

Language Learning With Some Negative Information ¹

Ganesh Baliga
Department of Computer Science
Rowan College of New Jersey
Glassboro, NJ 08028
USA

John Case
Department of Computer and Information Sciences
University of Delaware
Newark, DE 19716
USA

Sanjay Jain
Institute of Systems Science
National University of Singapore
Singapore 0511
Republic of Singapore

¹A preliminary version of this paper was presented at the *10th Symposium on Theoretical Aspects of Computer Science (STACS'93)*, Würzburg, Germany, Feb. 1993.

Abstract

Gold-style language learning is a formal theory of learning from examples by algorithmic devices called learning machines. Originally motivated by child language learning, it features the algorithmic synthesis (in the limit) of grammars for formal languages from information about those languages. In traditional Gold-style language learning, learning machines are not provided with negative information, i.e., information about the complements of the input languages. We investigate two approaches to providing small amounts of negative information and demonstrate in each case a strong resulting increase in learning power. Finally, we show that small packets of negative information also lead to increased *speed* of learning. This result agrees with a psycholinguistic hypothesis of McNeill correlating the availability of parental expansions with the speed of child language development.

1 Introduction

Gold-style formal language learning [12] features the algorithmic synthesis of generating procedures for formal languages from enumerations of positive information about the languages. Herein we consider Gold-style formal language learning augmented in various ways by some amount of negative information about the languages.

Chapter 6 of [9] treats the case of augmentation with various powerful forms of nearly complete grammatical information for the complement of the language. In [4] the report on this chapter is motivated by treating negative information as a more mathematically tractable substitute for semantic information. [18, 19] present evidence that semantics in addition to positive information may be essential to human language learning. [14] examines language learning with varying *densities* of negative information provided.

Fulk's chapter, of necessity, is about learning recursive languages, but we would like to consider the learning of r.e. not recursive languages too, where the positive information is supplemented by some negative information. For example, some recursively axiomatizable theories such as first order group theory are r.e. not recursive [17], yet we might want to study the learning of recursive axiomatizations (i.e., generators) for them.

In the present paper we consider the effects on learning power obtained by augmenting the positive information by apparently small, finite, *core* amounts of negative information. We consider two cases, one (Section 3) where only the finite core is supplied and one (Section 4) where negative information in addition to the finite core is allowed. This latter case was motivated in part by [20] which itself, in part, was motivated by [1, 29]. We refer to the latter style of negative information presentation as *open*.

In Section 3 we present results to the effect that tremendous gains in learning power are obtained from adding apparently very small, suitable, finite sets of negative information.

In Section 4 we present results showing that, with respect to learning power, in some cases, allowing more mistakes in final generators learned can more than compensate information theoretically for open negative information. In other cases, we show that small additions of open negative information can more than compensate, also information theoretically, for mistakes in final grammars. It is noted that, in almost all cases, additional open negative information results in strictly more learning power. We discuss the relation of this to the hypothesis in [16] that increasing certain forms of language correction leads to increased speed in language development. In Section 4 it is also noted that, regarding learning power for languages, supplying an unbounded amount of open negative information is equivalent to supplying *all* the negative (as well as the positive) information to a learning machine.

Finally in Section 5 we present a surprising preliminary result supportive of the hypothesis of [16] mentioned in the just previous paragraph. This result says that one does see a learning *speed* increase (as measured by mind-change complexity [8]) from a minimal, non-vacuous amount of open negative information.

Some open questions are presented in Section 6.

2 Preliminaries

2.1 Notation

Any unexplained recursion theoretic notation is from [28]. N denotes the set of natural numbers, $\{0, 1, 2, 3, \dots\}$. Unless otherwise specified, $e, i, j, k, m, n, p, s, w, x, y, z$, with or without decorations¹, range over N . $*$ denotes a non-member of N and is assumed to satisfy $(\forall n)[n < * < \infty]$. a, b and c , with or without decorations, range over $N \cup \{*\}$. \emptyset denotes the empty set. \subseteq denotes subset. \subset denotes proper subset. \supseteq denotes superset. \supset denotes proper superset. P and S , with or without decorations, range over sets. $\mathcal{P}(S)$ denotes the power set of S . $\text{card}(S)$ denotes the cardinality of S . $i^{\text{th}}\text{-min}(S)$ denotes the element x of S , if any, such that $\text{card}(\{y \in S \mid y \leq x\}) = i$. $S_1 \oplus S_2 \stackrel{\text{def}}{=} \{2x \mid x \in S_1\} \cup \{2x + 1 \mid x \in S_2\}$. $S_1 \Delta S_2$ denotes the symmetric difference between S_1 and S_2 . $S_1 \stackrel{\text{def}}{=}^n S_2$ denotes $\text{card}(\{x \mid x \in S_1 \Delta S_2\}) \leq n$; $S_1 \stackrel{\text{def}}{=}^* S_2$ means that $\text{card}(\{x \mid x \in S_1 \Delta S_2\})$ is finite. D_x denotes the finite set with canonical index x [28]. We sometimes identify finite sets with their canonical indices. We do this when we consider functions or machines which operate on complete knowledge of a finite set (equivalently, an argument which is a canonical index of the finite set), but when we want to display the argument simply as the set itself.

\uparrow denotes undefined. $\max(\cdot), \min(\cdot)$ denote the maximum and minimum of a set, respectively, where $\max(\emptyset) = 0$ and $\min(\emptyset) = \uparrow$.

η ranges over *partial* functions with arguments and values from N . $\eta(x)\downarrow$ denotes that $\eta(x)$ is defined; $\eta(x)\uparrow$ denotes that $\eta(x)$ is undefined.

f, g and F with or without decorations range over *total* functions with arguments and values from N . $\text{domain}(\eta)$ and $\text{range}(\eta)$ denote the domain and range of the function η , respectively.

$\langle i, j \rangle$ stands for an arbitrary, computable, one-to-one encoding of all pairs of natural numbers onto N [28]. Similarly we can define $\langle \cdot, \dots, \cdot \rangle$ for encoding multiple natural numbers onto N .

The quantifiers ‘ \forall^∞ ’, and ‘ \exists^∞ ’, essentially from [3], mean ‘for all but finitely many’ and ‘there exist infinitely many’, respectively. The quantifier ‘ $\exists!$ ’ means ‘there exists a unique’.

φ denotes a fixed *acceptable* programming system for the partial computable functions: $N \rightarrow N$ [27, 28, 15]. φ_i denotes the partial computable function computed by program i in the φ -system. Φ denotes an arbitrary fixed Blum complexity measure [3, 13] for the φ -system. The set of all total recursive functions of one variable is denoted by \mathcal{R} .

W_i denotes $\text{domain}(\varphi_i)$. W_i is, then, the r.e. set/language ($\subseteq N$) accepted (or equivalently, generated) by the φ -program i . \mathcal{E} will denote the set of all r.e. languages. L , with or without decorations, ranges over \mathcal{E} . \bar{L} denotes the complement of L . For language L , we use χ_L to denote the characteristic function of L . \mathcal{L} , with or without decorations, ranges over subsets of \mathcal{E} . $W_i^s \stackrel{\text{def}}{=} \{x \leq s \mid \Phi_i(x) \leq s\}$. $\text{MinGram}(L)$ denotes $\min(\{i \mid W_i = L\})$. $\mathcal{FIN} \stackrel{\text{def}}{=} \{L \mid \text{card}(L) < \infty\}$. $\mathcal{SVT} \stackrel{\text{def}}{=} \{L \mid (\forall x)(\exists!y)[\langle x, y \rangle \in L]\}$.

We sometimes consider partial computable functions with multiple arguments in the φ system. In such cases we implicitly assume that a $\langle \cdot, \dots, \cdot \rangle$ is used to code the arguments, so, for example, $\varphi_i(x, y)$ stands for $\varphi_i(\langle x, y \rangle)$.

¹Decorations are subscripts, superscripts and the like.

2.2 Learning Machines

We now consider language learning machines. Definition 1 below introduces a notion that facilitates discussion about elements of a language being fed to a learning machine.

Definition 1 A *sequence* σ is a mapping from an initial segment of N into $(N \cup \{\#\})$. The *content* of a sequence σ , denoted $\text{content}(\sigma)$, is the set of natural numbers in the range of σ . The *length* of σ , denoted by $|\sigma|$, is the number of elements in σ .

Intuitively, $\#$'s represent pauses in the presentation of data. We let σ and τ , with or without decorations, range over finite sequences. SEQ denotes the set of all finite sequences. The set of all finite sequences of natural numbers and $\#$'s, SEQ, can be coded onto N .

Definition 2 A *language learning machine* is an algorithmic device which computes a mapping from SEQ into N .

Later in Definition 13 and, again in Definition 49, we present variants of the language learning machines from Definition 2 just above. For convenience of exposition we avoid introducing these variants until we need them.

We let \mathbf{M} , with or without decorations, range over learning machines.

2.3 Fundamental Language Identification Paradigms

Definition 3 A *text* T for a language L is a mapping from N into $(N \cup \{\#\})$ such that L is the set of natural numbers in the range of T . The *content* of a text T , denoted $\text{content}(T)$, is the set of natural numbers in the range of T .

Intuitively, a text for a language is an enumeration or sequential presentation of all the objects in the language with the $\#$'s representing pauses in the listing or presentation of such objects. For example, the only text for the empty language is just an infinite sequence of $\#$'s.

We let T , with or without superscripts, range over texts. $T[n]$ denotes the finite initial sequence of T with length n . Hence, $\text{domain}(T[n]) = \{x \mid x < n\}$. For $n \leq |\sigma|$, $\sigma[n]$ denotes the finite initial sequence of σ with length n .

2.3.1 Explanatory Language Identification

In Definition 4 below we spell out what it means for a learning machine on a text to converge in the limit.

Definition 4 Suppose \mathbf{M} is a learning machine and T is a text. $\mathbf{M}(T)\downarrow$ (read: $\mathbf{M}(T)$ *converges*) $\Leftrightarrow (\exists i)(\forall^\infty n) [\mathbf{M}(T[n]) = i]$. If $\mathbf{M}(T)\downarrow$, then $\mathbf{M}(T)$ is defined = the unique i such that $(\forall^\infty n)[\mathbf{M}(T[n]) = i]$; otherwise, we say that $\mathbf{M}(T)$ *diverges* (written: $\mathbf{M}(T)\uparrow$).

We now introduce criteria for a learning machine to be considered *successful* on languages.

Definition 5 [12, 7, 24] Recall that a ranges over $N \cup \{*\}$.

