma}(\beta) = L_{\mathrm{E},\Sigma}(\delta)$ if and only if $L_{\mathrm{E},\Sigma}(\alpha) \subseteq L_{\mathrm{E},\Sigma}(\beta)$:

If $L_{\mathrm{E},\Sigma}(\alpha) \not\subseteq L_{\mathrm{E},\Sigma}(\beta)$, then, according to Theorem 1, there exists no morphism $\phi : X^* \to X^*$ satisfying $\phi(\beta) = \alpha$. Thus, there does not exist a substitution $\sigma$ with $\sigma(\beta) = r(\alpha)$, since otherwise the morphism $\phi := r^{-1} \circ \sigma$ would satisfy $\phi(\beta) = \alpha$. On the other hand, by definition, there exists a substitution $\tau$ with $\tau(\delta) = r(\alpha)$, and therefore $r(\alpha) \in L_{\mathrm{E},\Sigma}(\delta) \setminus L_{\mathrm{E},\Sigma}(\beta)$. This immediately implies $L_{\mathrm{E},\Sigma}(\beta) \neq L_{\mathrm{E},\Sigma}(\delta)$.

If $L_{\mathrm{E},\Sigma}(\alpha) \subseteq L_{\mathrm{E},\Sigma}(\beta)$, then there exist substitutions $\sigma, \sigma'$ with $\sigma(\beta) = r(\alpha)$ and $\sigma'(\beta) = r(\beta)$. Hence, $S \subseteq L_{\mathrm{E},\Sigma}(\beta)$. Furthermore, since $\delta$ is $\mathrm{ePAT}_\Sigma$-descriptive of $S$, there exists a substitution $\tau$ with $\tau(\delta) = r(\beta)$. Thus, $r^{-1} \circ \tau(\delta) = \beta$, and therefore, according to Theorem 1, $L_{\mathrm{E},\Sigma}(\beta) \subseteq L_{\mathrm{E},\Sigma}(\delta)$. If we now assume to the contrary that $L_{\mathrm{E},\Sigma}(\beta) \neq L_{\mathrm{E},\Sigma}(\delta)$, then this implies $L_{\mathrm{E},\Sigma}(\beta) \subset L_{\mathrm{E},\Sigma}(\delta)$. Consequently, $L_{\mathrm{E},\Sigma}(\delta) \supset L_{\mathrm{E},\Sigma}(\beta) \supseteq S$. This is a contradiction to the assumption that $\delta$ is $\mathrm{ePAT}_\Sigma$-descriptive of $S$. Hence, $L_{\mathrm{E},\Sigma}(\beta) = L_{\mathrm{E},\Sigma}(\delta)$.

Consequently, since $|r(\alpha)| + |r(\beta)|$ is polynomial in $|\alpha| + |\beta|$ and the runtimes of $\chi_\delta$ and $\chi_=$ are polynomial in $|r(\alpha)| + |r(\beta)|$, we have a polynomial-time algorithm deciding the inclusion problem for the class of terminal-free E-pattern languages over any alphabet $\Sigma$ with $|\Sigma| \geq 2$. Due to our initial remarks, this proves the theorem. □

Theorem 6 addresses a problem left open by Jiang et al. [16], and it provides a result that, apart from the fact that it depends on an unbounded terminal alphabet, is stronger than the corresponding statement by Angluin [1] on NE-descriptive patterns.

Since Theorem 6 can be proved using terminal-free patterns only, we can strengthen the corresponding result as follows:

**Corollary 7.** *If* P $\neq$ NP*, then there is no polynomial-time algorithm that, for every alphabet* $\Sigma$ *and every finite set* $S \subset \Sigma^*$ *of words, computes a pattern that is* $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$*-descriptive of* $S$*.*

## 4. Inferring Descriptive generalisations

In the present section, we formally introduce our notion of inferring descriptive generalisations, establish some of its basic properties (mainly by characterising, for any class of pattern languages determining the set of valid hypotheses, those indexed families that can be generalised in our model) and, finally, present a much more general inference paradigm that captures the essence of our approach. If we wish to compare Gold's well-known model of language identification in the limit from positive data (cf. Gold [12]) with our model, then we refer to the former occasionally as LIM-TEXT. We use the same notation for the *class* of all classes of languages that can be *inferred* in that model; the meaning of this term shall therefore follow from the context.

### *4.1. The Inference Paradigm*

We formalise our explanations of the model given in Section 1 as follows: For any alphabet $\Sigma$ and any nonempty language $L \subseteq \Sigma^*$, we call a total function $t : \mathbb{N} \to \Sigma^*$ a *text* of $L$ if and only if it satisfies $\{t(i) \mid i \in \mathbb{N}\} = L$. Moreover, for every text $t$ and every $n \in \mathbb{N}$, $t^n$ encodes the first $n$ values of $t$ in a single string, i. e., $t^n := t(1) \triangledown t(2) \triangledown t(3) \triangledown \cdots \triangledown t(n)$ with $\triangledown \notin \Sigma$; additionally, we define $t[n] := \{t(i) \mid i \leq n\}$. Finally, $\text{text}(L)$ denotes the set of all (computable and non-computable, repetitive and non-repetitive) texts of a language $L$.

Let $\mathcal{L}$ be a class of nonempty languages over an alphabet $\Sigma$, and let $\text{PAT}_{\star,\Sigma}$ be a class of NE-pattern languages or a class of E-pattern languages over $\Sigma$. Then $\mathcal{L}$ is $\text{PAT}_{\star,\Sigma}$*-descriptively generalisable* (or, if $\text{PAT}_{\star,\Sigma}$ is understood, *(descriptively) generalisable* for short) if and only if there exists a computable function $S : (\Sigma \cup \{\triangledown\})^* \to (\Sigma \cup X)^+$ such that, for every $L \in \mathcal{L}$ and for every $t \in \text{text}(L)$, $S(t^n)$ is defined for every $n \in \mathbb{N}$, and there is a $\delta \in (\Sigma \cup X)^+$ with $\delta \in D_{\text{PAT}_{\star,\Sigma}}(L)$ and there is an $m \in \mathbb{N}$ with $S(t^n) = \delta$ for every $n \geq m$. We call $S$ a *(generalisation) strategy* and, for every $n \in \mathbb{N}$, $S(t^n)$ a *hypothesis* of $S$. The notation $\text{DG}_{\text{PAT}_{\star,\Sigma}}$ refers to the class of all classes of languages that are $\text{PAT}_{\star,\Sigma}$-descriptively generalisable.

Consequently, and as already mentioned in Section 1, we have an inference model where the class to be inferred and the *hypothesis space* (we shall use this term in a rather informal manner for both the class $\text{PAT}_{\star,\Sigma}$ and any set $\text{Pat}_\star$ of patterns satisfying $\text{PAT}_{\star,\Sigma} = \{L_\Sigma(\alpha) \mid \alpha \in \text{Pat}_\star\}$) may be entirely different objects. We feel that this feature precisely reflects our motivation as outlined in Section 1, and it establishes the difference of our approach from a number of related models.

### *4.2. Fundamental Insights into the Model*

We now discuss some basic properties of descriptive generalisation without considering a specific class of pattern languages determining the hypothesis space. At first glance, the definitions of descriptive generalisation and of the LIM-TEXT model are closely related, and our first observation states that they are indeed equivalent if they are applied to any class of pattern languages:

**Proposition 8.** *Let* $\mathrm{PAT}_{\star,\Sigma}$ *be a class of pattern languages. Then* $\mathrm{PAT}_{\star,\Sigma} \in$ LIM-TEXT *if and only if* $\mathrm{PAT}_{\star,\Sigma} \in \mathrm{DG}_{\mathrm{PAT}_{\star,\Sigma}}$.

*Proof.* Directly from the definitions of LIM-TEXT and $\mathrm{DG}_{\mathrm{PAT}_{\star,\Sigma}}$. □

Thus, in such a lineup, insights into our model can directly be derived from the research on inductive inference of pattern languages (see, e. g., Angluin [1], Shinohara [31], Lange and Wiehagen [21], Wiehagen and Zeugmann [32], Reischuk and Zeugmann [28], Reidenbach [25, 26], Ng and Shinohara [24]).

While descriptive generalisation and inductive inference from positive data, thus, seem to be very similar, there are major differences between these two models. In fact, there are classes that can be descriptively generalised, although neither the class nor the hypothesis space can be exactly inferred from positive data:

**Proposition 9.** *There exists a class* $\mathcal{L}$ *of languages and a class* $\mathrm{PAT}_{\star,\Sigma}$ *of pattern languages satisfying* $\mathcal{L} \notin$ LIM-TEXT, $\mathrm{PAT}_{\star,\Sigma} \notin$ LIM-TEXT, *and* $\mathcal{L} \in \mathrm{DG}_{\mathrm{PAT}_{\star,\Sigma}}$.

*Proof.* The statement follows from our Corollaries 26 and 31 in Section 5 and the fact that $\mathrm{ePAT}_{\mathrm{tf},\Sigma} \notin$ LIM-TEXT for $|\Sigma| = 2$ (cf. Reidenbach [25]). □

Since the definition of descriptive generalisation allows any class of pattern languages to be chosen as a hypothesis space, we can even devise a maximally powerful (yet utterly useless) generalisation strategy:

**Proposition 10.** *Let* $\Sigma$ *be an alphabet. There exists a class* $\mathrm{PAT}_{\star,\Sigma}$ *of pattern languages such that every class* $\mathcal{L}$ *of languages over* $\Sigma$ *satisfies* $\mathcal{L} \in \mathrm{DG}_{\mathrm{PAT}_{\star,\Sigma}}$.

*Proof.* Let $\mathrm{PAT}_{\star,\Sigma} := \{L_{\mathrm{E},\Sigma}(x_1)\}$. Since $x_1$ is $\mathrm{PAT}_{\star,\Sigma}$-descriptive of every language $L \subseteq \Sigma^*$, a strategy $S$ that constantly outputs $x_1$ generalises $\mathcal{L}$. □

Obviously, the substantial gap between the LIM-TEXT model and descriptive generalisation illustrated by Proposition 10 is based on a proof that uses a trivial notion of descriptiveness. In Section 5, we shall demonstrate that there are similarly deep differences between both models if a natural and nontrivial class of pattern languages, namely $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$, is used as admissible hypotheses for the generalisation process.

The main result of the present section is the following characterisation of descriptively generalisable indexed families of recursive languages. While our model as well as our studies in Section 5 consider descriptive generalisations of arbitrary classes of languages, this restriction facilitates an interesting comparison of our result to Angluin's characterisation of those indexed families that are inferable in the LIM-TEXT model (see [2]). It is also worth noting that the following argument cannot be based on strong insights into the descriptiveness of patterns, since we deal with arbitrary classes of pattern languages.

