Learning and Extending Sublanguages

Sanjay Jain [a,1] and Efim Kinber [b]

[a] *School of Computing, National University of Singapore, 3 Science Drive 2, Singapore 117543. Email: sanjay@comp.nus.edu.sg*

[b] *Department of Computer Science, Sacred Heart University, Fairfield, CT 06432-1000, U.S.A. Email: kinbere@sacredheart.edu*

Abstract

A number of natural models for learning in the limit is introduced to deal with the situation when a learner is required to provide a grammar covering the input even if only a part of the target language is available. Examples of language families are exhibited that are learnable in one model and not learnable in another one. Some characterizations for learnability of algorithmically enumerable families of languages for the models in question are obtained. Since learnability of any part of the target language does not imply *monotonicity* of the learning process, we consider also our models under additional monotonicity constraint.

1 Introduction

Models of algorithmic learning in the limit have been used for quite a while for study of learning potentially infinite languages. In the widely used mathematical paradigm of learning in the limit, as suggested by Gold in his seminal article [Gol67], the learner eventually gets all positive examples of the language in question, and the sequence of its conjectures converges in the limit to a correct description. However, in the Gold's original model, the learner is not required to produce any reasonable description for partial data — whereas real learning process of languages by humans is rather a sort of incremental process: the learner first actually finds grammatical forms — in the beginning, probably, quite primitive — that describe partial data, and refines conjectures when more data becomes available. (Incremental nature of the process here may be understood as purely *monotonic*: every new conjecture always extends

[1] Supported in part by NUS grant number R252–000–127–112 and R252-000-212-112.

the previous one. However, a learner can, in fact, sometimes choose conjectures that are too *wide* and later are refined to cover *smaller* languages; we discuss these aspects of sublanguage learning in Section 5.)

Moreover, if some data never becomes available, a successful learner can still eventually come up with a feasible useful description of the part of the language it has learned so far. This situation can be well understood by those who have been exposed to a foreign language for a long time, but then stopped learning it. For example, English has many common grammatical forms with Russian, which makes them relatively easy to learn. However, the system of tenses in English is much more complex than in Russian, and remains a tough nut to crack for many adult Russians who mastered English otherwise relatively well. A similar argument can be made for many other situations when even partial descriptions based on partial input data might be important: diagnosing the complete health status of a patient versus detecting only some of his/her deficiencies, forecasting weather for a whole region, or just for some small towns, etc.

In this paper, we introduce several variants of the Gold's model for learning languages in the limit requiring a learner to converge to a reasonable description for just a *sublanguage* if the data from this sublanguage only is available (this approach to learning recursive functions in the limit was studied in [JKW04]). In particular, we consider

(1) a model (called **AllSubEx**), where, for any input representing a part P of a language L from the learnable class $\mathcal{L}$, the learner converges to a grammar describing a part of L containing P;

(2) a model (called **AllWSubEx**; here W stands for *weak*), where for any input representing a part P of some language L in the learnable class $\mathcal{L}$, the learner converges to a grammar describing a part (containing P) of some (maybe other) language L' in $\mathcal{L}$. The reason for considering this model is that the first model, **AllSubEx**, may be viewed as too restrictive — partial data P seen by the learner can belong to several different languages, and in such a case, the learner, following the model **AllSubEx**, must produce a grammar describing a part containing P and being a part of ALL languages in $\mathcal{L}$ which contain P;

(3) a model (called **AllMWSubEx**; here M stands for *minimal*), similar to **AllWSubEx** above, but the language L', containing the part P, is required to be a *minimal* language in the class $\mathcal{L}$ which contains P.

Later in the paper (Theorem 10), we will feature the following example of a family of languages in **AllSubEx**: any language in the family contains pairs of a function $\langle x, f(x)\rangle$, where the function f has a finite range, and every value $f(x)$ encodes an index for the set of all those z that have the value $f(z) = f(x)$.

Once the learner has received a pair $\langle x, f(x)\rangle$, it can always describe the part of the target language defined by $f^{-1}(x)$, and, thus, will eventually build a grammar for the target language as a union of the grammars for its parts.

For all three models, we also consider the variant where the final conjecture itself is required to be a grammar describing a language in the class $\mathcal{L}$ (rather than being a subset of such a language, as in the original models (1) — (3)). A slightly different variants of the models (1) and (3), with a slightly different motivation, and in somewhat different forms, were introduced in [Muk94] and [KY95,KY97b,KY97a].

We also consider a weaker variant of all the above models: for a learner to be able to learn just a part of the language, the part must be *infinite*. Sometimes, we may be interested in learning just potentially infinite languages: often concepts are infinite, and it is unreasonable to say that one does any reasonable deduction from just finite data. Also, even though we may not obtain all data, it is reasonable to say that, over time in this universe, we will keep getting more and more data, infinite over infinite time (assuming the universe does not collapse). Similarly, for language learning, even though we may not hear every sentence, it is reasonable to assume that we will get infinitely many sentences over infinite time. In these and similar cases, correct learning of just a finite fragment of a target language may be inessential.

We compare all these models, examining when one model has advantages over the other. This gives us opportunity to build some interesting examples of learnable families of languages, for which learnability of a part is possible in one sense, but not possible in the other. In particular, we show that all three models are different (Theorems 11 and 12) and demonstrate how the requirement of the last (correct) conjecture being a member of the learnable class (Theorem 10), or requiring sublearning of only infinite sublanguages (Theorem 13) affects sublanguage learners. We also look at how the requirement of being able to learn all (or just infinite) parts fairs against other known models of learnability — in particular, the one that requires the learner to be *consistent* [Bārz74a,Ang80b,WZ95] with the input seen so far. It turns out that learners in all our models that are able to learn *all* sublanguages can be made consistent (Theorem 17). We obtain some characterizations for learnability within our models when the final conjecture is required to be a member of the learnable class of languages (Theorems 19, 21 and 22).

Some of our examples separating one model from another use the fact that, while, in general, learning increasing parts of an input language can be perceived as an incremental process, actual learning strategies can, in fact, be *nonmonotonic* — each next conjecture is not required to contain every data item covered by the prior conjecture. Consequently, we also consider how our models of learnability fair in the context where monotonicity [Jan91,Wie90] is explicitly required. It turns out that monotonicity requirement is, in fact, a severe limitation on sublanguage learners. In particular, **WSub**, **MWSub** and **Sub** variants collapse for strong monotonic learning when one does not require the final conjecture to be within the class being learnt (Theorem 27).

Overall, our results show that the requirement of being able to learn sub-languages, while being a serious limitation of learners' capabilities, adds an interesting and important insight on how learners can learn when only partial data may be provided.

Learning from incomplete texts (with just finite amount of data missing) has earlier been studied in the context where the final grammar still was required to be a correct (or nearly correct) description of the full target language (see, for example, [OSW86,FJ96]). There have also been studies when the input text may contain extra (noisy) data (see for example [OSW86,FJ96,Ste97,Sch85]). These notions, in general, are incomparable with our approach.

A preliminary version of this paper appeared in [JK06].

2 Notation and Preliminaries

Any unexplained recursion theoretic notation is from [Rog67]. N denotes the set of natural numbers, $\{0, 1, 2, 3, \ldots\}$. $\emptyset$ denotes the empty set. $\subseteq$, $\subset$, $\supseteq$, $\supset$ respectively denote subset, proper subset, superset and proper superset. D_x denotes the finite set with the canonical index x [Rog67]. We sometimes identify finite sets with their canonical indices. The quantifier '$\forall^\infty$' essentially from [Blu67], means 'for all but finitely many'.

$\uparrow$ denotes undefined. $\max(\cdot), \min(\cdot)$ denotes the maximum and minimum of a set, respectively, where $\max(\emptyset) = 0$ and $\min(\emptyset) =\uparrow$. $\langle \cdot, \cdot \rangle$ stands for an arbitrary computable one-to-one encoding of all pairs of natural numbers onto N [Rog67]. Similarly we can define $\langle \cdot, \ldots, \cdot \rangle$ for encoding tuples of natural numbers onto N. π_k^n denotes the k-th projection for the pairing function for n-tuples, i.e., $\pi_k^n(\langle x_1, \ldots, x_n \rangle) = x_k$.

φ denotes a fixed *acceptable* programming system for the partial computable functions: $N \to N$ [Rog58,Rog67,MY78]. φ_i denotes the partial computable function computed by program i in the φ-system. The set of all total recursive functions of one variable is denoted by $\mathcal{R}$. W_i denotes domain(φ_i). W_i is, then, the recursively enumerable (r.e.) set/language ($\subseteq N$) accepted (or equivalently, generated) by the φ-program i. $\mathcal{E}$ denotes the set of all r.e. languages. Any L, with or without subscripts and superscripts, is a member of $\mathcal{E}$. Any $\mathcal{L}$, with or without subscripts and superscripts, is a subset of $\mathcal{E}$.

By Φ we denote an arbitrary fixed Blum complexity measure [Blu67,HU79] for the φ-system. Intuitively, $\Phi_i(x)$ denotes the time taken to compute $\varphi_i(x)$. $W_{i,s} = \{x \mid x < s, \Phi_i(x) < s\}$.