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

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

Gold [12] introduced the criteria we call \mathbf{TxtEx}^0 . The generalization to the $a > 0$ case in Definition 5 was motivated by the observation that humans rarely learn a language perfectly. The $a > 0$ case is from [7], but [24], independently, introduced the $a = *$ case. The influence of Gold’s paradigm [12] to human language learning is discussed by Pinker [26], Wexler and Culicover [31], Wexler [30], and Osherson, Stob, and Weinstein [21, 22, 23].

We sometimes write \mathbf{TxtEx} for \mathbf{TxtEx}^0 including in the names of those learning classes introduced in later sections where ‘ \mathbf{TxtEx}^0 ’ is a proper substring of those names.

Next we define *order independence*, but not in the same way as in [2].

Definition 6 [9, 10] A machine, \mathbf{M} , is said to be *order independent* $\Leftrightarrow (\forall \text{ texts } T, T' \mid \text{content}(T) = \text{content}(T'))[\mathbf{M}(T)\downarrow = i \Leftrightarrow \mathbf{M}(T')\downarrow = i]$.

Theorem 7 [9, 10, 23] *For all a, \mathbf{M} , there exists an order independent machine \mathbf{M}' such that $\mathbf{TxtEx}^a(\mathbf{M}) \subseteq \mathbf{TxtEx}^a(\mathbf{M}')$.*

2.3.2 Behaviorally Correct Language Identification

Definition 8

- (a) \mathbf{M} \mathbf{TxtBc}^a -identifies L (written: $L \in \mathbf{TxtBc}^a(\mathbf{M})$) $\Leftrightarrow (\forall \text{ texts } T \text{ for } L)(\overset{\infty}{\forall} n)[W_{\mathbf{M}(T[n])} =^a L]$.
- (b) $\mathbf{TxtBc}^a = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{TxtBc}^a(\mathbf{M})]\}$.

Definition 8 is from [7]. The $a \in \{0, *\}$ cases were independently introduced in [24, 25].

We sometimes write \mathbf{TxtBc} for \mathbf{TxtBc}^0 including in the names of those learning classes introduced in later sections where ‘ \mathbf{TxtBc}^0 ’ is a proper substring of those names.

2.3.3 Language Learning on Characteristic Function Input

Let $f[n]$ denote the sequence $(\langle 0, f(0) \rangle, \langle 1, f(1) \rangle, \dots, \langle n-1, f(n-1) \rangle)$. We say that $\mathbf{M}(f)\downarrow = i \Leftrightarrow (\overset{\infty}{\forall} n)[\mathbf{M}(f[n]) = i]$.

Definition 9 [7]

- (a) \mathbf{M} \mathbf{ExGen}^a -identifies L (written: $L \in \mathbf{ExGen}^a(\mathbf{M})$) $\Leftrightarrow \mathbf{M}(\chi_L)\downarrow$ and $W_{\mathbf{M}(\chi_L)} =^a L$.
- (b) $\mathbf{ExGen}^a = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{ExGen}^a(\mathbf{M})]\}$.

Definition 10 [7]

- (a) \mathbf{M} \mathbf{BcGen}^a -identifies L (written: $L \in \mathbf{BcGen}^a(\mathbf{M})$) $\Leftrightarrow (\overset{\infty}{\forall} n)[W_{\mathbf{M}(\chi_{L[n]})} =^a L]$.
- (b) $\mathbf{BcGen}^a = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{BcGen}^a(\mathbf{M})]\}$.

2.3.4 Some Basic Results

Theorem 11 *For all n , the following hold.*

- (a) $(\mathbf{TxtEx}^{n+1} \cap \mathcal{P}(\mathcal{SVT})) - \mathbf{TxtEx}^n \neq \emptyset$.
- (b) $(\mathbf{TxtBc}^{n+1} \cap \mathcal{P}(\mathcal{SVT})) - \mathbf{TxtBc}^n \neq \emptyset$.
- (c) $\mathbf{TxtEx}^{2n+1} - \mathbf{TxtBc}^n \neq \emptyset$.
- (d) $(\mathbf{TxtEx}^* \cap \mathcal{P}(\mathcal{SVT})) - \bigcup_n \mathbf{TxtEx}^n \neq \emptyset$.
- (e) $(\mathbf{TxtBc} \cap \mathcal{P}(\mathcal{SVT})) - \mathbf{TxtEx}^* \neq \emptyset$.
- (f) $(\mathbf{TxtBc}^* \cap \mathcal{P}(\mathcal{SVT})) - (\bigcup_n \mathbf{TxtBc}^n \cup \mathbf{TxtEx}^*) \neq \emptyset$.
- (g) $\mathbf{TxtEx}^{2n} \subset \mathbf{TxtBc}^n$.
- (h) $\mathcal{E} \notin \mathbf{TxtBc}^*$.

Parts (a), (b), (d), (e) and (f) of the above theorem can be derived from theorems proved in [7] and [8]. Parts (c), (g) and (h) of the above theorem are directly from [7].

The following result from [7, 24], based on a result from [12], is used below.

Theorem 12 *For each infinite language L , $\mathcal{FIN} \cup \{L\} \notin \mathbf{TxtBc}^*$.*

3 Identification with Finite Negative Information

In this section we consider the effects on learning if an apparently small finite set of negative information is given in addition to text. For this purpose, we introduce a variant of learning machine (called *type 2*).

Definition 13 *A type 2 language learning machine is an algorithmic device which computes a mapping from $\text{SEQ} \times N$ into N .*

Intuitively the second argument is for a canonical index for a finite set of negative information about the language to be learned.² From now on we will drop the phrase *type 2*. Context will show which type of learning machine we have in mind. \mathbf{M} , with or without decorations, will range over both types of learning machine.

Definition 14 We say that $\mathbf{M}(T, i)$ *converges to j* (denoted $\mathbf{M}(T, i) \downarrow = j$) $\Leftrightarrow (\forall^\infty n)[\mathbf{M}(T[n], i) = j]$; we say that $\mathbf{M}(T, i) \uparrow \Leftrightarrow (\exists^\infty n)[\mathbf{M}(T[n], i) \neq \mathbf{M}(T[n+1], i)]$.

²The canonical index is convenient but not essential; see Remark 17 in Section 3.1 below.

3.1 Definitions

Recall from Section 2.1 that we sometimes identify finite sets with their canonical indices. In part (a) of both Definitions 15 and 16 just below, S is the *core* of negative information.

Definition 15

- (a) $\mathbf{M} \text{NegF}^b \text{TtxtEx}^a$ -identifies $L \in \mathcal{E}$ (written: $L \in \mathbf{NegF}^b \text{TtxtEx}^a(\mathbf{M})$) $\Leftrightarrow (\exists S \subseteq \bar{L} \mid \text{card}(S) \leq b)(\forall T \mid T \text{ is a text for } L)[\mathbf{M}(T, S) \downarrow \text{ and } W_{\mathbf{M}(T, S)} =^a L]$.
- (b) $\mathbf{NegF}^b \text{TtxtEx}^a = \{\mathcal{L} \subseteq \mathcal{E} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{NegF}^b \text{TtxtEx}^a(\mathbf{M})]\}$.

Definition 16

- (a) $\mathbf{M} \text{NegF}^b \text{TtxtBc}^a$ -identifies $L \in \mathcal{E}$ (written: $L \in \mathbf{NegF}^b \text{TtxtBc}^a(\mathbf{M})$) $\Leftrightarrow (\exists S \subseteq \bar{L} \mid \text{card}(S) \leq b)(\forall T \mid T \text{ is a text for } L)(\forall n)[W_{\mathbf{M}(T[n], S)} =^a L]$.
- (b) $\mathbf{NegF}^b \text{TtxtBc}^a = \{\mathcal{L} \subseteq \mathcal{E} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{NegF}^b \text{TtxtBc}^a(\mathbf{M})]\}$.

Remark 17 In part (b) of Definitions 15 and 16 above the learning classes defined are extensionally equivalent to those we would obtain if, instead, we fed an r.e. index or a characteristic index of S to the learning device instead of a canonical index. Actually the classes would be extensionally invariant if we merely fed an enumeration of S marked as negative.

We have the following straightforward proposition.

Proposition 18 For all a ,

- (a) $\mathbf{NegF}^0 \text{TtxtEx}^a = \text{TtxtEx}^a$ and
- (b) $\mathbf{NegF}^0 \text{TtxtBc}^a = \text{TtxtBc}^a$.

3.2 Results

The next six theorems illustrate the tremendous learning power obtained already from sets of negative information with cardinality less than or equal two.

Recall that $i^{\text{th}}\text{-min}(S)$ denotes the element x of S , if any, such that $\text{card}(\{y \in S \mid y \leq x\}) = i$.

Theorem 19 $\{L \in \mathcal{E} \mid \bar{L} \text{ is infinite}\} \in \mathbf{NegF}^1 \text{TtxtEx}$.

PROOF. For each $L \in \mathcal{E}$ such that \bar{L} is infinite, define $S_L = \{i_L^{\text{th}}\text{-min}(\bar{L})\}$, where $i_L = \text{MinGram}(L) + 1$.

Define \mathbf{M} as follows.

$$\mathbf{M}(\sigma, S) = \begin{cases} i - 1, & \text{if } (\exists j)[[S = \{j\}] \wedge [\text{card}(\{x \leq j \mid x \notin \text{content}(\sigma)\}) = i]]; \\ 0, & \text{otherwise.} \end{cases}$$

Fix L such that \bar{L} is infinite. Consider any text T for L . It is easy to see that $\mathbf{M}(T, S_L) \downarrow$ and $W_{\mathbf{M}(T, S_L)} = L$. ■

Corollary 20 $\mathcal{SVT} \cap \mathcal{E} \in \mathbf{NegF}^1\mathbf{TxtEx}$.

PROOF. Follows from Theorem 19 and the fact that $\mathcal{SVT} \subset \{L \mid \bar{L} \text{ is infinite}\}$. ■

Theorem 21 $\mathbf{NegF}^1\mathbf{TxtEx} - \mathbf{TxtBc}^* \neq \emptyset$.

PROOF. Let $\mathcal{L} = \{L \mid (\text{card}(\bar{L}) > 0) \text{ and } (\exists x \in \bar{L})[W_x = L]\}$. It is easy to see that $\mathcal{L} \in \mathbf{NegF}^1\mathbf{TxtEx}$. Note that, by a suitably padded recursion theorem [28], $\mathcal{FTN} \subset \mathcal{L}$. Hence, it follows (from Theorem 12) that $\mathcal{L} \notin \mathbf{TxtBc}^*$. ■

Theorem 22 $\mathbf{TxtEx}^1 \subset \mathbf{NegF}^1\mathbf{TxtEx}$.

PROOF. We prove that $\mathbf{TxtEx}^1 \subseteq \mathbf{NegF}^1\mathbf{TxtEx}$. Proper containment then follows from Theorem 21 above. The remainder of the present proof is nonconstructive.