**Theorem 11.** *Let $\Sigma$ be an alphabet, let $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ be an indexed family of nonempty recursive languages over $\Sigma$, and let $\mathrm{PAT}_{\star,\Sigma}$ be a class of pattern languages. $\mathcal{L} = (L_i)_{i\in\mathbb{N}} \in \mathrm{DG}_{\mathrm{PAT}_{\star,\Sigma}}$ if and only if there are effective procedures $d$ and $f$ satisfying the following conditions:*

*(i) For every $i \in \mathbb{N}$, there exists a $\delta_{d(i)} \in D_{\mathrm{PAT}_{\star,\Sigma}}(L_i)$ such that procedure $d$, on input $i$, enumerates a sequence of patterns $d_{i,0}, d_{i,1}, d_{i,2}, \dots$ satisfying, for all but finitely many $j \in \mathbb{N}$, $d_{i,j} = \delta_{d(i)}$.*

*(ii) For every $i \in \mathbb{N}$, procedure $f$, on input $i$, enumerates a finite set $F_i \subseteq L_i$ such that, for every $j \in \mathbb{N}$ with $F_i \subseteq L_j$, if $\delta_{d(i)} \notin D_{\mathrm{PAT}_{\star,\Sigma}}(L_j)$, then there is a $w \in L_j$ with $w \notin L_i$.*

Before we give a proof of Theorem 11, we wish to point out that the index $d(i)$ of the pattern $\delta_{d(i)}$ does not refer to an enumeration of the patterns in $D_{\mathrm{PAT}_{\star,\Sigma}}(L_i)$, but is merely used as an identifier of this particular pattern. Hence, in other words, $\delta_{d(i)}$ is nothing but the pattern that $d$ converges to when given input $i$.

*Proof.* We begin with the *if* direction. In our proof, $F_i^{(m)}$ refers to the subset of $F_i$ that is enumerated by $f$ in $m \in \mathbb{N}$ steps of the computation.

We define a generalisation strategy $S$ as follows: For any text $t$ and for any $m \in \mathbb{N}$, when given $t^m$ as an input, $S$ outputs the pattern $d_{i,m}$, where $i \in \mathbb{N}$ is the smallest index satisfying:

(a) $t[m] \subseteq L_i$ and

(b) $F_i^{(m)} \subseteq t[m]$.

If no such $i$ exists, then $S$ outputs $d_{0,0}$.

Since $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ is an indexed family, which means that the membership problem is uniformly decidable for all $i$ and for all $w \in \Sigma^*$, and $d$ and $f$ are effective, it is obvious that $S$ is computable and defined for every input $t^m$.

We now demonstrate that $S$ $\mathrm{PAT}_{\star,\Sigma}$-descriptively generalises $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ if (i) and (ii) are satisfied. Thus, we choose an arbitrary $n \in \mathbb{N}$ and an arbitrary text $t$ of $L_n$, and we show that $S$, when reading $t$, converges to a pattern that is $\mathrm{PAT}_{\star,\Sigma}$-descriptive of $L_n$. Before we start our actual reasoning, we determine a value $m_0 \in \mathbb{N}$ such that a number of vital parameters for the computation of $S(t^{m_0})$ have already stabilised:

Let $m_1 \in \mathbb{N}$ be sufficiently large such that, for every $k \in \mathbb{N}$ with $k \leq n$ and $L_k \not\supseteq L_n$, $t[m_1]$ contains a word $w$ satisfying $w \notin L_k$. The value $m_1$ must exist since $L_k \not\supseteq L_n$ and $t$ is a text of $L_n$.

Let $m_2 \in \mathbb{N}$ be sufficiently large such that, for every $k \in \mathbb{N}$ with $k \leq n$, $d_{k,m} = \delta_{d(k)}$ for every $m \geq m_2$. The value $m_2$ must exist due to (i).

Let $m_3 \in \mathbb{N}$ be sufficiently large such that, for every $k \in \mathbb{N}$ with $k \leq n$, $F_k^{(m_3)} = F_k$. The value $m_3$ must exist since, according to (ii), $F_k$ is finite.

Let $m_4 \in \mathbb{N}$ be sufficiently large such that $F_n \subseteq t[m_4]$. The value $m_4$ must exist since $t$ is a text of $L_n$ and, according to (ii), $F_n \subseteq L_n$.

Then $m_0 := \max\{m_1, m_2, m_3, m_4, n\}$.

Referring to these definitions, our proof of the *if* direction is based on the following Claims:

*Claim 1.* For every $m \geq m_0$, $n$ satisfies $t[m] \subseteq L_n$, and $F_n^{(m)} \subseteq t[m]$.

*Proof (Claim 1).* The first part of the statement holds since $t$ is a text of $L_n$; the second part holds because of $m \geq m_0 \geq m_4$. □ (Claim 1)

*Claim 2.* For every $m \geq m_0$, $S(t^m) \in D_{\mathrm{PAT}_{\star,\Sigma}}(L_n)$.

*Proof (Claim 2).* Let $S(t^m) = d_{k,m}$. By definition, $S$ outputs the pattern $d_{k,m}$ for the smallest index $k \leq m$ satisfying conditions (a) and (b), or it outputs the auxiliary hypothesis $d_{0,0}$ if such a $k$ does not exist. Due to Claim 1, and since $m \geq m_0$, we know that there exists at least one index (namely $n$) satisfying (a) and (b) for $t^m$. Thus, $m \geq m_0 \geq n$ implies that $S$ does not choose to output its auxiliary hypothesis. Therefore, the following statements hold true for $k$:

(1) $k \leq n$ (since $S$ outputs $d_{k,m}$ for the smallest $k$ satisfying conditions (a) and (b) of the definition of $S$),

(2) $d_{k,m} = \delta_{d(k)}$ (because of $m \geq m_0 \geq m_2$ in conjunction with statement (1)),

(3) $t[m] \subseteq L_k$ (because of condition (a)), and

(4) $F_k \subseteq t[m] \subseteq L_n$ (because of $m \geq m_0 \geq m_3$ in conjunction with statement (1), and due to condition (b)).

Now assume to the contrary that $S(t^m) = d_{k,m} \notin D_{\mathrm{PAT}_{\star,\Sigma}}(L_n)$. Due to statement (2), this means that $\delta_{d(k)} \notin D_{\mathrm{PAT}_{\star,\Sigma}}(L_n)$. Then statement (4) and condition (ii) of the Theorem imply that there exists a word $w \in L_n \setminus L_k$ which, due to $m \geq m_0 \geq m_1$ in conjunction with statement (1), satisfies $w \in t[m]$. Hence, $t[m] \not\subseteq L_k$. This contradicts statement (3). □ (Claim 2)

*Claim 3.* There is a pattern $\delta$ and an $m' \geq m_0$ such that, for every $m \geq m'$, $S(t^m) = \delta$.

*Proof (Claim 3).* Due to statement (1) in the proof of Claim 2 and $m \geq m_0 \geq m_2$, there is only a finite number of possible hypotheses – namely $\delta_{d(0)}, \delta_{d(1)}, \ldots, \delta_{d(n)}$ – that $S$ can output when reading $t^m$. Therefore, it is sufficient to show that a hypothesis, once it has been discarded, is not chosen by $S$ anymore. More precisely, we prove that if, for an $l_0 \geq m$, $S(t^{l_0}) = \delta_{d(k)}$ and $S(t^{l_0+1}) \neq \delta_{d(k)}$, then, for every $l \geq l_0 + 1$, there exists a $k' \neq k$ with $S(t^l) = \delta_{d(k')}$.

Since $l_0 \geq m \geq m_3$, $S(t^{l_0}) = \delta_{d(k)}$ implies

(A) $t[l_0] \subseteq L_k$ and

(B) $F_k \subseteq t[l_0]$.

By definition, $t[l_0+1] \supseteq t[l_0]$, and therefore (B) is satisfied for $t^{l_0+1}$, too. Thus, the only event that can trigger a change of the hypothesis when extending $t^{l_0}$ to $t^{l_0+1}$ is $t[l_0+1] \not\subseteq L_k$; this implies $M \not\subseteq L_k$ for all supersets $M$ of $t[l_0+1]$. Hence, for every $l \geq l_{0+1}$, there is a $k' \neq k$ with $S(t^l) = \delta_{d(k')}$. □ (Claim 3)

To summarise, Claim 3 shows that $S$, when reading $t$, converges to a pattern $\delta$, and Claim 2 demonstrates that $\delta$ is $\mathrm{PAT}_{\star,\Sigma}$-descriptive of $L_n$. This concludes the proof of the *if* direction.

We continue with the *only if* direction. Hence, let $S$ be a computable generalisation strategy that $\mathrm{PAT}_{\star,\Sigma}$-descriptively generalises $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$, i. e., for every $i$ and for every text $t$ of $L_i$, $S$ converges to a pattern that is $\mathrm{PAT}_{\star,\Sigma}$-descriptive of $L_i$. We show that this implies the existence of effective procedures $d$ and $f$ satisfying conditions (i) and (ii).

Since $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ is an indexed family, there is an effective procedure enumerating, for every $i \in \mathbb{N}$, all words $w_{i,0}, w_{i,1}, w_{i,2}, \dots$ in $L_i$. Furthermore, we can use this to define a second effective procedure which enumerates, for every $i \in \mathbb{N}$, all finite sequences $s_{i,0}, s_{i,1}, s_{i,2}, \dots$ of words in $L_i$. Note that every sequence $s_{i,j}$, $j \in \mathbb{N}$, may contain repetitions of words. Furthermore, if $L_i$ is finite, we can nevertheless easily make sure that the output of the above procedures is infinite for every $i$.

We now give a procedure that defines the behaviour of $d$ and $f$:

Procedure SIM_*S*

Let $i \in \mathbb{N}$, and let $w_{i,0}, w_{i,1}, w_{i,2}, \dots$ and $s_{i,0}, s_{i,1}, s_{i,2}, \dots$ be as given above. Go to Stage 0.

*Stage* 0. Define $t_0 := w_{i,0}$, $F_i := \{w_{i,0}\}$. Define $x := 0$ and $d_{i,x} := S(t_0)$. Go to Stage 1.

*Stage* $n$ $(n \geq 1)$. For every $j = 0, 1, 2, \dots$ proceed as follows: Consider $s_{i,j} = (\widehat{w}_{j,0}, \widehat{w}_{j,1}, \dots, \widehat{w}_{j,y})$, $y \in \mathbb{N}$, and define $t'_j := \widehat{w}_{j,0} \nabla \widehat{w}_{j,1} \nabla \cdots \nabla \widehat{w}_{j,y}$. Define $x := x + 1$ and $d_{i,x} = S(t_{n-1} \nabla t'_j)$. If $S(t_{n-1} \nabla t'_j) \neq S(t_{n-1})$, then define $t_n := t_{n-1} \nabla t'_j \nabla w_{i,n}$, $F_i := F_i \cup \{\widehat{w}_{j,0}, \widehat{w}_{j,1}, \dots, \widehat{w}_{j,y}, w_{i,n}\}$, and go to Stage $n+1$.

Since $S$ and the procedures enumerating the $w_{i,j}$ and $s_{i,j}$, $i, j \in \mathbb{N}$, are computable, the same holds for SIM_*S*. Consequently, effective procedures $d$ and $f$ which, for all $i \in \mathbb{N}$, uniformly produce sequences $d_{i,0}, d_{i,1}, d_{i,2}, \dots$ and enumerate $F_i$, respectively, can be directly derived from SIM_*S*.

We now show that $d$ and $f$ satisfy conditions (i) and (ii). Our corresponding reasoning makes use of the following fact:

*Claim 4.* For every $i \in \mathbb{N}$ there exists an $n_0$ such that procedure SIM_*S*, when given input $i$, enters Stage $n_0$, but it does not enter Stage $n_0 + 1$.

