Robust learning of automatic classes of languages*

Sanjay Jain
School of Computing,
NUS, Singapore 117417,
Republic of Singapore.
sanjay@comp.nus.edu.sg

Eric Martin
School of Computer Science
and Engineering,
UNSW, Sydney 2052, Australia.
emartin@cse.unsw.edu.au

Frank Stephan
Department of Mathematics,
NUS, Singapore 119076,
Republic of Singapore
fstephan@comp.nus.edu.sg

Abstract

This paper adapts and investigates the paradigm of robust learning, originally defined in the framework of inductive inference of classes of recursive functions, to learning languages from positive data. Robustness is a very desirable property, as it captures a form of invariance of learnability under admissible transformations of the object of study. The classes of languages of interest are automatic, that is, recognisable by finite automata. A class of first-order definable operators — called translators — is introduced as natural transformations that preserve automaticity of a class of languages and the inclusion relationships between languages in the class. For many learning criteria, we provide a characterization of the classes of languages all of whose translations are learnable under that criterion. The learning criteria have been chosen from the literature on both explanatory learning and query learning, and include consistent and conservative learning, strong-monotonic learning, strong-monotonic consistent learning, finite learning, learning from subset queries, learning from superset queries and learning from membership queries.

1 Introduction

The present work introduces and studies a notion of robust learning in the context of language learning. Essentially, robust learning guarantees that learnability of a class is invariant under any admissible transformation: that is, not only the class itself, but also each of its images under an admissible transformation, is learnable. The search for invariants is quite prominent in many fields of mathematics. For example, Hermann Weyl described Felix Klein's famous Erlangen programme on the algebraic foundation of geometry in these words [28]: "If you are to find deep properties of some object, consider all natural transformations that preserve your object." Bārzdiņš was interested in this question in relation to learning a class of recursive functions, and he conjectured the following, see [6, 30]. Let a class of recursive functions be given. Then every image of the class under a general recursive operator is learnable iff the class is a subclass of a recursively enumerable (uniformly recursive) class of functions. Recursively enumerable classes of functions can be easily identified by a technique called "learning by enumeration" [8]. Using this technique, one just conjectures the first function in the list which is consistent with all data seen so far. Learnability of classes of functions by such algorithms cannot be destroyed by general recursive operators. Bārzdiņš' conjecture, see [30], essentially says that the enumeration technique also captures all cases where robust learnability holds. Fulk [6] disproved the conjecture and this led to a rich field of exploration within the field of function learning [16, 18, 26]. Further refinements, such as uniform robust learnability [18] (where the learner for a transformed class has to be computable in a description of the transformation) and hyperrobust learnability [26] (learnability, by the same learner, of all the transformations of a certain kind — more precisely, transformations given by primitive recursive operators) have also been investigated.

It was natural to try and generalize robust learning to the case of language learning, which was the first object of study in inductive inference and has been more broadly investigated than function learning. However, all attempts so far have been disappointing. This work intends to remedy this lack of success based on a natural notion that enjoys good properties and is amenable to appealing characterizations. Our work is mostly restricted to automatic families of languages (this will be

*This work was supported in part by the NUS grants R252-000-308-112 and R146-000-114-112.

explained shortly), which are a particular kind of automatic structures [9, 10, 21, 22]. This restriction to automatic families was mainly driven by the aim of providing appealing characterizations; the notions themselves are meaningful and well motivated even for arbitrary families of languages. We just found it preferable to investigate the notions under assumptions that yield the most elegant set of results.

Automatic families of languages are classes of regular languages of the form $(L_i)_{i\in I}$ such that I is a regular set and $\{(i, x) : x \in L_i\}$ are regular sets. Recall that sets of finite strings over some finite alphabet are regular if they are recognizable by a finite state automaton. Sets of pairs of finite strings over respective alphabets are regular if they are recognizable by a finite state multi-input automaton that uses two different inputs to read both coordinates of the pair, with a special symbol (say $\star$) being used to pad a shorter coordinate. For instance, to accept the pair $(010, 45)$ an automaton should read 0 from the first input and 4 on the second input and change its state from the start state to some state q_1, then read 1 from the first input and 5 from the second input and change its state from q_1 to some state q_2, finally read 0 from the first input and $\star$ from the second input and change its state from q_2 to an accepting state. Note that it is essential that all inputs involved are read synchronically, that is, one character per input and cycle. Learnability of automatic families have recently been studied [11, 12], and these families can be viewed as a special case of indexed families — a well-studied concept in learning theory [1, 23, 24]. One major advantage of automatic families over indexed families is that their theory is first-order definable [10, 11, 12, 21]. Thus, some properties, such as inclusion structure among languages in the class, tend to be decidable (unlike in the case of indexed families). The inclusion structure of the class plays an important role in this paper.

By a transformation of a class $(L_i)_{i\in I}$ of automatic regular languages, we mean a mapping Φ from subsets of the source domain D to subsets of an image domain D' such that, implicitly, $(L_i)_{i\in I}$ is mapped to a class $(L'_i)_{i\in I} = (\Phi\langle L_i\rangle)_{i\in I}$ of languages. Here the operator Φ is definable by a first-order formula, Φ preserves inclusions of all sets and Φ preserves noninclusions of members of the class. The first-order formula defining Φ is denoted by the same name and chosen such that the formula Φ has a unique free variable x and for all $L \subseteq D$ and all terms $t \in D'$, $\Phi[t/x]$ holds iff $t \in \Phi\langle L\rangle$. The formula Φ defining the operator may make reference to a one-place predicate expressing the membership in L as well as to a two-place predicate expressing whether $y \in L_i$ for some y ranging over D and i ranging over I. A key result of the theory of automatic structures is that the image $(\Phi\langle L_i\rangle)_{i\in I}$ of an automatic family under such an operator is again an automatic family. We call operators of this type, which in addition preserve inclusions among the subsets of D and noninclusions among members of the class, a translation and study the impact of such translations on learnability.

We proceed as follows. In Sections 2 and 3, we introduce the notation and notions studied in the paper. In Section 4, we illustrate the notions with a few examples and provide a general characterization of robust learnability in the limit of automatic families of languages. In Sections 4 to 8, we provide many further characterizations of robust learnability for some of the learning criteria that have been studied in the literature: consistent and conservative learning, strong-monotonic learning, strong-monotonic consistent learning, finite learning, learning from subset queries, learning from superset queries, and learning from membership queries. The characterizations are natural as they express a particular constraint on the inclusion structure of the original class. In many cases, they deal not only with transformations of the original class under all possible translations, but also with transformations under continuous translations or text-preserving translations. A text-preserving translator Φ can be applied not only to the languages, but also to texts describing the languages. Thus they can be realised via general recursive operators translating each text for a set $L \subseteq D$ into texts for $\Phi\langle L\rangle$.

2 Automatic structures, languages and translations

The languages considered in inductive inference [13] consist of numbers meant to code some underlying structure, but where the coding does not need to be explicit. In the context of this work though, where languages have to be recognized by finite automata, a minimum of structure has to be given to the members of a language: they will be assumed to be strings over a finite vocabulary denoted with Σ. We will always assume that the *alphabet* Σ is finite but taken so large that the set $\Sigma^\star$ of strings over this alphabet is a superset of all the regular sets considered. For $x \in \Sigma^\star$, the length of x, denoted $|x|$, is the number of symbols occurring in x, for example, $|00121| = 5$. We write xy for the concatenation of two strings x and y. We denote the empty string by ε.

We denote by D and I two regular subsets of $\Sigma^\star$. We assume that the members of Σ are strictly ordered and given two members x and y of $\Sigma^\star$, we write $x <_{ll} y$ iff x is length-lexicographically smaller than y, that is, if either $|x| < |y|$ or $|x| = |y|$ and x comes lexicographically before y. We write $x \leq_{ll} y$ iff $x = y$ or $x <_{ll} y$.

In order to capture the constraint that a class of languages is uniformly recognizable by a finite automaton, we make use of a particular kind of automatic structures [22], that for simplicity, is still referred to as automatic structures with no extra qualifier. Basically, the structures under consideration offer enough expressive power to refer to the target language that a learner will be given a presentation of and has to eventually correctly identify, and to refer to the whole class of languages that are the object of learning. A unary predicate symbol and a binary predicate symbol are used to refer to the target language and the class of languages, respectively.

Definition 1. We call *automatic structure* any $\mathcal{V}$-structure $\mathfrak{M}$ whose domain is $\Sigma^\star$, with $\mathcal{V}$ some relational vocabulary with the following properties. (1) $\mathcal{V}$ contains the unary predicate symbol L^1 and the binary predicate symbol L^2 (and possibly more predicate symbols). (2) The interpretation of L^1 in $\mathfrak{M}$ is included in D and the interpretation of L^2 in $\mathfrak{M}$ is included in $I \times D$. (3) The interpretation of all predicate symbols in $\mathcal{V}$ in $\mathfrak{M}$ is regular (recognizable by a multi-input finite automaton).

Given two terms t and t', we write $t \in L$ for $L^1(t)$ and $t' \in L_t$ for $L^2(t, t')$. By *language* we mean a subset of $\Sigma^\star$; by *source language* we mean a subset of D.

Definition 2. By *automatic class* we mean a repetition free regular I-family $(L_i)_{i\in I}$ of source languages (so $\{(i, x) : i \in I, x \in L_i\}$ is recognizable by a multi-input automaton). Members of I are referred to as *indices*.

The constraint that automatic classes be repetition free is not standard when one considers learning of indexed families. However, this is at no loss of generality in the context of this work and allows one to substantially simplify the arguments in most proofs.

We consider transformations of languages that are definable in the language used to describe the target language and the class of languages to be learnt. In the next definition, x is meant to denote an arbitrary member of the former.

Definition 3. We call *automatic translator* any first-order formula over the vocabulary of some automatic structure with the distinguished variable x as unique free variable.

The key constraint on translators is that they have to preserve the inclusion structure of the original class:

Definition 4. Let an automatic class $\mathbf{I} = (L_i)_{i\in I}$ be given. We call *automatic* $\mathbf{I}$*-translator* any automatic translator Φ with the following properties. For all source language L, denote by $\Phi_{\mathbf{I}}\langle L\rangle$ the language consisting of all $s \in \Sigma^\star$ such that $\Phi[s/x]$ is true in all automatic structures in which the interpretation of L^1 is L and the interpretation of L^2 is $\{(i, s) : i \in I, s \in L_i\}$.

- For all source languages L and L', if $L \subseteq L'$, then $\Phi_{\mathbf{I}}\langle L\rangle \subseteq \Phi_{\mathbf{I}}\langle L'\rangle$.
- For all members i and j of I, if $L_i \nsubseteq L_j$, then $\Phi_{\mathbf{I}}\langle L_i\rangle \nsubseteq \Phi_{\mathbf{I}}\langle L_j\rangle$.

Given an automatic $\mathbf{I}$-translator Φ, we let $\Phi\langle\mathbf{I}\rangle$ denote $(\Phi_{\mathbf{I}}\langle L_i\rangle)_{i\in I}$; we refer to any such family as a *translation of* $\mathbf{I}$. Note that a translation is always defined with respect to $\mathbf{I}$-translators for a particular automatic class $\mathbf{I}$. We drop the reference to $\mathbf{I}$ for ease of notation.

The benefit that translators be definable by first-order formulas is that automaticity is preserved. Indeed, by results in [21]:

Property 5. *For all automatic classes* $\mathbf{I}$*, all translations of* $\mathbf{I}$ *are automatic.*

3 Texts and learnability

Let us recall the basic concepts in inductive inference as originally defined in [8] and fix some notation. The only difference with the classical framework of learning from positive data is that we consider languages over strings rather than natural numbers.

We denote by SEQ the set of finite sequences of members of $\Sigma^\star \cup \{\#\}$. Given $\sigma \in$ SEQ, we denote by range(σ) the set of members of $\Sigma^\star$ that occur in σ. Given $\sigma \in$ SEQ and a family $\mathbf{I} = (L_i)_{i\in I}$ of languages, we say that σ is *for* $\mathbf{I}$ iff range$(\sigma) \subseteq L_i$ for some $i \in I$. Given a language L, a *text* for L refers to an enumeration $(e_k)_{k\in\mathbb{N}}$ of $L \cup \{\#\}$ where # might not occur. The concatenation of a member $\sigma \in \mathrm{SEQ}$ with a member s of $\Sigma^\star \cup \{\#\}$ is written $\sigma \diamond s$. We also write $\sigma \diamond \tau$ for the concatenation of $\sigma, \tau \in \mathrm{SEQ}$. An initial segment of a member σ of SEQ is a member τ of SEQ such that σ is of the form $\tau \diamond s_1 \diamond \cdots \diamond s_n$ for some $n \in \mathbb{N}$ and $s_1, \ldots, s_n \in \Sigma^*$.

The notion of translator is quite general and it is worthwhile to examine to which extent it can be constrained to continuous transformations, that is, translators such that any member of the translation can be determined from a finite subset of the original language:

Definition 6. Given an automatic class $\mathbf{I} = (L_i)_{i \in I}$ and an automatic $\mathbf{I}$-translator Φ, we say that Φ is *text-preserving* just in case for all source languages L and for all $s \in \Phi_{\mathbf{I}}\langle L\rangle$, there is a finite subset F of L with $s \in \Phi_{\mathbf{I}}\langle F\rangle$.