Suppose \mathbf{M} is given. Without loss of generality assume that \mathbf{M} is order independent. For each $L \in \mathbf{TxtEx}^1(\mathbf{M})$, let G_L be the grammar converged to on L by \mathbf{M} . For $L \in \mathbf{TxtEx}^1(\mathbf{M})$, define S_L as follows

$$S_L = \begin{cases} \{m\}, & \text{if } [[L \neq N] \wedge [W_{G_L} \neq L]], \\ & \text{where } m = \min(\bar{L}); \\ \emptyset, & \text{otherwise.} \end{cases}$$

Intuitively, S_L is the negative information that will be supplied to machine \mathbf{M}' in the two cases below. The cardinality and contents of the set S_L together code whether or not language $L = N$, and, whether the final grammar output by \mathbf{M} on any text for L is indeed a grammar for L (Coding tricks of a similar nature will be used in the proofs of Theorems 23 through 25 below).

Let g be a function such that, for all i, x , $W_{g(i,x)} = W_i - \{x\}$. Let f be a function such that, for all i, x , $W_{f(i,x)} = W_i \cup \{x\}$.

Define \mathbf{M}' as follows.

Case 1: \mathbf{M} \mathbf{TxtEx}^1 -identifies N and $W_{G_N} \neq N$.

Let $x \in N - W_{G_N}$. Note that there is a unique such x . Let i_N be a grammar for N . Define \mathbf{M}' as follows.

$$\mathbf{M}'(\sigma, S) = \begin{cases} i_N, & \text{if } [[S = \emptyset] \wedge [\mathbf{M}(\sigma) = G_N] \wedge [x \in \text{content}(\sigma)]]; \\ G_N, & \text{if } [[S = \emptyset] \wedge [\mathbf{M}(\sigma) = G_N] \wedge [x \notin \text{content}(\sigma)]]; \\ \mathbf{M}(\sigma), & \text{if } [[S = \emptyset] \wedge [\mathbf{M}(\sigma) \neq G_N]]; \\ g(\mathbf{M}(\sigma), y), & \text{if } [[S \neq \emptyset] \wedge \\ & [y = \min(\{z \mid z \in \text{content}(\sigma) \Delta W_{\mathbf{M}(\sigma)}^{|\sigma|}\})] \\ & \wedge [y \notin \text{content}(\sigma)]]; \\ f(\mathbf{M}(\sigma), y), & \text{if } [[S \neq \emptyset] \wedge \\ & [y = \min(\{z \mid z \in \text{content}(\sigma) \Delta W_{\mathbf{M}(\sigma)}^{|\sigma|}\})] \\ & \wedge [y \in \text{content}(\sigma)]]; \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that \mathbf{M}' $\mathbf{NegF}^1\mathbf{TxtEx}$ -identifies $\mathbf{TxtEx}^1(\mathbf{M})$.

Case 2: Not Case 1.

Define \mathbf{M}' as follows.

$$\mathbf{M}'(\sigma, S) = \begin{cases} \mathbf{M}(\sigma), & \text{if } S = \emptyset; \\ g(\mathbf{M}(\sigma), y), & \text{if } [[S \neq \emptyset] \wedge \\ & [y = \min(\{z \mid z \in \text{content}(\sigma) \Delta W_{\mathbf{M}(\sigma)}^{|\sigma|})] \\ & \wedge [y \notin \text{content}(\sigma)]]; \\ f(\mathbf{M}(\sigma), y), & \text{if } [[S \neq \emptyset] \wedge \\ & [y = \min(\{z \mid z \in \text{content}(\sigma) \Delta W_{\mathbf{M}(\sigma)}^{|\sigma|})] \\ & \wedge [y \in \text{content}(\sigma)]]; \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that \mathbf{M}' $\text{NegF}^1\text{TtxtEx}$ -identifies $\text{TtxtEx}^1(\mathbf{M})$. ■

Theorem 23 $\mathcal{E} \in \text{NegF}^2\text{TtxtEx}$.

PROOF. For each $L \in \mathcal{E}$, define S_L as follows.

$$S_L = \begin{cases} \{\min(\bar{L}), i^{\text{th}}\text{-min}(\bar{L})\}, & \text{if } [\text{card}(\bar{L}) = \infty] \wedge [i = \min(\{p > 1 \mid W_p = L\})]; \\ \{\max(\bar{L})\}, & \text{if } \bar{L} \text{ is nonempty and finite;} \\ \emptyset, & \text{otherwise.} \end{cases}$$

Intuitively, the cardinality of S_L codes whether the language L equals N , is co-finite and not equal to N , or, is co-infinite. In the last two cases, the content of S_L as defined above enables \mathbf{M} below to compute an index for L .

Let g be a computable function such that for all finite sets S , $W_{g(S)} = N - \{x \mid x \in S\}$.

Define \mathbf{M} as follows.

$$\mathbf{M}(\sigma, S) = \begin{cases} g(\emptyset), & \text{if } S = \emptyset; \\ g(\{x \mid x \leq m \wedge x \notin \text{content}(\sigma)\}), & \text{if } S = \{m\}; \\ k, & \text{if } [\text{card}(S) = 2] \wedge \\ & [k = \text{card}(\{x \leq \max(S) \mid x \notin \text{content}(\sigma)\})] \\ 0, & \text{otherwise.} \end{cases}$$

Fix $L \in \mathcal{E}$. Consider any text T for L . It is easy to see that $\mathbf{M}(T, S_L) \downarrow$ and also that $W_{\mathbf{M}(T, S_L)} = L$. ■

Theorem 24 $\mathcal{E} \in \text{NegF}^1\text{TtxtEx}^1$.

PROOF. For each $L \in \mathcal{E}$, define S_L as follows.

$$S_L = \begin{cases} \{2i^{\text{th}}\text{-min}(\bar{L})\}, & \text{if } [\text{card}(\bar{L}) = \infty] \wedge [i = \min(\{p > 0 \mid W_p = L\})]; \\ \emptyset, & \text{if } L = N; \\ \{\max(\bar{L})\}, & \text{if } \text{card}(\bar{L}) \text{ is finite and odd;} \\ \{\max(\bar{L} - \{\max(\bar{L})\})\}, & \text{otherwise.} \end{cases}$$

Intuitively, the cardinality of S_L codes whether or not the language L equals N . The content of S_L codes, among other things, whether or not \bar{L} is finite or infinite.

Let g be a computable function such that for all finite sets S , $W_{g(S)} = N - \{x \mid x \in S\}$.

Define \mathbf{M} as follows.

$$\mathbf{M}(\sigma, S) = \begin{cases} g(\emptyset), & \text{if } S = \emptyset; \\ \text{card}(\{x \mid x \leq m \wedge x \notin \text{content}(\sigma)\})/2, & \text{if } S = \{m\} \wedge \\ & \text{card}(\{x \mid x \leq m \wedge x \notin \text{content}(\sigma)\}) \text{ is even;} \\ g(\{x \mid x \leq m \wedge x \notin \text{content}(\sigma)\}), & \text{if } S = \{m\} \wedge \\ & \text{card}(\{x \mid x \leq m \wedge x \notin \text{content}(\sigma)\}) \text{ is odd;} \\ 0, & \text{otherwise.} \end{cases}$$

Fix $L \in \mathcal{E}$. Consider any text T for L . It is easy to see that $\mathbf{M}(T, S_L) \downarrow$ and also that $W_{\mathbf{M}(T, S_L)} =^1 L$. ■

Theorem 25 $\mathcal{E} \in \text{NegF}^1\text{TxtBc}$.

PROOF. ³ For each $L \in \mathcal{E}$ define S_L as follows.

$$S_L = \begin{cases} \{i^{\text{th}}\text{-min}(\bar{L})\}, & \text{if } [\text{card}(\bar{L}) = \infty] \wedge [i = \min(\{j > 0 \mid W_j = L\})]; \\ \emptyset, & \text{if } L = N; \\ \{\max(\bar{L})\}, & \text{otherwise.} \end{cases}$$

Intuitively, the cardinality of S_L codes whether or not the language L equals N . The content of S_L as defined above enables \mathbf{M} below to compute an index for L .

Let match be a recursive function such that for each σ and i , $\text{match}(\sigma, i) = \max(\{s \leq |\sigma| \mid W_i^s \subseteq \text{content}(\sigma) \wedge \sigma[s] \subseteq W_i^{|\sigma|}\})$. Let g be a computable function such that for all finite sets S , $W_{g(S)} = N - \{x \mid x \in S\}$. Define \mathbf{M} as follows.

$$\mathbf{M}(T[n], S) = \begin{cases} g(\emptyset), & \text{if } S = \emptyset; \\ e, & \text{if } (\exists m, P)[[S = \{m\}] \wedge [\text{card}(\{x \leq m \mid x \notin \text{content}(T[n])\}) = e] \wedge \\ & [P = \{x \leq m \mid x \notin \text{content}(T[n])\}] \wedge \\ & [\text{match}(T[n], e) > \text{match}(T[n], g(P))]]; \\ g(P), & \text{if } (\exists m, P)[[S = \{m\}] \wedge [\text{card}(\{x \leq m \mid x \notin \text{content}(T[n])\}) = e] \wedge \\ & [P = \{x \leq m \mid x \notin \text{content}(T[n])\}] \wedge \\ & [\text{match}(T[n], e) \leq \text{match}(T[n], g(P))]]; \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that \mathbf{M} $\text{NegF}^1\text{TxtBc}$ -identifies \mathcal{E} . ■

By contrast to Theorem 23 we have the following result.

Theorem 26 $\mathcal{E} \notin \text{NegF}^1\text{TxtEx}$.

³If we extend the definition of TxtFex_b^g in [5, 6] in the obvious way to define $\text{NegF}^c\text{TxtFex}_b^g$, then our proof actually shows that $\mathcal{E} \in \text{NegF}^1\text{TxtFex}_b^g$.

PROOF. Suppose by way of contradiction that $\mathcal{E} \subseteq \mathbf{NegF}^1\mathbf{TxtEx}(\mathbf{M})$. We now describe a construction which will give us some recursively enumerable languages at least one of which is not in $\mathbf{NegF}^1\mathbf{TxtEx}(\mathbf{M})$. We especially concentrate our construction on one of the languages, called L , below; the other languages are the L_i 's also described below.

The construction maintains an infinite array, neg , with elements from $\{1, 2, 3, \dots\}$. neg_i^s is the value stored in neg_i , the i^{th} element of this array, just before the start of stage s . neg_i^s is a non-decreasing function of s . We let $\text{neg}_i^\infty = \lim_{s \rightarrow \infty} \text{neg}_i^s$. L is $(N - \{\text{neg}_i^\infty \mid \text{neg}_i^\infty < \infty\})$. We also maintain another array S , where, for all s , $S_0^s = \emptyset$ and, for all $i > 0$, $S_i^s = \{\text{neg}_i^s\}$. Lastly we also maintain an array σ with elements from SEQ . σ_i^s is the value stored in σ_i , the i^{th} element of this array, just before the start of stage s .