A class $\mathcal{L}$ is said to be an *indexed family* [Ang80b] of recursive languages iff there exists an indexing $(L_i)_{i \in N}$ (possibly with repetition) of languages in $\mathcal{L}$ and a recursive function f such that $f(i, x) = 1$ iff $x \in L_i$. When learning indexed families $\mathcal{L}$ (with indexing $(L_i)_{i \in N}$ as above), we often consider the hypotheses space being $(L_i)_{i \in N}$. In such cases, $\mathcal{L}$-grammar i is a grammar for L_i.

We now consider some basic notions in language learning. We first introduce the concept of data that is presented to a learner. A *text* T is a mapping from N into $(N \cup \{\#\})$ (see [Gol67]). The *content* of a text T, denoted content(T), is the set of natural numbers in the range of T. T is a text for L iff content$(T) = L$. $T[n]$ denotes the initial segment of T of length n. We let T, with or without superscripts, range over texts. Intuitively, #'s in the text denote pauses in the presentation of data. For example, the only text for the empty language is just an infinite sequence of #'s.

A finite sequence σ is an initial segment of a text. Λ denotes the empty sequence. content(σ) is the set of natural numbers in the range of σ. $|\sigma|$ denotes the length of σ, and if $n \leq |\sigma|$, then $\sigma[n]$ denotes the initial segment of σ of length n. $\sigma\tau$ denotes the concatenation of σ and τ. $\sigma \subseteq \tau$ ($\sigma \subseteq T$) denotes that σ is an initial segment of τ (initial sequence of T).

A *language learning machine* is an algorithmic device which computes a mapping from finite initial segments of texts into $N \cup \{?\}$. (Here ? intuitively denotes the fact that $\mathbf{M}$ does not wish to output a conjecture on a particular input). We let $\mathbf{M}$, with or without subscripts and superscripts, range over learning machines. Intuitively, learning machines process a text T as infinite sequences of initial segments $T[n]$, $n \in N$. Thus, one considers $\mathbf{M}(T[n])$, $n \in N$, as a sequence of conjectures made by learning machine $\mathbf{M}$ on the input text T. We say that $\mathbf{M}(T)\downarrow = i \Leftrightarrow (\forall^\infty n)[\mathbf{M}(T[n]) = i]$. Thus, $\mathbf{M}(T)\downarrow = i$ denotes that the infinite sequence of hypotheses, $\mathbf{M}(T[n])$, $n \in N$, output by $\mathbf{M}$, when processing text T, converges to i.

We now introduce criteria for a learning machine to be considered *successful* on languages. Our first criterion requires that a learner, given a text for some language in the class being learnt, converges to a grammar for that language.

Definition 1 [Gol67] (a) $\mathbf{M}$ $\mathbf{TxtEx}$-*identifies* L (written: $L \in \mathbf{TxtEx}(\mathbf{M})$) $\Leftrightarrow$ ($\forall$ texts T for L)$(\exists i \mid W_i = L)[\mathbf{M}(T)\downarrow = i]$.

(b) $\mathbf{M}$ $\mathbf{TxtEx}$-identifies $\mathcal{L}$, if it $\mathbf{TxtEx}$-identifies each $L \in \mathcal{L}$.

(c) $\mathbf{TxtEx} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{TxtEx}(\mathbf{M})]\}$.

The influence of the Gold's paradigm [Gol67] to analyze human language learning is discussed by various authors, for example [Pin79,WC80,OSW86].

Note that the hypotheses space used for interpreting the conjectures of the learner in the above definition is the acceptable numbering $W_0, W_1, \ldots$. In some cases we use special hypotheses spaces (for example, when learning indexed families $\mathcal{L}$, we often use the indexing $(L_i)_{i \in N}$ for $\mathcal{L}$ as the hypotheses space). We will make it explicit when we use such a hypotheses space.

Note that in **TxtEx** learning, the learner does not know when it has arrived at its final conjecture. Motivated by the need to know when the learner has arrived at its final hypothesis, Gold [Gol67] also considered the case when the learner is required to learn a language without making any mind changes (which is equivalent to knowing when the learner has arrived at its final conjecture).

Definition 2 [Gol67] (a) $\mathbf{M}$ $\mathbf{TxtFin}$-*identifies* L (written: $L \in \mathbf{TxtFin}(\mathbf{M})$) $\Leftrightarrow$ ($\forall$ texts T for L)$(\exists i \mid W_i = L)(\exists n)[\mathbf{M}(T[n]) = i \ \wedge \ (\forall m < n)[\mathbf{M}(T[m]) = ?]]$.

(b) $\mathbf{M}$ $\mathbf{TxtFin}$-identifies $\mathcal{L}$, if it $\mathbf{TxtFin}$-identifies each $L \in \mathcal{L}$.

(c) $\mathbf{TxtFin} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{TxtFin}(\mathbf{M})]\}$.

The following definition is based on a learner semantically, rather than syntactically, converging to a grammar (or grammars) for an input language. Here note that equivalence of grammars is non-computable. The corresponding notion for learning functions was introduced by [Bār74b,CS83].

Definition 3 [CL82,OW82a].

(a) $\mathbf{M}$ $\mathbf{TxtBc}$-*identifies* L (written: $L \in \mathbf{TxtBc}(\mathbf{M})$) $\Leftrightarrow$ ($\forall$ texts T for L)$(\forall^\infty n)[W_{\mathbf{M}(T[n])} = L]$.

(b) $\mathbf{M}$ $\mathbf{TxtBc}$-identifies $\mathcal{L}$, if it $\mathbf{TxtBc}$-identifies each $L \in \mathcal{L}$.

(c) $\mathbf{TxtBc} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{TxtBc}(\mathbf{M})]\}$.

It can be shown that $\mathbf{TxtEx} \subset \mathbf{TxtBc}$ (for example, see [CL82,OW82a]).

The following concept is useful for proving some of our results.

Definition 4
(a) [Ful85] σ is a $\mathbf{TxtEx}$-*stabilizing sequence for* $\mathbf{M}$ *on* L just in case $\text{content}(\sigma) \subseteq L$ and $(\forall \tau \mid \text{content}(\tau) \subseteq L \wedge \sigma \subseteq \tau)[\mathbf{M}(\tau) = \mathbf{M}(\sigma)]$.

(b) [BB75,OW82b] σ is a **TxtEx**-*locking sequence for* $\mathbf{M}$ *on* L just in case σ is a **TxtEx**-stabilizing sequence for $\mathbf{M}$ on L and $W_{\mathbf{M}(\sigma)} = L$.

Lemma 5 *[BB75] If* $\mathbf{M}$ **TxtEx**-*identifies* L*, then there exists a* **TxtEx**-*locking sequence for* $\mathbf{M}$ *on* L*. Furthermore, all stabilizing sequences for* $\mathbf{M}$ *on* L *are locking sequences for* $\mathbf{M}$ *on* L.

Similarly one can define **TxtBc**-stabilizing sequences and **TxtBc**-locking sequences for $\mathbf{M}$ on L. A lemma similar to Lemma 5 can be established for **TxtBc**-learning as well as other criteria of inference considered below. We often drop **TxtEx**- (**TxtBc**-, etc.) from **TxtEx**-(**TxtBc**-, etc.)-stabilizing sequence, when it is clear from context.

Definition 6 [Bār74a,Ang80b] (a) $\mathbf{M}$ is *consistent* on text T iff, for all n, $\text{content}(T[n]) \subseteq W_{\mathbf{M}(T[n])}$.

(b) $\mathbf{M}$ is *consistent* on L iff it is consistent on all texts for L.

(c) $\mathbf{M}$ is *consistent* on $\mathcal{L}$ iff it is consistent on all $L \in \mathcal{L}$.

(d) $\mathbf{M}$ **TxtCons**-identifies $\mathcal{L}$ iff it is consistent on $\mathcal{L}$ and **TxtEx**-identifies $\mathcal{L}$.

TxtCons $= \{\mathcal{L} \mid$ some $\mathbf{M}$ **TxtCons**-identifies $\mathcal{L}\}$.

3 Learning Sublanguages: Definitions and Separations

Below we define our three models for learning sublanguages, as explained in the Introduction, as well as their variants reflecting the requirement of the final correct conjecture describing a language in the learnable class. We give our definitions for the **Ex** and **Bc** paradigms of learnability in the limit.

Intuitively, we vary three parameters in our learning criteria (in addition to the base criterion such as **Ex** or **Bc**): (a) whether we want the extensions to be subsets of *every* language in the class of which the input is a subset (denoted by **Sub** in the name of the criterion), or of a *minimal* language in the class of which the input is a subset (denoted by **MWSub** in the name of the criterion), or *only one* of the languages in the class of which the input is a subset (denoted by **WSub** in the name of the criterion), (b) whether *all* sublanguages are to be extended (denoted by **All** in the name of the criterion), or only the *infinite* ones (denoted by **Inf** in the name of the criterion), and (c) whether we require the final hypothesis extending the input to be *within the class* or not (denoted by presence or absence of **Res** in the name of the criterion).