We talk about *text-preserving translation of* $\mathbf{I}$ to refer to any family of the form $\Phi\langle\mathbf{I}\rangle$ where Φ is a text-preserving automatic $\mathbf{I}$-translator.

Example 7. Given an automatic class $\mathbf{I} = (L_i)_{i \in I}$, let Φ^{nc} be a formula with x as unique free variable that expresses that $x \in I \wedge L \not\subseteq L_x$. Then for all languages L, $\Phi^{nc}_{\mathbf{I}}\langle L\rangle$ is the set of all $i \in I$ with $L \not\subseteq L_i$. Moreover, Φ^{nc} is text-preserving.

Almost all results will involve recursive learners, with one exception where we had to drop the recursiveness requirement. This result will be expressed in terms of general learners. Both kinds of learners are defined next.

Definition 8. A *general learner* is any partial function from SEQ into I. A *learner* is any partial recursive function from SEQ into I where the set of σ on which the learner is defined is also recursive.

In the context of automatic structures, the fact that a learner is undefined on some input signals that the learner cannot make a reasonable guess, rather than the learner not being unable to make a guess due to computational unfeasibility. This justifies the constraint on the domain of a learner in the definition above. We could also let a learner output some special symbol like ? rather than being undefined.

Definition 9 (Gold [8]). Let $\mathbf{I} = (L_i)_{i \in I}$ be an automatic class.

Given a learner M, we say that M *learns* $\mathbf{I}$ just in case for all $i \in I$ and for all *texts* $(e_k)_{k \in \mathbb{N}}$ for L_i, $M\big((e_0 \ldots, e_k)\big)$ is defined and equal to i for cofinitely many $k \in \mathbb{N}$.

We say that $\mathbf{I}$ *is learnable* iff some learner learns $\mathbf{I}$.

Note that for simplicity, we talk about "learning" for a notion that in the literature, is more precisely referred to as *explanatory learning*. Furthermore, note that the definition above takes advantage of the one-one indexing of the automatic families considered.

We now recall some of the restrictions on learnability that have been investigated in the literature [19, 29, 20] and that will be considered in this paper, individually or combined.

Definition 10. Let $\mathbf{I} = (L_i)_{i \in I}$ be an automatic class. Let M, a learner that learns $\mathbf{I}$, be given.

M is *consistent* iff for all $\sigma \in \mathrm{SEQ}$, if σ is for $\mathbf{I}$, then $M(\sigma)$ is defined and $\mathrm{range}(\sigma) \subseteq L_{M(\sigma)}$.

M is *strong-monotonic* iff for all members σ and τ of SEQ, if σ is an initial segment of τ, τ is for $\mathbf{I}$ and both $M(\sigma)$ and $M(\tau)$ are defined, then $L_{M(\sigma)} \subseteq L_{M(\tau)}$.

M is *conservative* iff for all $\sigma, \tau \in \mathrm{SEQ}$, if $\sigma \diamond \tau$ is for $\mathbf{I}$, both $M(\sigma)$ and $M(\sigma \diamond \tau)$ are defined and $L_{M(\sigma \diamond \tau)} \neq L_{M(\sigma)}$, then some member of $\mathrm{range}(\sigma \diamond \tau)$ does not belong to $L_{M(\sigma)}$.

M is *confident* iff for all enumerations e of members of $\Sigma^\star \cup \{\#\}$, there exists $m \in \mathbb{N}$ such that, for all $n \geq m$, $M\big((e(0) \ldots e(n))\big)$ is undefined or $M\big((e(0) \ldots e(n))\big)$ is equal to $M\big((e(0) \ldots e(m))\big)$.

Definition 11. An automatic class $\mathbf{I}$ is said to be *consistently, strong-monotonically, conservatively or confidently learnable* iff some consistent, strong-monotonic, conservative or confident learner learns $\mathbf{I}$, respectively.

For robust learning, one requires that each translation $\Phi\langle\mathbf{I}\rangle$ of the family $\mathbf{I}$ is learnable (according to the given criterion), where Φ ranges over all automatic $\mathbf{I}$-translators. Note that requiring the learnability of each translation implies requiring the learnability of $\mathbf{I}$ itself, as the translator can be the identity-operator. In some cases, we consider learnability only of $\Phi\langle\mathbf{I}\rangle$, for all text-preserving automatic $\mathbf{I}$-translators Φ.

The proof of the learnability of indexed families of languages in terms of tell-tales given by Angluin [1] can easily be adapted to the current setting, with indexed families replaced by automatic classes. The characterization is simpler as the tell-tales do not have to be assumed to be computable from the languages in the class, but are necessarily so.

Definition 12. Let $\mathbf{I} = (L_i)_{i \in I}$ be an automatic class. Given $i \in I$, a *tell-tale set for L_i* is a finite subset F of L_i such that, for all $i' \in I$, if $F \subseteq L_{i'} \subseteq L_i$ then $L_i = L_{i'}$.

Proposition 13 (based on Angluin [1]). *Let $\mathbf{I} = (L_i)_{i \in I}$ be an automatic class. Then $\mathbf{I}$ is learnable iff for all $i \in I$, there exists a tell-tale set for L_i.*

Alternatively, one could also describe the tell-tale set by an upper bound in order to get a first-order definable formula which expresses learnability. An automatic class $\mathbf{I} = (L_i)_{i \in I}$ is learnable iff, for every $i \in I$, there is a bound $b_i \in D$ (the domain of the source languages) such that no $j \in I$ satisfies $\{x \in L_i : x \leq_{ll} b_i\} \subseteq L_j \subset L_i$. This is equivalent to

$$\forall i \in I \, \exists b_i \in D \, \forall j \in I \Big[\big(\exists x \in D[x \leq_{ll} b_i \wedge x \in L_i \setminus L_j]\big) \quad \vee \quad \big(\exists y \in D[y \in L_j \setminus L_i]\big) \quad \vee \quad \big(\forall z \in D[z \in L_i \Rightarrow z \in L_j]\big) \Big].$$

In order not to clutter notation, we will from now on abstain from breaking subset-relations down into first-order formulas as exemplified with the previous formula; we leave it to the reader to formalize subset-relations via quantified predicates using membership.

Example 14. Let an automatic class $\mathbf{I} = (L_i)_{i \in I}$ be given. There are two learners, M_{smon} and M_{ex} (which use automatic description of $\mathbf{I}$ as a parameter), which learn $\mathbf{I}$ whenever $\mathbf{I}$ is strong-monotonically and explanatorily learnable, respectively. These two learners are defined as follows.

- In response to $\sigma \in \mathrm{SEQ}$, M_{smon} outputs the unique $i \in I$ such that $\mathrm{range}(\sigma) \subseteq L_i$ and for all $j \in I$, if $\mathrm{range}(\sigma) \subseteq L_j$ then $L_i \subseteq L_j$; if such an i does not exist then M_{smon} is undefined.
- In response to $\sigma \in \mathrm{SEQ}$, M_{ex} outputs the unique $i \in I$ such that $\mathrm{range}(\sigma) \subseteq L_i$ and there is no $j \in I$ with (1) $\mathrm{range}(\sigma) \subseteq L_j$ and (2) either $L_j \subset L_i$ or $j <_{ll} i$ and $L_i \not\subset L_j$; if such an index i does not exist then M_{ex} is undefined.

Proof: Clearly, the learner M_{smon} is partial recursive and has a recursive domain. Note that by definition M_{smon} outputs a hypothesis only if it is the smallest one (with respect to $\subseteq$) which is consistent with the data observed so far; hence any further hypothesis output by M_{smon} is a superset of the current one and therefore M_{smon} is strong-monotonic. Now it is shown that M_{smon} succeeds on $\mathbf{I}$ if $\mathbf{I}$ is learnable by some strong-monotonic learner N, even if N is not partial recursive. In particular, it is shown that, for all σ for $\mathbf{I}$, whenever $N(\sigma) = i$ then either $\mathrm{range}(\sigma) \not\subseteq L_i$ or $M_{smon} = i$; this implies learnability as whenever N converges to i on a text for L_i, so does M_{smon}. So assume that N outputs i and L_i is consistent with all the data seen so far. Let j be any further index with $\mathrm{range}(\sigma) \subseteq L_j$. There is a text for L_j which starts with σ; hence N outputs j on some τ extending σ. It follows from the strong monotonicity of N that $L_i \subseteq L_j$. In other words, i is the index of the $\subseteq$-minimal language consistent with all the data seen so far and hence $M_{smon}(\sigma) = i$. Therefore M_{smon} is a strong-monotonic learner for the class $\mathbf{I}$ whenever $\mathbf{I}$ has a strong-monotonic learner at all.

Clearly, the learner M_{ex} is partial recursive and has a recursive domain. Now consider that the class $\mathbf{I}$ is explanatorily learnable. Therefore it satisfies Angluin's tell-tale condition. Fix $i \in L_i$ and a text for L_i; now it is needed to show that M_{ex} converges on this text to i. For every sufficiently long initial segment σ of the given text for L_i it holds that (a) $\mathrm{range}(\sigma)$ contains the tell-tale set of L_i and (b) $\mathrm{range}(\sigma)$ contains some element of $L_i \setminus L_j$ for every $j <_{ll} i$ with $L_i \not\subseteq L_j$. Condition (a) implies that there is no j with $\mathrm{range}(\sigma) \subseteq L_j \subset L_i$ and condition (b) implies that there is no $j <_{ll} i$ with $\mathrm{range}(\sigma) \subseteq L_j \subset L_i$. Hence, for every $i \in I$ and every text for L_i, $M_{ex}(\sigma)$ conjectures i on every sufficiently long initial segment of the text. ■

Note that not all explanatorily learnable classes are strong-monotonically learnable, hence the learner M_{smon} is not so powerful as M_{ex}. An example of a class which is explanatorily learnable but not strong-monotonically learnable is the class of all sets $\{0,1\}^\star \setminus \{x\}$ with $x \in \{0,1\}^\star$. Furthermore, the above learners can of course also run on classes of the form $\Phi\langle\mathbf{I}\rangle$ (using $\Phi\langle\mathbf{I}\rangle$ as a parameter instead of $\mathbf{I}$) whenever such a class is learnable under the corresponding condition.

4 General characterization

We start our investigation with two examples of conditions that guarantee robust learnability.

Proposition 15. *Let an automatic class $\mathbf{I} = (L_i)_{i \in I}$ be given.*

- *If $(\{L_i : i \in I\}, \supset)$ is well ordered (for the superset relation, not the subset relation), then all translations of $\mathbf{I}$ are learnable.*
- *If for all $i, j \in I$, $L_i \subseteq L_j$ and $L_i = L_j$ are equivalent, then all translations of $\mathbf{I}$ are learnable.*

Proof: Let an automatic $\mathbf{I}$-translator Φ be given.

Suppose that $(\{L_i : i \in I\}, \supset)$ is well ordered. Let ordinal κ and $(L'_\lambda)_{\lambda<\kappa}$ be a well ordering of $(\{L_i : i \in I\}, \supset)$. Let M_{smon} be the learner defined in Example 14. Let $i \in I$ and $\lambda < \kappa$ be such that $L_i = L'_\lambda$. Let $s \in \Phi_\mathbf{I}\langle L'_\lambda\rangle$ be such that if $\lambda+1 < \kappa$ then $s \notin \Phi_\mathbf{I}\langle L'_{\lambda+1}\rangle$. Clearly, for all members σ of SEQ with $|\sigma| \geq |i|$, $s \in \mathrm{range}(\sigma)$ and $\mathrm{range}(\sigma) \subseteq \Phi_\mathbf{I}\langle L_i\rangle$, $M_{smon}(\sigma) = i$, hence M_{smon} learns $\Phi\langle\mathbf{I}\rangle$.

Suppose that for all members i and j of I, $L_i \subseteq L_j$ and $L_i = L_j$ are equivalent. Let M be a learner such that for all $\sigma \in$ SEQ for $\mathbf{I}$, $M(\sigma)$ is the $\leq_{ll}$-least $i \in I$ such that $\mathrm{range}(\sigma) \subseteq \Phi_\mathbf{I}\langle L_i\rangle$. Obviously, M learns $\Phi\langle\mathbf{I}\rangle$. ■

As can be expected, learning does not imply robust learning, even if we wanted to restrict to text-preserving translations:

Proposition 16. *Some text-preserving translation of some learnable automatic class is not learnable.*

Proof: Take I equal to $\{0,1\}^\star$. Clearly, there exists an automatic class $\mathbf{I} = (L_i)_{i\in I}$ such that $L_\varepsilon = \{0,1\}^\star$ and for all $i \in I$ with $i \neq \varepsilon$, $L_i = \{i\}$. Define Φ as $\exists v(v \in L \wedge v \neq x)$. Obviously, $\mathbf{I}$ is learnable, Φ is an automatic $\mathbf{I}$-translator, and $\Phi\langle\mathbf{I}\rangle$ consists of I and all cosingletons of I except for $I \setminus \{\varepsilon\}$, implying that $\Phi\langle\mathbf{I}\rangle$ is not learnable, see [8]. ■

The next proposition offers a general characterization of robust learning in this framework along the lines of Proposition 13.