If $(\forall s)[S_i^s = S_i^{s+1}]$, then we let $S_i^\infty = \lim_{s \rightarrow \infty} S_i^s$; otherwise S_i^∞ is undefined. Similarly, if $(\forall s)[\sigma_i^s = \sigma_i^{s+1}]$, then we let $\sigma_i^\infty = \lim_{s \rightarrow \infty} \sigma_i^s$; otherwise σ_i^∞ is undefined. Intuitively, the sets S_i^∞ are the possible finite cores of information regarding \bar{L} that could be used by \mathbf{M} . $L_i = N - \bigcup_{j \leq i} S_j^\infty$, and σ_i^∞ , when finite, functions as a locking sequence [23] for \mathbf{M} on L_i and negative information $\{\text{neg}_i^\infty\}$.

For each s and i , it will be the case that,

- (a) $\sigma_i^s \subset \sigma_{i+1}^s$,
- (b) $\text{neg}_{i+1}^s > \max(\text{content}(\sigma_i^s))$,
- (c) $(\{x \mid x \leq \text{neg}_i^s\} - \{\text{neg}_j^s \mid 1 \leq j \leq i\}) \subseteq \text{content}(\sigma_i^s)$,
- (d) $\{\text{neg}_j^s \mid 1 \leq j \leq i\} \cap \text{content}(\sigma_i^s) = \emptyset$.

For each $i > 0$, let $\text{neg}_i = 2i$. Let $\sigma_0 = \emptyset$. For $i > 0$, let σ_i be an extension of σ_{i-1} such that $\text{content}(\sigma_i) = \{x \mid x \leq \text{neg}_i\} - \{\text{neg}_j \mid 1 \leq j \leq i\}$. Go to stage 0.

Begin stage s

(* The inequality ' $\tau < s$ ' just below means that for some fixed canonical indexing, the canonical index of τ is less than s . *)

1. Let $P = \{\langle i, \tau \rangle \mid i < s \wedge \tau < s \wedge \sigma_i^s \subseteq \tau \wedge \text{content}(\tau) \subseteq N - \bigcup_{j \leq i} S_j^s \wedge \mathbf{M}(\tau, S_i^s) \neq \mathbf{M}(\sigma_i^s, S_i^s)\}$.
2. **if** $P \neq \emptyset$ **then**
 - 2.1 Let $i_0 = \min(\{i \mid (\exists \tau)[\langle i, \tau \rangle \in P]\})$.
 - 2.2 Let τ_0 be such that $\langle i_0, \tau_0 \rangle \in P$.
 - 2.3 Let σ_{i_0} be an extension of τ_0 such that $\text{content}(\sigma_{i_0}) = \{x \mid x \leq \max(\text{content}(\tau_0)) + 1\} - \{\text{neg}_j^s \mid 1 \leq j \leq i_0\}$.
 - 2.4 For each $j > i_0$, let $\text{neg}_j = 2j + \max(\text{content}(\tau_0)) + 1$ and σ_j be an extension of σ_{j-1} such that $\text{content}(\sigma_j) = \{x \mid x \leq 2j + 1 + \max(\text{content}(\tau_0))\} - \{\text{neg}_k^s \mid 1 \leq k \leq j\}$.

endif

3. Go to stage $s + 1$.

End stage s

Now we consider the following cases.

Case 1: $(\forall i)[\sigma_i^\infty \text{ is defined}]$.

Since for all i, s , if $S_i^s \neq S_i^{s+1}$, then $\sigma_i^s \neq \sigma_i^{s+1}$, we have that for all i , S_i^∞ is defined.

Let $L = N - \bigcup_i S_i^\infty$. Note that $T = \bigcup_i \sigma_i^\infty$ is a text for L . For each i , define $L_i = N - \bigcup_{j \leq i} S_j^\infty$. Note that, for all i , for all τ such that $\sigma_i^\infty \subseteq \tau$ and $\text{content}(\tau) \subseteq L_i$, $\mathbf{M}(\tau, S_i^\infty) = \mathbf{M}(\sigma_i^\infty, S_i^\infty)$.

Now for each $i < j$, the following three assertions hold.

- (a) $\text{neg}_j^\infty > \max(\text{content}(\sigma_i^\infty))$,
- (b) $L \subset L_j \subset L_i$, and
- (c) $\sigma_i^\infty \subset \sigma_j^\infty \subset T$.

For each i , let T_i be a text for L_i such that $\sigma_i^\infty \subset T_i$. It is clear that $\mathbf{M}(T_i, S_i^\infty) = \mathbf{M}(\sigma_i^\infty, S_i^\infty)$.

For each j , let S'_j be such that,

$$\text{card}(S'_j) \leq 1, S'_j \subseteq \overline{L}_j, \mathbf{M}(T_j, S'_j) \downarrow, \text{ and } W_{\mathbf{M}(T_j, S'_j)} = L_j. \quad (1)$$

(if such an S'_j does not exist then \mathbf{M} does not $\mathbf{NegF}^1\mathbf{TxtEx}$ -identify L_j).

Claim 27 For all j , $S'_j = S_j^\infty$.

PROOF. Suppose by way of contradiction otherwise. Let j be the least number such that $S'_j \neq S_j^\infty$. Since $S_0 = S'_0 = \emptyset$, we have that $j > 0$. Let $i < j$ be such that $S_i^\infty = S_j^\infty$ (clearly, such an i exists). Now $\mathbf{M}(T_i, S_i^\infty) = \mathbf{M}(\sigma_i^\infty, S_i^\infty) = \mathbf{M}(T_j, S_j^\infty)$ (since $S_i^\infty = S_j^\infty$, $\sigma_i^\infty \subseteq T_i$, $\sigma_i^\infty \subseteq \sigma_j^\infty \subseteq T_j$, and for all τ such that $\sigma_i^\infty \subseteq \tau$ and $\text{content}(\tau) \subseteq L_i$, $\mathbf{M}(\tau, S_i^\infty) = \mathbf{M}(\sigma_i^\infty, S_i^\infty)$). It follows that $\{L_i, L_j\} \not\subseteq \mathbf{NegF}^1\mathbf{TxtEx}(\mathbf{M})$ contradicting the assumption that $\mathcal{E} \subseteq \mathbf{NegF}^1\mathbf{TxtEx}(\mathbf{M})$. ■ (Claim 27)

Now suppose S' is such that $S' \subseteq \overline{L} \wedge \mathbf{M}(T, S') \downarrow \wedge W_{\mathbf{M}(T, S')} = L$. Let i be such that $S_i^\infty = S'$ (clearly, such an i exists, since $S_0^\infty = \emptyset$ and for each $x \in \overline{L}$ there exists a j such that $S_j^\infty = \{x\}$). Now $\mathbf{M}(T_i, S_i^\infty) = \mathbf{M}(\sigma_i^\infty, S_i^\infty) = \mathbf{M}(T, S')$ (since $S_i^\infty = S'$ and for all τ such that $\sigma_i^\infty \subseteq \tau \subset T$ and $\text{content}(\tau) \subseteq L \subseteq L_i$, $\mathbf{M}(\tau, S_i^\infty) = \mathbf{M}(\sigma_i^\infty, S_i^\infty)$). It follows that $\{L, L_i\} \not\subseteq \mathbf{NegF}^1\mathbf{TxtEx}(\mathbf{M})$.

Therefore $\mathcal{E} \not\subseteq \mathbf{NegF}^1\mathbf{TxtEx}(\mathbf{M})$.

Case 2: Not *Case 1*.

Let j_0 denote the least j such that the value of σ_j^∞ is undefined.

Case 2.1: $j_0 = 0$.

In this case, let T be such that for all s , $\sigma_0^s \subseteq T$. It is easy to see that $\mathbf{M}(T, \emptyset) \uparrow$.

Case 2.2: $j_0 > 0$.

For $j \leq j_0$, let $L_j = N - \bigcup_{k \leq j} S_k^\infty$. For $j < j_0$, let T_j be an arbitrary text for L_j such that $\sigma_j^\infty \subset T_j$. Let s' be such that for all $s \geq s'$, for each $j < j_0$, $\sigma_j^s = \sigma_j^{s+1}$. Let T_{j_0} be such that for all $s > s'$, $\sigma_{j_0}^s \subseteq T_{j_0}$.

As in *Case 1*, it can be argued that for all $j \leq j_0$, if $S'_j \subseteq L_j$, $\mathbf{M}(T_j, S'_j) \downarrow$ and $W_{\mathbf{M}(T_j, S'_j)} = L_j$, then $S'_j = S_j^\infty$. But, $\mathbf{M}(T_{j_0}, S_{j_0}^\infty) \uparrow$. Thus $L_{j_0} \notin \mathbf{NegF}^1\mathbf{TxtEx}(\mathbf{M})$.

From the above cases it follows that $\mathcal{E} \not\subseteq \mathbf{NegF}^1\mathbf{TxtEx}(\mathbf{M})$. ■ (Theorem 26)

4 Identification with Open Negative Information

In this section we introduce a different way of presenting some negative information to learning machines. Here the negative information is supplied in a manner reminding one of the basic open sets for the topology with respect to which enumeration operators are continuous. This is the first topology described in [28, Exercise 11-35, page 217]. The basic definitions in this section were suggested to us in part by those in [20] and those in Section 3 above. Basically, in this section, we allow the possibility of more negative information being supplied in addition to the finite cores of negative information; whereas, in Section 3 we considered supplying only the finite cores.

4.1 Definitions

For a segment σ , let $\text{PosInfo}(\sigma) = \{x \mid 2x \in \text{content}(\sigma)\}$, and $\text{NegInfo}(\sigma) = \{x \mid 2x + 1 \in \text{content}(\sigma)\}$. The functions $\text{PosInfo}()$ and $\text{NegInfo}()$ are defined to enable the provision of both positive and negative information, respectively, as part of a single text. For a text T , let $\text{PosInfo}(T) = \{x \mid 2x \in \text{content}(T)\}$, and $\text{NegInfo}(T) = \{x \mid 2x + 1 \in \text{content}(T)\}$. Thus, T is a text for $\text{PosInfo}(T) \oplus \text{NegInfo}(T)$.