*Proof (Claim 4).* Assume to the contrary that procedure SIM_*S* enters an infinite number of stages. This implies that $S$ does not converge to a fixed pattern, since SIM_$S$ goes to the next stage if and only if $S$ changes its hypothesis for the given input. However, since all considered words are contained in $L_i$, each transition from Stage $n$ to Stage $n+1$ adds the word $w_{i,n}$ to $t_n$, and $\{w_{i,j} \mid j \in \mathbb{N}\} = L_i$, the string $\lim_{n\to\infty} t_n$ is an encoding of a text $t$ of $L_i$. Since $S$ $\mathrm{PAT}_{\star,\Sigma}$-descriptively generalises $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$, this means that $S$, when reading $t$, must converge to a pattern. This is a contradiction. □ (Claim 4)

By definition, for every $i \in \mathbb{N}$, SIM_$S$ produces an infinite sequence of patterns. It outputs a pattern $d_{i,x+1}$ that differs from $d_{i,x}$ only if it moves from one stage to another. Thus, due to Claim 4 and for the corresponding $x_0 \in \mathbb{N}$, the sequence of patterns $d_{i,x_0}, d_{i,x_0+1}, d_{i,x_0+2}, \dots$ produced in Stage $n_0$ satisfies, for every $j \in \mathbb{N}$, $d_{i,x_0+j} = d_{i,x_0}$. Furthermore, due to fact that the constructed input is a text of $L_i$, $S$ needs to converge to a $\mathrm{PAT}_{\star,\Sigma}$-descriptive pattern of $L_i$. This implies that $d_{i,x_0} = \delta_{d(i)}$ for a $\delta_{d(i)} \in D_{\mathrm{PAT}_{\star,\Sigma}}(L_i)$. Consequently, the sequence of patterns $d_{i,0}, d_{i,1}, d_{i,2}, \dots$ satisfies condition (i).

SIM_$S$ adds a finite number of words to $F_i$ if and only if it moves to the next stage. Hence, Claim 4 shows that every $F_i$ is finite. Now assume to the contrary that, for an $i \in \mathbb{N}$, $F_i$ does not satisfy condition (ii), i. e., there exists a $j \in \mathbb{N}$ with $F_i \subseteq L_j$, $\delta_{d(i)} \notin D_{\mathrm{PAT}_{\star,\Sigma}}(L_j)$ and $L_j \subseteq L_i$. Let $t_{<j>}$ be an arbitrary text of $L_j$. Since $F_i \subseteq L_j$ and $t_{n_0-1}$ encodes the words in $F_i$, for every $m \in \mathbb{N}$, $t_{n_0-1} \nabla t^m_{<j>}$ is an encoding of initial values of a text of $L_j$. Thus, for $m \to \infty$, $S$ must, when reading $t_{n_0-1} \nabla t^m_{<j>}$, converge to a pattern that is $\mathrm{PAT}_{\star,\Sigma}$-descriptive of $L_j$. According to Claim 4, when $t_{n_0-1}$ is continued with the encoding of any finite sequence of words from $L_i$, SIM_$S$ does not leave Stage $n_0$. Since $L_j \subseteq L_i$, this implies that SIM_$S$ does not leave Stage $n_0$ for $t_{n_0-1}$ being continued with the encoding of any finite sequence of initial values of $t_{<j>}$. Therefore $S$ converges, when given $t_{n_0-1} \nabla t^m_{<j>}$ for $m = 0, 1, 2, \dots$, by definition to $\delta_{d(i)}$. This contradicts $\delta_{d(i)} \notin D_{\mathrm{PAT}_{\star,\Sigma}}(L_j)$. Consequently, $F_i$ satisfies condition (ii), and this concludes the proof of the *only if* direction.

Hence, $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ is $\mathrm{PAT}_{\star,\Sigma}$-descriptively generalisable if and only if there are effective procedures $d$ and $f$ satisfying conditions (i) and (ii). □

As briefly mentioned above, Theorem 11 shows natural connections to the seminal characterisation of learnable indexed families given by Angluin [2], and therefore it is not surprising that some elements of our proof do not need to differ from hers. Most of these similarities result from the fact that each successful inductive inference process requires the existence of so-called *locking sequences* (see [22] for a detailed discussion), and this is reflected by Angluin's *telltale* $T_i$ and our comparable concept $F_i$. Nevertheless, there are crucial differences between the two characterisations. First, we need to define an enumeration of an appropriate subset of our hypothesis space (this is done by the procedure $d$), whereas this is automatically given in Angluin's model. In this context, it is important to note that we have to attune the set $F_i$ to the pattern $\delta_{d(i)}$, $i \in \mathbb{N}$, which leads to $d$ and $f$ being defined by the same procedure SIM_$S$. Second, while Angluin's $T_i$ must, for *every* $j$ with $L_j \subset L_i$, contain a word from $L_i \setminus L_j$, our equivalent $F_i$ only needs to do so if $\delta_{d(i)}$ is not an acceptable hypothesis for $L_j$. This fits with the requirement of inductive inference from positive examples to distinguish between *all* languages $L_i$ and $L_j$ with $L_i \neq L_j$, whereas descriptive generalisation only has to distinguish between some of them, and this requisite might be *asymmetric*, i. e., a strategy $S$ might have to discover that a text of a language $L_i$ is not a text of a language $L_j$, but it might not need to figure out that a text of $L_j$ is not a text of $L_i$. The explanation of why descriptive generalisation, for many cases, is more powerful than inductive inference from positive data directly follows from this observation; further considerations on this topic are given in Section 4.3. Thirdly, and finally, the strategy $S$ we deploy in our proof is, in a sense, not optimal, as it might discard a correct hypothesis – i. e. pattern $\delta_{d(j)}$ that incidentally is descriptive of the language $L_i$ the text of which is read – simply because $L_i$ contains a word that is not contained in $L_j$.

Our generic strategy $S$ of course is not very efficient; furthermore, it has the bothersome property described in the third point in the preceding paragraph. However, it is worth mentioning that $S$ does not test whether the given words are contained in the language of the hypothesis pattern, and it does not check the inclusion of pattern languages, either. Thus, it circumvents two decision problems that, for many natural classes of pattern languages, are known to be NP-complete or even undecidable (see, e. g., [1] and [8]), although these decision problems are essential elements of the definition of descriptiveness. Instead, $S$ infers descriptive patterns purely based on membership tests for the languages in the indexed family. Thus, if indexed families with a fast membership test are to be generalised, then our strategy raises hope that it might be possible to do this efficiently in spite of using a hypothesis space with an NP-complete membership test. On the other hand, it might be difficult to find rich classes of pattern languages where the procedure $d$ introduced by Theorem 11 is efficient (even though it should normally be possible to devise a $d$ that, for every $i \in \mathbb{N}$, directly outputs the pattern $\delta_{d(i)}$ instead of enumerating the sequence $d_{i,j}$). This expectation is substantiated by Theorem 4.2 in [1] and our Theorem 6 and Corollary 7 given in Section 2.

### *4.3. A More General View*

While an application of Theorem 11 might require profound knowledge of the descriptiveness of patterns, a closer look confirms our above remark that the actual characterisation and its proof do not at all. More precisely, neither the Theorem nor our reasoning deal with the properties of the descriptive patterns $\delta_{d(i)}$, $i \in \mathbb{N}$, but they merely make use of a notion of the *validity* of a hypothesis for a given language, i. e., a hypothesis is valid for a language if it is descriptive, but we do not check for descriptiveness. This view is quite convenient to study the difference between descriptive generalisation and inductive inference from positive data. In the LIM-TEXT model when applied to indexed families, a hypothesis $i$ – i. e., the index of the language $L_i$ – is valid for a language $L_j$, $j \neq i$, if and only if the hypothesis $j$ is valid for the language $L_i$ (if and only if $L_i = L_j$). In our model, this symmetry does not necessarily exist, as demonstrated by the following example:

**Example 12.** *Let $\Sigma := \{\mathtt{a}, \mathtt{b}\}$. Let $L_1 := \{\mathtt{ababa}, \mathtt{babab}\}$ and $L_2 := \{\mathtt{ababa}, \mathtt{babab}, \mathtt{abaaba}\}$. We state without proof that $\delta_1 := x_1\,\mathtt{abab}\,x_2$ is $\mathrm{ePAT}_\Sigma$-descriptive of $L_1$ and $\delta_2 := x_1x_2x_1x_2x_1$ is $\mathrm{ePAT}_\Sigma$-descriptive of $L_2$. While $\delta_2$ is also $\mathrm{ePAT}_\Sigma$-descriptive of $L_1$, $\delta_1$ is not $\mathrm{ePAT}_\Sigma$-descriptive of $L_2$. Hence, a strategy $S$ that $\mathrm{ePAT}_\Sigma$-descriptively generalises a class including $L_1$ and $L_2$ can output $\delta_1$ or $\delta_2$ when reading a text for $L_1$, but it must not output $\delta_1$ when reading a text for $L_2$.*

Referring to this phenomenon and restricted to indexed families, we can now give a much more general model of inference than the one of descriptive generalisation, and we can still characterise those indexed families that can be inferred according to this model in exactly the same way as we have done in Theorem 11. Hence, let $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ be an indexed family. Furthermore, for any $i \in \mathbb{N}$, let HYP be a function that maps $i$ to a subset of $\mathbb{N}$ that consists of all *valid hypotheses* for $L_i$. Here it is important to note that the numbers in $\mathrm{HYP}(i)$ do normally not refer to indices of the indexed family $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$; e. g., in our model of descriptive generalisation they would stand for indices in an arbitrary enumeration of a set of patterns. We then say that $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ is *inductively inferable with hypotheses validity relation* HYP if and only if there exists a computable function $S : (\Sigma \cup \{\nabla\})^* \to \mathbb{N}$ such that, for every $i \in \mathbb{N}$ and for every $t \in \mathrm{text}(L_i)$,

1. $S(t^n)$ is defined for every $n \in \mathbb{N}$ and
2. there is a $j \in \mathrm{HYP}(i)$ and there is an $m \in \mathbb{N}$ with $S(t^n) = j$ for every $n \geq m$.

Our notion of descriptive generalisation demonstrates that there are natural instances of the model of inductive inference with hypotheses validity relation HYP. Nevertheless, to the best of our knowledge, its properties have not been explicitly studied so far.