Proposition 17. *Let $\mathbf{I} = (L_i)_{i\in I}$ be an automatic class. The following conditions are equivalent.*

(1) *All translations of $\mathbf{I}$ are learnable.*

(2) *All text-preserving translations of $\mathbf{I}$ are learnable.*

(3) *For all $i \in I$, there exists $b_i \in I$ such that for all $j \in I$, either $L_j \not\subset L_i$ or there exists $k \in I$ with $k \leq_{ll} b_i$, $L_j \subseteq L_k$ and $L_i \nsubseteq L_k$.*

Proof: It suffices to prove that (3) implies (1) and (2) implies (3). Assume that (3) holds, and let Φ be an automatic $\mathbf{I}$-translator. Let a learner M be such that in response to $\sigma \in$ SEQ, M outputs the $\leq_{ll}$-least $i \in I$, if any, such that $\mathrm{range}(\sigma) \subseteq \Phi_\mathbf{I}\langle L_i\rangle$ and for all $j \in I$, if $L_j \subset L_i$ then $\mathrm{range}(\sigma)$ contains a member of $\Phi_\mathbf{I}\langle L_i\rangle \setminus \Phi_\mathbf{I}\langle L_k\rangle$ for some $k \leq_{ll} b_i$ with $L_j \subseteq L_k$. It is easily verified that M learns $\Phi_\mathbf{I}\langle\mathbf{I}\rangle$.

We show that (2) implies (3). For a contradiction, assume that there exists $i \in I$ for which there exists no $b_i \in I$ such that for all $j \in I$, either $L_j \not\subset L_i$ or there exists $k \in I$ with $k \leq_{ll} b_i$, $L_j \subseteq L_k$ and $L_i \nsubseteq L_k$. Thus, I is infinite. We have to exhibit a text-preserving automatic $\mathbf{I}$-translator Φ such that $\Phi\langle\mathbf{I}\rangle$ is not learnable. Define Φ as follows: given a source language L, $\Phi_\mathbf{I}\langle L\rangle$ consists of all $(p, n) \in I \times I$ such that at least one of the following conditions holds.

- For all s with $s \leq_{ll} n$, if $s \in L_p$, then $s \in L$.
- For all s with $s \leq_{ll} n$, if $s \in L_p$, then $s \in L_i$. Furthermore, for all $k \in I$ with $k \leq_{ll} \max(p, n)$, either $L_i \subseteq L_k$ or $L \nsubseteq L_k$.

This is a first-order definition. Let $H_j = \Phi_\mathbf{I}\langle L_j\rangle$.

It follows from definition of Φ that if $L \subseteq L'$, then $\Phi(L) \subseteq \Phi(L')$.

Let members j, j' and n of I be such that there exists s with $s \leq_{ll} n$, $s \in L_j$ and $s \notin L_{j'}$. Then at least one of the following conditions holds.

- For all s' with $s' \leq_{ll} n$, if $s' \in L_i$, then $s' \in L_{j'}$. Hence $s \in L_j \setminus (L_i \cup L_{j'})$ and $s \leq n$; therefore (j, n) belongs to $H_j \setminus H_{j'}$.
- There exists s' such that $s' \leq_{ll} n$, $s' \in L_i$ and $s' \notin L_{j'}$. Let $j'' = \max(j', n)$. Clearly, $(j, j'') \in H_j$. Furthermore, by the existence of s' and since $L_{j'} \subseteq L_{j'}$, it follows that if k is chosen as j', then the condition for checking whether $(j, j'') \in H_{j'}$ is not satisfied. Hence (j, j'') belongs to $H_j \setminus H_{j'}$.

Hence by case distinction, $H_j \not\subseteq H_{j'}$.

We conclude that Φ is an automatic $\mathbf{I}$-translator. Moreover, it is immediately verified that Φ is text-preserving. Now for every $b_i \in I$, there is a $j \in I$ with $L_j \subset L_i$ such that for all $k \in I$ with $k \leq_{ll} b_i$, either $L_i \subseteq L_k$ or $L_j \not\subseteq L_k$. It follows that for all elements (p, n) of H_i, if both p and n are $\leq_{ll}$-smaller than b_i, then (p, n) is also in H_j. But H_j is still a proper subset of H_i and hence H_i does not have a finite tell-tale set. Hence $\Phi\langle\mathbf{I}\rangle$ is not learnable. ∎

5 Consistent and conservative learning

Consistency is a rather weak constraint on learners, and is often combined with other desirable properties. In Section 7, we will combine it with strong-monotonicity. In this section, we combine it with conservativeness. Note that here "a class is consistently and conservatively learnable" means that the class is learnable by a learner which is both consistent and conservative (rather than having two different learners, one satisfying consistency and the other satisfying conservativeness). Similar convention applies to combining other constraints on learners. Let us first illustrate the notion with an example.

Example 18. Take I equal to $\{1^n, 2^n : n \in \{1, 2, 3, \dots\}\}$. Let $\mathbf{I} = (L_i)_{i \in I}$ be defined by $L_{1^n} = \{0^m : m > n\}$ and $L_{2^n} = \{0^m : m < n\}$ for all nonzero $n \in \mathbb{N}$. Note that $\mathbf{I}$ is an automatic class that is neither $\subset$- nor $\supset$-well founded. Let Φ be a text-preserving automatic $\mathbf{I}$-translator.

Some consistent and conservative learner M learns $\Phi\langle\mathbf{I}\rangle$, proceeding as follows in response to $\sigma \in \mathrm{SEQ}$.

- If σ is of the form $\tau \diamond s$, $M(\tau)$ is defined and $\mathrm{range}(\sigma) \subseteq L_{M(\tau)}$, then M outputs $M(\tau)$.
- Else if $\mathrm{range}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_{1^n}\rangle$ and $\mathrm{range}(\sigma) \not\subseteq \Phi_{\mathbf{I}}\langle L_{1^{n+1}}\rangle$ for some $n \in \mathbb{N}$, then M outputs 1^n.
- Else M conjectures 2^n for the least $n \in \mathbb{N} \setminus \{0\}$ with $\mathrm{range}(\sigma)$ included in $\Phi_{\mathbf{I}}\langle L_{2^n}\rangle$.

Since Φ is text-preserving, for all $n \in \mathbb{N} \setminus \{0\}$, every finite subset of $\Phi_{\mathbf{I}}\langle L_{1^n}\rangle$ is contained in $\Phi_{\mathbf{I}}\langle L_{2^m}\rangle$ for some $m \in \mathbb{N}$; hence M is consistent. By the first clause in the definition of M, M is conservative.

We now show that M learns $\Phi\langle\mathbf{I}\rangle$. Let $n \in \mathbb{N} \setminus \{0\}$ be given. Presented with a text for $\Phi_{\mathbf{I}}\langle L_{1^n}\rangle$, M eventually observes a datum outside $\Phi_{\mathbf{I}}\langle L_{1^{n+1}}\rangle$, at which point M either conjectures 1^n or outputs the last hypothesis — of the form 2^m for some $m \in \mathbb{N} \setminus \{0\}$ — until $\Phi_{\mathbf{I}}\langle L_{2^m}\rangle$ becomes inconsistent with the data observed, at which point M makes a mind change to 1^n. Presented with a text for $\Phi_{\mathbf{I}}\langle L_{2^n}\rangle$, M eventually conjectures 2^n as soon as the data observed become inconsistent with $\Phi_{\mathbf{I}}\langle L_1\rangle$ and $\Phi_{\mathbf{I}}\langle L_{2^m}\rangle$ for all nonzero $m < n$, which is guaranteed to happen as there are only finitely many languages of the latter type.

Combined with Proposition 17, the next result characterizes robust learnability by consistent, conservative learners.

Proposition 19. *Let $\mathbf{I}$ be a learnable automatic class all of whose translations are learnable. Then every translation of $\mathbf{I}$ is consistently and conservatively learnable iff the members of $\mathbf{I}$ are well founded under inclusion.*

Proof: Suppose $\mathbf{I} = (L_i)_{i \in \mathbb{N}}$. First assume that the members of $\mathbf{I}$ are well founded under inclusion. Note that the inclusion structure is preserved under all translations and it is therefore sufficient to let a learner not exploit any other information on $\mathbf{I}$ but the inclusion structure of $\mathbf{I}$ and its automaticity. Let Φ be an automatic $\mathbf{I}$-translator.

Now define the learner M on input σ for $\mathbf{I}$ as follows. If there are τ, s with $\sigma = \tau \diamond s$ and $\mathrm{range}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_{M(\tau)}\rangle$, then let $M(\sigma) = M(\tau)$, else let $M(\sigma)$ be the length-lexicographically least $i \in I$, if any, such that

(a) $\mathrm{range}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_i\rangle$ and

(b) no $j \in I$ satisfies $\mathrm{range}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle \subset \Phi_{\mathbf{I}}\langle L_i\rangle$.

Now it is shown that M is consistent and conservative. Consistency needs that M is defined on all relevant input. To see this, consider any input σ from the class. Due to the well foundedness of the structure under inclusion, there is some i satisfying (a) and (b) from the definition of M. Hence, M is defined on σ. Furthermore, M is consistent, as a mind change is done whenever the old hypothesis becomes inconsistent. On the other hand, a consistent old hypothesis will not be withdrawn, hence the learner is conservative.

So it remains to show that M actually learns the class. Let i be the index of the set $\Phi_{\mathbf{I}}\langle L_i\rangle$ to be learnt and assume that a text for this set is given. Then there is an initial segment σ of this text such that the tell-tale of $\Phi_{\mathbf{I}}\langle L_i\rangle$ is contained in range(σ) and range(σ) $\not\subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$ for every $k <_{ll} i$ with $\Phi_{\mathbf{I}}\langle L_i\rangle \not\subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$. Note that i is then the length-lexicographically least index satisfying (a) and (b) in the definition of M above. Now let $j = M(\sigma)$. If $j = i$, then the learner has converged and will never give up the correct index. If $j \neq i$, then $\Phi_{\mathbf{I}}\langle L_j\rangle$ cannot be a superset of $\Phi_{\mathbf{I}}\langle L_i\rangle$ due to the definition of the learner. Thus the learner will eventually observe a datum inconsistent with the current hypothesis. So let τ be the least initial segment of the given text with range(τ) $\not\subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$. Then τ extends σ and M updates its hypothesis on τ to i, as i is the length-lexicographic least index satisfying (a) and (b) above with range(τ) in place of range(σ). Hence M converges to i and M is indeed a learner as required.

Conversely, assume for a contradiction that there exists a sequence $(i_n)_{n\in\mathbb{N}}$ of members of I such that $(L_{i_n})_{n\in\mathbb{N}}$ is a $\subseteq$-descending chain. Let $(j_n)_{n\in\mathbb{N}}$ be a sequence of members of I such that for all $n \in \mathbb{N}$, the following holds.

- $|j_n| < |j_{n+1}|$ and $L_{j_{n+1}} \subset L_{j_n}$;
- infinitely many members of $\{i_0, i_1, i_2, \ldots\}$ extend the initial segment of j_{n+1} of length $|j_n|$;
- j_{n+2} extends the initial segment of j_{n+1} of length $|j_n|$.

The existence of $(j_n)_{n\in\mathbb{N}}$ can be shown by induction. Suppose we have defined $j_0, j_1, \ldots, j_{n+1}$ (where we take j_0 to be i_0, and for purposes of definition, take $j_{-1} = \varepsilon$). To define j_{n+2} for $n \geq -1$, note that, by induction, there exists an $h \in \Sigma^*$ such that h extends the initial segment of j_{n+1} of length $|j_n|$, $|j_{n+1}| < |h|$, and h is an initial segment of infinitely many i_r's. Choose j_{n+2} to be one of these i_r's such that $L_{j_{n+2}} \subset L_{j_{n+1}}$. We now describe a coding of $(j_n)_{n\in\mathbb{N}}$ in three ω-words α, β and γ. The ω-word α is such that, for all $n \in \mathbb{N}$, its initial segment of length $|j_n|$ is an initial segment of j_{n+1}. The ω-word β consists of j_0 followed by the $|j_1| - |j_0|$ last symbols of j_1 followed by the last $|j_2| - |j_1|$ symbols of j_2 and so on. The ω-word γ is an ω-word over $\{0, 1\}$ such that, for all $m \in \mathbb{N}$, $\gamma(m) = 1$ iff there is $n \in \mathbb{N}$ with $|j_n| = m$. Clearly, $(j_n)_{n\in\mathbb{N}}$ can be retrieved from these three ω-words, and so there exists a Rabin automaton which recognizes all triples of ω-words which code an infinite descending chain of indices in I, see [27]. It follows that there exists an infinite regular language $R \subseteq I$, such that R consists of indices of an infinite descending chain of sets. Consider a first-order formula $\Phi(L)$ with x as unique free variable which expresses the following condition:

if there exists $j \in R$ with $L_j \subseteq L$,
then x is either the empty string or of the form $0y$ for some $y \in L$,
else x is of the form $0y$ for some $y \in L$.

It is easily verified that Φ is an automatic $\mathbf{I}$-translator.