Definition 28

- (a) $\mathbf{M} \mathbf{NegO}^b\mathbf{TxtEx}^a$ -identifies $L \in \mathcal{E}$ (written: $L \in \mathbf{NegO}^b\mathbf{TxtEx}^a(\mathbf{M})$) $\Leftrightarrow (\exists S \subseteq \bar{L} \mid \text{card}(S) \leq b)(\forall L' \mid S \subseteq L' \subseteq \bar{L})(\forall T \mid \text{content}(T) = L \oplus L')[\mathbf{M}(T) \downarrow \text{ and } W_{\mathbf{M}(T)} =^a L]$.
- (b) $\mathbf{NegO}^b\mathbf{TxtEx}^a = \{\mathcal{L} \subseteq \mathcal{E} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{NegO}^b\mathbf{TxtEx}^a(\mathbf{M})]\}$.

Definition 28 just above contrasts interestingly with Definition 15 (from Section 3). For each definition there need only *exist* a S , but for Definition 28 this S must satisfy the strong constraint that it work for *all* L' such that $S \subseteq L' \subseteq \bar{L}$. Similar remarks apply to the next definition and Definition 16 in Section 3.

Definition 29

- (a) $\mathbf{M} \mathbf{NegO}^b\mathbf{TxtBc}^a$ -identifies $L \in \mathcal{E}$ (written: $L \in \mathbf{NegO}^b\mathbf{TxtBc}^a(\mathbf{M})$) $\Leftrightarrow (\exists S \subseteq \bar{L} \mid \text{card}(S) \leq b)(\forall L' \mid S \subseteq L' \subseteq \bar{L})(\forall T \mid \text{content}(T) = L \oplus L')(\forall n)[W_{\mathbf{M}(T[n])} =^a L]$.
- (b) $\mathbf{NegO}^b\mathbf{TxtBc}^a = \{\mathcal{L} \subseteq \mathcal{E} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{NegO}^b\mathbf{TxtBc}^a(\mathbf{M})]\}$.

A quantificational variant of our $\mathbf{NegO}^b\mathbf{TxtEx}^0$ -identification (from Definition 28 above) is quite close to \mathbf{PP}^b -identification from Section 5.3 of [29]: for \mathbf{PP}^b -identification different finite sets can be used for different texts for the same language (and the finite sets are required to be of size at least b). We can show there are language classes $\mathbf{NegO}^b\mathbf{TxtEx}^0$ -identifiable but not \mathbf{PP}^b -identifiable with \mathbf{PP}^b -identification generalized to the \mathbf{TxtBc}^* case (and even if the finite sets are required to be of size less than or equal to b).

We have the following straightforward proposition.

Proposition 30 *For all a ,*

(a) $\mathbf{NegO}^0\mathbf{TxtEx}^a = \mathbf{TxtEx}^a$ and

(b) $\mathbf{NegO}^0\mathbf{TxtBc}^a = \mathbf{TxtBc}^a$.

4.2 Results

Proposition 31 *For all a , $\mathbf{NegO}^*\mathbf{TxtEx}^a \cap \mathcal{P}(\mathcal{SVT}) = \mathbf{TxtEx}^a \cap \mathcal{P}(\mathcal{SVT})$.*

PROOF. It is easy to see that for each $L \in \mathcal{SVT}$, a text for $L \oplus \bar{L}$ can be effectively obtained from a text for L . The Proposition follows. \blacksquare

Similarly,

Proposition 32 *For all a , $\mathbf{NegO}^*\mathbf{TxtBc}^a \cap \mathcal{P}(\mathcal{SVT}) = \mathbf{TxtBc}^a \cap \mathcal{P}(\mathcal{SVT})$.*

The following two theorems provide classes of languages which can be learned with $n + 1$ mistakes, but not with n , no matter how much open negative information is provided in the n mistake case. The mechanism partly responsible is that the gap left by the possible extra anomaly can be greater in information content than the information provided by open negative information.

Theorem 33 *For all n , $\mathbf{TxtEx}^{n+1} - \mathbf{NegO}^*\mathbf{TxtEx}^n \neq \emptyset$.*

PROOF. Follows using Proposition 31 and Theorem 11. \blacksquare

Theorems 34 and 35 can be shown similarly.

Theorem 34 *For all n , $\mathbf{TxtBc}^{n+1} - \mathbf{NegO}^*\mathbf{TxtBc}^n \neq \emptyset$.*

Our class witnessing Theorem 34 immediately above is also not in $\mathbf{NegO}^*\mathbf{TxtEx}^*$. Additionally, we have the following.

Theorem 35 $\mathbf{TxtBc} - \mathbf{NegO}^*\mathbf{TxtEx}^* \neq \emptyset$.

The separation results in Theorems 34 and 35 are witnessed by subclasses of \mathcal{SVT} . Thus, by Corollary 20, we have the following corollary.

Corollary 36 *For all n ,*

(a) $\mathbf{NegF}^1\mathbf{TxtEx} - \mathbf{NegO}^*\mathbf{TxtBc}^n \neq \emptyset$ and

(b) $\mathbf{NegF}^1\mathbf{TxtEx} - \mathbf{NegO}^*\mathbf{TxtEx}^* \neq \emptyset$.

The following theorem presents another *interesting* connection between some \mathbf{NegF} and \mathbf{NegO} learning criteria.

Theorem 37 $\mathbf{NegO}^*\mathbf{TxtEx} \subset \mathbf{NegF}^1\mathbf{TxtEx}$.

PROOF. We prove that $\mathbf{NegO}^*\mathbf{TxtEx} \subseteq \mathbf{NegF}^1\mathbf{TxtEx}$. Proper containment follows from part (b) of Corollary 36 above. Suppose \mathbf{M} $\mathbf{NegO}^*\mathbf{TxtEx}$ -identifies \mathcal{L} . Fix an arbitrary $L \in \mathcal{L}$. Let $S_L \subseteq \bar{L}$ be the finite set with smallest canonical index such that $(\forall L' \mid S_L \subseteq L' \subseteq \bar{L})(\forall T \mid \text{content}(T) = L \oplus L')[\mathbf{M}(T) \downarrow \text{ and } W_{\mathbf{M}(T)} = L]$. Such a S_L clearly exists. Let $S'_L = \{\max(S_L)\}$, if $S_L \neq \emptyset$; $S'_L = \emptyset$ otherwise.

Let F be a computable mapping from $\text{SEQ} \times \mathcal{FIN}$ to SEQ such that, for all $S \in \mathcal{FIN}$ and for all $\sigma \subset \tau$ $[[F(\sigma, S) \subseteq F(\tau, S)] \wedge [\text{content}(\sigma) = \text{PosInfo}(F(\sigma, S))] \wedge [S = \text{NegInfo}(F(\sigma, S))]]$.

Now define \mathbf{M}' as follows.

$$\mathbf{M}'(\sigma, S) = \begin{cases} \mathbf{M}(F(\sigma, S)), & \text{if } [S = \emptyset]; \\ \mathbf{M}(F(\sigma, X)), & \text{if } [[S = \{m\}] \wedge [X = \{x \leq m \mid x \notin \text{content}(\sigma)]]]. \end{cases}$$

Consider any text T for L . It is easy to see that $\mathbf{M}'(T, S'_L) \downarrow$ and $W_{\mathbf{M}'(T, S'_L)} = L$. ■

Theorem 38 $\mathcal{E} \in \mathbf{NegO}^*\mathbf{TxtBc}^*$.

PROOF. For each $L \in \mathcal{E}$, let $S_L = \{x \mid (\exists i < \text{MinGram}(L))[x = \min(W_i \cap \bar{L})]\}$. Define \mathbf{M} as follows. $\mathbf{M}(T[n]) = i$ such that, $W_i = \bigcup_s W_{\min(\{n\} \cup \{j \mid \text{PosInfo}(T[n]) \subseteq W_j^s \wedge \text{NegInfo}(T[n]) \subseteq \bar{W}_j^s\})}$.

For all $L \in \mathcal{E}$, for all L' such that $S_L \subseteq L' \subseteq \bar{L}$, let T be a text such that $\text{content}(T) = L \oplus L'$. Let n_0 be so large that,

- (i) $\text{MinGram}(L) < n_0$,
- (ii) $S_L \subseteq \text{NegInfo}(\text{content}(T[n_0]))$.
- (iii) $(\forall i < \text{MinGram}(L) \mid W_i \subseteq L)[\text{PosInfo}(T[n_0]) \not\subseteq W_i]$.
- (iv) $(\forall i < \text{MinGram}(L) \mid W_i \not\subseteq L)[S_L \cap W_i^{n_0} \neq \emptyset]$.

Now for all $n > n_0$, $\min(\{n\} \cup \{j \mid \text{PosInfo}(T[n]) \subseteq W_j \wedge \text{NegInfo}(T[n]) \subseteq \bar{W}_j\}) \geq \text{MinGram}(L)$. Thus, for all $n > n_0$, for all but finitely many s , $\min(\{n\} \cup \{j \mid \text{PosInfo}(T[n]) \subseteq W_j^s \wedge \text{NegInfo}(T[n]) \subseteq \bar{W}_j^s\}) = \text{MinGram}(L)$ (since $\text{PosInfo}(T) = L$ and $S_L \subseteq \text{NegInfo}(T) \subseteq \bar{L}$).

It follows that, for all $n > n_0$, $W_{\mathbf{M}(T[n])} =^* L$. ■

The next three theorems contrast nicely with Theorems 33 and 34 above. They provide classes of languages which can be learned with $n + 1$ pieces of open negative information, but not with n , no matter how many anomalies are permitted in the n piece case. The mechanism partly responsible is that the extra possible negative information can be greater in information content than the information that may be omitted by the anomalies.

Theorem 39 $\mathbf{NegO}^1\mathbf{TxtEx} - \mathbf{NegO}^0\mathbf{TxtBc}^* \neq \emptyset$.

PROOF. Let $\mathcal{L}_0 = \mathcal{FIN} \cup \{N\}$. It is easy to see $\mathcal{L}_0 \in \mathbf{NegO}^1\mathbf{TxtEx}$. By Theorem 12 and Proposition 30, $\mathcal{L}_0 \notin \mathbf{NegO}^0\mathbf{TxtBc}^*$. \blacksquare

Theorem 40 For all n , $\mathbf{NegO}^{n+1}\mathbf{TxtEx} - \mathbf{NegO}^n\mathbf{TxtEx}^* \neq \emptyset$.

PROOF. The $n = 0$ case follows from Theorem 39 above.
Suppose $n > 0$.

We say that L is n -nice iff the following four conditions hold.

- a. $(\forall x > n)(\forall y)[\langle x, y \rangle \in L]$.
- b. $(\exists y)[\langle 0, y \rangle \notin L]$.
- c. $[(\exists y)(\forall x \leq n)[\langle x, y \rangle \notin L]] \Rightarrow W_{\min(\{y | (\forall x \leq n)[\langle x, y \rangle \notin L\})} = L$.
- d. $[\neg(\exists y)(\forall x \leq n)[\langle x, y \rangle \notin L]] \Rightarrow W_{\min(\{y | \langle 0, y \rangle \notin L\})} = L$.