A language $L \in \mathcal{L}$ is said to be a *minimal language* [Muk94] containing S in $\mathcal{L}$, iff $S \subseteq L$, and no $L' \in \mathcal{L}$ satisfies $S \subseteq L' \subset L$.

Below, **Sub** denotes learning subsets, **WSub**, denotes weak learning of subsets, and **MWSub** denotes minimal weak learning of subsets. We first consider extending all subsets.

Definition 7 (a) **M AllSubEx**-*identifies* $\mathcal{L}$, iff for all $L \in \mathcal{L}$, for all texts T such that content$(T) \subseteq L$, $\mathbf{M}(T)$ converges to a grammar i such that content$(T) \subseteq W_i \subseteq L$.

(b) **M AllWSubEx**-*identifies* $\mathcal{L}$ iff **M TxtEx**-identifies $\mathcal{L}$ and for all $L \in \mathcal{L}$, for all texts T such that content$(T) \subseteq L$, $\mathbf{M}(T)$ converges to a grammar i such that content$(T) \subseteq W_i \subseteq L'$, for some $L' \in \mathcal{L}$.

(c) **M AllMWSubEx**-*identifies* $\mathcal{L}$ iff for all $L \in \mathcal{L}$, for all texts T such that content$(T) \subseteq L$, $\mathbf{M}(T)$ converges to a grammar i such that content$(T) \subseteq W_i \subseteq L'$, for some $L' \in \mathcal{L}$, such that L' is a minimal language containing content(T) in $\mathcal{L}$.

(d) For $\mathbf{I} \in \{\mathbf{AllSubEx}, \mathbf{AllWSubEx}, \mathbf{AllMWSubEx}\}$, we say that **M ResI**-identifies $\mathcal{L}$ iff **M I**-identifies $\mathcal{L}$, and for all $L \in \mathcal{L}$, for all texts T such that content$(T) \subseteq L$, $W_{\mathbf{M}(T)} \in \mathcal{L}$.

As for the latter part of the above definition, it must be noted that Mukouchi [Muk94] considered a variation of **ResAllMWSubEx** for indexed families and provided some sufficient conditions for learnability in the model. Essentially his model allowed a learner to diverge if the input language did not have any minimal extension in $\mathcal{L}$. Kobayashi and Yokomori [KY95] considered a variation of **ResAllSubEx** learning (and briefly also **ResAllMWSubEx** learning) for indexed families of recursive languages and provided some characterizations. Essentially, they required a learner to learn on all inputs, even those which may not be contained in any language in the class (in other words, they required N to be a member of the class). Mukouchi and Kobayashi and Yokomori arrived at their definitions via a slightly different motivation (to find minimal extensions within the class), and, thus, had definitions somewhat different from ours. Here note that Kobayashi and Yokomori's technique also gives that the class of pattern languages [Ang80a] belongs to **AllSubEx**.

Note also that learning from incomplete texts (with just finite amount of data missing) has been studied in the context where the final grammar still was required to be a correct (or nearly correct) description of the full target language (see, for example, [OSW86,FJ96]). In general, this is incomparable with our approach.

In part (b) of the above definition, we explicitly added **TxtEx**-identifiability as the rest of the definition in part (b) does not imply **TxtEx**-identifiability (for parts (a) and (c), this was not needed, as the conditions imply **TxtEx**-identifiability).

We now consider **Bc**-learnability of sublanguages.

Definition 8 (a) **M AllSubBc**-*identifies* $\mathcal{L}$ iff for all $L \in \mathcal{L}$, for all texts T such that content$(T) \subseteq L$, for all but finitely many n, content$(T) \subseteq W_{\mathbf{M}(T[n])} \subseteq L$.

(b) **M AllWSubBc**-*identifies* $\mathcal{L}$ iff **M TxtBc**-identifies $\mathcal{L}$ and for all $L \in \mathcal{L}$, for all texts T such that content$(T) \subseteq L$, for all but finitely many n, for some $L' \in \mathcal{L}$, content$(T) \subseteq W_{\mathbf{M}(T[n])} \subseteq L'$.

(c) **M AllMWSubBc**-*identifies* $\mathcal{L}$ iff for all $L \in \mathcal{L}$, for all texts T such that content$(T) \subseteq L$, for all but finitely many n, for some $L' \in \mathcal{L}$ such that L' is a minimal language containing content(T) in $\mathcal{L}$, content$(T) \subseteq W_{\mathbf{M}(T[n])} \subseteq L'$.

(d) For $\mathbf{I} \in \{\mathbf{AllSubBc}, \mathbf{AllWSubBc}, \mathbf{AllMWSubBc}\}$, we say that **M ResI**-identifies $\mathcal{L}$ iff **M I**-identifies $\mathcal{L}$, and for all $L \in \mathcal{L}$, for all texts T such that content$(T) \subseteq L$, for all but finitely many n, $W_{\mathbf{M}(T[n])} \in \mathcal{L}$.

In the above definitions, when we only require extending infinite subsets, then we replace **All** by **Inf** in the name of the criterion (for example, **InfSubEx**).

Based on [OSW86], one can show that there exists a recursive sequence $\mathbf{M}_0, \mathbf{M}_1, \ldots,$ of total learning machines such that, for all the learning criteria **I** discussed in this paper (except for those involving consistent learning), if $\mathcal{L} \in \mathbf{I}$, then some $\mathbf{M}_i$ in the sequence witnesses that $\mathcal{L} \in \mathbf{I}$. From now on we fix such a recursive sequence $\mathbf{M}_0, \mathbf{M}_1, \ldots$ of learning machines.

Our first proposition establishes a number of simple relationships between our different models that easily follow from the definitions. For example, $\mathbf{AllSubEx} \subseteq \mathbf{AllMWSubEx} \subseteq \mathbf{AllWSubEx}$. This is so, since any language $X \supseteq L$, which is a subset of all languages in $\mathcal{L}$ containing L, is also a subset of any minimal language in $\mathcal{L}$ containing L. Similarly, any language $X \supseteq L$, which is a subset of a minimal language, A, in $\mathcal{L}$ containing L, is also a subset of some language (in particular A) in $\mathcal{L}$ containing L.

Proposition 9 *Suppose* $\mathbf{I} \in \{\mathbf{All}, \mathbf{Inf}\}$, $\mathbf{J} \in \{\mathbf{Sub}, \mathbf{WSub}, \mathbf{MWSub}\}$, $\mathbf{K} \in \{\mathbf{Ex}, \mathbf{Bc}\}$.

(a) $\mathbf{ResIJK} \subseteq \mathbf{IJK}$.

(b) $\mathbf{AllJK} \subseteq \mathbf{InfJK}$.

(c) $\mathbf{ISubK} \subseteq \mathbf{IMWSubK} \subseteq \mathbf{IWSubK}$.

(d) **IJEx** $\subseteq$ **IJBc**.

(b), (c), (d) above hold for **Res** *versions too.*

PROOF. Follows directly from definitions. ∎

Results below will show that the above inclusions are proper. They give the advantages of having a weaker restriction, such as the final conjecture not being required to be within the class (Theorem 10), **WSub** vs **MWSub** vs **Sub** (Theorems 12 and 11) and **Inf** vs **All** (Theorem 13).

First we show that the requirement of the last correct conjecture being a member of the learnable class makes a difference for the sublanguage learners: there are classes of languages learnable in our most restrictive model, **AllSubEx**, and not learnable in the least restrictive model, **ResInfWSubBc**, satisfying this requirement.

Theorem 10 **AllSubEx** $-$ **ResInfWSubBc** $\neq \emptyset$.

PROOF. For any function $f : N \to N$, let $L_f = \{\langle x, f(x)\rangle \mid x \in N\}$. Let $\mathcal{L} = \{L_f \mid f \in \mathcal{R} \ \wedge \ \text{card}(\text{range}(f)) < \infty \ \wedge \ (\forall e \in \text{range}(f))[W_e = f^{-1}(e)]\}$. It is easy to verify that $\mathcal{L} \in$ **AllSubEx**: a learner can just form the set S of all (the finitely many) e such that $\langle x, e\rangle$ appears in the input text for some x, and then output a grammar (depending only on S) for $\bigcup_{e \in S}\{\langle x, e\rangle \mid x \in W_e\}$. However $\mathcal{L}$ is not in **ResInfWSubBc** (proof of Theorem 23 in [JKW04] can be easily adapted to show this). ∎

The above proof also shows **AllSubCons** $-$ **ResInfWSubBc** $\neq \emptyset$.

On the other hand, an **AllMWSubEx**-learner, even satisfying the **Res** variant of sublanguage learnability, can sometimes do more than any **SubBc**-learner even if just learnability of only infinite sublanguages is required.

Theorem 11 **ResAllMWSubEx** $-$ **InfSubBc** $\neq \emptyset$.

PROOF. Let $Y = \{\langle 1, x\rangle \mid x \in N\}$.

Let $Z_e = \{\langle 1, x\rangle \mid x \leq e\} \cup \{\langle 1, 2x\rangle \mid x \in N\} \cup \{\langle 0, 0\rangle\}$.

Let $\mathcal{L} = \{Y\} \cup \{Z_e \mid e > 0\}$.

Note that Y is not contained in any other language in the class, nor contains any other language of the class.