Suppose a consistent learner M learns $\Phi\langle\mathbf{I}\rangle$. In response to the empty string, M must output some $i \in I$ for which there is $j \in R$ with $L_j \subseteq L_i$ and hence $\Phi_{\mathbf{I}}\langle L_j\rangle \subseteq \Phi_{\mathbf{I}}\langle L_i\rangle$. But by the choice of R, the empty word can be extended to a text for a language L_h with $L_h \subset L_j$, hence $\Phi_{\mathbf{I}}\langle L_h\rangle \subset \Phi_{\mathbf{I}}\langle L_j\rangle$. This implies that M cannot be conservative, completing the proof of the proposition. ■

6 Strong-Monotonic Learning

For the learning criterion of strong monotonicity, we first consider the concept by itself and then, in next section, combined with consistency. Again, we are able to characterize robust learnability under these restrictions, and to provide further insights.

Proposition 20. *Let $\mathbf{I} = (L_i)_{i\in\mathbb{N}}$ be an automatic class. The following conditions are equivalent.*

(1) *For all $i \in I$, there exists $b \in I$ such that for all $j \in I$ with $L_i \not\subseteq L_j$, there exists $k \in I$ with $k \leq_{ll} b$, $L_i \not\subseteq L_k$ and $L_j \subseteq L_k$.*

(2) *Every translation of $\mathbf{I}$ is strong-monotonically learnable.*

(3) *Every text-preserving translation of $\mathbf{I}$ is strong-monotonically learnable.*

Proof: First it is shown that (1) implies (2) and (3), by exhibiting that the learner M_{smon} from Example 14 learns the class. Note that M_{smon} only exploits the inclusion-structure and automaticity of the class $\Phi\langle\mathbf{I}\rangle$, making the same algorithm (using parameter Φ) work on all translations of $\mathbf{I}$. For all $i \in I$, let b_i be the $\leq_{ll}$-least $b \in I$ that satisfies the first condition of the proposition. Note that, for all $i \in I$, b_i is first-order definable from i; therefore the mapping $i \mapsto b_i$ is recursive.

Let Φ be an automatic $\mathbf{I}$-translator. Recall that M_{smon} from Example 14 works as follows: $M_{smon}(\sigma) = i$ iff i is the unique index such that $\text{range}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_i \rangle \subseteq \Phi_{\mathbf{I}}\langle L_j \rangle$ for all j with $\text{range}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L_j \rangle$. If such an index i does not exist, then $M_{smon}(\sigma)$ is undefined. It is clear that M_{smon} is partial recursive. Furthermore, if i and j are subsequent hypotheses of M_{smon}, then $\Phi_{\mathbf{I}}\langle L_i \rangle \subseteq \Phi_{\mathbf{I}}\langle L_j \rangle$ and hence M_{smon} is strong-monotonic.

So the main task is to show that M_{smon} indeed learns the class $\Phi\langle \mathbf{I} \rangle$. Let $i \in I$ and a text for $\Phi_{\mathbf{I}}\langle L_i \rangle$ be given. Let σ be any initial segment of the text which is so long that $\text{range}(\sigma) \nsubseteq \Phi_{\mathbf{I}}\langle L_k \rangle$ for every $k \leq_{ll} b_i$ with $\Phi_{\mathbf{I}}\langle L_i \rangle \nsubseteq \Phi_{\mathbf{I}}\langle L_k \rangle$. Then every j, with $\Phi_{\mathbf{I}}\langle L_i \rangle \nsubseteq \Phi_{\mathbf{I}}\langle L_j \rangle$, satisfies that there is a $k \leq_{ll} b_i$ with $\Phi_{\mathbf{I}}\langle L_i \rangle \nsubseteq \Phi_{\mathbf{I}}\langle L_k \rangle$ and $\Phi_{\mathbf{I}}\langle L_j \rangle \subseteq \Phi_{\mathbf{I}}\langle L_k \rangle$. By assumption, $\text{range}(\sigma) \nsubseteq \Phi_{\mathbf{I}}\langle L_k \rangle$ and hence $\text{range}(\sigma) \nsubseteq \langle L_j \rangle$. Thus i is the index of the language in the class which is minimal with respect to set inclusion and consistent with $\text{range}(\sigma)$; so $M_{smon}(\sigma) = i$. It follows that M_{smon} strong-monotonically learns $\Phi\langle \mathbf{I} \rangle$. Hence (1) implies (2) and (3).

Now it is shown that (3) implies (1), from which it follows that (2) implies (1). Let Φ^{nc} be the text-preserving automatic $\mathbf{I}$-translator defined in Example 7. Let M be a strong-monotonic learner that learns $\Phi^{nc}\langle \mathbf{I} \rangle$. Let $i \in I$ be given, and let b_i be the $\leq_{ll}$-least member of I for which there exists $\sigma \in \text{SEQ}$ such that $M(\sigma) = i$ and for all $s \in \text{range}(\sigma)$, $s \leq_{ll} b_i$. Then, for any $j \in I$ with $\Phi^{nc}_{\mathbf{I}}\langle L_i \rangle \nsubseteq \Phi^{nc}_{\mathbf{I}}\langle L_j \rangle$, there exists $k \in \text{range}(\sigma)$ with $k \leq_{ll} b_i$ and $k \in \Phi^{nc}_{\mathbf{I}}\langle L_i \rangle \setminus \Phi^{nc}_{\mathbf{I}}\langle L_j \rangle$, implying that $L_j \subseteq L_k$ and $L_i \nsubseteq L_k$. Hence condition (1) holds, completing the proof of the proposition. ■

Proposition 21. *Every automatic class has a strong-monotonically learnable translation.*

Proof: Let $\mathbf{I} = (L_i)_{i \in I}$ be an automatic class, and let Φ be the formula with a unique free variable x defined as $\forall z (z \in L_x \Rightarrow z \in L)$. So for all source languages L, $\Phi_{\mathbf{I}}\langle L \rangle$ is the set of all $j \in I$ with $L_j \subseteq L$. Clearly, Φ is an automatic $\mathbf{I}$-translator. Let M be a learner such that for all $k \in \mathbb{N}$ and members $i, i_0, \ldots, i_k$ of I, $M\big((i_0, \ldots, i_k)\big)$ is defined and equal to i iff $L_{i_0}, \ldots, L_{i_k}$ are all subsets of L_i and i is the index of the $\subseteq$-minimal member of $\mathbf{I}$ that contains $L_{i_0}, \ldots, L_{i_k}$. It is easily verified that M learns $\Phi\langle \mathbf{I} \rangle$ and that M is strong monotonic. ■

Proposition 22. *Let an automatic class* $\mathbf{I}$ *be such that some text-preserving translation of* $\mathbf{I}$ *is strong-monotonically learnable. Then the class* $\mathbf{I}$ *is strong-monotonically learnable.*

Proof: Set $\mathbf{I} = (L_i)_{i \in I}$. Let Φ be a text-preserving $\mathbf{I}$-translator such that $\Phi\langle \mathbf{I} \rangle$ is strong-monotonically learnable. Then there exists, for all $i \in I$, a finite subset F_i of $\Phi_{\mathbf{I}}\langle L_i \rangle$ such that, for all $j \in I$, if $F_i \subseteq \Phi_{\mathbf{I}}\langle L_j \rangle$ then $\Phi_{\mathbf{I}}\langle L_i \rangle \subseteq \Phi_{\mathbf{I}}\langle L_j \rangle$. Since Φ is text-preserving, there exists, for all $i \in I$, a finite subset E_i of L_i with $F_i \subseteq \Phi_{\mathbf{I}}\langle E_i \rangle$. For all members i and j of I, if $E_i \subseteq L_j$ then $\Phi_{\mathbf{I}}\langle E_i \rangle \subseteq \Phi_{\mathbf{I}}\langle L_j \rangle$; thus $F_i \subseteq \Phi_{\mathbf{I}}\langle L_j \rangle$, and hence $\Phi_{\mathbf{I}}\langle L_i \rangle \subseteq \Phi_{\mathbf{I}}\langle L_j \rangle$ and $L_i \subseteq L_j$. Thus, for every $i \in I$, there is a finite subset E_i of L_i such that, for all $j \in I$, if $E_i \subseteq L_j$ then $L_i \subseteq L_j$. As $\mathbf{I}$ is automatic, one can determine E_i effectively from i. Thus there is a learner M learning $\mathbf{I}$ which outputs in response to $\sigma \in \text{SEQ}$ the $\leq_{ll}$-least $i \in I$, if it exists, such that $E_i \subseteq \text{range}(\sigma) \subseteq L_i$ — if such an i does not exist, then the learner repeats the previous conjecture or remains undefined, depending on which case applies. It is easily verified that M is strong-monotonic. ■

7 Strong-Monotonic Consistent Learning

In the present section, the emphasis is on learners which are both strong-monotonic and consistent. The following example shows that consistency adds a genuine constraint to strong-monotonicity.

Example 23. Take I equal to $\{0,1\} \cup \{2\}^\star$. Let $\mathbf{I} = (L_i)_{i \in I}$ be defined by $L_0 = \{0\}$, $L_1 = \{1\}$ and $L_{2^n} = \{0,1\} \cup \{2^m : m \geq n\}$ for all $n \in \mathbb{N}$. Note that $\mathbf{I}$ is an automatic class. Let Φ be an automatic $\mathbf{I}$-translator. Note that M_{smon} from Example 14 (using $\Phi\langle \mathbf{I} \rangle$ as a parameter) learns $\Phi\langle \mathbf{I} \rangle$. This is due to the following behaviour on input $\sigma \in \text{SEQ}$: If $\text{range}(\sigma)$ is contained in exactly one of $\Phi_{\mathbf{I}}\langle L_0 \rangle$ and $\Phi_{\mathbf{I}}\langle L_1 \rangle$, then M_{smon} outputs 0 or 1, respectively. If there exists $n \in \mathbb{N}$ such that $\text{range}(\sigma)$ is contained in $\Phi_{\mathbf{I}}\langle L_{2^n} \rangle$ but not in $\Phi_{\mathbf{I}}\langle L_{2^{n+1}} \rangle$, then $M_{smon}(\sigma)$ outputs 2^n. In any other case, $M_{smon}(\sigma)$ is undefined. Clearly, M_{smon} is a strong-monotonic learner that learns $\Phi\langle \mathbf{I} \rangle$.

On the other hand, no consistent, strong-monotonic learner M learns $\Phi\langle \mathbf{I} \rangle$. Indeed, suppose otherwise, and let $\sigma \in \text{SEQ}$ be such that $\text{range}(\sigma)$ is a subset of the union of $\Phi_{\mathbf{I}}\langle L_0 \rangle$ with $\Phi_{\mathbf{I}}\langle L_1 \rangle$, but not a subset of any of both sets. Then $M(\sigma)$ is equal to 2^n for some $n \in \mathbb{N}$, and $M(\tau)$ remains equal to 2^n for all $\tau \in \text{SEQ}$ that extend σ and that are initial segments of a text for $\Phi_{\mathbf{I}}\langle L_{2^{n+1}} \rangle$; therefore M fails to learn $\Phi\langle \mathbf{I} \rangle$.

Proposition 24. *Given an automatic class $\mathbf{I} = (L_i)_{i \in I}$, the following conditions are equivalent.*

(1) $\{L_i : i \in I\}$ *is $\subset$-well-ordered and of type ω at most.*

(2) *Every translation of* $\mathbf{I}$ *is strong-monotonically consistently learnable.*

(3) *Every text-preserving translation of* $\mathbf{I}$ *is strong-monotonically consistently learnable.*

Proof: Assume that the first condition holds. It is easily verified that given an automatic $\mathbf{I}$-translator Φ, a learner that, on input $\sigma \in \text{SEQ}$, outputs the index of the $\subseteq$-least member L of $\{L_i : i \in I\}$ with $\text{range}(\sigma) \subseteq \Phi_{\mathbf{I}}\langle L \rangle$, if such an L exists, is a consistent strong-monotonic learner that learns $\mathbf{I}$. Hence (1) implies both (2) and (3).

To complete the proof of the proposition, it suffices to show that (3) implies (1). So assume that (3) holds. We first show that, for any members i and j of I, L_i and L_j are $\subseteq$-comparable. Suppose otherwise for a contradiction, and fix i, j such that $L_i \not\subseteq L_j$ and $L_j \not\subseteq L_i$. Consider a first-order formula $\Phi(L)$ with x as unique free variable which expresses both conditions that follow:

- If L intersects $L_i \setminus L_j$ or $L_j \setminus L_i$, then either x is the empty string or x is of the form $0y$ for some $y \in L$;
- If L intersects neither $L_i \setminus L_j$ nor $L_j \setminus L_i$, then x is of the form $0y$ for some $y \in L$.

It is easily verified that Φ is of an automatic $\mathbf{I}$-translator and is text-preserving. Let M be a consistent and strong-monotonic learner that learns $\Phi\langle \mathbf{I} \rangle$.

Since M is consistent, M must output, in response to the input sequence consisting of the empty string ε, a member k of I with $\varepsilon \in \Phi_{\mathbf{I}}\langle L_k \rangle$. By definition of Φ, L_k contains an element outside L_i or an element outside L_j. Hence $\Phi_{\mathbf{I}}\langle L_k \rangle \not\subseteq \Phi_{\mathbf{I}}\langle L_i \rangle$ or $\Phi_{\mathbf{I}}\langle L_k \rangle \not\subseteq \Phi_{\mathbf{I}}\langle L_j \rangle$. It follows that M cannot be strong-monotonic as it must be able to switch its hypotheses from k to i or j when it is presented with a text for $\Phi_{\mathbf{I}}\langle L_i \rangle$ or $\Phi_{\mathbf{I}}\langle L_j \rangle$, respectively.