As announced above, we now rephrase Theorem 11 so that it characterises those indexed families that are inductively inferable with hypotheses validity relation HYP:

**Theorem 13.** *Let $\Sigma$ be an alphabet, let $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ be an indexed family of nonempty languages over $\Sigma$, and let $\mathrm{HYP} : \mathbb{N} \to \mathcal{P}(\mathbb{N})$ be a function. $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ is inductively inferable with hypotheses validity relation* HYP *if and only if there are effective procedures $h$ and $f$ satisfying the following conditions:*

*(i) For every $i \in \mathbb{N}$, there exists a $\eta_i \in \mathrm{HYP}(i)$ such that procedure $h$, on input $i$, enumerates a sequence of natural numbers $i_0, i_1, i_2, \ldots$ satisfying, for all but finitely many $k \in \mathbb{N}$, $i_k = \eta_i$.*

*(ii) For every $i \in \mathbb{N}$, procedure $f$, on input $i$, enumerates a finite set $F_i \subseteq L_i$ such that, for every $j \in \mathbb{N}$ with $F_i \subseteq L_j$, if $\eta_i \notin \mathrm{HYP}(j)$, then there is a $w \in L_j$ with $w \notin L_i$.*

*Proof.* Minor and straightforward editing of the proof of Theorem 11 – mainly substituting $h$ for $d$, $i_k$ for $d_{i,k}$, $\eta_i$ for $\delta_{d(i)}$, and $\mathrm{HYP}(i)$ for $D_{\mathrm{PAT}_{\star,\Sigma}}(L_i)$ – turns it into a reasoning suitable for Theorem 13. □

To conclude this section on basic properties of our model, we wish to mention that descriptive generalisation can alternatively be interpreted as inductive inference of classes of pattern languages from *partial texts*. Hence, we can understand any language $L_i$ as a tool to define texts that do not contain all words in $L(\delta_{d(i)})$, but nevertheless can be used to infer $\delta_{d(i)}$. Within the scope of the present paper, we do not explicitly discuss such a view, but we expect that it might be a worthwhile topic for further considerations. There exist some studies of related approaches, for example by Fulk and Jain [11] and Jain and Kinber [14]. However, the corresponding paradigm $\mathit{In}^a\,\mathit{TxtEx}^b$ in the paper [11] uses the generic parameter of the *number a* of words of the target language missing from the text given to the learner, and – as briefly described in Section 1 – the model *ResAllMWSubEx* in the paper [14] postulates inferrability for *all* sublanguages of the target language. Hence, unlike the motivation of our inference paradigm, these models do not directly provide a strong control of the nature of the languages to be inferred, and therefore we anticipate that the interpretation of our model as inference from incomplete texts might involve substantial conceptual challenges that cannot be solved using the established approaches in the literature.

## 5. Inferring $\mathbf{ePAT_{tf,\Sigma}}$-Descriptive Patterns

We now study our model for a specific hypothesis space, namely the class $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$. The decidability of the inclusion problem for this class (see Theorem 1) allows us to develop a set of powerful tools.

This section is divided into three parts. In the first part, we consider some questions on the existence of $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive patterns for various classes of languages and develop a set of basic concepts and insights in order to simplify corresponding proofs.

The second part deals with a generalisation strategy that is based on the procedure that is described in Proposition 5, which we deem so natural that we call it the canonical strategy Canon for $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive generalisations. Most importantly, we give a characterisation of the class $\mathcal{TSL}_\Sigma$ of languages that can be descriptively generalised with Canon.

In the final part of this section, we examine the relationship of various classes of languages to $\mathcal{TSL}_\Sigma$ in order to gain further insights into $\mathrm{DG}_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}$ and the power of Canon.

### *5.1. Basic tools*

We begin the present part of Section 5 with a short remark illustrating that there are finite classes of languages which are not contained in $\mathrm{DG}_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}$:

**Proposition 14.** *Let $\Sigma$ be an alphabet, $|\Sigma| \geq 2$. There exists a class $\mathcal{L}$ of nonempty languages over $\Sigma$ with*

- $|\mathcal{L}| = 1$ *and*
- $\mathcal{L} \notin \mathrm{DG}_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}$.

*Proof.* As stated in Theorem 4, there is a language $L_\Sigma$ with $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L_\Sigma) = \emptyset$. If we choose $\mathcal{L} = \{L_\Sigma\}$, then $|\mathcal{L}| = 1$ holds, and no strategy will be able to compute any hypothesis that is $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive of $L_\Sigma$. □

Before we proceed to a closer examination of $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive generalisation in the next part of this section, we develop some tools and techniques that simplify the work with $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive patterns, and gather some results on the existence and nonexistence of such patterns for some classes of languages. We begin with the following result:

**Lemma 15.** *Let $\Sigma$ be an alphabet with $|\Sigma| \geq 2$, and let $L_1, L_2 \subseteq \Sigma^*$ with $L_1 \supseteq L_2$. If there is a $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L_2)$ with $L_{\mathrm{E},\Sigma}(\delta) \supseteq L_1$, then $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L_1)$.*

*Proof.* Assume to the contrary that there is a $\gamma \in X^+$ with $L_{\mathrm{E},\Sigma}(\delta) \supset L_{\mathrm{E},\Sigma}(\gamma) \supseteq L_1$. Due to $L_1 \supseteq L_2$, this would imply $L_{\mathrm{E},\Sigma}(\delta) \supset L_{\mathrm{E},\Sigma}(\gamma) \supseteq L_2$ and contradict $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L_2)$. Therefore, $\delta$ is descriptive of $L_1$. □

This observation might seem to be elementary, but together with Lemma 18, it forms the basis of the proofs of almost all results in this section. The technical base of that Lemma derives from a phenomenon that often arises when dealing with $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive patterns. We consider the following example:

**Example 16.** *Let $\Sigma := \{\mathtt{a}, \mathtt{b}\}$ and let*

$$\begin{aligned} L_1 &:= \{\mathtt{a}^2\}, \\ L_2 &:= \{(\mathtt{a}\,\mathtt{b}^1\,\mathtt{a}\;\mathtt{a}\,\mathtt{b}^2\,\mathtt{a}\cdots\mathtt{a}\,\mathtt{b}^n\,\mathtt{a})^2 \mid n \geq 2\}, \\ L_3 &:= L_{\mathrm{E},\Sigma}(x_1^2) \setminus \{\mathtt{a}^2, \mathtt{b}^2\}. \end{aligned}$$

*It is easy to see that all three languages are included in $L_{\mathrm{E},\Sigma}(x_1^2)$. However, in addition to this, for every $\alpha \in X^+$ with $L_{\mathrm{E},\Sigma}(\alpha) \supseteq L_i$ (with $1 \leq i \leq 3$), $L_{\mathrm{E},\Sigma}(\alpha) \supseteq L_{\mathrm{E},\Sigma}(x_1^2)$ holds as well. For $L_1$, this is obvious. For $L_2$, assume that $L_{\mathrm{E},\Sigma}(\alpha) \supseteq L_2$ for some $\alpha \in X^+$, let $n := |\alpha|$ and $w := (\mathtt{a}\,\mathtt{b}^1\,\mathtt{a}\,\mathtt{a}\,\mathtt{b}^2\,\mathtt{a}\cdots\mathtt{a}\,\mathtt{b}^n\,\mathtt{a})^2 \in L_2$, and choose any morphism $\phi$ with $\phi(\alpha) = w$. As $w$ contains $n$ distinct factors of the form $\mathtt{a}\,\mathtt{b}^+\,\mathtt{a}$, each occurring exactly twice, there must be an $x \in \mathrm{var}(\alpha)$ such that $\phi(x)$ contains at least one complete occurrence of such a segment, which implies $|\alpha|_x \in \{1, 2\}$. In both cases, we can construct a morphism $\psi$ with $\psi(\alpha) = x_1^2$ (by mapping $x$ to $x_1$ or $x_1^2$ and erasing all other variables), which (according to Theorem 1) leads to $L_{\mathrm{E},\Sigma}(\alpha) \supseteq L_{\mathrm{E},\Sigma}(x_1^2)$. Finally, as $L_3 \supset L_2$, this also proves the claim for $L_3$.*

*As $L_{\mathrm{E},\Sigma}(x_1^2)$ and all three $L_i$ have exactly the same "superpatterns", we are able to conclude that, for every $i \in \{1, 2, 3\}$, $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L_{\mathrm{E},\Sigma}(x_1^2)) = D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L_i)$. In other words, although the four languages might seem rather different, they have exactly the same set of* $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$*-descriptive patterns.*

When generalising languages using $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive patterns, every language has a certain superset that is covered by every descriptive generalisation of this language, and cannot be avoided. In order to formalise this line of reasoning (and in order to use this phenomenon), we introduce the set of *superpatterns* $\mathrm{Super}(L)$, and the *superpattern hulls* $\text{S-Hull}_\Sigma(L)$, which are defined as

$$\mathrm{Super}(L) := \{\alpha \in X^+ \mid \text{for every } w \in L, \text{ there is a morphism } \sigma \text{ with } \sigma(\alpha) = w\},$$

$$\text{S-Hull}_\Sigma(L) := \bigcap_{\alpha \in \mathrm{Super}(L)} L_{\mathrm{E},\Sigma}(\alpha)$$

for all alphabets $\Sigma, \Pi$ and any language $L \subseteq (\Pi)^*$. Note that, in order to acquire a better understanding of these definitions, Example 19 can be consulted, which describes the superpatterns and superpattern hulls of an example language.

By Theorem 1, for every pair of patterns $\alpha, \beta \in X^+$ and every $\Sigma$ with $\Sigma \geq 2$, the following three conditions are equivalent:

1. $L_{\mathrm{E},\Sigma}(\alpha) \subseteq L_{\mathrm{E},\Sigma}(\beta)$,
2. $\beta \in \mathrm{Super}(L_{\mathrm{E},\Sigma}(\alpha))$,
3. $\beta \in \mathrm{Super}(\{\alpha\})$.

This allows us to state the following corollary:

**Corollary 17.** *Let $\Sigma, \Pi$ be alphabets with $|\Sigma|, |\Pi| \geq 2$. Then $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L) = D_{\mathrm{ePAT}_{\mathrm{tf},\Pi}}(L)$ for every $L \subseteq (\Sigma \cap \Pi)^*$.*

Although $\mathrm{Super}(L)$ and $\text{S-Hull}_\Sigma(L)$ might appear to be rather simple concepts, they can be used to establish most of the results in this section. Using Lemma 15, we can develop one of our main tools:

**Lemma 18.** *Let $\Sigma$ be an alphabet with $|\Sigma| \geq 2$. For every $L \subseteq \Sigma^*$, $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L) = D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(\text{S-Hull}_\Sigma(L))$.*

*Proof.* Let $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$. Then $L_{\mathrm{E},\Sigma}(\delta) \supseteq L$, and $L_{\mathrm{E},\Sigma}(\delta) \supseteq \text{S-Hull}_\Sigma(L)$ by definition of $\text{S-Hull}_\Sigma$. Thus, $L_{\mathrm{E},\Sigma}(\delta) \supseteq \text{S-Hull}_\Sigma(L) \supseteq L$. By Lemma 15, $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(\text{S-Hull}_\Sigma(L))$ follows.