Let $\mathcal{L}_n = \{L \mid L \text{ is } n\text{-nice}\}$.

Claim 41 For each n , $\mathcal{L}_n \in \mathbf{NegO}^{n+1}\mathbf{TxtEx}$.

PROOF. Fix n . For each $L \in \mathcal{L}_n$, define S_L as follows.

$$S_L = \begin{cases} \{\langle 0, x \rangle \mid x = \min(\{z \mid \langle 0, z \rangle \notin L\})\}, & \text{if } [\neg(\exists y)(\forall x \leq n)[\langle x, y \rangle \notin L]]; \\ \{\langle x, w \rangle \mid x \leq n \wedge w = \min(\{y \mid (\forall z \leq n)[\langle z, y \rangle \notin L\})\}, & \text{otherwise.} \end{cases}$$

Consider learning machine \mathbf{M} defined as follows.

$$\mathbf{M}(\sigma) = \begin{cases} \min(\{y \mid (\forall x \leq n)[\langle x, y \rangle \in \mathbf{NegInfo}(\sigma)]\}), & \text{if } (\exists y)(\forall x \leq n)[\langle x, y \rangle \in \mathbf{NegInfo}(\sigma)]; \\ \min(\{y \mid \langle 0, y \rangle \in \mathbf{NegInfo}(\sigma)\}), & \text{otherwise.} \end{cases}$$

It is easy to see that $\mathcal{L}_n \in \mathbf{NegO}^{n+1}\mathbf{TxtEx}(\mathbf{M})$. \blacksquare (Claim 41)

We show that $\mathcal{L}_1 \notin \mathbf{NegO}^1\mathbf{TxtEx}^*$. This proof can be generalized to show that $\mathcal{L}_n \notin \mathbf{NegO}^n\mathbf{TxtEx}^*$.

Suppose by way of contradiction that \mathbf{M} $\mathbf{NegO}^1\mathbf{TxtEx}^*$ -identifies \mathcal{L}_1 . Then, by the Kleene recursion theorem [28, Page 214], there exists an e such that W_e can be defined in stages as follows.

Enumerate $\{\langle 0, x \rangle, \langle 1, x \rangle \mid x < e\} \cup \{1, e\}$ into W_e and let σ^0 be the lexicographically least segment such that $\mathbf{PosInfo}(\sigma^0) = \{\langle 0, x \rangle, \langle 1, x \rangle \mid x < e\} \cup \{1, e\}$ and $\mathbf{NegInfo}(\sigma^0) = \{\langle 0, e \rangle\}$.

Go to stage 0.

Begin stage s

1. Search for σ extending σ^s such that the following four conditions are satisfied. If and when such a σ is found, go to step 2.
 - (a) $\neg(\exists x)[\{\langle 0, x \rangle, \langle 1, x \rangle\} \subseteq \mathbf{NegInfo}(\sigma)]$,
 - (b) $\mathbf{PosInfo}(\sigma) \cap \mathbf{NegInfo}(\sigma) = \emptyset$,
 - (c) $\mathbf{NegInfo}(\sigma) \subseteq \{\langle 0, x \rangle, \langle 1, x \rangle \mid x \in N\}$, and
 - (d) $\mathbf{M}(\sigma) \neq \mathbf{M}(\sigma^s)$.

2. Let σ be as found in step 1.

Let $m = 1 + \max(\{x \mid \langle 1, x \rangle \in \text{PosInfo}(\sigma) \cup \text{NegInfo}(\sigma)\} \vee \{\langle 0, x \rangle \in \text{PosInfo}(\sigma) \cup \text{NegInfo}(\sigma)\})$.

Let S be

$$\begin{aligned} & (\{\langle x, y \rangle \mid 2 \leq x \leq s \wedge y \leq s\} \cup \text{PosInfo}(\sigma) \cup \\ & \{\langle 0, x \rangle \mid x \leq m \wedge \langle 0, x \rangle \notin \text{NegInfo}(\sigma)\} \cup \{\langle 1, x \rangle \mid x \leq m \wedge \langle 1, x \rangle \notin \text{NegInfo}(\sigma)\}). \end{aligned}$$

Enumerate S in W_e .

Let σ^{s+1} be an extension of σ such that

- (a) $\text{PosInfo}(\sigma^{s+1}) = S$, and,
- (b) $\text{NegInfo}(\sigma^{s+1}) = \{\langle 0, x \rangle, \langle 1, x \rangle \mid x \leq m\} - S$.

Go to stage $s + 1$.

End stage s .

This completes the definition of W_e . Now we consider the following cases.

Case 1: Infinitely many stages are executed.

In this case, let $L = W_e \in \mathcal{L}_1$. Let $T = \bigcup_{s \in \mathbb{N}} \sigma^s$. Clearly, $\text{PosInfo}(T) = L$ and $\text{NegInfo}(T) = \bar{L}$. However, $\mathbf{M}(T) \uparrow$.

Case 2: Stage s starts but does not terminate.

Let $m = \max(\{x \mid \langle 1, x \rangle \in \text{PosInfo}(\sigma^s) \cup \text{NegInfo}(\sigma^s) \vee \langle 0, x \rangle \in \text{PosInfo}(\sigma^s) \cup \text{NegInfo}(\sigma^s)\})$. Let

$$\mathcal{L}' = \{L \mid \text{PosInfo}(\sigma^s) \subseteq L \wedge \text{NegInfo}(\sigma^s) \subseteq \bar{L} \wedge \{\langle x, y \rangle \mid x \geq 2\} \subseteq L\}. \quad (2)$$

Claim 42 *For each $L \in \mathcal{L}'$, and for each $w \in \bar{L}$, there exists a text T satisfying $\sigma^s \subseteq T$, $L = \text{PosInfo}(T)$, $w \in \text{NegInfo}(T)$ and $\text{PosInfo}(T) \cap \text{NegInfo}(T) = \emptyset$, for which $\mathbf{M}(T) = \mathbf{M}(\sigma^s)$.*

PROOF. Otherwise step 1, in stage s , would succeed. ■ (Claim 42)

By suitably padded applications of the Kleene recursion theorem, there exist e_1 and e_2 (each greater than m) such that,

$$W_{e_1} = \text{PosInfo}(\sigma^s) \cup \{\langle x, y \rangle \mid x \geq 2\} \cup \{\langle 0, x \rangle \mid x > m \wedge x \neq e_1\} \quad (3)$$

and

$$W_{e_2} = \text{PosInfo}(\sigma^s) \cup \{\langle x, y \rangle \mid x \geq 2\} \cup \{\langle 1, x \rangle \mid x > m \wedge x \neq e_2\}. \quad (4)$$

It is easy to see that $\{W_{e_1}, W_{e_2}\} \subseteq \mathcal{L}_1 \cap \mathcal{L}'$, and $W_{e_1} \neq^* W_{e_2}$.

We claim that $\{W_{e_1}, W_{e_2}\} \not\subseteq \mathbf{NegO}^1\mathbf{TxtEx}^*(\mathbf{M})$. Suppose by way of contradiction that $S_1 \subseteq W_{e_1}$ and $S_2 \subseteq W_{e_2}$ are sets of cardinality at most one, such that,

$$\begin{aligned} & (\forall T \mid [W_{e_1} = \text{PosInfo}(T)] \wedge [S_1 \subseteq \text{NegInfo}(T)] \wedge [\text{PosInfo}(T) \cap \text{NegInfo}(T) = \emptyset]) \\ & [W_{\mathbf{M}(T)} =^* W_{e_1}] \end{aligned} \quad (5)$$

and

$$(\forall T \mid [W_{e_2} = \text{PosInfo}(T)] \wedge [S_2 \subseteq \text{NegInfo}(T)] \wedge [\text{PosInfo}(T) \cap \text{NegInfo}(T) = \emptyset])$$

$$[W_{\mathbf{M}(T)} =^* W_{e_2}]. \quad (6)$$

Without loss of generality assume $\text{card}(S_1) = \text{card}(S_2) = 1$. Let w_1, w_2 be such that $S_1 = \{w_1\}$ and $S_2 = \{w_2\}$. Let T_1 and T_2 be the texts, as claimed in Claim 42, for $L = W_{e_1}, w = w_1$ and $L = W_{e_2}, w = w_2$ respectively.

Now $W_{e_1} =^* W_{\mathbf{M}(T_1)} = W_{\mathbf{M}(\sigma^s)} = W_{\mathbf{M}(T_2)} =^* W_{e_2}$. But this is not possible, since $W_{e_1} \neq^* W_{e_2}$. It follows that $\{W_{e_1}, W_{e_2}\} \not\subseteq \text{NegO}^1\text{TxtEx}^*(\mathbf{M})$.

From the above cases, it follows that $\mathcal{L}_1 \not\subseteq \text{NegO}^1\text{TxtEx}^*(\mathbf{M})$. ■ (Theorem 40)

The language classes which witness the previous theorem also witness the following theorem.

Theorem 43 For all n , $\text{NegO}^{n+1}\text{TxtEx} - \bigcup_j \text{NegO}^n\text{TxtBc}^j \neq \emptyset$.

PROOF. The $n = 0$ case follows from Theorem 39 above.

Suppose $n > 0$. Fix j . Let \mathcal{L}_n be as defined in the proof of Theorem 40. We show that $\mathcal{L}_1 \not\subseteq \text{NegO}^1\text{TxtBc}^j$. This proof can be generalized to show that $\mathcal{L}_n \not\subseteq \text{NegO}^n\text{TxtBc}^j$. This proof is similar to the proof of Theorem 40.

Suppose by way of contradiction that $\mathbf{M}\text{NegO}^1\text{TxtBc}^j$ -identifies \mathcal{L}_1 . Then, by the Kleene recursion theorem, there exists an e such that W_e can be defined in stages as follows.

Enumerate $\{\langle 0, x \rangle, \langle 1, x \rangle \mid x < e\} \cup \{1, e\}$ in W_e and let σ^0 be the lexicographically least segment such that $\text{PosInfo}(\sigma^0) = \{\langle 0, x \rangle, \langle 1, x \rangle \mid x < e\} \cup \{1, e\}$ and $\text{NegInfo}(\sigma^0) = \{\langle 0, e \rangle\}$.

Go to stage 0.

Begin stage s

1. Search for σ extending σ^s and P such that the following seven conditions are satisfied. If and when such a σ is found go to step 2.