Next we show that, for all $i \in I$ such that L_i is not $\subseteq$-minimal, there exists $j \in I$ with $L_j \subset L_i$ such that there exists no $k \in I$ with $L_j \subset L_k \subset L_i$. Assume otherwise for a contradiction, and choose $i \in I$ for which this property does not hold. Consider a first-order formula $\Phi'(L)$ with x as unique free variable which expresses that

$$x \in L \text{ and } [x \notin L_i \text{ or there exists } j \in I \text{ with } x \in L_j \text{ and } L_j \subset L_i].$$

We now show that, for all $j \in I$ and $k \in I$, if $L_j \not\subseteq L_k$, then $\Phi'_{\mathbf{I}}\langle L_j \rangle \not\subseteq \Phi'_{\mathbf{I}}\langle L_k \rangle$. Clearly, it suffices to verify that, for all $j \in I$, if $L_j \subset L_i$, then $\Phi'_{\mathbf{I}}\langle L_j \rangle \subset \Phi'_{\mathbf{I}}\langle L_i \rangle$. So let $j \in I$ be such that $L_j \subset L_i$. By the choice of i, there exists $k \in I$ with $L_j \subset L_k \subset L_i$. Let $x \in L_k \setminus L_j$ be given. Then x belongs to $\Phi'_{\mathbf{I}}\langle L_i \rangle$, hence $x \in \Phi'_{\mathbf{I}}\langle L_i \rangle \setminus \Phi'_{\mathbf{I}}\langle L_j \rangle$. Hence Φ' is noninclusion preserving, and it follows easily that Φ' is of an automatic $\mathbf{I}$-translator, and is text-preserving. However, as $\Phi'_{\mathbf{I}}\langle L_i \rangle$ is the ascending union of the sets of the form $\Phi'_{\mathbf{I}}\langle L_j \rangle$ with $L_j \subset L_i$, $\Phi'\langle \mathbf{I} \rangle$ cannot be learnable [1, 8], a contradiction.

To complete the proof of the proposition, it suffices to show that for all $i \in I$, there exist only finitely many $j \in I$ with $L_j \subset L_i$. For a contradiction, assume otherwise. Let R be the set of all $i \in I$ for which there exists infinitely many $j \in I$ with $L_j \subset L_i$. By assumption, R is not empty and by the previous paragraph, there is no $i \in R$ such that $L_i \subseteq L_j$ for all $j \in R$. Consider a first-order formula $\Phi''(L)$ with x as unique free variable which expresses that

- x is either the empty string or of the form $0y$ for some $y \in L$ if there exists $z \in L$ and $j \in R$ with $z \notin L_j$.
- x is of the form $0y$ for some $y \in L$ otherwise.

It is easily verified that Φ'' is of an automatic $\mathbf{I}$-translator, and is text-preserving.

Since M is consistent, M must output, in response to the input sequence consisting of the empty string ε, a member i of I with $\varepsilon \in \Phi''_{\mathbf{I}}\langle L_i \rangle$. So there is a $j \in R$ and $y \in L_i$ with $y \notin L_j$, hence $L_j \subset L_i$. Furthermore, there is $k \in R$ with $L_k \subset L_j$; so there exists $z \in L_j \setminus L_k \neq \varnothing$, implying that $\varepsilon \in \Phi''_{\mathbf{I}}\langle L_j \rangle$. We infer that M overgeneralized in response to (ε), in contradiction with the assumption that M is strongly-monotonic and learns $\Phi''\langle \mathbf{I} \rangle$. We conclude that R is empty, completing the proof of the proposition. ∎

Proposition 25. *Let an automatic class $\mathbf{I} = (L_i)_{i \in I}$ be given. Consider the following clause $(\star)$, which expresses that every finite set of members of $\mathbf{I}$, which is $\subseteq$-bounded in $\mathbf{I}$, has a (necessarily unique) $\subseteq$-least upper bound in $\mathbf{I}$.*

($\star$) *For all finite $F \subseteq I$, if there exists $i \in I$ such that $L_k \subseteq L_i$ for all $k \in F$, then there exists $i \in I$ such that $L_k \subseteq L_i$ for all $k \in F$ and for all $j \in I$*

$$L_i \subseteq L_j \Leftrightarrow \forall k \in F(L_k \subseteq L_j).$$

Then both following statements hold.

(1) *There exists an automatic $\mathbf{I}$-translator Φ such that $\Phi\langle\mathbf{I}\rangle$ is consistently and strongly-monotonically learnable iff ($\star$) holds.*

(2) *Suppose that $\mathbf{I}$ is strongly-monotonically learnable. Then some text-preserving automatic translation of $\mathbf{I}$ is consistently and strongly-monotonically learnable iff ($\star$) holds.*

Proof: It is convenient to prove both results together. For this, some notation is needed. If $\mathbf{I}$ is not strong-monotonically learnable, then for all $i \in I$, let $E_i = L_i$. If $\mathbf{I}$ is strong-monotonically learnable, then for all $i \in I$, let $E_i = \{y \in L_i : y \leq_{ll} x\}$ for the $\leq_{ll}$-least $x \in D$ such that for all $j \in I$, if $\{y \in L_i : y \leq_{ll} x\} \subseteq L_j$ then $L_i \subseteq L_j$ — such an x exists as a strong-monotonic learner that learns $\mathbf{I}$ conjectures i based on a finite set and then makes mind changes only when it is presented with texts for sets which are supersets of L_i. Note that for all $i \in I$, E_i is in both cases first-order definable from L_i.

Now sufficiency is shown. So assume that ($\star$) holds and, in case of (2), that $\mathbf{I}$ is also strong-monotonically learnable and the sets E_i are therefore finite. Consider a first-order formula $\Phi(L)$ with i as unique free variable which expresses that E_i is a subset of L. Hence $\Phi_{\mathbf{I}}\langle L\rangle = \{i \in I : E_i \subseteq L\}$. It is easily verified that Φ is an automatic $\mathbf{I}$-translator. Furthermore, in the case of (2), the sets E_i, $i \in I$, are finite and Φ is text-preserving.

Define a learner M as follows. Presented with a finite set F of data, M outputs the member i of I such that L_i is the $\subseteq$-least upper bound of the sets L_k with $k \in F$, which exists by ($\star$). Note that in case $F = \varnothing$, M still can output a conjecture as ($\star$) implies that there is a $\subseteq$-least language in $\mathbf{I}$. To see that M learns $\Phi\langle\mathbf{I}\rangle$, note that whenever M is presented with a text for $\Phi_{\mathbf{I}}\langle L_i\rangle$, then i occurs in the text and from that point onwards, M outputs i since L_i is the $\subseteq$-least upper bound of any class of languages which contains L_i and which only contains sets L_j satisfying $E_j \subseteq L_i$. Hence M learns $\Phi_{\mathbf{I}}\langle L_i\rangle$. Clearly, M is consistent. Furthermore, M is strong-monotonic: if $F \subseteq F'$, then also the $\subseteq$-least upper bound of $\{L_j : j \in F\}$ is a subset of the $\subseteq$-least upper bound of $\{L_j : j \in F'\}$. This completes the proof on the sufficiency of the conditions given in (1) and (2), respectively.

For necessity, note that, by Proposition 22, it is necessary in the case of (2) that $\mathbf{I}$ be strong-monotonically learnable. Hence it suffices to show ($\star$) in both cases. Let Φ be an automatic $\mathbf{I}$-translator that in the case of (2), is text-preserving. Let M be a consistent and strong monotonic learner that learns $\Phi\langle\mathbf{I}\rangle$. Let F be a finite subset of I. For each $j \in F$, there is a finite subset G_j of $\Phi_{\mathbf{I}}\langle L_j\rangle$ such that $\Phi_{\mathbf{I}}\langle L_j\rangle \subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$ whenever $G_j \subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$. Now assume that M is presented with data that include the union of all sets G_j with $j \in F$ — it is a finite set. Then M outputs a conjecture i such that $\Phi_{\mathbf{I}}\langle L_i\rangle$ contains all data seen so far. As M is strong-monotonic, it holds that $\Phi_{\mathbf{I}}\langle L_i\rangle \subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$ for all k where $\Phi_{\mathbf{I}}\langle L_k\rangle$ contains the data seen so far. Hence, $\Phi_{\mathbf{I}}\langle L_i\rangle$ is a $\subseteq$-least upper bound, in $\Phi\langle\mathbf{I}\rangle$, of the sets $\Phi_{\mathbf{I}}\langle L_j\rangle$ with $j \in F$. As Φ preserves inclusions and noninclusions within $\mathbf{I}$, it follows that L_i is the $\subseteq$-least upper bound of all the L_j with $j \in F$. Hence ($\star$) holds. ∎

Note that a class that contains an ascending infinite chain is not confidently learnable. Furthermore, a strong-monotonic learner of a class without an infinite ascending chain is confident as it cannot output an infinite ascending chain of hypotheses.

Property 26. *The four statements below hold.*

(1) *A strong-monotonically learnable automatic class is confidently learnable iff it does not contain infinite ascending chains.*

(2) *Every translation of an automatic class $\mathbf{I} = (L_i)_{i\in I}$ is strong-monotonically and confidently learnable iff $\mathbf{I}$ does not contain infinite ascending chains and for all $i \in I$, there exists $b \in I$ such that for all $j \in I$, if $L_i \not\subseteq L_j$, then there is $k \leq_{ll} b$ with $L_j \subseteq L_k$ and $L_i \not\subseteq L_k$.*

(3) *Some translation of an automatic class $\mathbf{I}$ is strong-monotonically and confidently learnable iff $\mathbf{I}$ does not have an ascending infinite chain.*

(4) *If some text-preserving translation of an automatic class $\mathbf{I}$ is strong-monotonically and confidently learnable, then $\mathbf{I}$ itself is strong-monotonically and confidently learnable.*

An immediate corollary of this property is the following.

Corollary 27. *The three statements below hold.*

(1) *Every translation of an automatic class* $\mathbf{I}$ *is consistently, confidently and strong-monotonically learnable iff* $\mathbf{I}$ *is a finite chain of languages.*

(2) *Some translation of an automatic class* $\mathbf{I}$ *is consistently, confidently and strong-monotonically learnable iff* $\mathbf{I}$ *has no infinite ascending chain and every* $\subseteq$*-bounded finite subclass of* $\mathbf{I}$ *has a* $\subseteq$*-least upper bound, that is, for all finite* $F \subseteq I$*, if there is* $i \in I$ *with* $L_k \subseteq L_i$ *for all* $k \in F$*, then there is* $i \in I$ *with* $L_k \subseteq L_i$ *for all* $k \in F$ *and* $L_i \subseteq L_h$ *for all* $h \in I$ *for which all* $k \in F$ *satisfy* $L_k \subseteq L_h$*.*

(3) *Some text-preserving translation of an automatic class* $\mathbf{I}$ *is consistently, confidently and strong-monotonically learnable iff* $\mathbf{I}$ *satisfies the conditions in* (2) *and* $\mathbf{I}$ *itself is strong-monotonically learnable.*

These results give a full characterization on how confident learnability combines with strong-monotonic learning. One might ask whether there is a connection between these two in the sense that whenever every translation is confidently learnable then the class is already strong-monotonically learnable. The answer to this question is negative and such an implication does not hold.

Example 28. Consider the automatic class $\mathbf{I}$ which contains $\{0\}^\star$ and for every $n \in \{1, 2, 3, \dots\}$, the language $\{0^m : m < n\} \cup \{1^n\}$ as well as the language $\{0\}^\star \cup \{1^m : m \geq n\}$. This class is not strong-monotonically learnable as any learner that learns $\mathbf{I}$ must output the index for $\{0\}^\star$ after seeing a finite sequence of suitable examples (say σ). But then, for sufficiently large n, a mind change to the index for the language $\{0^m : m < n\} \cup \{1^n\}$ would be necessary to learn it from any text which extends σ.

Still for every automatic $\mathbf{I}$-translator Φ, $\Phi\langle\mathbf{I}\rangle$ is confidently learnable by a learner M that can proceed as follows. As long as the data is consistent with $\Phi_{\mathbf{I}}\langle\{0\}^\star\rangle$, M conjectures the index for $\Phi_{\mathbf{I}}\langle\{0\}^\star\rangle$. If there exists (a necessarily unique) $n \in \{1, 2, 3, \dots\}$ such that the data seen so far is consistent with the set $\Phi_{\mathbf{I}}\langle\{0^m : m < n\} \cup \{1^n\}\rangle$ but not with $\Phi_{\mathbf{I}}\langle\{0\}^\star \cup \{1^m : m > n\}\rangle$, then M outputs the index for $\Phi_{\mathbf{I}}\langle\{0^m : m < n\} \cup \{1^n\}\rangle$. Otherwise, if M is presented with some data that is consistent with $\Phi_{\mathbf{I}}\langle 0^\star \cup \{1^m : m \geq n\}\rangle$ for all $n \in \mathbb{N}$, but not with $\Phi_{\mathbf{I}}\langle 0^\star\rangle$, then M outputs its previous hypothesis. Otherwise, M conjectures the index for $\Phi_{\mathbf{I}}\langle\{0\}^\star \cup \{1^m : m \geq n\}\rangle$, where n is the largest member of $\{1, 2, 3, \dots\}$ such that the former set is consistent with the data; that n might go down as more data are presented, but it will have eventually to stabilize. Hence $\Phi\langle\mathbf{I}\rangle$ is confidently learnable.