$\mathcal{L} \in$ **ResAllMWSubEx** as, on input σ, a learner can output as follows. If content$(\sigma) \subseteq Y$, then output a (standard) grammar for Y. If content(σ)

contains just $\langle 0,0\rangle$, then output a standard grammar for $\{\langle 0,0\rangle\}$. Otherwise output Z_e, where e is the maximum odd number such that $\langle 1,e\rangle \in \text{content}(\sigma)$ (if there is no such odd number, then one takes e to be 1).

On the other hand, suppose by way of contradiction that $\mathcal{L} \in$ **InfSubBc** as witnessed by **M**. Let σ be a **Bc**-locking sequence for **M** on Y (that is, $\text{content}(\sigma) \subseteq Y$, and on any τ such that $\sigma \subseteq \tau$ and $\text{content}(\tau) \subseteq Y$, **M** outputs a grammar for Y). Now, let e be the largest odd number such that $\langle 1,x\rangle \in \text{content}(\sigma)$ (we assume without loss of generality that there does exist such an odd number). Now let $L' = Y \cap Z_e$. So **M**, on any text for L' extending σ, should output (in the limit) grammars for L' rather than Y, a contradiction. ∎

The above proof also shows $\mathbf{ResAllMWSubCons} - \mathbf{InfSubBc} \neq \emptyset$.

Similarly to the above result, a **ResAllWSubEx**-learner can learn sometimes more than any **MWSubBc**-learner even if learnability for just infinite sub-languages is required.

Theorem 12 $\mathbf{ResAllWSubEx} - \mathbf{InfMWSubBc} \neq \emptyset$.

Proof. Let $L_0^k = \{\langle k,i,x\rangle \mid i > 0, x \in N\} \cup \{\langle k,0,0\rangle\}$.

For $j \in N$, let $L_{j+1}^k = \{\langle k,i,x\rangle \mid i > 0, x \leq j\} \cup \{\langle k,0,j+1\rangle\}$.

Let $\mathcal{L} = \{N\} \cup \{L_{r_k}^k \mid k \in N\}$, where we will determine r_k below.

First we show that, irrespective of the values of r_k, $\mathcal{L} \in$ **ResAllWSubEx**. Let **M** be defined as follows. Let g_N be a grammar for N, and let g_j^k be a grammar for L_j^k.

$$\mathbf{M}(\sigma) = \begin{cases} g_j^k, & \text{if content}(\sigma) \cap \{\langle x,0,y\rangle \mid x,y \in N\} = \{\langle k,0,j\rangle\} \text{ and} \\ & \text{content}(\sigma) \subseteq L_j^k; \\ g_N, & \text{otherwise.} \end{cases}$$

M witnesses that $\mathcal{L} \in$ **ResAllWSubEx**, as except for N, all languages in the class are minimal languages in the class, containing exactly one element from $\{\langle x,0,y\rangle \mid x,y \in N\}$.

For any given k, we now select r_k appropriately to show that $\mathbf{M}_k$ does not **InfMWSubBc**-identify $\mathcal{L}$. Consider the behaviour of $\mathbf{M}_k$ on inputs being $S_j^k = L_j^k - \{\langle k,0,j\rangle\}$. Note that $\mathbf{M}_k$ cannot $\mathbf{TxtBc}^1$-identify the class $\{S_j^k \mid j \in N\}$ (based on [Gol67]; here $\mathbf{TxtBc}^1$-identification is similar to **TxtBc**-identification except that on texts for a language L, **M** is allowed to output grammars which enumerate L, except for upto one error (of either omission or commission)). Pick r_k such that $\mathbf{M}_k$ does not $\mathbf{TxtBc}^1$-identify $S^k_{r_k}$. Now, if the input text is for the language $S^k_{r_k}$, then $\mathbf{M}_k$, in the limit, is supposed to output grammars for either $S^k_{r_k}$ or $L^k_{r_k}$, and thus $\mathbf{TxtBc}^1$-identify $S^k_{r_k}$, a contradiction as r_k was picked so that $\mathbf{M}_k$ does not $\mathbf{TxtBc}^1$-identify $S^k_{r_k}$. Since k was arbitrary, the theorem follows. ∎

The above proof also shows $\mathbf{ResAllWSubCons} - \mathbf{InfMWSubBc} \neq \emptyset$.

Now we show that limiting learnability to just infinite sublanguages, even in the most restrictive model, can give us sometimes more than learners in the least restrictive model, **WSub**, required to learn descriptions for *all* sublanguages.

Theorem 13 *$\mathbf{ResInfSubEx} - \mathbf{AllWSubBc} \neq \emptyset$.*

Proof. Using Kleene's Recursion Theorem [Rog67], for any i, let e_i be such that $W_{e_i} = \{\langle i, e_i, x\rangle \mid x \in N\}$. If $\mathbf{M}_i$ does not $\mathbf{TxtBc}$-identify W_{e_i}, then let $L_i = W_{e_i}$. Otherwise, let σ^i be a $\mathbf{TxtBc}$-locking sequence for $\mathbf{M}_i$ on W_{e_i}. Without loss of generality assume that $\text{content}(\sigma^i) \neq \emptyset$. Using Kleene's Recursion Theorem [Rog67], let ${e_i}' > e_i$ be such that $W_{e_i'} = \text{content}(\sigma^i) \cup \{\langle i, e_i', x\rangle \mid x \in N\}$, and then let $L_i = W_{e_i'}$. (Note that e_i' need not be uniformly constructed from e_i).

Let $\mathcal{L} = \{L_i \mid i \in N\}$. Now clearly, $\mathcal{L}$ is in $\mathbf{ResInfSubEx}$, as the learner can just output the maximum value of $\pi_2^3(x)$, where x is in the input language.

Now we show that $\mathcal{L} \notin \mathbf{AllWSubBc}$. For any i either $\mathbf{M}_i$ does not $\mathbf{TxtBc}$-identify $W_{e_i} = L_i$ or on any text extending σ^i for $\text{content}(\sigma^i) \subseteq L_i$, beyond σ^i, $\mathbf{M}_i$ outputs only grammars for W_{e_i} — which is not contained in any language in $\mathcal{L}$.

It follows that $\mathbf{M}_i$ does not $\mathbf{AllWSubBc}$-identify $\mathcal{L}$. Since i was arbitrary, the theorem follows. ∎

The above proof also shows $\mathbf{ResInfSubCons} - \mathbf{AllWSubBc} \neq \emptyset$.

Now we note that not all classes learnable within the traditional paradigm of algorithmic learning — even without requirement of providing the right conjecture in the limit — are learnable in our weakest model even if learnability of infinite sublanguages only is required.

Theorem 14 *$\mathbf{TxtFin} - \mathbf{InfWSubBc} \neq \emptyset$.*

Proof. Let $\mathbf{SVT} = \{L \mid (\forall x \in N)(\exists \text{ a unique } y \in N)[\langle x, y\rangle \in L]\}$. Let $L_e = \{\langle 1, e\rangle\} \cup \{\langle 0, x\rangle \mid x \in W_e\}$. Let $\mathcal{L} = \{L_e \mid W_e \in \mathbf{SVT}\}$. It is easy to verify that $\mathcal{L} \in \mathbf{TxtFin}$. However $\mathcal{L} \in \mathbf{InfWSubBc}$ implies that for any text T for $L_e - \{\langle 1, e\rangle\}$, the learner must output grammars for $L_e - \{\langle 1, e\rangle\}$ (except, maybe, for an extra element of form $\langle 1, e'\rangle$, for some e'), on almost all initial segments of T. This learner can be modified to $\mathbf{TxtBc}$ identify $\mathbf{SVT}$ as follows. For any grammar i, let $g(i)$ be a grammar for $\{x \mid \langle 0, x\rangle \in W_i\}$. Given a text T, define $T'(i) = \langle 0, T(i)\rangle$, if $T(i) \in N$; $T'(i) = \#$, if $T(i) = \#$. Now, let T be a text for an $\mathbf{SVT}$ language W_e. Then, T' is a text for $L_e - \{\langle 1, e\rangle\}$. Thus, the learner from the above must output grammars for $L_e - \{\langle 1, e\rangle\}$ (except, maybe, for an extra element of form $\langle 1, e'\rangle$, for some e'), on almost all initial segments of T'. One can convert these grammars to grammar for W_e using g defined above. This would give us $\mathbf{SVT} \in \mathbf{TxtBc}$, a contradiction to a result from [CS83]. ∎

We next show another cost of learning sublanguages: increase in mind changes [CS83].

Theorem 15 *There exists an $\mathcal{L}$ such that*

*(a) $\mathcal{L} \in$ **AllSubEx**.*

*(b) $\mathcal{L} \in$ **TxtFin**.*

*(c) $\mathcal{L} \in$ **ResAllMWSubEx**.*

*(d) For all $n \in N$, no learner which makes at most n mind changes can witness $\mathcal{L} \in$ **InfSubEx**.*

Proof. Let $L_e = \{\langle 0, e\rangle\} \cup \{\langle 1, x\rangle \mid x \in N\} \cup \{\langle 2, x\rangle \mid x \in W_e\}$. $\mathcal{L} = \{L_e \mid \text{card}(W_e) < \infty\}$.