- (a) $\neg(\exists x)[\{\langle 0, x \rangle, \langle 1, x \rangle\} \subseteq \text{NegInfo}(\sigma)]$,
- (b) $\text{PosInfo}(\sigma) \cap \text{NegInfo}(\sigma) = \emptyset$,
- (c) $\text{NegInfo}(\sigma) \subseteq \{\langle 0, x \rangle, \langle 1, x \rangle \mid x \in N\}$,
- (d) $P \subseteq \{\langle 0, x \rangle \mid x \in N\}$,
- (e) $\text{card}(P) = j + 1$,
- (f) $(P \cup \{\langle 1, x \rangle \mid \langle 0, x \rangle \in P\}) \cap (\text{PosInfo}(\sigma) \cup \text{NegInfo}(\sigma)) = \emptyset$, and
- (g) $P \subseteq W_{\mathbf{M}(\sigma)}$.

2. Let σ, P be as found in step 1.

Let $P' = \{\langle 1, x \rangle \mid \langle 0, x \rangle \in P\}$.

Let $m = 1 + \max(\{x \mid [\langle 1, x \rangle \in \text{PosInfo}(\sigma) \cup \text{NegInfo}(\sigma)] \vee [\langle 0, x \rangle \in \text{PosInfo}(\sigma) \cup \text{NegInfo}(\sigma) \cup P]\})$.

Let S be

$$\begin{aligned} & (\{\langle x, y \rangle \mid 2 \leq x \leq s \wedge y \leq s\} \cup \text{PosInfo}(\sigma) \cup \\ & \{\langle 0, x \rangle \mid x \leq m \wedge \langle 0, x \rangle \notin \text{NegInfo}(\sigma) \cup P\} \cup \{\langle 1, x \rangle \mid x \leq m \wedge \langle 1, x \rangle \notin \text{NegInfo}(\sigma)\}). \end{aligned}$$

Enumerate S in W_e .

Let σ^{s+1} be an extension of σ such that

- (a) $\text{PosInfo}(\sigma^{s+1}) = S$, and,
- (b) $\text{NegInfo}(\sigma^{s+1}) = \{\langle 0, x \rangle, \langle 1, x \rangle \mid x \leq m\} - S$.

(* Note that $W_{\mathbf{M}(\sigma)} \supseteq P$ and $P \subseteq \text{NegInfo}(\sigma^{s+1})$. *)

Go to stage $s + 1$.

End stage s .

This completes the definition of W_e . Now we consider the following cases.

Case 1: Infinitely many stages are executed.

In this case, let $L = W_e \in \mathcal{L}_1$. Let $T = \bigcup_s \sigma^s$. Clearly, T is a text for $L \oplus \bar{L}$. However, for infinitely many n , $W_{\mathbf{M}(T[n])} \neq^j L$ (due to the success of step 1 at each stage).

Case 2: Stage s starts but does not terminate.

Let $m = \max(\{x \mid \langle 1, x \rangle \in \text{PosInfo}(\sigma^s) \cup \text{NegInfo}(\sigma^s) \vee \langle 0, x \rangle \in \text{PosInfo}(\sigma^s) \cup \text{NegInfo}(\sigma^s)\})$. By a suitably padded version of the Kleene recursion theorem, there exists an $e_1 > m$ such that

$$W_{e_1} = \{\text{PosInfo}(\sigma^s)\} \cup \{\langle x, y \rangle \mid x \geq 2\} \cup \{\langle 0, x \rangle \mid x > m \wedge x \neq e_1\} \quad (7)$$

It is easy to see that $W_{e_1} \in \mathcal{L}_1$.

However, for each $S_1 \subseteq \overline{W_{e_1}}$ of cardinality at most one, $(\forall T \supseteq \sigma^s \mid W_{e_1} = \text{PosInfo}(T) \wedge S_1 \subseteq \text{NegInfo}(T) \wedge \text{PosInfo}(T) \cap \text{NegInfo}(T) = \emptyset)$,

$$(\forall^\infty n)[(W_{\mathbf{M}(T[n])} \cap \langle 0, x \rangle \mid x \in N) \text{ is finite}]$$

(otherwise step 1 in stage s will succeed). It follows that $W_{e_1} \notin \mathbf{NegO}^1\mathbf{TxtBc}^j(\mathbf{M})$.

From the above cases, it follows that $\mathcal{L}_1 \not\subseteq \mathbf{NegO}^1\mathbf{TxtBc}^j(\mathbf{M})$. ■

The previous three theorems have the following straightforward corollary.

Corollary 44 *For all a, j and n ,*

(a) $\mathbf{NegO}^{n+1}\mathbf{TxtEx}^a - \mathbf{NegO}^n\mathbf{TxtEx}^a \neq \emptyset$ and

(b) $\mathbf{NegO}^{n+1}\mathbf{TxtBc}^j - \mathbf{NegO}^n\mathbf{TxtBc}^j \neq \emptyset$.

McNeill [16] posits that there is *faster* learning of language for children in homes in which more corrections (usually in the form of *possibly exemplary* expansions) are given. These corrections are, in part, a form of negative information. The previous corollary says with more core open negative information (bigger S), there is more *learning power*. It doesn't, however, directly inform us about more *speed* of learning, but is, nonetheless, quite interesting. In Section 5 below we present a preliminary result (Theorem 51) showing that an improvement in *speed* (measured by mind-changes) can result from the presence of minimal, non-vacuous, open negative information.

Theorem 45 $\mathbf{TxtEx}^* \subset \mathbf{NegO}^1\mathbf{TxtBc}$.

PROOF. We will prove that $\mathbf{TxtEx}^* \subseteq \mathbf{NegO}^1\mathbf{TxtBc}$. Proper containment follows from Theorem 35.

Suppose \mathbf{M} is given. We will construct a machine \mathbf{M}' which satisfies $\mathbf{TxtEx}^*(\mathbf{M}) \subseteq \mathbf{NegO}^1\mathbf{TxtBc}(\mathbf{M}')$. Without loss of generality we assume that \mathbf{M} is order independent (Theorem 7). For each $L \in \mathbf{TxtEx}^*(\mathbf{M})$, let i_L be such that \mathbf{M} on any text for L converges to i_L , and then, let $S_L = \{\max(\{x \mid x \in W_{i_L} - L\})\}$, if $W_{i_L} - L$ is not empty; otherwise, let $S_L = \emptyset$.

Let F be a recursive function mapping SEQ to SEQ such that

- (a) for all σ and τ , $\sigma \subset \tau \Rightarrow F(\sigma) \subset F(\tau)$ and
- (b) For all σ , $\text{content}(F(\sigma)) = \text{PosInfo}(\sigma)$.

Define \mathbf{M}' as follows.

Let $\mathbf{M}'(\sigma) = i$, such that $W_i = (W_{\mathbf{M}(F(\sigma))} \cup \text{PosInfo}(\sigma)) - \{x \notin \text{PosInfo}(\sigma) \mid x \leq \max(\{y \mid y \in \text{NegInfo}(\sigma) \wedge y \in W_{\mathbf{M}(F(\sigma))}^{\lfloor \sigma \rfloor}\})\}$.

It is easy to see that, for each $L \in \mathbf{TxtEx}^*(\mathbf{M})$, for each L' such that $S_L \subseteq L' \subseteq \bar{L}$, for each text T for $L \oplus L'$, $(\forall n)[W_{\mathbf{M}'(T[n])} = L]$. The theorem follows. \blacksquare

Theorem 46 For all a and j , $[\mathbf{NegO}^a\mathbf{TxtEx}^{2j} \subseteq \mathbf{NegO}^a\mathbf{TxtBc}^j]$.

PROOF. Fix a, j . We will prove that $[\mathbf{NegO}^a\mathbf{TxtEx}^{2j} \subseteq \mathbf{NegO}^a\mathbf{TxtBc}^j]$. Proper containment follows from Theorem 35.

This proof is similar to the proof used by Case and Lynes [7] (see also [6]) to show that $\mathbf{TxtEx}^{2j} \subseteq \mathbf{TxtBc}^j$. Suppose \mathbf{M} is given. We give a machine \mathbf{M}' such that $\mathbf{NegO}^a\mathbf{TxtEx}^{2j}(\mathbf{M}) \subseteq \mathbf{NegO}^a\mathbf{TxtBc}^j(\mathbf{M}')$.

$\mathbf{M}'(\sigma) = i$, such that $W_i = (W_{\mathbf{M}(\sigma)} \cup \text{PosInfo}(\sigma)) - \{x \in W_{\mathbf{M}(\sigma)} - \text{PosInfo}(\sigma) \mid \text{card}(\{y \leq x \mid y \in W_{\mathbf{M}(\sigma)} - \text{PosInfo}(\sigma)\}) \leq j\}$. Now fix L and T such that $\text{PosInfo}(T) = L$ and $W_{\mathbf{M}(T)} = {}^{2j}L$.

Let $S_1 = \{x \mid x \in W_{\mathbf{M}(T)} - L\}$. Let $S_2 = \{x \mid x \in L - W_{\mathbf{M}(T)}\}$.

We now consider the following two cases:

Case 1: $\text{card}(S_1) \geq j$.

Let $S'_1 = \{x \in S_1 \mid \text{card}(\{y \leq x \mid y \in S_1\}) \leq j\}$. It is easy to see that for all but finitely many n , $W_{\mathbf{M}'(T[n])} = (W_{\mathbf{M}(T)} \cup S_2) - S'_1$. Thus for all but finitely many n , $W_{\mathbf{M}'(T[n])} = {}^jL$.

Case 2: $\text{card}(S_1) < j$.

It is easy to see that for all but finitely many n , $W_{\mathbf{M}'(T[n])} \subseteq (W_{\mathbf{M}(T)} \cup S_2) - S_1$. Moreover, for all but finitely many n , $\text{card}(((W_{\mathbf{M}(T)} \cup S_2) - S_1) - W_{\mathbf{M}'(T[n])}) \leq j$. Thus for all but finitely many n , $W_{\mathbf{M}'(T[n])} = {}^jL$.

It follows that $\mathbf{NegO}^a\mathbf{TxtEx}^{2j}(\mathbf{M}) \subseteq \mathbf{NegO}^a\mathbf{TxtBc}^j(\mathbf{M}')$. \blacksquare

We refer the reader to Section 2.3.3 for the definitions of \mathbf{ExGen}^a and \mathbf{BcGen}^a .

Theorem 47 For all a ,

(a) $\mathbf{NegO}^*\mathbf{TxtEx}^a = \mathbf{ExGen}^a$ and

(b) $\mathbf{NegO}^*\mathbf{TxtBc}^a = \mathbf{BcGen}^a$.

PROOF. We prove part (a). The proof of part (b) is similar.

It is easy to convert \mathbf{M} which $\mathbf{NegO}^*\mathbf{TxtEx}^a$ -identifies \mathcal{L} into a machine \mathbf{M}' which \mathbf{ExGen}^a -identifies \mathcal{L} .