It is an open problem whether there exists an appealing characterization of the automatic classes for which every translation is confidently learnable (by a computable learner). A characterization will be provided for general learners in the next section.

Proposition 29. *Every translation of an automatic class* $\mathbf{I}$ *is confidently, conservatively and consistently learnable iff* $\mathbf{I}$ *is finite.*

Proof: Let $\mathbf{I}$ be an automatic class, and assume that every translation of $\mathbf{I}$ is confidently, conservatively and consistently learnable. Confidence implies that $\mathbf{I}$ contains no infinite ascending chain of languages. Conservativeness and consistency imply that $\mathbf{I}$ has no infinite descending chain of languages. For a contradiction, assume that $\mathbf{I}$ contains an infinite antichain.

By arguments similar to those in the proof of Proposition 19, there is an infinite regular set R which consists of indices of an antichain. Consider a first-order formula $\Phi(L)$ with x as unique free variable which expresses that

> x is either of the form $0y$ for some $y \in L$ or of the form $1^{|i|+1}01^n$ for some $i \in R$ and $n \in \mathbb{N}$ such that either $L_i \subseteq L$ or there is $j \in R$ with $|j| > |i| + 2 + n$ and $L_j \subseteq L$.

It is easily verified that Φ is an automatic $\mathbf{I}$-translator.

Note that if a language L is a superset of L_j for infinitely many $j \in I$ with $j \in R$, then $\Phi_{\mathbf{I}}\langle L\rangle$ contains all strings of the form $1^{|i|+1}01^n$. Furthermore, for every finite set E of such strings and almost all $j \in R$, $E \subseteq \Phi_{\mathbf{I}}\langle L_j\rangle$. These two facts will now be used to disprove that $\mathbf{I}$ is confidently, conservatively and consistently learnable.

Present all strings of the form $1^{|i|+1}01^n$ to a consistent learner M that learns $\Phi\langle\mathbf{I}\rangle$. If M converges to an index k, then $\Phi_{\mathbf{I}}\langle L_k\rangle$ contains infinitely many sets of the form $\Phi_{\mathbf{I}}\langle L_i\rangle$ with $i \in R$. As k is output after finitely many data have been received, there exists $i \in R$ such that $L_i \subset L_k$ and k has been output after only data from $\Phi_{\mathbf{I}}\langle L_i \rangle$ have been received; hence M cannot be conservative. If M does not converge to an index k, then M is not confident. Hence there is no consistent, conservative and confident learner that learns $\Phi\langle \mathbf{I} \rangle$. We conclude that $\mathbf{I}$ does not have an infinite antichain. As every infinite class contains an infinite ascending chain or an infinite descending chain or an infinite antichain, $\mathbf{I}$ must be finite.

For the sufficiency, assume that $\mathbf{I}$ is finite. Then for all automatic $\mathbf{I}$-translators Φ, $\Phi\langle \mathbf{I} \rangle$ is finite, and by Proposition 19, there exists a consistent and conservative learner M that learns $\Phi\langle \mathbf{I} \rangle$. This learner never returns to the index for a language that has been conjectured and then abandoned. As there are only finitely many indices, M is also confident. ■

A more restrictive notion of learning is finite learning where the very first conjecture output by the learner has to correctly identify the set to be learnt. Obviously, finitely learnable classes are antichains as otherwise one could see the data for a set L_i and conjecture an index for this set only to find out later that the set to be learnt is actually a superset of L_i. So a key question is to characterize the size of these antichains.

Proposition 30. *Let an automatic class* $\mathbf{I}$ *be given. The three statements below hold.*

(1) *Every text-preserving translation of* $\mathbf{I}$ *is finitely learnable iff* $\mathbf{I}$ *is a finite antichain.*

(2) *Some translation of* $\mathbf{I}$ *is finitely learnable iff* $\mathbf{I}$ *is an antichain.*

(3) *If* $\mathbf{I}$ *has a finitely learnable text-preserving translation, then* $\mathbf{I}$ *itself is finitely learnable.*

Proof: Let $\mathbf{I} = (L_i)_{i \in I}$ be an automatic class.

(1) Finite antichains are clearly finitely learnable by a learner M that waits until there is a unique $i \in I$ such that, for each $j \in I$ distinct to i, some member of $L_i \setminus L_j$ is part of the input, at which point M correctly issues i.

For the converse direction, assume that $\mathbf{I}$ is an infinite antichain. Let Φ^{nc} be the text-preserving automatic $\mathbf{I}$-translator defined in Example 7. For all $i \in I$, $\Phi^{nc}_{\mathbf{I}}\langle L_i \rangle$ is equal to $\{j \in I : j \neq i\}$. Thus, $\Phi^{nc}\langle \mathbf{I} \rangle$ is not finitely learnable.

(2) Suppose that $\mathbf{I}$ is an antichain. Consider the first-order formula with x as unique free variable which expresses that $L_x \subseteq L$. It is easily verified that Φ is an automatic $\mathbf{I}$-translator. Then for all $i \in I$, $\Phi\langle L_i \rangle$ is the singleton $\{i\}$. Obviously, $\Phi\langle \mathbf{I} \rangle$ is finitely learnable.

(3) Assume that Φ is a text-preserving automatic $\mathbf{I}$-translator which maps $\mathbf{I}$ to a finitely learnable class. Let M be a learner that finitely learns $\Phi\langle \mathbf{I} \rangle$. Then there is, for every $i \in I$, a finite subset E_i of $\Phi_{\mathbf{I}}\langle L_i \rangle$ such that M outputs i once it has been presented with all members of E_i. Let $i \in I$ be given. Then there exists a finite subset F_i of L_i with $E_i \subseteq \Phi_{\mathbf{I}}\langle F_i \rangle$. Now $F_i \nsubseteq L_j$ for all $j \in I \setminus \{i\}$. One can give the following first-order definition of a finite set G_i with the same property:

$$G_i = \{x \in L_i : \exists j \in I \setminus \{i\}\, \forall y \leq_{ll} x\, [y \in L_i \Rightarrow y \in L_j]\}.$$

Hence there is a finite learner which outputs i iff i is the $\leq_{ll}$-least member of I such that G_i is contained in the data observed. Thus $\mathbf{I}$ is finitely learnable. ■

8 A characterization of confident learning for nonrecursive learners

The following result is the only one that involves general learners rather than computable learners. So our characterization of robust learnability constrained to confident learners is open to improvement. Recall the definition of Φ^{nc} in Example 7.

Proposition 31. *Let* $\mathbf{I} = (L_i)_{i \in I}$ *be an automatic class all of whose translations are learnable. Then both conditions below are equivalent.*

- *Every translation of* I *is confidently learnable by some general learner.*
- *There exists no nonempty set* $J \subseteq I$ *such that, for all* $i \in J$ *and finite subsets* F *of* $\Phi^{nc}_{\mathbf{I}}\langle L_i \rangle$, *there exists* $j \in J$ *with* $F \cup \{i\} \subseteq \Phi^{nc}_{\mathbf{I}}\langle L_j \rangle$.

Proof: First one tries to build by induction along the ordinals α below ω_1 a sequence of indices i_α and bounds b_α such that, for each L_{i_α} and every $j \in I \setminus \{i_\beta : \beta < \alpha\}$, $\Phi^{nc}_{\mathbf{I}}\langle L_j \rangle$ does not contain the union of $\{i_\alpha\}$ with $\{k \in \Phi^{nc}_{\mathbf{I}}\langle L_{i_\alpha} \rangle : k \leq_{ll} b_\alpha\}$. This induction stops at some ordinal $\gamma < \omega_1$ (that is, i_γ, b_γ don't get defined, but i_α and b_α get defined for all $\alpha < \gamma$). Now let $J = I \setminus \{i_\alpha : \alpha < \gamma\}$. There are two cases:

(a) J is empty. Let Φ be any automatic $\mathbf{I}$-translator. Then one can build the following learner M which is a restriction of the learner M_{ex} defined in Example 14 on the class $\Phi\langle\mathbf{I}\rangle$. After seeing some data, M conjectures i_α iff M_{ex} conjectures i_α on the same data and for each $k \leq_{ll} b_\alpha$ with $k \in \Phi_{\mathbf{I}}^{nc}\langle L_{i_\alpha}\rangle$ some member of $\Phi_{\mathbf{I}}\langle L_{i_\alpha}\rangle \setminus \Phi_{\mathbf{I}}\langle L_k\rangle$ has been observed in the data seen so far; otherwise M is undefined. Then M learns every language $\Phi_{\mathbf{I}}\langle L_{i_\alpha}\rangle$ as on a text for L_{i_α}, after some finite time M_{ex} has converged to i_α and for all $k \leq_{ll} b_\alpha$ with $k \in \Phi_{\mathbf{I}}^{nc}\langle L_{i_\alpha}\rangle$, some datum in $\Phi_{\mathbf{I}}\langle L_{i_\alpha}\rangle \setminus \Phi_{\mathbf{I}}\langle L_k\rangle$ has been observed; hence M outputs i_α from then onward as well. To see that M is confident, consider any two hypotheses i_α and i_β consecutively output by M on some text. Assume for a contradiction that $\beta > \alpha$. By definition of i_α and b_α, there exists $k \in I$ with $k \leq_{ll} b_\alpha$ such that $k \in \{i_\alpha\} \cup \Phi_{\mathbf{I}}^{nc}\langle L_{i_\alpha}\rangle \setminus \Phi_{\mathbf{I}}^{nc}\langle L_{i_\beta}\rangle$. If $k = i_\alpha$, then $L_{i_\beta} \subseteq L_{i_\alpha}$, hence $\Phi_{\mathbf{I}}\langle L_{i_\beta}\rangle \subseteq \Phi_{\mathbf{I}}\langle L_{i_\alpha}\rangle$, which is in contradiction with both following facts taken together:

- i_α and i_β are consecutively output by M_{ex};
- M_{ex} never makes a mind change from a hypothesis for a set to a hypothesis for a proper subset of that set.

Otherwise, if $k \in \Phi_{\mathbf{I}}^{nc}\langle L_{i_\alpha}\rangle$, then $L_{i_\beta} \subseteq L_k$, $\Phi_{\mathbf{I}}\langle L_{i_\beta}\rangle \subseteq \Phi_{\mathbf{I}}\langle L_k\rangle$ and i_α had been output only after observing some data outside $\Phi_{\mathbf{I}}\langle L_k\rangle$; hence M would not output the hypothesis i_β which is then inconsistent with data observed before. From this contradiction it follows that $\beta < \alpha$ and hence the ordinals by which the indices in I are indexed go only down with every new hypothesis. It follows that the learner M is confident.

(b) In case J is not empty, one can show that every general learner M that learns $\{\Phi_{\mathbf{I}}^{nc}\langle L_i\rangle : i \in I\}$ is not confident. So consider such a general learner and pick iteratively some members $j_0, j_1, \ldots$ of J as follows. The inductive definition has the following invariants for $j_0 \in J$ and $n \in \mathbb{N}$:

- σ_n is the sequence of data observed from $\Phi_{\mathbf{I}}^{nc}\langle L_{j_n}\rangle$ until M outputs a hypothesis for j_n;
- $\sigma_n \subseteq \sigma_{n+1}$ and $j_n \in \mathrm{range}(\sigma_{n+1})$;
- $\mathrm{range}(\sigma_n) \cup \{j_n\} \subseteq \Phi_{\mathbf{I}}^{nc}\langle L_{j_{n+1}}\rangle$.

Note that j_{n+1} can always be picked from J as otherwise one could take $i_\gamma = j_n$ and $b_\gamma = \max_{ll}(\mathrm{range}(\sigma_n) \cup \{j_n\})$ in the above definition and extend the induction which by assumption cannot be extended at γ. As $\Phi_{\mathbf{I}}^{nc}\langle L_{j_{n+1}}\rangle$ contains $\mathrm{range}(\sigma_n)$ and $j_n \in \Phi_{\mathbf{I}}^{nc}\langle L_{j_{n+1}}\rangle$, one can always feed M with a data sequence starting with $\sigma_n \diamond j_n$ first and therefore assume that σ_{n+1} extends $\sigma_n \diamond j_n$. This completes the inductive definition. It follows that M is not a confident learner.

We conclude that either every translation of I is confidently learnable by a general learner or the subset J of I considered in the proof witnesses that for all $i \in J$ and finite subsets F of $\Phi_{\mathbf{I}}^{nc}\langle L_i\rangle$, there exists $j \in J$ with $F \cup \{i\} \subseteq \Phi_{\mathbf{I}}^{nc}\langle L_j\rangle$. ∎

9 Learning from Queries

Whereas learnability in the limit offers a model of passive learning, learning from queries allows agents to play an active role by questioning an oracle on some properties that the target language might have, and sometimes getting further clues [2]. Four kinds of queries are usually used, alone or in combination. In a given context, the selection of queries that the learner is allowed to make is dictated by the desire to obtain natural, elegant and insightful characterizations. Some work has been done that compares both the passive and active models of learning, which usually turn out to be different [25]. Interestingly, in our model both families of paradigms bind tightly as learnability of a class of languages from superset queries is equivalent to learnability from positive data of every translation of the class (Corollary 42 below).