(a) Consider a learner which, on input σ, outputs a (canonical) grammar for $(\text{content}(\sigma) \cap \{\langle 0, x\rangle, \langle 2, x\rangle \mid x \in N\}) \cup \{\langle 1, x\rangle \mid x \in N\}$. It is easy to verify that the learner **AllSubEx**-identifies $\mathcal{L}$.

(b) Consider a learner which, on input σ, outputs ?, if content(σ) does not contain any element of form $\langle 0, e\rangle$. Otherwise, the learner outputs a grammar for L_e for the least e such that $\langle 0, e\rangle \in \text{content}(\sigma)$. It is easy to verify that the learner **TxtFin**-identifies $\mathcal{L}$.

(c) Consider a learner which, on input σ, outputs a grammar for L_e for the least e such that $\langle 0, e\rangle \in \text{content}(\sigma)$, if there exists such an e. Otherwise the learner outputs a grammar for L_e, where e is obtained using effective version of Kleene's Recursion Theorem [Rog67] such that $W_e = \{x \mid \langle 2, x\rangle \in \text{content}(\sigma)\}$. It is easy to verify that the learner **ResAllMWSubEx**-identifies $\mathcal{L}$.

(d) Suppose by way of contradiction that $\mathbf{M}$ **InfSubEx**-identifies $\mathcal{L}$, and makes at most n mind changes on any input. Let σ be such that content$(\sigma) \subseteq \{\langle 1, x\rangle, \langle 2, x\rangle \mid x \in N\}$, and the number of mind changes done by $\mathbf{M}$ on σ is maximal (note that such a σ exists, as $\mathbf{M}$ makes at most n mind changes on any input). It follows that $W_{\mathbf{M}(\sigma)} \cap \{\langle 2, x\rangle \mid x \in N\}$ is finite. But then, consider an e such that W_e is finite and $\{\langle 2, x\rangle \mid x \in W_e\} \not\subseteq W_{\mathbf{M}(\sigma)}$ (note that there exists such an e). Let $L' = \{\langle 1, x\rangle \mid x \in N\} \cup \{\langle 2, x\rangle \mid x \in W_e\}$, and $L = L' \cup \{\langle 0, e\rangle\}$. Clearly, $L' \subseteq L$ and $L \in \mathcal{L}$. However, $\mathbf{M}$, on any text for L' extending σ, converges to $W_{\mathbf{M}(\sigma)}$, which is not an extension of L'. ∎

On the other hand, **Bc**-learners in the most restrictive model of sublanguage learnability can sometimes learn more than traditional **Ex**-learners that are not required to learn sublanguages.

Theorem 16 **ResAllSubBc** $-$ **TxtEx** $\neq \emptyset$.

Proof. Let $L_i = \{\langle i, x\rangle \mid x \in N\}$.

Let $\mathcal{L} = \{\emptyset\} \cup \{S_i \mid i \in N\}$, where S_i would be defined below. S_i will satisfy the following properties.

There exists an e_i such that

A) $\emptyset \subset S_i \subseteq L_{e_i}$,

B) W_{e_i} enumerates an infinite set of elements such that all but finitely many of these are grammars for S_i.

It follows immediately from the above that $\mathcal{L} \in$ **ResAllSubBc**, as on an input being a nonempty subset of L_{e_i}, a learner can just output an increasing sequence of elements from W_{e_i}.

We now define S_i such that $\mathbf{M}_i$ does not **TxtEx**-identify S_i. By implicit use of Kleene's Recursion Theorem [Rog67], there exists an e_i such that W_{e_i} can be defined as follows.

Let $X = \{\sigma \mid \text{content}(\sigma) \subseteq L_{e_i} \ \wedge\ \emptyset \subset \text{content}(\sigma) \subset W_{\mathbf{M}_i(\sigma)}\}$.

Let $Y = \{\sigma \mid \text{content}(\sigma) \subseteq L_{e_i} \ \wedge\ (\exists \tau \mid \sigma \subseteq \tau)[\text{content}(\tau) \subseteq L_{e_i} \ \wedge\ \mathbf{M}_i(\sigma) \neq \mathbf{M}_i(\tau)]\}$.

Note that both X and Y are recursively enumerable.

We assume without loss of generality that X is not empty. Let $\tau_0, \tau_1, \ldots$ be an infinite recursive sequence such that $\{\tau_j \mid j \in N\} = X$.

Let $Y_0, Y_1, \ldots$ be a sequence of recursive approximations to Y such that $Y_j \subseteq Y_{j+1}$ and $\bigcup_{j \in N} Y_j = Y$.

We now define W_{e_i} as follows.

Let g_j be defined such that

$$W_{g_j} = \begin{cases} \text{content}(\tau_j), & \text{if } \tau_j \notin Y; \\ L_{e_i}, & \text{otherwise.} \end{cases}$$

Let $s_r = \max(\{j \leq r \mid (\forall j' < j)[\tau_{j'} \in Y_r]\})$.

Now, if $\mathbf{M}_i$ does not have a stabilizing sequence, belonging to X, for L_{e_i}, then every g_r is a grammar for L_{e_i}, which is not **TxtEx**-identified by $\mathbf{M}_i$. In this case, let $S_i = L_{e_i}$. On the other hand, if j is the least number such that τ_j is a stabilizing sequence for $\mathbf{M}_i$ on L_{e_i}, then $\lim_{r \to \infty} s_r = j$, and W_{g_j} is a grammar for content(τ_j), which is not **TxtEx**-identified by $\mathbf{M}_i$. In this case let $S_i = \text{content}(\tau_j)$. Clearly, (A) is satisfied, and $\mathbf{M}_i$ does not **TxtEx**-identify S_i.

Let *pad* be a 1–1 recursive function such that $W_{pad(i,j)} = W_i$, for all i, j. Let $W_{e_i} = \{pad(g_{s_r}, r) \mid r \in N\}$. It is easy to verify that (B) is satisfied. ∎

Our next result, following a similar result from [JKW04], shows that learners in all our models that are required to learn *all* sublanguages can be made *consistent* (with the input seen so far).

Theorem 17 *Suppose* $\mathbf{I} \in \{\mathbf{Sub}, \mathbf{WSub}, \mathbf{MWSub}\}$.

(a) **AllIEx** $\subseteq$ **AllICons**.

(b) **ResAllIEx** $\subseteq$ **ResAllICons**.

Proof. Can be shown in a way similar to proof of Theorem 28 in [JKW04]. ∎

On the other hand, similarly to a corresponding fact in [JKW04], if learnability of infinite sublanguages only is required, consistency cannot be achieved sometimes.

Theorem 18 **ResInfSubEx** $-$ **TxtCons** $\neq \emptyset$.

Proof. Let $\mathcal{L} = \{L \mid \text{card}(L) = \infty \text{ and } (\exists e)[W_e = L \text{ and } (\forall^\infty x \in L)[\pi_1^2(x) = e]]\}$. It is easy to verify that $\mathcal{L} \in$ **ResInfSubEx**. A straightforward modification of the proof of Proposition 29 in [JKW04] can be adapted to show that $\mathcal{L} \notin$ **TxtCons**. ∎

4 Some Characterizations

In this section, we suggest some characterizations for sublanguage learnability of indexed classes.

First, we get a characterization of **ResAllSubEx** in terms of requirements that must be imposed on regular **TxtEx**-learnability.

For any set S, $\mathrm{Min}_{\mathcal{L}}(S)$ denotes a language $X \in \mathcal{L}$, if any, such that (a) $S \subseteq X$, and (b) for all $Y \in \mathcal{L}$ such that $S \subseteq Y$, $X \subseteq Y$. If there is no such X, then $\mathrm{Min}_{\mathcal{L}}(S)$ is undefined. Note that $\mathrm{Min}_{\mathcal{L}}(S)$, if defined, is the unique minimal language in $\mathcal{L}$ containing S. Also note that if $\mathcal{L}$ is closed under infinite intersections, then $\mathrm{Min}_{\mathcal{L}}(S)$ is defined for all S which are contained in some $L \in \mathcal{L}$.

Theorem 19 *Suppose $\mathcal{L}$ is an indexed family of recursive languages (with indexing $(L_i)_{i\in N}$). Then $\mathcal{L} \in$ **ResAllSubEx** iff (a) to (d) below hold.*

*(a) $\mathcal{L} \in$ **TxtEx**;*

(b) $\mathcal{L}$ is closed under non-empty infinite intersections (that is for any non-empty $\mathcal{L}' \subseteq \mathcal{L}$, $\bigcap_{L\in\mathcal{L}'} L \in \mathcal{L}$);

(c) For all finite S such that, for some $L \in \mathcal{L}$, $S \subseteq L$, one can effectively find in the limit an $\mathcal{L}$-grammar for $\mathrm{Min}_{\mathcal{L}}(S)$;

(d) For all infinite S which are contained in some $L \in \mathcal{L}$, $\mathrm{Min}_{\mathcal{L}}(S) = \mathrm{Min}_{\mathcal{L}}(X)$, for some finite subset X of S.