Suppose \mathbf{M} \mathbf{ExGen}^a -identifies \mathcal{L} . Fix an arbitrary $L \in \mathcal{L}$. Let i_L be the least value such that for all $i \geq i_L$, $\mathbf{M}(\chi_L[i]) = \mathbf{M}(\chi_L[i_L])$. Since $L \in \mathbf{ExGen}^a(\mathbf{M})$, such an i_L clearly exists. Let $S_L = \{x \in \bar{L} \mid x \leq i_L\}$.

Let $\text{FullInfoPt}(\sigma) = \max(\{x \mid (\forall x' < x)[(x' \in \text{PosInfo}(\sigma)) \vee (x' \in \text{NegInfo}(\sigma))]\})$.

Define \mathbf{M}' as follows.

$$\mathbf{M}'(\sigma) = \begin{cases} \mathbf{M}(\chi_L[\text{FullInfoPt}(\sigma)]), & \text{if } \text{FullInfoPt}(\sigma) \downarrow; \\ 0, & \text{otherwise;} \end{cases}$$

It can now be observed that for all L' such that $S_L \subseteq L' \subseteq \bar{L}$ and for each text T for $L \oplus L'$, $\mathbf{M}'(T) \downarrow = \mathbf{M}(\chi_L) \downarrow$. The theorem follows. \blacksquare

Corollary 48 For all i ,

(a) $\mathbf{NegO}^*\mathbf{TxtEx}^* \subset \mathbf{NegO}^*\mathbf{TxtBc}$,

(b) $\mathbf{NegO}^*\mathbf{TxtEx}^i \subset \mathbf{NegO}^*\mathbf{TxtEx}^{i+1}$ and

(c) $\mathbf{NegO}^*\mathbf{TxtBc}^i \subset \mathbf{NegO}^*\mathbf{TxtBc}^{i+1}$.

PROOF. This corollary follows from Theorem 47 above and the results from [7] that $\mathbf{ExGen}^* \subset \mathbf{BcGen}$, $(\forall i)[\mathbf{ExGen}^i \subset \mathbf{ExGen}^{i+1}]$ and $(\forall i)[\mathbf{BcGen}^i \subset \mathbf{BcGen}^{i+1}]$. \blacksquare

Theorem 47 just above shows us that the \mathbf{NegO}^* criteria are equivalent to supplying *all* the negative (as well as the positive) information to a learning machine.

5 Complexity Advantages of Open Negative Information

For this section it is convenient to change slightly the meaning of the first kind of learning machine (introduced in Definition 2) to the following.

Definition 49 A *language learning machine* is an algorithmic device which computes a mapping from SEQ into $N \cup \{?\}$.

Intuitively the outputted ?s represent the machine not yet committing to an output. The reason we want the ?s is so we can avoid biasing the number of *mind changes* before a learning machine converges: if we allow initial outputs ?s before, if ever, the first program is output, then we can learn more things within n mind changes than if we had to begin with a program (numerical) output.

In the next definition, the subscript b represents a bound on the number of mind changes allowed before convergence.

Definition 50 We say that \mathbf{M} \mathbf{TxtEx}_b^a -identifies $L \Leftrightarrow [[L \in \mathbf{TxtEx}^a(\mathbf{M})] \wedge (\forall \text{ texts } T \text{ for } L)[\text{card}(\{x \mid [? \neq \mathbf{M}(T[x])]\} \wedge [\mathbf{M}(T[x]) \neq \mathbf{M}(T[x+1])]) \leq b]]$.

In a similar fashion, we can expand the definition of $\mathbf{NegO}^c \mathbf{TxtEx}^a$ into the obvious definition of $\mathbf{NegO}^c \mathbf{TxtEx}_b^a$.

Just below is the theorem we promised showing a speed advantage from minimal, non-vacuous open negative information.

Theorem 51 *There exists a class of languages \mathcal{L} such that,*

- (a) $\mathcal{L} \in \mathbf{TxtEx}$,
- (b) $\mathcal{L} \in \mathbf{NegO}^1 \mathbf{TxtEx}_0$, and
- (c) $\mathcal{L} \notin \bigcup_n \mathbf{TxtEx}_n^*$.

PROOF. For each m , let $L_m = \{\langle i, x \rangle \mid i, x \in N \wedge i \neq m\}$. Let $\mathcal{L} = \{L_m \mid m \in N\}$.

It is easy to see that $\mathcal{L} \in \mathbf{TxtEx} \cap \mathbf{NegO}^1 \mathbf{TxtEx}_0$. Suppose by way of contradiction that n, \mathbf{M} are such that $\mathcal{L} \in \mathbf{TxtEx}_n^*(\mathbf{M})$. Without loss of generality, assume that for all texts T , $\text{card}(\{m \mid ? \neq \mathbf{M}(T[m]) \neq \mathbf{M}(T[m+1])\}) \leq n$. Let $\sigma_0 = ()$.

For $i \leq n$, let

$$\sigma_{i+1} = \begin{cases} \tau', & \text{if } (\exists \tau \supseteq \sigma_i)[\mathbf{M}(\sigma_i) \neq \mathbf{M}(\tau) \wedge \mathbf{M}(\tau) \neq ?] \wedge \\ & \tau' = \min(\{\tau \supseteq \sigma_i \mid \mathbf{M}(\sigma_i) \neq \mathbf{M}(\tau) \wedge \mathbf{M}(\tau) \neq ?\}); \\ \sigma_i, & \text{otherwise.} \end{cases}$$

Clearly, $(\forall \tau \supseteq \sigma_{n+1})[\mathbf{M}(\tau) = \mathbf{M}(\sigma_{n+1}) \vee \mathbf{M}(\tau) = ?]$. However, for all $x > \max(\{i \mid (\exists y)[\langle i, y \rangle \in \text{content}(\sigma_{n+1})])$, there exists a text T such that $\sigma_{n+1} \subseteq T$ and $\text{content}(T) = L_x$. Since, for all $x \neq x'$, $L_x \neq^* L_{x'}$, the theorem follows. \blacksquare

6 Open Problems

We list some of the open problems.

- (a) For $i \geq 2$, is $\mathbf{TxtEx}^i \subset \mathbf{NegF}^1 \mathbf{TxtEx}$?
- (b) For $i \geq 1$, $\mathcal{E} \in \mathbf{NegO}^i \mathbf{TxtBc}^*$? We draw the attention of the reader to Theorem 38.
- (c) In [7], it was shown that $\mathbf{TxtEx}^{2j+1} - \mathbf{TxtBc}^j \neq \emptyset$. Similarly, can it be shown that, for $i \geq 1$, $\mathbf{NegO}^i \mathbf{TxtEx}^{2j+1} - \mathbf{NegO}^i \mathbf{TxtBc}^j \neq \emptyset$?
- (d) For $i \geq 1$, is $\mathbf{NegO}^i \mathbf{TxtEx}^* \subset \mathbf{NegO}^{i+1} \mathbf{TxtBc}$? So far we know that $\mathbf{NegO}^* \mathbf{TxtEx}^* \subset \mathbf{NegO}^* \mathbf{TxtBc}$.

Many of the results in the present paper are proved by self-referential and other direct coding techniques. It would be interesting to see if these results can be robustly extended in the important sense of [11].

References

- [1] D. Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.
- [2] L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.
- [3] M. Blum. A machine-independent theory of the complexity of recursive functions. *Journal of the ACM*, 14:322–336, 1967.
- [4] J. Case. Learning machines. In W. Demopoulos and A. Marras, editors, *Language Learning and Concept Acquisition*, pages 83–102. Ablex Publishing Company, 1986.
- [5] J. Case. The power of vacillation. In D. Haussler and L. Pitt, editors, *Proceedings of the Workshop on Computational Learning Theory*, pages 133–142. Morgan Kaufmann, 1988.
- [6] J. Case. The power of vacillation in language learning. *SIAM Journal on Computing*, 28(6):1941–1969, 1999.
- [7] 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.
- [8] J. Case and C. Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science*, 25:193–220, 1983.
- [9] M. Fulk. *A Study of Inductive Inference Machines*. PhD thesis, SUNY/Buffalo, 1985.
- [10] M. Fulk. Prudence and other conditions on formal language learning. *Information and Computation*, 85:1–11, 1990.
- [11] M. Fulk. Robust separations in inductive inference. In *31st Annual IEEE Symposium on Foundations of Computer Science*, pages 405–410. IEEE Computer Society Press, 1990.
- [12] E. M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
- [13] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979.
- [14] S. Jain and A. Sharma. Learning in the presence of partial explanations. *Information and Computation*, 95:162–191, 1991.
- [15] M. Machtey and P. Young. *An Introduction to the General Theory of Algorithms*. North Holland, New York, 1978.
- [16] D. McNeill. Developmental psycholinguistics. In F. Smith and G. Miller, editors, *The Genesis of Language*, pages 15–84. MIT Press, 1966.

- [17] E. Mendelson. *Introduction to Mathematical Logic*. Brooks-Cole, San Francisco, 1986. 3rd Edition.
- [18] D. Moeser and A. Bregman. The role of reference in the acquisition of a miniature artificial language. *Journal of Verbal Learning and Verbal Behavior*, 11:759–769, 1972.
- [19] D. Moeser and A. Bregman. Imagery and language acquisition. *Journal of Verbal Learning and Verbal Behavior*, 12:91–98, 1973.
- [20] T. Motoki. Inductive inference from all positive and some negative data. *Information Processing Letters*, 39(4):177–182, 1991.
- [21] D. Osherson, M. Stob, and S. Weinstein. Ideal learning machines. *Cognitive Science*, 6:277–290, 1982.
- [22] D. Osherson, M. Stob, and S. Weinstein. Learning theory and natural language. *Cognition*, 17:1–28, 1984.
- [23] D. Osherson, M. Stob, and S. Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists*. MIT Press, 1986.
- [24] D. Osherson and S. Weinstein. Criteria of language learning. *Information and Control*, 52:123–138, 1982.
- [25] D. Osherson and S. Weinstein. A note on formal learning theory. *Cognition*, 11:77–88, 1982.
- [26] S. Pinker. Formal models of language learning. *Cognition*, 7:217–283, 1979.
- [27] H. Rogers. Gödel numberings of partial recursive functions. *Journal of Symbolic Logic*, 23:331–341, 1958.
- [28] H. Rogers. *Theory of Recursive Functions and Effective Computability*. McGraw-Hill, 1967. Reprinted, MIT Press 1987.
- [29] T. Shinohara. *Studies on Inductive Inference from Positive Data*. PhD thesis, Kyushu University, Kyushu, Japan, 1986.
- [30] K. Wexler. On extensional learnability. *Cognition*, 11:89–95, 1982.
- [31] K. Wexler and P. Culicover. *Formal Principles of Language Acquisition*. MIT Press, 1980.