Definition 32. Let T be the set of queries of at least one of the following types:

Membership query: is $x \in L$?	Superset query: is $L_e \supseteq L$?
Subset query: is $L_e \subseteq L$?	Equivalence query: is $L_e = L$?

Let an automatic class $\mathbf{I} = (L_i)_{i \in I}$ be given.

- An $\mathbf{I}$-*query learner of type* T is a machine M such that, for all $i \in I$, when learning L_i, M makes finitely many queries from T, possibly taking into account the answers to earlier queries, with all queries answered correctly w.r.t. $L = L_i$, and eventually outputs a member of I.

- An $\mathbf{I}$-query learner of type T is said to *learn* $\mathbf{I}$ iff for all $i \in I$, i is the member of $\mathbf{I}$ that M eventually outputs when learning L_i. A *query learner of type T for* $\mathbf{I}$ is an $\mathbf{I}$-query learner of type T that learns $\mathbf{I}$.
- $\mathbf{I}$ is said to be *learnable from queries of type* T iff there exists a query learner of type T for $\mathbf{I}$.

When T is clear from the context, we omit to mention "of type T" when referring to a particular query learner of type T for $\mathbf{I}$.

Remark 33. *Let an automatic class $\mathbf{I} = (L_i)_{i\in I}$ and an automatic $\mathbf{I}$-translator Φ be given. Since Φ preserves inclusion and no counterexamples are involved, "learnability of $\mathbf{I}$ from queries of type T" and "learnability of $\Phi\langle\mathbf{I}\rangle$ from queries of type T" are equivalent notions as long as T does not include the membership queries. Observe that subset, superset and equivalence queries are only with reference to languages in $\mathbf{I}$, as $\mathbf{I}$-query learners are given too much power otherwise.*

Since no complexity bound is imposed, we immediately have the following:

Property 34. *Every automatic class is learnable from equivalence queries.*

We illustrate the notion of query learning with a few examples.

Example 35. All translations of the automatic classes below are learnable from membership queries.

- The class of all sets of the form $\{x \in \{0,1\}^\star : x \preceq y \vee x \succeq y\}$ where $y \in \{0,1\}^\star$.
- The class of all sets of the form $\{0\}^\star \cup \{1^m : m \leq n\}$ or $\{0^m : m \geq n\}$ with $n \in \mathbb{N} \setminus \{0\}$.
- Any finite class.

Example 36. Let an automatic class $\mathbf{I}$ be given, and let Φ^{nc} be the text-preserving automatic $\mathbf{I}$-translator defined in Example 7. Then $\Phi^{nc}\langle\mathbf{I}\rangle$ is learnable from membership and subset queries by searching for the unique $i \in I$ such that $i \notin \Phi^{nc}_{\mathbf{I}}\langle L\rangle \wedge \Phi^{nc}_{\mathbf{I}}\langle L_i\rangle \subseteq \Phi^{nc}_{\mathbf{I}}\langle L\rangle$. Indeed, a negative answer to the membership query for i implies that $\Phi^{nc}_{\mathbf{I}}\langle L\rangle \subseteq \Phi^{nc}_{\mathbf{I}}\langle L_i\rangle$ and so $\Phi^{nc}_{\mathbf{I}}\langle L_i\rangle$ is equal to $\Phi^{nc}_{\mathbf{I}}\langle L\rangle$.

Example 37. Let an automatic class $\mathbf{I}$ be given and let Φ be a, not necessarily text-preserving, automatic $\mathbf{I}$-translator that satisfies

$$\Phi_{\mathbf{I}}\langle L\rangle = \{i \in I : L_i \subseteq L\}$$

for all languages L. Then $\Phi\langle\mathbf{I}\rangle$ is learnable from membership and superset queries as a $\Phi\langle\mathbf{I}\rangle$-query learner can search for the unique $i \in I$ with $i \in \Phi_{\mathbf{I}}\langle L\rangle$ and $\Phi_{\mathbf{I}}\langle L\rangle \subseteq \Phi_{\mathbf{I}}\langle L_i\rangle$. This i satisfies $\Phi_{\mathbf{I}}\langle L_i\rangle = \Phi_{\mathbf{I}}\langle L\rangle$ and can be found using both kinds of queries.

Example 38. Consider the automatic class $\mathbf{I}$ consisting of $\{0,1\}^\star$ and all co-singletons of the form $\{0,1\}^\star \setminus \{x\}$ with $x \in \{0,1\}^\star$. Then none of $\mathbf{I}$'s text-preserving translations is learnable from superset and membership queries. Let Φ be a text-preserving $\mathbf{I}$-translator, and assume for a contradiction that a query learner M for $\Phi\langle\mathbf{I}\rangle$ outputs an index for $\Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$ after finitely many superset and membership queries on $x_1, x_2, \ldots, x_n$. If L is any member of $\Phi\langle\mathbf{I}\rangle$, then the superset query "is $\Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle \supseteq L$?" necessarily receives the answer "yes", and for all $i \in I$ with $L_i \neq \{0,1\}^\star$, the superset query "is $\Phi_{\mathbf{I}}\langle L_i\rangle \supseteq L$?" necessarily receives the answer "no". Furthermore, the membership queries "is $x_k \in L$?" necessarily receive the answer "yes" when answered with respect to $L = \Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$. Now for each $x_k \in \Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$, there is a finite subset E_k of $\{0,1\}^\star$ with $x_k \in \Phi_{\mathbf{I}}\langle E_k\rangle$. Consider any $y \in \{0,1\}^\star$ such that:

- for all $k \in I$ such that M has queried the membership of x_k to the target language when learning $\Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$, $y \notin E_k$;
- the superset query "is $\Phi_{\mathbf{I}}\langle L\rangle \subseteq \Phi_{\mathbf{I}}\langle\{0,1\}^\star \setminus \{y\}\rangle$?" has not been asked by M when learning $\Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$.

Then all queries would have received the same answer if the language to be learnt would have been $\Phi_{\mathbf{I}}\langle\{0,1\}^\star \setminus \{y\}\rangle$; therefore M cannot distinguish that language from $\Phi_{\mathbf{I}}\langle\{0,1\}^\star\rangle$. Hence M is incorrect and $\Phi\langle\mathbf{I}\rangle$ is not learnable from subset and membership queries.

Proposition 39. *Every automatic class has a translation which is learnable using membership queries.*

Proof: Let $\mathbf{I} = (L_i)_{i\in I}$ be an automatic class. Consider a formula $\Phi(L)$ with x as unique free variable which expresses that either x is of the form $i0$ for some $i \in I$ with $L \nsubseteq L_i$, or x is of the form $i1$ for some $i \in I$ with $L_i \subseteq L$. It is easy to verify that Φ is an automatic $\mathbf{I}$-translator; note that Φ is not text-preserving. In order to learn $\Phi\langle\mathbf{I}\rangle$, a $\Phi\langle\mathbf{I}\rangle$-query learner can search for the first $i \in I$ such that $i0 \notin \Phi_{\mathbf{I}}\langle L\rangle \wedge i1 \in \Phi_{\mathbf{I}}\langle L\rangle$. Since $i0 \notin \Phi_{\mathbf{I}}\langle L\rangle$, $L \subseteq L_i$. Since $i1 \in \Phi_{\mathbf{I}}\langle L\rangle$, $L_i \subseteq L$. Hence i is uniquely determined and is such that $\Phi_{\mathbf{I}}\langle L\rangle = \Phi_{\mathbf{I}}\langle L_i\rangle$. ■

The following result offers a characterization of learnability from subset queries.

Proposition 40. *Let an automatic class $\mathbf{I} = (L_i)_{i\in I}$ be given. Then $\mathbf{I}$ is learnable from subset queries iff for all $i \in I$, there exists $b_i \in I$ such that, for all $j \in I$ with $L_i \subset L_j$, there exists $k \in I$ with $k \leq_{ll} b_i$ and $L_k \subseteq L_j \wedge L_k \nsubseteq L_i$.*

Proof: Suppose that, for all $i \in I$, there exists $b_i \in I$ that satisfies the condition of the proposition. Note that there exists a computable function that maps any $i \in I$ to a member b_i of I that satisfies the condition of the proposition. Hence an $\mathbf{I}$-query learner can, using subset queries $L_j \subseteq L$ where L is the language to be learnt, find and output the first $i \in I$ such that $L_i \subseteq L$ and for all $k \in I$ with $k \leq_{ll} b_i$, $L_k \subseteq L$ iff $L_k \subseteq L_i$. Obviously $L_i = L$. Note that testing whether $L_k \subseteq L_i$ is recursive as the structure is automatic.

Conversely, assume that there exists $i \in I$ such that no $b_i \in I$ satisfies the condition of the proposition. For a contradiction, suppose that M is a query learner for $\mathbf{I}$ that uses subset queries. Then there exists $b_i \in I$ such that M outputs i after asking subset queries of the form $L_k \subseteq L$ only for L_k with $k \leq_{ll} b_i$, answered w.r.t. $L = L_i$. By the choice of i, there exists $j \in I$ such that $L_i \subset L_j$ and there exists no member k of I with $k \leq_{ll} b_i$, $L_k \subseteq L_j$ and $L_k \nsubseteq L_i$. Hence, all queries involving indices $k \leq_{ll} b_i$ are answered in the same way when learning $L = L_i$ and when learning $L = L_j$. Hence the algorithm would give the same answer i when learning L_j and thus cannot be correct. ■

A similar result can be obtained when using superset queries only:

Corollary 41. *Let an automatic class $\mathbf{I} = (L_i)_{i\in I}$ be given. Then $\mathbf{I}$ is learnable from superset queries iff for all $i \in I$, there exists $b_i \in I$ such that for all $j \in I$ with $L_i \supset L_j$, there exists $k \in I$ with $k \leq_{ll} b_i$ and $L_k \supseteq L_j \wedge L_k \nsupseteq L_i$.*

From the previous corollary and Proposition 17, we derive the corollary that follows.

Corollary 42. *An automatic class $\mathbf{I}$ is learnable from superset queries iff every translation of $\mathbf{I}$ is learnable from positive data.*

Furthermore, one can show that, given an automatic class $\mathbf{I}$ of languages all of whose text-preserving translations are learnable from superset and membership queries, $\mathbf{I}$-query learners that ask superset queries do not benefit from also asking membership queries:

Proposition 43. *Let an automatic class $\mathbf{I}$ be such that every text-preserving translation of $\mathbf{I}$ is learnable from membership and superset queries. Then $\mathbf{I}$ is learnable from superset queries alone.*

Proof: Suppose $\mathbf{I} = (L_i)_{i\in I}$. Let Φ^{nc} be the text-preserving automatic $\mathbf{I}$-translator defined in Example 7. When learning $\Phi^{nc}\langle\mathbf{I}\rangle$, a $\Phi^{nc}\langle\mathbf{I}\rangle$-query learner can replace every membership query of the form "is $i \in \Phi^{nc}\langle L\rangle$?" by the superset query "is $\Phi^{nc}\langle L\rangle \subseteq \Phi^{nc}\langle L_i\rangle$?" and reverse the answer. Hence membership queries can be simulated and $\Phi^{nc}\langle\mathbf{I}\rangle$ can be learnt by using superset queries alone. As learnability from superset queries is invariant under translations, $\mathbf{I}$ can also be learnt from superset queries alone. ■

One has an analogous result for subset queries, but considering all translations rather than all text-preserving translations of the class, thanks to a (not text-preserving) automatic $\mathbf{I}$-translator Φ that satisfies $\Phi_{\mathbf{I}}\langle L\rangle = \{i \in I : L_i \subseteq L\}$ for all languages L. Indeed a membership query of the form "is $i \in \Phi_{\mathbf{I}}\langle L\rangle$?" is then equivalent to the subset query "is $\Phi_{\mathbf{I}}\langle L_i\rangle \subseteq \Phi_{\mathbf{I}}\langle L\rangle$?":

Property 44. *If every translation of an automatic class $\mathbf{I}$ is learnable from membership queries and subset queries then $\mathbf{I}$ is also learnable from subset queries only.*

In the previous result, it is not possible to restrict to text-preserving translations:

Proposition 45. *Let $\mathbf{I}$ be the automatic class that consists of $\varnothing$ and $\{0,1\}^\star \setminus \{x\}$ for all $x \in \{0,1\}^\star$.*

- *Every text-preserving translation of* **I** *is learnable using membership and subset queries only.*
- *Some translation of* **I** *is not learnable using membership queries only.*
- **I** *is not learnable using subset queries only.*

Proof: Given an automatic **I**-translator Φ, the translation $\Phi\langle\mathbf{I}\rangle$ can be learnt from membership queries and subset queries as follows. There is a finite subset D of $\{0,1\}^\star \setminus \{0\}$ such that $\Phi_\mathbf{I}\langle D\rangle$ contains an element y outside $\Phi_\mathbf{I}\langle\varnothing\rangle$. Now for every $x \notin D$, y belongs to $\Phi_\mathbf{I}\langle\{0,1\}^\star \setminus \{x\}\rangle$. Hence a $\Phi\langle\mathbf{I}\rangle$-query learner can first issue the membership query "is $y \in \Phi_\mathbf{I}\langle L\rangle$?". If the answer is "yes", then the query learner goes on querying whether $\Phi_\mathbf{I}\langle\{0,1\}^\star \setminus \{x\}\rangle \subseteq L$ until the answer is again "yes", for some x, and then the correct language is found. If the answer is "no" then the query learner knows that $\Phi_\mathbf{I}\langle L\rangle$ is either $\Phi_\mathbf{I}\langle\varnothing\rangle$ or $\Phi_\mathbf{I}\langle\{0,1\}^\star \setminus \{x\}\rangle$ for one of the finitely many members x of D and these finitely many cases can be distinguished using membership queries.