Assume to the contrary that $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(\text{S-Hull}_\Sigma(L))$, and there is a $\gamma \in X^+$ with $L_{\mathrm{E},\Sigma}(\delta) \supset L_{\mathrm{E},\Sigma}(\gamma) \supseteq L$. Then, $L_{\mathrm{E},\Sigma}(\gamma) \supseteq \text{S-Hull}_\Sigma(L)$ holds (again by definition of $\text{S-Hull}_\Sigma$), and this contradicts the initial assumption. □

In a sense, $\text{S-Hull}_\Sigma(L)$ captures the whole essence of $L$ with respect to $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive patterns, as every pattern that is $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive of $L$ is unable to distinguish between these two languages. This is illustrated by the following example:

**Example 19.** *Let $|\Sigma| \geq 2$ and define $L := L_{\mathrm{E},\Sigma}(x_1^2) \cup L_{\mathrm{E},\Sigma}(x_1^3)$. Furthermore, let*

$$\begin{array}{llll} \delta_1 := x_1^2 x_2^3, & \delta_2 := x_1 x_2 x_1 x_2^2, & \delta_3 := x_1 x_2^2 x_1 x_2, & \delta_4 := x_1 x_2^3 x_1, \\ \delta_5 := x_1 x_2^2 x_1^2, & \delta_6 := x_1 x_2 x_1 x_2 x_1, & \delta_7 := x_1 x_2 x_1^2 x_2, & \delta_8 := x_1^2 x_2^2 x_1, \\ \delta_9 := x_1^2 x_2 x_1 x_2, & \delta_{10} := x_1^3 x_2^2. & & \end{array}$$

*Recalling Theorem 1, it is easy to see that, for every $\alpha \in \mathrm{Super}(L)$, there is a $\delta_i$, $1 \leq i \leq 10$, with $L_{\mathrm{E},\Sigma}(\alpha) \supseteq L_{\mathrm{E},\Sigma}(\delta_i)$ (as, for every $\alpha$, there must be morphisms mapping $\alpha$ to both $x_1^2$ and $x_1^3$). By a convention common in the literature, all patterns are given in* canonical form *(cf. [27]), where variables names are introduced in increasing lexicographic order.*

*This example illustrates two important phenomena. First, note that $\delta_i \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$ for $1 \leq i \leq 10$, and for every $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$, there is a $\delta_i$ with $L_{\mathrm{E},\Sigma}(\delta) = L_{\mathrm{E},\Sigma}(\delta_i)$, but $L_{\mathrm{E},\Sigma}(\delta) \neq L_{\mathrm{E},\Sigma}(\delta_j)$ for every $j \neq i$. Thus, $L$ has ten distinct* $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$*-descriptive patterns.*

*Second, the previous observation leads to* $\text{S-Hull}_\Sigma(L) = \bigcap_{i=1}^{10} L_{\mathrm{E},\Sigma}(\delta_i)$*. For every $n \geq 2$, there are $j, k \geq 0$ with $n = 2j + 3k$, and therefore,* $\text{S-Hull}_\Sigma(L) \supseteq \bigcup_{n=2}^{\infty} L_{\mathrm{E},\Sigma}(x_1^n)$*. Thus, every* $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$*-descriptive generalisation of $L$ is unable to exclude any language $L_{\mathrm{E},\Sigma}(x_1^n)$ with $n \geq 2$. In this sense,* $\text{S-Hull}_\Sigma(L)$ *provides information on the coarseness of all descriptive generalisations.*

Observe that $L$ in the previous example is a finite union of languages from $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$ that has a descriptive pattern, and recall that, according to Proposition 5, every finite set of words has an $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive pattern, while (by Theorem 4) there are infinite unions of languages from $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$ that have no descriptive pattern.

Using Lemma 18, we can extend Proposition 5 to show that not only every finite set of words, but every finite union of languages from $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$ has an $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive pattern:

**Proposition 20.** *Let $\Sigma$ be an alphabet with $|\Sigma| \geq 2$, let $A := \{\alpha_1, \dots, \alpha_n\} \subset X^+$ and let $L := \bigcup_{i=1}^{n} L_{\mathrm{E},\Sigma}(\alpha_i)$. Then $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L) \neq \emptyset$.*

*Proof.* By Theorem 1, the equation $\mathrm{Super}(\{\alpha\}) = \mathrm{Super}(L_{\mathrm{E},\Sigma}(\alpha))$ holds for every $\alpha \in X^+$. Thus,

$$\mathrm{Super}(\{\alpha_1, \dots, \alpha_n\}) = \mathrm{Super}(L_{\mathrm{E},\Sigma}(\alpha_1) \cup \dots \cup L_{\mathrm{E},\Sigma}(\alpha_n)),$$

and therefore $\text{S-Hull}_\Sigma(A) = \text{S-Hull}_\Sigma(L)$. By Lemma 18, this is equivalent to $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(A) = D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$.

As $A$ is a finite set, according to Proposition 5, $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(A)$ is nonempty, and thus, $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$ is nonempty as well. □

Basically, Example 19 and Proposition 20 are based on the fact that words in languages from $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$ and the generating patterns of these languages can often be used interchangeably by defining a morphism that maps the words back to their generating pattern. We proceed to develop this approach into another tool that allows us to make further statements on the (non-)existence of $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive patterns. Let $\nu : \Sigma^* \to X^*$ an arbitrary renaming. We define

$$\text{V-Hull}_\Sigma(L) := \bigcup_{w \in L} L_{\mathrm{E},\Sigma}(\nu(w)).$$

Note that, for every pair of renamings $\nu, \nu' : \Sigma^* \to X^*$ and every $w \in \Sigma^*$, $L_{\mathrm{E},\Sigma}(\nu(w)) = L_{\mathrm{E},\Sigma}(\nu'(w))$ holds by definition, which means that $\text{V-Hull}_\Sigma(L)$ does not depend on the choice of $\nu$.

Like $\text{S-Hull}_\Sigma(L)$, $\text{V-Hull}_\Sigma(L)$ is equivalent to $L$ with respect to Super and $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}$:

**Lemma 21.** *Let $\Sigma$ be an alphabet, $|\Sigma| \geq 2$. For every language $L$ over $\Sigma$,* $\mathrm{Super}(L) = \mathrm{Super}(\text{V-Hull}_\Sigma(L))$*, and* $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L) = D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(\text{V-Hull}_\Sigma(L))$*.*

*Proof.* We first show $\mathrm{Super}(L) = \mathrm{Super}(\text{V-Hull}_\Sigma(L))$.

Let $L \subseteq \Sigma^*$ and let $\nu : \Sigma^* \to X^*$ be a renaming. Naturally, there is an inverse renaming $\nu^{-1} : (\nu(\Sigma))^* \to \Sigma^*$ with $(\nu^{-1} \circ \nu)(w) = w$ for every $w \in L$. Thus, $w \in L_{\mathrm{E},\Sigma}(\nu(w))$ holds for every $w \in L$, which implies $L \subseteq \text{V-Hull}_\Sigma(L)$, and $\mathrm{Super}(L) \supseteq \mathrm{Super}(\text{V-Hull}_\Sigma(L))$.

For the other direction, consider any $\alpha \in \mathrm{Super}(L)$. This is equivalent to $L_{\mathrm{E},\Sigma}(\alpha) \supseteq L$, and thus, for every $w \in L$, there is a morphism $\sigma : X^* \to \Sigma^*$ with $\sigma(\alpha) = w$. Accordingly, $(\nu \circ \sigma)(\alpha) = \nu(w)$, and thus, $L_{\mathrm{E},\Sigma}(\alpha) \supseteq L_{\mathrm{E},\Sigma}(\nu(w))$ for every $w \in L$. This immediately implies $L_{\mathrm{E},\Sigma}(\alpha) \supseteq \text{V-Hull}_\Sigma(L)$, and therefore, $\alpha \in \mathrm{Super}(\text{V-Hull}_\Sigma(L))$.

As $\mathrm{Super}(L) = \mathrm{Super}(\text{V-Hull}_\Sigma(L))$, $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L) = D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(\text{V-Hull}_\Sigma(L))$ follows by definition of $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}$. □

This leads us to the following insight into the existence of $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive patterns for infinite unions of languages from $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$:

**Proposition 22.** *Let $\Sigma$ be an alphabet, $|\Sigma| \geq 2$. Then there is a set of patterns $A \subset \{x_1, x_2\}^*$ such that no pattern in $X^+$ is $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive of $\bigcup_{\alpha \in A} L_{\mathrm{E},\Sigma}(\alpha)$.*

*Proof.* Choose any alphabet $\Pi \subseteq \Sigma$ with $|\Pi| = 2$. By Theorem 4, there exists a language $L' \subset (\Pi)^*$ such that $D_{\mathrm{ePAT}_{\mathrm{tf},\Pi}}(L') = \emptyset$. Furthermore, let $\nu : (\Pi)^* \to \{x_1, x_2\}^*$ be a renaming. We claim that $A := \nu(L')$ fulfils the given requirements. As $A \subset \{x_1, x_2\}^*$ holds by definition, we only need to show that no pattern in $X^+$ is $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive of the language $L := \bigcup_{\alpha \in A} L_{\mathrm{E},\Sigma}(\alpha)$.

Assume to the contrary that $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L) \neq \emptyset$. Note that $L = \text{V-Hull}_\Sigma(L')$ holds, which allows us to use Lemma 21 to conclude $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L) = D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L')$. Furthermore, Corollary 17 implies $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L') = D_{\mathrm{ePAT}_{\mathrm{tf},\Pi}}(L')$. Combining these two equations and the initial assumption of $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L) \neq \emptyset$, we arrive at $D_{\mathrm{ePAT}_{\mathrm{tf},\Pi}}(L') \neq \emptyset$, which contradicts our choice of $L'$. □

Thus, unlike in the case of *finite* unions of languages from $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$ (cf. Proposition 20), even restricting the number of variables in the generating patterns does not ensure that *infinite* unions of languages from $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$ have a descriptive pattern. The renaming $\nu$ that maps terminals to variables can also be used to obtain the following technical result:

**Lemma 23.** *Let $\Sigma$ be an alphabet with $|\Sigma| \geq 2$. For every nonempty language $L \subseteq \Sigma^*$ with $L \neq \{\lambda\}$, $\text{S-Hull}_\Sigma(L)$ is infinite.*

*Proof.* Let $\nu : \Sigma^+ \to X^+$ be a nonerasing morphism and choose any $w \in L$ such that $w \neq \lambda$. For every $\alpha \in \mathrm{Super}(L)$, there is a morphism $\sigma : X^+ \to \Sigma^*$ with $\sigma(\alpha) = w$. Thus, $\nu(\sigma(\alpha)) = \nu(w)$, which implies $L_{\mathrm{E},\Sigma}(\alpha) \supseteq L_{\mathrm{E},\Sigma}(\nu(w))$. As this holds for every $\alpha \in \mathrm{Super}(L)$, we conclude $\text{S-Hull}_\Sigma(L) \supseteq L_{\mathrm{E},\Sigma}(\nu(w))$; and as $\nu(w) \in X^+$, this language is infinite. □

This insight shall be used in Section 5.3.

### *5.2. The Canonical Strategy and Telling Sets*

According to Proposition 5, every finite set has a computable $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive pattern. We consider it the canonical strategy of descriptive inference on any text $t$ of a given language $L$ to compute a descriptive pattern of every initial segment $t^n$, in the hope that the hypotheses will converge to a pattern that is descriptive of $L$. As evidenced by the language $L := L_{\mathrm{E},\Sigma}(x_1^2) \cup L_{\mathrm{E},\Sigma}(x_1^3)$ (cf. Example 19), there are languages with more than one descriptive pattern. Furthermore, this applies also to finite languages, as for the set $S := \{\mathtt{a}^2, \mathtt{b}^3\}$ (for arbitrary letters $\mathtt{a}, \mathtt{b} \in \Sigma$), $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(S) = D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$ holds. Although $S$ already contains all the information that is needed to compute a descriptive generalisation of $L$, the ten distinct patterns $\delta_1$ to $\delta_{10}$ from Example 19 are all valid hypotheses. In order to allow our strategy to converge to one single hypothesis, we impose a total and well-founded order $<_{\mathrm{LLO}}$ on $X^+$ and let our strategy return the $<_{\mathrm{LLO}}$-minimal hypothesis.

Let $<_{\mathrm{LLO}}$ denote the length-lexicographic order on $X^+$ (i. e., $\alpha <_{\mathrm{LLO}} \beta$ if $|\alpha| < |\beta|$, or if $|\alpha| = |\beta|$, and $\alpha$ precedes $\beta$ in the lexicographic order). Note that $<_{\mathrm{LLO}}$ is total and does not contain infinite decreasing chains. Thus, every set has exactly one element that is minimal with respect to $<_{\mathrm{LLO}}$.