Proof. ($\Longrightarrow$) Suppose $\mathcal{L} \in$ **ResAllSubEx** as witnessed by $\mathbf{M}$.

(a) and (b) follow from definition of **ResAllSubEx**.

(c): Given any finite set S which is contained in some language in $\mathcal{L}$, for any text T_S for S, $\mathbf{M}(T_S)$ converges to a (r.e.) grammar for the minimal language in $\mathcal{L}$ containing S. This r.e. grammar can now be easily converted to a $\mathcal{L}$-grammar using **TxtEx**-identifiability of $\mathcal{L}$ (note that for an indexed family of recursive languages, **TxtEx**-learnability implies learnability using the hypotheses space $(L_i)_{i\in N}$).

(d): Suppose by way of contradiction that (d) does not hold. We then construct a text for S on which M does not converge to $\mathrm{Min}_{\mathcal{L}}(S)$. Let $(X_i)_{i\in N}$ be a family of non-empty and finite sets such that $\bigcup_{i\in N} X_i = S$ and $X_i \subseteq X_{i+1}$ for all i. Define $\sigma_0 = \Lambda$. Let σ_{i+1} be an extension of σ_i such that $\mathrm{content}(\sigma_{i+1}) = X_i$, and $\mathbf{M}(\sigma_{i+1})$ is a grammar for $\mathrm{Min}_{\mathcal{L}}(X_i)$ (note that there exists such a σ_{i+1} as $\mathbf{M}$ on any text for X_i converges to a grammar for $\mathrm{Min}_{\mathcal{L}}(X_i)$). Now let

$T = \bigcup_{i \in N} \sigma_i$. Clearly, T is a text for S. However, $\mathbf{M}(T)$ does not converge to a grammar for $\mathrm{Min}_{\mathcal{L}}(S)$, as $\mathrm{Min}_{\mathcal{L}}(X_i) \neq \mathrm{Min}_{\mathcal{L}}(S)$, for all i (by the assumption about (d) not holding). A contradiction to **M ResAllSubEx**-identifying $\mathcal{L}$. Thus, (d) must hold.

($\Longleftarrow$) Suppose (a) to (d) are satisfied. Let f be a recursive function such that for all finite S, $\lim_{t\to\infty} f(S,t)$ is an $\mathcal{L}$-grammar for $\mathrm{Min}_{\mathcal{L}}(S)$ (by the clause (c), there exists such an f). Then, define $\mathbf{M}'$ as follows. $\mathbf{M}'$ on any input $T[n]$, computes $i_j^n = f(\mathrm{content}(T[j]), n)$, for $j \leq n$. Then it outputs i_j^n, for minimal j such that $\mathrm{content}(T[n]) \subseteq L_{i_j^n}$. By definition of f, for each j, $i_j = \lim_{n\to\infty} i_j^n$ is defined and is an $\mathcal{L}$-grammar for $\mathrm{Min}_{\mathcal{L}}(\mathrm{content}(T[j]))$. As for all but finitely many j, $\mathrm{Min}_{\mathcal{L}}(\mathrm{content}(T[j])) = \mathrm{Min}_{\mathcal{L}}(\mathrm{content}(T))$ (by the clause (d)), we have that $\mathbf{M}'$ will converge on T to i_k, where k is minimal such j. It follows that $\mathbf{M}'(T)$ converges to an $\mathcal{L}$-grammar for $\mathrm{Min}_{\mathcal{L}}(\mathrm{content}(T))$. Note that this also implies **TxtEx**-identifiability of $\mathcal{L}$ by $\mathbf{M}'$. ∎

Our next theorem shows that, if an indexed class is learnable within the models **WSub** or **MWSub** under the requirement that the last (correct) conjecture is a member of the learnable class $\mathcal{L}$, then the learner can use conjectures from the class $\mathcal{L}$ itself. In particular, this result will be used in our next characterizations.

Theorem 20 *Suppose $\mathcal{L}$ (with indexing $(L_i)_{i\in N}$) is an indexed family of recursive languages. Then $\mathcal{L} \in$* **ResAllWSubEx** *(***ResInfWSubEx***,* **ResAllMWSubEx***,* **ResInfMWSubEx***) iff there exists a machine* **M** *such that* **M ResAllWSubEx***-identifies (***ResInfWSubEx***-identifies,* **ResAllMWSubEx***-identifies,* **ResInfMWSubEx***-identifies) $\mathcal{L}$ using indexing $(L_i)_{i\in N}$ of $\mathcal{L}$ as its hypotheses space.*

PROOF. We only show the case of **ResAllWSubEx**. The same proof applies for **ResInfWSubEx**, **ResAllMWSubEx**, and **ResInfMWSubEx**.

Suppose **M ResAllWSubEx**-identifies $\mathcal{L}$. Then, for all L' such that $L' \subseteq L$ for some $L \in \mathcal{L}$, there exists a σ, τ, i such that

(a) σ is a stabilizing sequence for $\mathbf{M}$ on L';

(b) τ is the least stabilizing sequence for $\mathbf{M}$ on $W_{\mathbf{M}(\sigma)}$ as well as on L_i;

(c) for any τ', i' such that τ' is the least stabilizing sequence for $\mathbf{M}$ on $W_{\mathbf{M}(\sigma)}$ as well as on $W_{i'}$, we must have that $W_{\mathbf{M}(\sigma)} = L_{i'}$.

Note that (a) holds by definition of **ResAllWSubEx**-identifiability. So fix one such σ. (b) holds as $W_{\mathbf{M}(\sigma)}$ must be a member of $\mathcal{L}$, and thus $W_{\mathbf{M}(\sigma)} = L_i$ for some i, and τ is then the least stabilizing sequence for $\mathbf{M}$ on L_i. (c) holds as

M has different least stabilizing sequences for different languages in $\mathcal{L}$ (since **M** **TxtEx**-learns $\mathcal{L}$).

Define $\mathbf{M}'$ as follows. On any input text T, search for σ, τ, i such that (a) and (b) above hold for $L' = \text{content}(T)$. Then, in the limit on T, output i.

Now from **M** **ResAllWSubEx**-identifying $\mathcal{L}$ and (c), we immediately have that $\mathbf{M}'$ **ResAllWSubEx**-identifies $\mathcal{L}$ using hypotheses space $(L_i)_{i \in N}$. ∎

Now we show that learnability within the model **ResAllWSubEx** is equivalent to regular learnability **TxtEx** if a learner just stabilizes on every input sublanguage of every language in the learnable indexed family $\mathcal{L}$.

Theorem 21 *Suppose $\mathcal{L}$ (with indexing $(L_i)_{i \in N}$) is an indexed family of recursive languages. Then $\mathcal{L} \in$* **ResAllWSubEx** *iff there exists a machine* **M** *such that:*

(a) **M** **TxtEx**-*identifies $\mathcal{L}$ using indexing $(L_i)_{i \in N}$ as the hypotheses space.*

(b) For all texts T such that, for some $L \in \mathcal{L}$, $\text{content}(T) \subseteq L$, we have: $\mathbf{M}(T)\downarrow$.

Proof. ($\Longrightarrow$) If $\mathcal{L} \in$ **ResAllWSubEx**, then (a) and (b) follow from the definition of **ResAllWSubEx** and Theorem 20.

($\Longleftarrow$) Suppose **M** is given such that (a) and (b) hold. Define $\mathbf{M}'$ as follows:

$$\mathbf{M}'(\sigma) = \begin{cases} \mathbf{M}(\sigma), & \text{if content}(\sigma) \subseteq L_{M(\sigma)}; \\ j, & \text{otherwise, where } j = \min(\{|\sigma|\} \cup \{i : \text{content}(\sigma) \subseteq L_i\}). \end{cases}$$

The first clause ensures **TxtEx** learnability of $\mathcal{L}$ by $\mathbf{M}'$ using the hypotheses space $(L_i)_{i \in N}$. Now consider any text T for $L' \subseteq L$, where $L \in \mathcal{L}$. Since **M** converges on T, let i be such that $\mathbf{M}(T) = i$. If $\text{content}(T) \subseteq L_i$, then clearly $\mathbf{M}'(T) = i$ too. On the other hand, if $\text{content}(T) \not\subseteq L_i$, then by the second clause in the definition of $\mathbf{M}'$, $\mathbf{M}'(T)$ will converge to the least j such that $\text{content}(T) \subseteq L_j$. It follows that $\mathbf{M}'$ **ResAllWSubEx**-identifies $\mathcal{L}$ using the hypotheses space $(L_i)_{i \in N}$. ∎

Proof technique used for Theorem 21 can also be used to show the following.

Theorem 22 *Suppose $\mathcal{L}$ (with indexing $(L_i)_{i \in N}$) is an indexed family of recursive languages. Then $\mathcal{L} \in$* **ResInfWSubEx** *iff there exists a machine* **M** *such that:*

(a) $\mathbf{M}$ $\mathbf{TxtEx}$*-identifies* $\mathcal{L}$ *using the hypotheses space* $(L_i)_{i \in N}$.

(b) For all texts T *such that* $\mathrm{content}(T)$ *is infinite and* $\mathrm{content}(T) \subseteq L$ *for some* $L \in \mathcal{L}$, $\mathbf{M}(T)\downarrow$.