For the second item, consider an automatic **I**-translator Φ with the property that $\Phi_\mathbf{I}\langle\{0,1\}^\star \setminus \{x\}\rangle = \{x\}$ and $\Phi_\mathbf{I}\langle\varnothing\rangle = \varnothing$. In order to learn $\varnothing$ from queries only, a $\Phi\langle\mathbf{I}\rangle$-query learner M can make only finitely many membership queries before it concludes that $\varnothing$ is the language L to be learnt. The answers to these queries are consistent with L being one of infinitely many singletons (the individuals whose membership to the target language has been queried are excluded) rather than $\varnothing$. Hence M cannot learn $\Phi\langle\mathbf{I}\rangle$.

Finally, **I** cannot be learnt from subset queries only: if finitely many queries of the form "is $\{0,1\}^\star \setminus \{x\} \subseteq L$?" have all been answered negatively, then an **I**-query learner still does not know whether $L = \varnothing$ or whether $L = \{0,1\}^\star \setminus \{y\}$ for some y such that no corresponding query has been made yet. ■

We end this section with a characterization of the automatic classes all of whose translations are learnable from membership queries.

Proposition 46. *Let* $\mathbf{I} = (L_i)_{i\in I}$ *be an automatic class. Then every translation of* **I** *is learnable from membership queries iff*

$$\forall i\, \exists b_i\, \forall j \neq i\, \exists k \leq_{ll} b_i\, [[L_j \subseteq L_k \wedge L_i \nsubseteq L_k] \vee [L_k \subseteq L_j \wedge L_k \nsubseteq L_i]].$$

Proof: Assume that the condition of the proposition holds. We exhibit a query learner M for **I** that uses membership queries. Since translations of automatic classes preserve inclusion between languages, we have that for all **I**-translators Φ, the condition of the proposition also holds for $\Phi\langle\mathbf{I}\rangle$, and M can be modified into a query learner for $\Phi\langle\mathbf{I}\rangle$ that uses membership queries.

Let M ask membership queries for individuals taken in length lexicographic order until it finds the $\leq_{ll}$-minimal $i \in I$ such that L_i is consistent with the answers to the queries asked so far and for all $k \leq_{ll} b_i$, both the following conditions hold.

- If $L_i \nsubseteq L_k$, then M got the answer "Yes" to some query of the form "Is $x \in L$" with $x \notin L_k$.
- If $L_k \nsubseteq L_i$, then M got the answer "No" to some query of the form "Is $x \in L$" with $x \in L_k$.

By the assumed condition, M is well defined. To see that M learns **I**, let $i \in I$ be given and assume that L_i is the language to be learnt. Let $j \in I \setminus \{i\}$ be given. If there is a member k of I with $k \leq_{ll} b_i$ and $L_j \subseteq L_k \wedge L_i \nsubseteq L_k$, then M has discovered a member x of $L_i \setminus L_k$, hence has not eventually issued j. Otherwise there is a member k of I with $k \leq_{ll} b_i$ and $L_k \subseteq L_j \wedge L_k \nsubseteq L_i$, and M has discovered a member x of $L_k \setminus L$, and again has not eventually issued j. Hence M eventually issues i, and learns **I**.

For the converse, suppose that the condition of the proposition does not hold. Let $i \in I$ be such that for all $b_i \in I$, it is not true that

$$\forall j \neq i\, \exists k \leq_{ll} b_i\, [[L_j \subseteq L_k \wedge L_i \nsubseteq L_k] \vee [L_k \subseteq L_j \wedge L_k \nsubseteq L_i]].$$

Consider a formula $\Phi(L)$ with x as unique free variable which expresses that one the following holds:

(1) x is of the form (α, β) for $\alpha, \beta \in I$ with $L_\alpha \subseteq L$ and $\alpha \leq_{ll} \beta$,

(2) there exists a $\leq_{ll}$-least member j of I such that $L \subseteq L_j$ and $L_i \setminus L_j \neq \varnothing$, and x is of the form (α, β) for $\alpha, \beta \in I$ with $\alpha \leq_{ll} \beta \leq_{ll} j$ and $L_\alpha \subseteq L_i$,

(3) there exists no member j of I such that $L \subseteq L_j$ and $L_i \setminus L_j \neq \varnothing$, and x is of the form (α, β) for $\alpha, \beta \in I$ with $\alpha \leq_{ll} \beta$ and $L_\alpha \subseteq L_i$.

Note that, for all members j, k of I, if $L_j \not\subseteq L_k$, then $\Phi_{\mathbf{I}}\langle L_k\rangle$ contains no pair of the form (j,β) with $\beta > k$, hence L_k contains only finitely many elements of the form (j,β). This implies that the second item in Definition 4 holds, and the first item easily follows from the definition of Φ.

For a contradiction, assume that M is a query learner for $\Phi\langle\mathbf{I}\rangle$. Let $b_i \in I$ be such that M makes membership queries only about elements (α,β) with $\alpha,\beta \le_{ll}$-smaller than b_i when learning $\Phi_{\mathbf{I}}\langle L_i\rangle$. Let $j \in I \setminus \{i\}$ be such that

$$\forall k \le_{ll} b_i\,[[L_j \not\subseteq L_k \vee L_i \subseteq L_k] \wedge [L_k \not\subseteq L_j \vee L_k \subseteq L_i]].$$

holds. We claim that $\Phi_{\mathbf{I}}\langle L_j\rangle$ agrees with $\Phi_{\mathbf{I}}\langle L_i\rangle$ on all elements $\le_{ll}$-smaller than b_i, hence cannot be distinguished from $\Phi_{\mathbf{I}}\langle L_i\rangle$ by M, contrary to the assumption that M learns $\Phi\langle\mathbf{I}\rangle$. First, $\Phi_{\mathbf{I}}\langle L_j\rangle \setminus \Phi_{\mathbf{I}}\langle L_i\rangle$ contains no element of the form (α,β) with $\alpha \le_{ll} b_i$ and $\beta \le_{ll} b_i$: indeed, only (1) could introduce such an element, but that condition is satisfied for no $\alpha \le_{ll} b_i$. Second, consider a member (α,β) of $\Phi_{\mathbf{I}}\langle L_i\rangle$ with $\alpha,\beta \le_{ll}$-smaller than b_i. By item 2 above in the definition of Φ, $\Phi_{\mathbf{I}}\langle L_j\rangle$ also contains (α,β) as otherwise, there would exist a $k \in I$ with $k \le_{ll} b_i$ such that $L_j \subseteq L_k$ and $L_i \not\subseteq L_k$. Hence $\Phi_{\mathbf{I}}\langle L_j\rangle$ agrees with $\Phi_{\mathbf{I}}\langle L_i\rangle$ on all elements $\le_{ll}$-smaller than b_i, as needed. ■

10 Conclusion

A notion of learnability is robust if it immune to natural transformations of the class of objects to be learned. Whereas such a notion had been successfully discovered when the objects to be learned are functions, no satisfactory concept of robust learning of languages from positive data had been found so far. This paper has addressed this deficiency. The associated notion of transformation of languages has been defined as a function that maps a language into another language and that preserves the inclusion structure of the languages in the original class; such functions had been called translators. Our study has focused on automatic classes of languages, as automaticity is invariant under translation and as this restriction allows one to obtain appealing characterizations of robust learning under many classical learning criteria, namely: consistent and conservative learning, strong-monotonic learning, strong-monotonic consistent learning, finite learning, learning from subset queries, learning from superset queries, and learning from membership queries. The characterizations are natural as they express a particular constraint on the inclusion structure of the original class. In many cases, they are especially strong as they deal not only with learnability of the original class under all possible translations that preserve the inclusion structure, but also with learnability under those translations that are text-preserving, that is, that can be generated from an enumeration of the languages rather than necessitating the languages to be "seen as a whole." Open questions remain. For instance, for confident learning, we only found a characterization with respect to general, that is, nonrecursive learners. Other learning criteria can also be the subject of further study.

References

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

[2] Dana Angluin. Learning Regular Sets From Queries and Counterexamples. *Information and Computation*, 75, pp. 87–106, 1987.

[3] Dana Angluin. Finding Patterns Common to a Set of Strings. *Journal of Computer and System Sciences*, 21, pp. 46–62, 1980.

[4] Janis Bārzdiņš. Two theorems on the limiting synthesis of functions. *Theory of Algorithms and Programs*, 1, pp. 82–88, Latvian State University, Riga, Latvia, 1974.

[5] John Case. The power of vacillation in language learning. *SIAM Journal on Computing*, 28(6), pp. 1941–1969, 1999.

[6] Mark Fulk. Robust separations in inductive inference. *Proceedings of the 31st Annual Symposium on Foundations of Computer Science* (FOCS), pp. 405–410, St. Louis, Missouri, 1990.

[7] Rūsiņš Freivalds, Efim Kinber and Carl H. Smith. On the intrinsic complexity of learning. *Information and Computation*, 123, pp. 64–71, 1995.

[8] Mark E. Gold. Language identification in the limit. *Inf. and Control*, 10(5), pp. 447–474, 1967.

[9] Bernard R. Hodgson. *Théories décidables par automate fini.* Ph.D. thesis, U. of Montréal, 1976.

[10] Bernard R. Hodgson. Décidabilité par automate fini. *Annales des sciences mathématiques du Québec*, 7(1), pp. 39–57, 1983.

[11] Sanjay Jain, Qinglong Luo and Frank Stephan. Learnability of automatic classes. Technical Report TRA1/09, School of Computing, National University of Singapore, 2009.

[12] Sanjay Jain, Yuh Shin Ong, Shi Pu and Frank Stephan. *On automatic families.* Technical Report TRB1/10, School of Computing, National University of Singapore, 2010.

[13] Sanjay Jain, Daniel Osherson, James S. Royer and Arun Sharma. *Systems That Learn*, 2nd Edition. MIT Press, 1999.

[14] Sanjay Jain and Arun Sharma. The intrinsic complexity of language identification. *Journal of Computer and System Sciences*, 52, pp. 393–402, 1996.

[15] Sanjay Jain and Arun Sharma. The Structure of Intrinsic Complexity of Learning. *The Journal of Symbolic Logic*, 62, pp. 1187–1201, 1997.

[16] Sanjay Jain and Frank Stephan. A tour of robust learning. In S. Barry Cooper and Sergei S. Goncharov, editors. *Computability and Models.* Perspectives East and West. Kluwer Academic / Plenum Publishers, University Series in Mathematics, pp. 215–247, 2003.

[17] Sanjay Jain and Frank Stephan. Mitotic classes in inductive inference. *SIAM Journal on Computing*, 38, pp. 1283–1299, 2008.

[18] Sanjay Jain, Carl H. Smith and Rolf Wiehagen. Robust learning is rich. *Journal of Computer and System Sciences*, 62(1), pp. 178–212, 2001.

[19] Daniel N. Osherson and Scott Weinstein. Criteria of language learning. *Information and Control* 52, pp. 123–138, 1982.

[20] Klaus P. Jantke. Monotonic and non-monotonic inductive inference. *New Generation Computing* 8, pp. 349–360, 1991.

[21] Bakhadyr Khoussainov and Anil Nerode. Automatic Presentations of Structures. In Daniel Leivant, editor. *Selected Papers from LCC'94, the International Workshop on Logical and Computational Complexity*, pp. 367–392, 1994.

[22] Bakhadyr Khoussainov and Sasha Rubin. Automatic structures: overview and future directions. *Journal of Automata, Languages and Combinatorics*, 8(2), pp. 287–301, 2003.

[23] Steffen Lange and Thomas Zeugmann. Language learning in dependence on the space of hypotheses. *Proceedings of the Sixth Annual Conference on Computational Learning Theory* (COLT), pp. 127–136. ACM Press, 1993.

[24] Steffen Lange, Thomas Zeugmann and Sandra Zilles. Learning indexed families of recursive languages from positive data: a survey. *Theoretical Computer Science*, 397, pp. 194–232, 2008.

[25] Steffen Lange and Sandra Zilles. Formal language identification: Query learning vs. Gold-style learning. *Information Processing Letters*, 91(6), pp. 251–304, 2004.

[26] Matthias Ott and Frank Stephan. Avoiding coding tricks by hyperrobust learning. *Theoretical Computer Science*, 284(1), pp. 161–180, 2002.

[27] Michael Oser Rabin. *Automata on Infinite Objects and Church's Problem.* AMS, 1972.

[28] Herman Weyl. *Symmetry.* Princeton University Press, 1952.

[29] Rolf Wiehagen and Thomas Zeugmann. Learning and consistency. *Algorithmic Learning for Knowledge-Based Systems*, Lecture Notes in Artificial Intelligence 961, pp. 1–24, Springer, 1995.

[30] Thomas Zeugmann. On Bārzdiņš' Conjecture. *Proceedings of the International Workshop on Analogical and Inductive Inference* (AII'86), Lecture Notes in Computer Science 265, pp. 220–227, Springer, 1986.