The strategy $\mathrm{Canon} : (\Sigma \cup \{\nabla\})^* \to (\Sigma \cup X)^+$ is defined by, for every text $t$,

$$\mathrm{Canon}(t^n) := \delta,$$

such that $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(t[n])$ and $\delta <_{\mathrm{LLO}} \gamma$ for every other $\gamma \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(t[n])$.

The computability of Canon follows immediately from the proof of Proposition 5, as all that remains is to sort the finite search space by $<_{\mathrm{LLO}}$. We say that Canon *converges* on a text $t \in \mathrm{text}(L)$ (of some language $L$ over some alphabet $\Sigma$) if there is a pattern $\alpha \in X^+$ with $\mathrm{Canon}(t^n) = \alpha$ for all but finitely many values of $n$. If, in addition to this, $\alpha \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$, Canon is said to *converge correctly* on $t$. Now, when considering the languages $L$ and $S$ given in the example above, for every text $t \in \mathrm{text}(L)$, there is an $n \geq 0$ with $S \subseteq t[n]$. From this point onwards, Canon($t[n]$) will return the pattern $\delta_{10} = x_1^3 x_2^2$, as $\delta_{10}$ is an element of $(D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(S) \cap D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L))$ and the $<_{\mathrm{LLO}}$-minimum of the canonical forms of the patterns $\delta_1$ to $\delta_{10}$ (see Example 19). This phenomenon leads to the definition of what we call *telling sets*, which are of crucial importance for the study of descriptive generalisability with the strategy Canon:

**Definition 24.** Let $L \subseteq \Sigma^*$. A finite set $S \subseteq L$ is a *telling set* for $L$ if $(D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(S) \cap D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)) \neq \emptyset$.

Note that telling sets have some similarity to the concept of telltales that is used in the model of learning in the limit. For a comparison of telltales and telling sets, see our comments after Corollary 32.

Using Lemma 15, we are now able to show that the existence of a telling set is characteristic for the correct convergence of Canon on any text:

**Theorem 25.** *Let $\Sigma$ an alphabet with $|\Sigma| \geq 2$. For every language $L \subseteq \Sigma^*$ and every text $t \in$* text($L$)*,* Canon *converges correctly on $t$ if and only if $L$ has a telling set.*

*Proof.* We begin by proving the *only if* direction via its contraposition. Assume that $L$ has no telling set. Then, for every text $t$ of $L$ and every $n \geq 0$, $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(t[n])$ and $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$, by definition, are disjoint, as otherwise $t[n]$ would be a telling set. Therefore, Canon($t^n$) $\notin D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$, which means that even if Canon converges on $t$ to a pattern $\delta$, this pattern is not $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive of $L$.

For the *if* direction, assume to the contrary that $S$ is a telling set of $L$ and that there exists a text $t$ of $L$ such that Canon does not converge correctly on $t$. We first show that if Canon converges, it always converges correctly. Assume to the contrary that there exist some pattern $\delta$ and some $n \geq 0$ such that $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(t[m])$ for every $m \geq n$, but $\delta \notin D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$. As $L_{\mathrm{E},\Sigma}(\delta) \supseteq t[m]$ for every $m \geq n$, it follows that

$$\begin{aligned} L_{\mathrm{E},\Sigma}(\delta) &\supseteq \bigcup_{m \geq n} t[m] \\ &= \{t[i] \mid i \in \mathbb{N}\} = L. \end{aligned}$$

As $L_{\mathrm{E},\Sigma}(\delta) \supseteq L \supseteq t[n]$ and $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(t[n])$, Lemma 15 gives $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$, which contradicts the initial assumption.

Next, assume that Canon does not converge on $t$; i. e., there is an infinite sequence $(\delta_n)_{n \geq 0}$ over $X^+$ with

1. $\delta_n =$ Canon($t^n$) for every $n \geq 0$,
2. for every $n \geq 0$, there is an $m > n$ with $\delta_m \neq \delta_n$.

We first show that at most one pattern occurs at least twice in $(\delta_n)_{n \geq 0}$. Assume that there are $m_1 < n_1 < m_2 < n_2$ with $\delta_{m_1} = \delta_{m_2} \neq \delta_{n_1} = \delta_{n_2}$, and observe that

$$t[m_1] \subseteq t[n_1] \subseteq t[m_2] \subseteq t[n_2]$$

holds. As $\delta_{m_1}$ is $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive of $t[m_1]$, and

$$L_{\mathrm{E},\Sigma}(\delta_{m_1}) = L_{\mathrm{E},\Sigma}(\delta_{m_2}) \supseteq t[m_2] \supseteq t[n_1] \supseteq t[m_1],$$

Lemma 15 implies $\delta_{m_1} \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(t[n_1])$. However, due to $\mathrm{Canon}(t[n_1]) = \delta_{n_1} \neq \delta_{m_1}$, $\delta_{n_1} <_{\mathrm{LLO}} \delta_{m_1}$ must hold. Analogously, one can use Lemma 15 and $L_{\mathrm{E},\Sigma}(\delta_{n_1}) \supseteq t[n_2] \supseteq t[m_2] \supseteq t[n_1]$ to conclude $\delta_{m_1} <_{\mathrm{LLO}} \delta_{n_1}$, which leads to the contradictory statement $\delta_{m_1} <_{\mathrm{LLO}} \delta_{n_1} <_{\mathrm{LLO}} \delta_{m_1}$.

We now consider two cases. First, assume there is some $\delta$ such that $\delta_n = \delta$ for infinitely many $n \geq 0$ (as we have seen, there can be at most one such a $\delta$). Then $L_{\mathrm{E},\Sigma}(\delta) \supseteq t[n]$ holds for infinitely many $n \in \mathbb{N}$, which implies $L_{\mathrm{E},\Sigma}(\delta) \supseteq t[m]$ for every $m \geq 0$. As above, this leads to

$$L_{\mathrm{E},\Sigma}(\delta) \supseteq \{t(i) \mid i \in \mathbb{N}\} = L.$$

Especially, if $\delta_n = \delta$,

$$L_{\mathrm{E},\Sigma}(\delta) \supseteq L \supseteq t[n],$$

and the initial assumption $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(t[n])$ allow us to use Lemma 15 to conclude $\delta \in D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$. But only a finite number of $\delta_m$ can satisfy $\delta_m <_{\mathrm{LLO}} \delta$, which means that Canon converges to $\delta$, and this contradicts our initial assumption.

For the other case, assume that no pattern occurs infinitely often in $(\delta_n)_{n\geq 0}$. Then for every $\delta \in X^+$, there is a $k \geq 0$ such that $\delta <_{\mathrm{LLO}} \delta_n$ for all $n \geq k$. This holds especially for that pattern $\delta_S \in (D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(S) \cap D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L))$ that is minimal with respect to $<_{\mathrm{LLO}}$ (such a pattern has to exist, as we required $S$ to be a telling set of $L$). Choose some $k$ such that

1. $\delta_S <_{\mathrm{LLO}} \delta_n$ for all $n \geq k$, and
2. $S \subseteq t[k]$,

and observe that, due to Lemma 15,

$$(D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(S) \cap D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)) \subseteq D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(t[n])$$

for every $n \geq k$. Thus, $\mathrm{Canon}(t[n]) = \delta_S$ for every $n \geq k$. This contradicts our initial assumption that Canon does not converge correctly on $t$. □

In the final part of this section, we shall demonstrate that this is a strong result, by investigating the existence and nonexistence of telling sets for various languages.

### *5.3. Examination of the Class $\mathcal{TSL}_\Sigma$*

As stated by Theorem 25, the existence of telling sets is a strong sufficient criterion for $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive generalisability. Furthermore, generalisability of a class $\mathcal{L} \subseteq \mathcal{P}(\Sigma^*)$ using Canon does not depend on the properties of the whole class, but only on the existence of a telling set for every single language

$L \in \mathcal{L}$. Thus, we consider the largest possible class that can be generalised by Canon and define

$$\mathcal{TSL}_\Sigma := \{L \subseteq \Sigma^* \mid L \text{ has a telling set}\}.$$

Theorem 25 immediately leads to the following corollary:

**Corollary 26.** *For every alphabet* $\Sigma$ *with* $|\Sigma| \geq 2$, $\mathcal{TSL}_\Sigma \in \mathrm{DG}_{\mathrm{ePAT}_{\mathbf{tf},\Sigma}}$.

Thus, by examining $\mathcal{TSL}_\Sigma$, we gain insights into the power of Canon and of the whole model of descriptive generalisation. Before we proceed to an examination of the relation of various classes of languages to $\mathcal{TSL}_\Sigma$, we show that it is not required to choose $\Sigma$ as small as possible, a result that is similar to Corollary 17, which states that $D_{\mathrm{ePAT}_{\mathbf{tf},\Sigma}}(L)$ is largely independent of the choice of $\Sigma$. The same holds for telling sets:

**Corollary 27.** *Let* $\Sigma, \Pi$ *be alphabets with* $|\Sigma|, |\Pi| \geq 2$. *Then, for every* $L \subseteq (\Sigma \cap \Pi)$, $L \in \mathcal{TSL}_\Sigma$ *if and only if* $L \in \mathcal{TSL}_\Pi$.

*Proof.* As $L \subseteq (\Sigma \cap \Pi)^*$, the same holds for every telling set $S \subseteq L$. According to Corollary 17, $D_{\mathrm{ePAT}_{\mathbf{tf},\Sigma}}(S) = D_{\mathrm{ePAT}_{\mathbf{tf},\Pi}}(S)$. □

This also implies that, for every $\Pi \supseteq \Sigma$, $\mathcal{TSL}_\Pi \supseteq \mathcal{TSL}_\Sigma$. We begin our examination of $\mathcal{TSL}_\Sigma$ by expanding finite languages without losing their telling set properties:

**Lemma 28.** *Let* $\Sigma$ *be an alphabet with* $|\Sigma| \geq 2$. *Every nonempty* $S \in \mathrm{FIN}_\Sigma$ *is a telling set of* $\text{S-Hull}_\Sigma(S)$ *and of every* $L$ *with* $S \subseteq L \subseteq \text{S-Hull}_\Sigma(S)$.

*Proof.* Due to Lemma 18, $D_{\mathrm{ePAT}_{\mathbf{tf},\Sigma}}(S) = D_{\mathrm{ePAT}_{\mathbf{tf},\Sigma}}(\text{S-Hull}_\Sigma(S))$ holds; therefore $S$ is a telling set of $\text{S-Hull}_\Sigma(S)$. Now choose any $L$ with $S \subseteq L \subseteq \text{S-Hull}_\Sigma(S)$ and any $\delta \in D_{\mathrm{ePAT}_{\mathbf{tf},\Sigma}}(S)$. According to Lemma 15, $\delta \in D_{\mathrm{ePAT}_{\mathbf{tf},\Sigma}}(L)$, which means that $S$ is a telling set of $L$. □

In addition to showing that $\mathrm{FIN}_\Sigma \subseteq \mathcal{TSL}_\Sigma$, this result allows us to make the following statement on the cardinality of $\mathcal{TSL}_\Sigma$:

**Proposition 29.** $\mathcal{TSL}_\Sigma$ *is uncountable for every alphabet* $\Sigma$ *with* $|\Sigma| \geq 2$.

*Proof.* Select a finite nonempty language $S \subseteq \Sigma^+$. Let

$$\begin{aligned}\mathcal{U} :=& \{L \mid S \subseteq L \subseteq \text{S-Hull}_\Sigma(S)\}\\ =& \{L \cup S \mid L \in \mathcal{P}(\text{S-Hull}_\Sigma(S) \setminus S)\}.\end{aligned}$$

Due to Lemma 28, $S$ is a telling set of every $L \in \mathcal{U}$, and thus, $\mathcal{U} \subseteq \mathcal{TSL}_\Sigma$.