The next theorem presents a simple natural condition sufficient for learnability of indexed classes in the model **ResAllWSubEx**.

Theorem 23 *Suppose* $\mathcal{L}$ *is an indexed family of recursive languages (with indexing* $(L_i)_{i \in N}$*) such that for any distinct languages* L, L' *in* $\mathcal{L}$, $L \not\subseteq L'$. *Then,* $\mathcal{L} \in$ **ResAllWSubEx**.

PROOF. $\mathbf{M}$, on input σ, outputs the least i such that $\mathrm{content}(\sigma) \subseteq L_i$. It is easy to verify that $\mathbf{M}$ **ResAllWSubEx**-identifies $\mathcal{L}$. ∎

Theorem 14 proof shows that the condition of indexed family in the above theorem cannot be dropped.

Our main characterizations show how classes learnable within our models of sublanguage learning can be described in terms of some different aspects of learnability of the languages (or their sublanguages) in these classes. These characterizations may be useful when the aspects in question are known, but sublanguage learnability is yet to be established. It would be interesting to obtain similar characterizations for **ResAllMWSubEx**, sublearning without **Res** (i.e., for **AllSubEx**, **AllWSubEx** **AllMWSubEx**), and for the case when sublearning only in the presence of infinite inputs is considered (the **Inf** versions).

5 Monotonicity Constraints

In this section we consider sublanguage learnability satisfying monotonicity constraints. Our primary goal is to explore how so-called *strong monotonicity* ([Jan91]) affects sublanguage learnability: the learners are strongly monotonic for the criteria discussed in this paper in the sense that, when we get more data in the text, then the languages conjectured are larger.

Definition 24 [Jan91] (a) $\mathbf{M}$ is said to be *strong-monotonic* on L just in case $(\forall \sigma, \tau \mid \sigma \subseteq \tau \ \wedge \ \mathrm{content}(\tau) \subseteq L)[\mathbf{M}(\sigma) =? \ \vee \ W_{\mathbf{M}(\sigma)} \subseteq W_{\mathbf{M}(\tau)}]$.

(b) $\mathbf{M}$ is said to be *strong-monotonic* on $\mathcal{L}$ just in case $\mathbf{M}$ is strong-monotonic on each $L \in \mathcal{L}$.

(c) $\mathbf{SMon} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathbf{M}$ is strong-monotonic on $\mathcal{L}$ and $\mathcal{L} \subseteq \mathbf{TxtEx}(\mathbf{M})]\}$.

Now, following [Jan91], we will also define a much weaker notion of monotonicity — and will show that general learners in our most restrictive model **ResAllSubEx** do not satisfy even this requirement.

Definition 25 [Jan91] (a) $\mathbf{M}$ is said to be *weak-monotonic* on L just in case $(\forall \sigma, \tau \mid \sigma \subseteq \tau \ \wedge \ \text{content}(\tau) \subseteq L)[\mathbf{M}(\sigma) = ? \ \vee \ [\text{content}(\tau) \subseteq W_{\mathbf{M}(\sigma)} \Rightarrow W_{\mathbf{M}(\sigma)} \subseteq W_{\mathbf{M}(\tau)}]]$.

(b) $\mathbf{M}$ is said to be *weak-monotonic* on $\mathcal{L}$ just in case $\mathbf{M}$ is weak-monotonic on each $L \in \mathcal{L}$.

(c) $\mathbf{WMon} = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathbf{M}$ is weak-monotonic on $\mathcal{L}$ and $\mathcal{L} \subseteq \mathbf{TxtEx}(\mathbf{M})]\}$.

Theorem 26 **ResAllSubEx** − **WMon** $\neq \emptyset$.

PROOF. Let $L_j = \{\langle j, x\rangle \mid x \in N\}$. $L_j^m = \{\langle j, x\rangle \mid x < m\}$. Let T_j be a text for L_j such that content$(T_j[m]) = L_j^m$. Let $S_j = \{\langle m, n\rangle \mid m > 0 \ \wedge \ \{\langle j, x\rangle \mid x \leq m\} \subseteq W_{\mathbf{M}_j(T_j[m]),n}\}$.

Let $\mathcal{L} = \{\emptyset\} \cup \{L_j \mid S_j = \emptyset\} \cup \{L_j^m \mid S_j \neq \emptyset \ \wedge \ (\exists n)[\langle m, n\rangle = \min(S_j)]\}$.

It is easy to verify that $\mathcal{L} \in$ **ResAllSubEx**. It was shown in [JS98] that $\mathcal{L} \notin$ **WMon**. ∎

Let **AllWSubSMon**, etc., denote the corresponding learning criteria. In those criteria, **Ex**-type of learnability is assumed by default, unless **Bc** is explicitly added at the end.

Unlike the general case of sublanguage learning, strong monotonicity requirement forces all variants of the least restrictive model, **MWSub**, to collapse to the most restrictive model **Sub**. For **Bc**-learning, it can also be shown that there is no difference whether only infinite sublanguages are required to be learned, or all sublanguages. This latter result, though, does not hold when we consider **Ex**-learning, or require the learners to converge to grammars for a language within the class.

Theorem 27 *(a)* **AllWSubSMon** $\subseteq$ **AllSubSMon**.

(b) **InfWSubSMon** $\subseteq$ **InfSubSMon**.

(c) **AllWSubSMonBc** $\subseteq$ **AllSubSMonBc**.

(d) **InfWSubSMonBc** $\subseteq$ **InfSubSMonBc**.

(e) **InfSubSMonBc** $\subseteq$ **AllSubSMonBc**.

(a) to (d) above hold for **Res** *versions too.*

PROOF. We show (a). (b) to (e) (and **Res** versions for (a) to (d)) can be proved similarly. Suppose **M AllWSubSMon**-identifies $\mathcal{L}$. We first note that for all $L \in \mathcal{L}$, for all σ such that content$(\sigma) \subseteq L$, $W_{\mathbf{M}(\sigma)} \subseteq L$. This is so, since otherwise for any text T for L which extends σ, $\mathbf{M}$ does not output a grammar contained in L for any extension of σ, due to strong monotonicity of $\mathbf{M}$. This, along with **AllWSubSMon**-identifiability of $\mathcal{L}$ by $\mathbf{M}$, implies **AllSubSMon**-identifiability of $\mathcal{L}$ by $\mathbf{M}$. ∎

A result similar to Theorem 27 holds (essentially by definition) if, instead of requiring strong monotonicity of the learner, one requires that for all $L \in \mathcal{L}$, for all σ such that content$(\sigma) \subseteq L$, $W_{\mathbf{M}(\sigma)} \subseteq L$.

Note that the proof of Theorem 27 is not able to show **InfSubSMon** $\subseteq$ **AllSubSMon**, as an **InfSubSMon**-learner may not converge on finite sets. Similarly, we do not get **ResInfSubSMonBc** $\subseteq$ **ResAllSubSMonBc** using the above proof. The following two theorems show that the above failure is not avoidable.

Theorem 28 **ResInfSubSMon** $-$ **AllSubSMon** $\neq \emptyset$*.*

PROOF. Let $X_{i,j} = \{\langle i, j, x\rangle \mid x \in N\}$. Using Kleene's Recursion Theorem [Rog67], for any i, let e_i be such that W_{e_i} is defined as follows. If there is no **TxtEx**-stabilizing sequence for $\mathbf{M}_i$ on X_{i,e_i}, then $W_{e_i} = X_{i,e_i}$. Otherwise, W_{e_i} is a finite set such that content$(\sigma^i) \subseteq W_{e_i} \subseteq X_{i,e_i}$, where σ^i is the least **TxtEx**-stabilizing sequence for $\mathbf{M}_i$ on X_{i,e_i} (here, without loss of generality we assume that content$(\sigma^i) \neq \emptyset$). Note that one can define such W_{e_i} as one can find the least **TxtEx**-stabilizing sequence, if any, in the limit.

If $\mathbf{M}_i$ does not have a **TxtEx**-stabilizing sequence on X_{i,e_i}, then let $L_i = W_{e_i}$. Otherwise, let σ^i be the least **TxtEx**-stabilizing sequence for $\mathbf{M}_i$ on X_{i,e_i}. Define S_i based on following two cases.

Case 1: $W_{\mathbf{M}_i(\sigma^i)}$ contains an infinite subset of X_{i,e_i}. In this case let $S_i =$ content(σ^i).

Case 2: Not case 1. In this case, let S_i be a finite set such that content$(\sigma^i) \subseteq S_i \subseteq X_{i,e_i}$ and $S_i \not\subseteq W_{\mathbf{M}_i(\sigma^i)}$.

Using Kleene's Recursion Theorem [Rog67], let $e_i' > e_i$ be such that $W_{e_i'} = S_i \cup W_{e_i} \cup \{\langle i, e_i', x\rangle \mid x \in N\}$, and then let $L_i = W_{e_i'}$.

Let $\mathcal{L} = \{L_i \mid i \in N\}$. Now clearly, $\mathcal{L}$ is in **ResInfSubSMon**, as (on an input with non-empty content) the learner can just output the maximum value of $\pi_2^3(x)$, where x is in the input language.