As $S$ is nonempty, $\text{S-Hull}_\Sigma(S)$ must be infinite according to Lemma 23. Therefore, $\mathcal{U}$ is uncountable, which means that $\mathcal{TSL}_\Sigma$ is uncountable as well. □

This is an uncommon property, as inference from positive data is normally considered for classes consisting of countably many languages from some countable domain. Nonetheless, inferrability of uncountable classes has been studied before, see [15].

Next, we shall see that $\mathcal{TSL}_\Sigma$ contains a rich and natural class of languages, the DTF0L languages. A DTF0L language $L$ over $\Sigma$ is defined through a finite set of axioms $w_1, \dots, w_m \in \Sigma^*$ and a finite set of morphisms $\phi_1, \dots, \phi_n : \Sigma^* \to \Sigma^*$. Then $L$ is the smallest language that satisfies these two conditions:

1. $w_i \in L$ for every $i \in \{1, \dots, m\}$,
2. if $w \in L$, then $\phi_i(w) \in L$ for every $i \in \{1, \dots, n\}$.

We denote the class of all DTF0L languages over $\Sigma$ by $\mathrm{DTF0L}_\Sigma$. Apart from $\mathrm{FIN}_\Sigma$, the most prominent subclass of $\mathrm{DTF0L}_\Sigma$ is the class of D0L languages, where every language is defined through a single axiom and a single morphism (i. e., $m = n = 1$). The class D0L has been widely studied, for details, see [18].

**Theorem 30.** *Let $\Sigma$ be an alphabet with $|\Sigma| \geq 2$. Then* $\mathrm{DTF0L}_\Sigma \subseteq \mathcal{TSL}_\Sigma$.

*Proof.* Let $L \subseteq \Sigma^*$ be a DTF0L-language, where $F$ is a nonempty set of axioms and $\Phi$ a nonempty set of morphisms generating $L$ from $F$. It suffices to show that $\mathrm{Super}(F) = \mathrm{Super}(L)$, as this shall allow us to use Lemma 18 to obtain the desired result.

First, observe that, by definition, $F \subseteq L$, and thus, $\mathrm{Super}(F) \supseteq \mathrm{Super}(L)$. For the other direction, consider any $\alpha \in \mathrm{Super}(F)$. For every $w \in L$, there is a $v \in F$ and a finite sequence of morphisms in $\Phi$ that can be composed to a single morphism $\phi$ with $\phi(v) = w$. As $v \in F$ and $\alpha \in \mathrm{Super}(F)$, there is a morphism $\sigma$ with $\sigma(\alpha) = v$. The composition of these morphisms leads to $(\phi \circ \sigma)(\alpha) = w$, and (as $w$ was chosen arbitrarily) $\alpha \in \mathrm{Super}(L)$.

Now, $\mathrm{Super}(F) = \mathrm{Super}(L)$ results in $\text{S-Hull}_\Sigma(F) = \text{S-Hull}_\Sigma(L)$ and, by Lemma 18, in $D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(F) = D_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}(L)$. As $F$ is finite, it has at least one $\mathrm{ePAT}_{\mathrm{tf},\Sigma}$-descriptive pattern (according to Proposition 5), and is a telling set of $L$. □

Lemma 28 and Theorem 30 both imply that $\mathrm{FIN}_\Sigma \subseteq \mathcal{TSL}_\Sigma$. Furthermore, Proposition 29 and Theorem 30 both demonstrate that $\mathcal{TSL}_\Sigma$ contains at least one infinite language, which leads to the following observation:

**Corollary 31.** *The class $\mathcal{TSL}_\Sigma$ is superfinite for every alphabet $\Sigma$ with $|\Sigma| \geq 2$.*

Together with Proposition 14, this allows us to describe the relation between $\mathrm{DG}_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}$ and LIM-TEXT:

**Corollary 32.** *Let $\Sigma$ be an alphabet, $|\Sigma| \geq 2$. Then* $\mathrm{DG}_{\mathrm{ePAT}_{\mathrm{tf},\Sigma}}$ *and* LIM-TEXT *are incomparable.*

*Proof.* Directly from Proposition 14 and Corollary 31, as LIM-TEXT contains every finite class, but no superfinite classes (cf. [12]). □

We now briefly discuss the relation between telling sets and the notion of telltales. As already mentioned in Section 4.2, according to [2], an indexed family $\mathcal{L} = (L_i)_{i \in \mathbb{N}}$ of non-empty recursive languages is in LIM-TEXT if and only if there exists an effective procedure which, for every $j \geq 0$, enumerates a set $T_j$ such that

- $T_j$ is finite,
- $T_j \subseteq L_j$, and
- there does not exist a $j' \geq 0$ with $T_j \supseteq L_{j'} \supset L_j$.

If there exists a set $T_j$ satisfying these conditions, it is called a *telltale* for $L_j$ with respect to $\mathcal{L} = (L_i)_{i \in \mathbb{N}}$. Thus, the concepts of telltales and telling sets are incomparable, as the former refers to a language and the class of languages it is contained in, whereas the latter relates to a language and certain properties of the class $\mathrm{ePAT}_{\mathtt{tf},\Sigma}$. Nevertheless, for every language $L$ in $\mathrm{ePAT}_{\mathtt{tf},\Sigma}$, a set $S$ is a telling set for $L$ if and only if $S$ is a telltale for $L$ with respect to $\mathrm{ePAT}_{\mathtt{tf},\Sigma}$ (for more details on the existence of telltales for languages in $\mathrm{ePAT}_{\mathtt{tf},\Sigma}$, see [26]).

As Proposition 33 and Proposition 34 below show, Lemma 25 by [26] and Lemma 7 by Reidenbach [25] on the existence and nonexistence of telltales lead to the corresponding results for telling sets:

**Proposition 33.** *Let $\Sigma$ be an alphabet with $|\Sigma| \geq 2$. For every $\alpha \in X^+$, $L_{\mathrm{E},\Sigma}(\alpha)$ has a telling set.*

*Proof.* This follows immediately from Lemma 25 in [26]. □

On the other hand, it is impossible to encode the structure of comparatively simple patterns in their languages with only two letters, which leads to the following negative result:

**Proposition 34.** *Let $\Sigma$ be an alphabet with $|\Sigma| \geq 2$, and let $\mathtt{a}$, $\mathtt{b}$ be two distinct letters from $\Sigma$. Then $L_{\mathrm{E},\{\mathtt{a},\mathtt{b}\}}(x_1^2 x_2^2 x_3^2) \notin \mathcal{TSL}_\Sigma$.*

*Proof.* This follows immediately from Lemma 7 by Reidenbach [25], which states that, for every finite $S \subset L_{\mathrm{E},\{\mathtt{a},\mathtt{b}\}}(x_1^2 x_2^2 x_3^2)$, there is a pattern $\alpha \in X^+$ with $L_{\mathrm{E},\Sigma}(x_1^2 x_2^2 x_3^2) \supset L_{\mathrm{E},\Sigma}(\alpha) \supset S$. □

In contrast to this, Lemma 21 can be used to show that restricting the number of variables in the patterns leads to telling sets not only for languages from $\mathrm{ePAT}_{\mathtt{tf},\Sigma}$, but also for their finite unions:

**Proposition 35.** *Let $\Sigma$ be an alphabet with $|\Sigma| \geq 2$, let $\alpha_1, \dots, \alpha_n \in \{x_1, \dots, x_{|\Sigma|}\}^+$, and let $L := \bigcup_{i=1}^n L_{\mathrm{E},\Sigma}(\alpha_i)$. Then $L \in \mathcal{TSL}_\Sigma$.*

*Proof.* Let $X_{|\Sigma|} := \{x_1, \dots, x_{|\Sigma|}\}$ and let $\alpha_1, \dots, \alpha_n \in X_{|\Sigma|}^+$. Choose any renaming $\nu : \Sigma^* \to X_{|\Sigma|}^*$, let $\nu^{-1}$ be the corresponding inverse renaming and let

$$S := \{\nu^{-1}(\alpha_1), \dots, \nu^{-1}(\alpha_n)\}.$$

It is easily seen that $\text{V-Hull}_\Sigma(S) = L$ holds. By Lemma 21, $D_{\text{ePAT}_{\text{tf},\Sigma}}(S) = D_{\text{ePAT}_{\text{tf},\Sigma}}(\text{V-Hull}_\Sigma(S))$, and therefore $D_{\text{ePAT}_{\text{tf},\Sigma}}(S) = D_{\text{ePAT}_{\text{tf},\Sigma}}(L)$. As $S$ is finite, Proposition 5 leads to $D_{\text{ePAT}_{\text{tf},\Sigma}}(S) \neq \emptyset$. Therefore, $S$ is a telling set for $L$. □

Proposition 35 is especially interesting when compared to Proposition 22, which tells us that infinite unions of languages from $\text{ePAT}_{\text{tf},\Sigma}$ might not only have no telling set, but not even a descriptive pattern.

Furthermore, we state that the infinite sequence $(\beta_n)_{n\geq 0}$ that is used in the definition of the languages $L_\Sigma$ for the proof of Theorem 4 describes an infinite ascending chain of languages from $\text{ePAT}_{\text{tf},\Sigma}$; i.e., $L_{\text{E},\Sigma}(\beta) \subset L_{\text{E},\Sigma}(\beta_{n+1})$ for every $n \geq 0$. Although the presence of such a chain in $\text{S-Hull}_\Sigma(L)$ for a language $L$ does not necessarily imply emptiness of $D_{\text{ePAT}_{\text{tf},\Sigma}}(L)$, it is a sufficient criterion for $L \notin \mathcal{TSL}_\Sigma$:

**Lemma 36.** *Let $\Sigma$ be an alphabet with $|\Sigma| \geq 2$ and let $L \subseteq \Sigma^*$. If there is an infinite chain $(\beta_n)_{n\geq 0}$ over $X^+$ with*

- *$L_{\text{E},\Sigma}(\beta_n) \subseteq \text{S-Hull}_\Sigma(L)$ for every $n \geq 0$,*
- *$L_{\text{E},\Sigma}(\beta_n) \subset L_{\text{E},\Sigma}(\beta_{n+1})$ for every $n \geq 0$, and*
- *$\bigcup_{n\geq 0} L_{\text{E},\Sigma}(\beta_n) \supseteq L$,*

*then $L$ has no telling set.*

*Proof.* Let $L \subseteq \Sigma^*$ and $(\beta_n)_{n\geq 0}$ a strictly ascending infinite chain over $X^+$ that satisfies the above criteria. Assume to the contrary there are a finite set $S \subseteq L$ and a pattern $\delta \in D_{\text{ePAT}_{\text{tf},\Sigma}}(L) \cap D_{\text{ePAT}_{\text{tf},\Sigma}}(S)$.

As $S \subseteq L \subseteq \bigcup_{n\geq 0} L_{\text{E},\Sigma}(\beta_n)$, there is an $m \geq 0$ such that $L_{\text{E},\Sigma}(\beta_m) \supseteq S$, and therefore,

$$\text{S-Hull}_\Sigma(L) \supseteq L_{\text{E},\Sigma}(\beta_m) \supseteq S.$$

As, due to Lemma 18, $\delta \in D_{\text{ePAT}_{\text{tf},\Sigma}}(\text{S-Hull}_\Sigma(L))$, we conclude

$$L_{\text{E},\Sigma}(\delta) \supseteq \text{S-Hull}_\Sigma(L) \supseteq L_{\text{E},\Sigma}(\beta_m) \supseteq S.$$

By definition, $\beta_m$ is part of an infinite ascending chain, and therefore

$$\text{S-Hull}_\Sigma(L) \supseteq L_{\text{E},\Sigma}(\beta_{m+1}) \supset L_{\text{E},\Sigma}(\beta_m) \supseteq S$$

holds. This contradicts $\delta \notin D_{\text{ePAT}_{\text{tf},\Sigma}}(S)$. Thus, $L$ either has no $\text{ePAT}_{\text{tf},\Sigma}$-descriptive pattern, or it has an $\text{ePAT}_{\text{tf},\Sigma}$-descriptive pattern, but still no telling set. □

As a direct application of this result, we can prove that there are regular languages that have no telling set:

**Theorem 37.** *For every alphabet $\Sigma$ with $|\Sigma| \geq 2$, there is a regular language $L \subseteq \Sigma^*$ with $L \notin \mathcal{TSL}_\Sigma$.*

*Proof.* Let $\mathtt{a}, \mathtt{b}$ two distinct letters from $\Sigma$ and define $L := \{\mathtt{a}\,\mathtt{a}, \mathtt{b}\,\mathtt{b}\}^*$. Clearly, $L$ is regular. Next, we shall show that $\text{S-Hull}_\Sigma(L) = \Sigma^*$. For $\alpha \in \text{Super}(L)$, let $n := |\alpha|$ and

$$w := \mathtt{a\,a\,b\,b\,a\,a}\;\mathtt{a\,a}(\mathtt{b\,b})^2\,\mathtt{a\,a}\cdots\mathtt{a\,a}(\mathtt{b\,b})^{n+1}\,\mathtt{a\,a}\,.$$

As $w \in L$, there is a morphism $\sigma$ with $\sigma(\alpha) = w$. Furthermore, as $\alpha$ has at most $n$ distinct variables, there is at least one variable $x \in \text{var}(\alpha)$ such that $\sigma(x)$ contains a factor $\mathtt{a\,a}(\mathtt{b\,b})^i\,\mathtt{a\,a}$ for some $i$ with $1 \leq i \leq n+1$. As this factor occurs exactly once in $w$, $x$ occurs exactly once in $\alpha$, and thus, $L_{\text{E},\Sigma}(\alpha) = \Sigma^*$.

Therefore, $\text{S-Hull}_\Sigma(L) = \Sigma^*$, and $L_{\text{E},\Sigma}(\delta) = \Sigma^*$ for every $\delta \in D_{\text{ePAT}_{\text{tf},\Sigma}}(L)$. The chain $(\beta_n)_{n\geq 0}$ with $\beta_1 := x_1^2$, $\beta_2 := x_1^2x_2^2$, …, $\beta_n := x_1^2\cdots x_n^2$ satisfies the criteria of Lemma 36, which implies that $L$ has no telling set. □

Note that the language $\{\mathtt{a\,a}, \mathtt{b\,b}\}^*$ considered in the above proof is not only an example of a language that has no telling set, although it has a descriptive pattern, but also an example of a language $L$ that has no telling set, although $\text{S-Hull}_\Sigma(L)$ has a telling set.

## 6. Conclusion and Further Directions

In this paper, we have introduced a new inference paradigm, called descriptive generalisation, and we have showed that the loss of precision (in comparison to Gold's model) that comes with the use of descriptive patterns as hypotheses can lead to greater learning power. In particular, using the natural inference strategy Canon, we have demonstrated that the well-known superfinite language class DTF0L can be descriptively generalised. We have also provided a characterisation of those indexed families that can be generalised in our model – a result that employs a topological condition similar to Angluin's telltales.

In Section 4.3, we have demonstrated that our model is a natural instance of a generic paradigm of identification in the limit, which we refer to as *inductive inference with hypotheses validity relation* HYP and which covers Gold's model as well. We consider this paradigm a worthwhile topic for further examination. For example, one could study the impact of the choice of the relation HYP on the inferrability of classes of languages; in other words, this might reveal the nature of those learning tasks that can successfully be accomplished for a given class. In a more applied approach, one could investigate the existence of natural instances of this model that differ from standard Gold-style identification and from our notion of descriptive generalisation.

Regarding the descriptive generalisation of terminal-free E-pattern languages, we first note that the language $L$ from the proof of Theorem 37 can be used to define a class $\mathcal{L} \in (\text{DG}_{\text{ePAT}_{\text{tf},\Sigma}} \setminus \mathcal{TSL}_\Sigma)$, by $\mathcal{L} := \{L\}$. While Canon does not converge on any text of $L$, the trivial strategy that outputs $x$ on every input yields $\mathcal{L} \in \text{DG}_{\text{ePAT}_{\text{tf},\Sigma}}$ (analogously to Proposition 14). This answer might be considered somewhat artificial, which leads to the following question: Are there rich and natural classes in $\text{DG}_{\text{ePAT}_{\text{tf},\Sigma}}$ that cannot be inferred with Canon?

Another question we consider worth studying is in what cases Canon can be computed efficiently. In particular, the proof of Corollary 7 depends on the fact that the alphabet $\Sigma$ is unbounded. It remains open whether this Corollary still holds if the alphabet $\Sigma$ is fixed (and non-unary).

Furthermore, an alternative characterisation of $\mathcal{TSL}_\Sigma$ or additional sufficient criteria for the non-existence of telling sets as in Lemma 36 should be very interesting.

Finally, another promising direction for further work is the use of hypotheses other than pattern languages. One example is studied by Freydenberger and Kötzing [7], who examine the canonical strategy using subclasses of regular expressions as hypothesis space.

*Acknowledgements*

The authors wish to thank the anonymous referees for their helpful remarks and suggestions. Furthermore the enlightening comments by Steffen Lange on the inference model and on some results presented in this paper are gratefully acknowledged.

## References

[1] D. Angluin, Finding patterns common to a set of strings, Journal of Computer and System Sciences 21 (1980) 46–62.

[2] D. Angluin, Inductive inference of formal languages from positive data, Information and Control 45 (1980) 117–135.

[3] H. Arimura, T. Shinohara, S. Otsuki, Finding minimal generalizations for unions of pattern languages and its application to inductive inference from positive data, in: Proc. 11th Annual Symposium on Theoretical Aspects of Computer Science, STACS 1994, volume 775 of *LNCS*, pp. 649–660.

[4] C. Câmpeanu, K. Salomaa, S. Yu, A formal study of practical regular expressions, International Journal of Foundations of Computer Science 14 (2003) 1007–1018.

[5] A. Ehrenfeucht, G. Rozenberg, Finding a homomorphism between two words is NP-complete, Information Processing Letters 9 (1979) 86–88.

[6] H. Fernau, Identification of function distinguishable languages, Theoretical Computer Science 290 (2003) 1679–1711.

[7] D.D. Freydenberger, T. Kötzing, Fast learning of restricted regular expressions and DTDs. Submitted.

[8] D.D. Freydenberger, D. Reidenbach, Bad news on decision problems for patterns, Information and Computation 208 (2010) 83–96.

[9] D.D. Freydenberger, D. Reidenbach, Existence and nonexistence of descriptive patterns, Theoretical Computer Science 411 (2010) 3274–3286.

[10] D.D. Freydenberger, D. Reidenbach, Inferring descriptive generalisations of formal languages, in: Proc. 23rd Annual Conference on Learning Theory, COLT 2010, pp. 194–206.

[11] M. Fulk, S. Jain, Learning in the presence of inaccurate information, Theoretical Computer Science 161 (1996) 235–261.

[12] E. Gold, Language identification in the limit, Information and Control 10 (1967) 447–474.

[13] Š. Holub, Polynomial-time algorithm for fixed points of nontrivial morphisms, Discrete Mathematics 309 (2009) 5069–5076.

[14] S. Jain, E. Kinber, Learning and extending sublanguages, Theoretical Computer Science 397 (2008) 233–246.

[15] S. Jain, Q. Luo, P. Semukhin, F. Stephan, Uncountable automatic classes and learning, in: Proc. 20th International Conference on Algorithmic Learning Theory, ALT 2009, pp. 293–307.

[16] T. Jiang, E. Kinber, A. Salomaa, K. Salomaa, S. Yu, Pattern languages with and without erasing, International Journal of Computer Mathematics 50 (1994) 147–163.

[17] T. Jiang, A. Salomaa, K. Salomaa, S. Yu, Decision problems for patterns, Journal of Computer and System Sciences 50 (1995) 53–63.

[18] L. Kari, G. Rozenberg, A. Salomaa, L systems, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, volume 1, Springer, 1997, pp. 253–328.

[19] S. Kobayashi, T. Yokomori, On approximately identifying concept classes in the limit, in: Proc. 6th International Workshop on Algorithmic Learning Theory, ALT 1995, volume 997 of *LNAI*, pp. 298–312.

[20] S. Kobayashi, T. Yokomori, Learning approximately regular languages with reversible languages, Theoretical Computer Science 174 (1997) 251–257.

[21] S. Lange, R. Wiehagen, Polynomial-time inference of arbitrary pattern languages, New Generation Computing 8 (1991) 361–370.

[22] S. Lange, T. Zeugmann, S. Zilles, Learning indexed families of recursive languages from positive data: A survey, Theoretical Computer Science 397 (2008) 194–232.

[23] Y. Mukouchi, Inductive inference of an approximate concept from positive data, in: Proc. 5th International Workshop on Algorithmic Learning Theory, ALT 1994, volume 872 of *LNAI*, pp. 484–499.

[24] Y. Ng, T. Shinohara, Developments from enquiries into the learnability of the pattern languages from positive data, Theoretical Computer Science 397 (2008) 150–165.

[25] D. Reidenbach, A non-learnable class of E-pattern languages, Theoretical Computer Science 350 (2006) 91–102.

[26] D. Reidenbach, Discontinuities in pattern inference, Theoretical Computer Science 397 (2008) 166–193.

[27] D. Reidenbach, J. Schneider, Morphically primitive words, Theoretical Computer Science 410 (2009) 2148–2161.

[28] R. Reischuk, T. Zeugmann, An average-case optimal one-variable pattern language learner, Journal of Computer and System Sciences 60 (2000) 302–335.

[29] H. Rogers, Theory of Recursive Functions and Effective Computability, MIT Press, Cambridge, MA, 1992. 3rd print.

[30] G. Rozenberg, A. Salomaa, Handbook of Formal Languages, volume 1, Springer, Berlin, 1997.

[31] T. Shinohara, Polynomial time inference of extended regular pattern languages, in: Proc. RIMS Symposia on Software Science and Engineering, Kyoto, volume 147 of *LNCS*, pp. 115–127.

[32] R. Wiehagen, T. Zeugmann, Ignoring data may be the only way to learn efficiently, Journal of Experimental and Theoretical Artificial Intelligence 6 (1994) 131–144.