Now suppose by way of contradiction that $\mathbf{M}_i$ **AllSubSMon**-identifies $\mathcal{L}$. If $\mathbf{M}_i$ does not have a **TxtEx**-stabilizing sequence on X_{i,e_i}, then $\mathbf{M}_i$ does not **TxtEx**-identify $L_i = W_{e_i} = X_{i,e_i} \in \mathcal{L}$. Thus $\mathbf{M}_i$ cannot **AllSubSMon**-identify $\mathcal{L}$.

On the other hand, if $\mathbf{M}_i$ has σ^i as the least **TxtEx**-stabilizing sequence on X_{i,e_i}, then: in Case 1 above, $\mathbf{M}_i$ cannot **SMon**-identify L_i, as $W_{\mathbf{M}_i(\sigma^i)}$ is not a subset of L_i; in Case 2 above, $\mathbf{M}_i$ on any text for S_i, which extends σ^i, converges to $W_{\mathbf{M}_i(\sigma^i)}$, which is not a superset of S_i.

It follows that $\mathcal{L} \notin$ **AllSubSMon**. ∎

Theorem 29 **ResInfSubSMon** − **ResAllSubBc** $\neq \emptyset$.

PROOF. Define L_i as follows. Let T_i be a text for $\{\langle i, 0\rangle\}$. If $\mathbf{M}_i(T_i)$ infinitely often outputs a grammar containing $\langle i, 2x\rangle$, for some $x > 0$, then let $L_i = \{\langle i, 0\rangle\} \cup \{\langle i, 2x+1\rangle \mid x \in N\}$. Otherwise, let $L_i = \{\langle i, 0\rangle\} \cup \{\langle i, 2x\rangle \mid x \in N\}$.

Let $\mathcal{L} = \{L_i \mid i \in N\}$.

By construction of L_i, $\mathbf{M}_i$ on T_i infinitely often outputs a grammar different from the grammar for L_i, the only language in $\mathcal{L}$ which contains content(T_i). Thus, $\mathcal{L} \notin$ **ResAllSubBc**.

On the other hand, it is easy to verify that $\mathcal{L} \in$ **ResInfSubSMon** (as one can easily determine L_i from a text for any subset of L_i, which contains at least one element other than $\langle i, 0\rangle$). ∎

Note that proof of Theorem 10 also shows that

Theorem 30 **AllSubSMon** − **ResInfWSubBc** $\neq \emptyset$.

6 Conclusion

We introduced, discussed, and compared a number of different natural models for correct learning in the limit from partial data. In particular, we established how these models differ, and what is their relationship with other traditional models of inductive inference. We also studied the effect of strong monotonicity on learning sublanguages.

There are many different aspects of this topic that we have not addressed. In particular, an interesting issue is how the requirement of being able to correctly learn from partial data can affect complexity of learning (e.g., number of mind changes, long-term memory [FKS95], etc.; we considered just one example showing the cost of learning sublanguages in terms of mind changes).

We obtained some characterizations for learnability within our models, however, this issue is far from being closed: it would be interesting to find some more, possibly, structural characterizations of the classes learnable within the given models. Yet another possible direction for future work can include exploring learnability of some other reasonable types of sublanguages, for example, *dense* ones (we considered only infinite sublanguages as a reasonable restriction).

7 Acknowledgements

We thank the anonymous referees of the conference ALT'2006 and this journal for several helpful comments.

References

[Ang80a] D. Angluin. Finding patterns common to a set of strings. *Journal of Computer and System Sciences*, 21:46–62, 1980.

[Ang80b] D. Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.

[Bār74a] J. Bārzdiņš. Inductive inference of automata, functions and programs. In *Int. Math. Congress, Vancouver*, pages 771–776, 1974.

[Bār74b] J. Bārzdiņš. Two theorems on the limiting synthesis of functions. In *Theory of Algorithms and Programs, vol. 1*, pages 82–88. Latvian State University, 1974. In Russian.

[BB75] L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.

[Blu67] M. Blum. A machine-independent theory of the complexity of recursive functions. *Journal of the ACM*, 14:322–336, 1967.

[CL82] J. Case and C. Lynes. Machine inductive inference and language identification. In M. Nielsen and E. M. Schmidt, editors, *Proceedings of the 9th International Colloquium on Automata, Languages and Programming*, volume 140 of *Lecture Notes in Computer Science*, pages 107–115. Springer-Verlag, 1982.

[CS83] J. Case and C. Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science*, 25:193–220, 1983.

[FJ96] M. Fulk and S. Jain. Learning in the presence of inaccurate information. *Theoretical Computer Science A*, 161:235–261, 1996.

[FKS95] R. Freivalds, E. Kinber, and C. Smith. On the impact of forgetting on learning machines. *Journal of the ACM*, 42(6):1146–1168, 1995.

[Ful85] M. Fulk. *A Study of Inductive Inference Machines*. PhD thesis, SUNY/Buffalo, 1985.

[Gol67] E. M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[HU79] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979.

[Jan91] K. Jantke. Monotonic and non-monotonic inductive inference. *New Generation Computing*, 8:349–360, 1991.

[JK06] S. Jain and E. Kinber. Learning and extending sublanguages. In J. L. Balcazar, P. Long, and F. Stephan, editors, *Algorithmic Learning Theory: Seventeenth International Conference (ALT' 2006)*, volume 4264 of *Lecture Notes in Artificial Intelligence*, pages 139–153. Springer-Verlag, 2006.

[JKW04] S. Jain, E. Kinber, and R. Wiehagen. Learning all subfunctions of a function. *Information and Computation*, 192(2):185–215, August 2004.

[JS98] S. Jain and A. Sharma. Generalization and specialization strategies for learning r.e. languages. *Annals of Mathematics and Artificial Intelligence*, 23:1–26, 1998.

[KY95] S. Kobayashi and T. Yokomori. On approximately identifying concept classes in the limit. In K. Jantke, T. Shinohara, and T. Zeugmann, editors, *Algorithmic Learning Theory: Sixth International Workshop (ALT '95)*, volume 997 of *Lecture Notes in Artificial Intelligence*, pages 298–312. Springer-Verlag, 1995.

[KY97a] S. Kobayashi and T. Yokomori. Identifiability of subspaces and homomorphic images of zero-reversible languages. In Ming Li and Akira Maruoka, editors, *Algorithmic Learning Theory: Eighth International Workshop (ALT '97)*, volume 1316 of *Lecture Notes in Artificial Intelligence*, pages 48–61. Springer-Verlag, 1997.

[KY97b] S. Kobayashi and T. Yokomori. Learning approximately regular languages with reversible languages. *Theoretical Computer Science A*, 174:251–257, 1997.

[Muk94] Y. Mukouchi. Inductive inference of an approximate concept from positive data. In S. Arikawa and K. Jantke, editors, *Algorithmic Learning Theory: Fourth International Workshop on Analogical and Inductive Inference (AII '94) and Fifth International Workshop on Algorithmic Learning Theory (ALT '94)*, volume 872 of *Lecture Notes in Artificial Intelligence*, pages 484–499. Springer-Verlag, 1994.

[MY78] M. Machtey and P. Young. *An Introduction to the General Theory of Algorithms*. North Holland, New York, 1978.

[OSW86] D. Osherson, M. Stob, and S. Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists*. MIT Press, 1986.

[OW82a] D. Osherson and S. Weinstein. Criteria of language learning. *Information and Control*, 52:123–138, 1982.

[OW82b] D. Osherson and S. Weinstein. A note on formal learning theory. *Cognition*, 11:77–88, 1982.

[Pin79] S. Pinker. Formal models of language learning. *Cognition*, 7:217–283, 1979.

[Rog58] H. Rogers. Gödel numberings of partial recursive functions. *Journal of Symbolic Logic*, 23:331–341, 1958.

[Rog67] H. Rogers. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, 1967. Reprinted by MIT Press in 1987.

[Sch85] G. Schäfer-Richter. Some results in the theory of effective program synthesis - learning by defective information. In W. Bibel and K. Jantke, editors, *Mathematical Methods of Specification and Synthesis of Software Systems, Wendisch-Rietz, GDR*, volume 215 of *Lecture Notes in Computer Science*, pages 219–225. Springer-Verlag, 1985.

[Ste97] F. Stephan. Noisy inference and oracles. *Theoretical Computer Science A*, 185:129–157, 1997.

[WC80] K. Wexler and P. Culicover. *Formal Principles of Language Acquisition.* MIT Press, 1980.

[Wie90] R. Wiehagen. A thesis in inductive inference. In J. Dix, K. Jantke, and P. Schmitt, editors, *Nonmonotonic and Inductive Logic, 1st International Workshop*, volume 543 of *Lecture Notes in Artificial Intelligence*, pages 184–207. Springer-Verlag, 1990.

[WZ95] R. Wiehagen and T. Zeugmann. Learning and consistency. In K. P. Jantke and S. Lange, editors, *Algorithmic Learning for Knowledge-Based Systems*, volume 961 of *Lecture Notes in Artificial Intelligence*, pages 1–24. Springer-Verlag, 1995.