Prudence in Vacillatory Language Identification

Sanjay Jain
Institute of Systems Science
National University of Singapore
Singapore 0511
Republic of Singapore
Email: sanjay@iss.nus.sg

Arun Sharma
School of Computer Science and Engineering
The University of New South Wales
Sydney, NSW 2033, Australia
Email: arun@cse.unsw.edu.au

Abstract

The present paper settles a question about 'prudent' 'vacillatory' identification of languages.

Consider a scenario in which an algorithmic device **M** is presented with all and only the elements of a language L, and **M** conjectures a sequence, possibly infinite, of grammars. Three different criteria for success of **M** on L have been extensively investigated in formal language learning theory. If **M** converges to a single correct grammar for L, then the criterion of success is Gold's seminal notion of **TxtEx**-identification. If **M** converges to a finite number of correct grammars for L, then the criterion of success is called **TxtFex**-identification. And, if **M**, after a finite number of incorrect guesses, outputs only correct grammars for L (possibly infinitely many distinct grammars), then the criterion of success is known as **TxtBc**-identification.

A learning machine is said to be *prudent* according to a particular criterion of success just in case the only grammars it ever conjectures are for languages that it can learn according to that criterion. This notion was introduced by Osherson, Stob, and Weinstein with a view to investigate certain proposals for characterizing natural languages in linguistic theory. Fulk showed that prudence does not restrict **TxtEx**-identification, and later Kurtz and Royer showed that prudence does not restrict **TxtBc**-identification. The present paper shows that prudence does not restrict **TxtFex**-identification.

1 Introduction

Languages are sets of sentences and a sentence is a finite object; the set of all possible sentences can be coded into N — the set of natural numbers. Hence, languages may be construed as subsets of N. A grammar for a language is a set of rules that accepts (or equivalently, generates [HU79]) the language. Essentially, any computer program may be viewed as a grammar. Languages for which a grammar exists are called *recursively enumerable*. Henceforth, we work under the assumption that natural languages fall in the class of recursively enumerable languages.

A text for a language L is any infinite sequence that lists all and only the elements of L; repetitions are permitted.

Motivated by psycholinguistic studies which suggest that children are rarely, if ever, informed of grammatical errors[1], Gold [Gol67] introduced the seminal notion of *identification in the limit* (which we refer to as **TxtEx**-identification following [CL82]) as a model for first language acquisition. According to this paradigm, a child (modeled as a machine) receives a text for a language L, and simultaneously conjectures a succession of grammars. A criterion of success is for the child to eventually conjecture a correct grammar for L and never to change its conjecture thereafter. If, in this scenario for success, the child machine is replaced by an algorithmic machine **M**, then we say that **M** **TxtEx**-*identifies* L. The reader is directed to [Pin79, WC80, Wex82, OSW84, OW92] for a discussion of the influence of this paradigm on contemporary theories of natural language.

Now, a major concern of linguistic theory is to characterize the class of natural languages. Any such characterization must account for the fact that children master natural languages in a few years time on the basis of rather casual and unsystematic exposure to it. Formal language learning theory provides a tool to evaluate proposals for characterizing the collection of natural languages by modeling the salient features of a proposal in the above paradigm (see Osherson, Stob, Weinstein [OSW84] for discussion of these issues.)

A collection of such proposals, are known as "prestorage models" of linguistic development. A prestorage model assumes that an internal list of candidate grammars that coincides exactly with the collection of natural languages is available to a child. Language acquisition is thus a process of selecting a grammar from this list in response to linguistic input. Motivated by such models and with a view to investigate the effect of such a restriction, Osherson, Stob, and Weinstein [OSW82] introduced the notion of "prudent" learning machines. According to their definition, prudent learners only conjecture grammars for languages they are prepared to learn. In other words, every incorrect grammar emitted by a prudent learner in response to any linguistic input is for some language that can be learned by the learner. Osherson, Stob, and Weinstein raised the natural question: "Does prudence restrict **TxtEx**-identification?" Fulk [Ful85, Ful90] provided the answer by establishing a surprising result that prudence does not restrict **TxtEx**-identification. He showed that given any learning machine **M**, a prudent learning machine $\mathbf{M}'$ can be constructed which **TxtEx**-identifies every language **TxtEx**-identified by **M**.

[1] See [BH70, HPTS84, DPS86, Pen87] for further studies discussing this hypothesis.

However, the investigation of prudence for more general learning criteria, notably **TxtBc**-identification and **TxtFex**-identification, was left open by Fulk. Below, we informally describe these criteria.

A learning machine **M** is said to **TxtBc**-identify a language L just in case **M**, fed any text for L, outputs an infinite sequence of grammars such that after a finite number of incorrect guesses, **M** outputs only grammars for L. This criterion was first studied by Osherson and Weinstein [OW82a] and by Case and Lynes [CL82], and is also referred to as "extensional" identification. A machine **M** is said to be **TxtBc**-prudent just in case any grammar conjectured by **M** in response to any linguistic input is for a language which **M** can **TxtBc**-identify.

Let b be a positive integer. A learning machine **M** is said to $\mathbf{TxtFex}_b$-identify a language L just in case **M**, fed any text for L, converges in the limit to a finite set, with cardinality $\leq b$, of grammars for L. In other words, for any text T for L, there exists a set D of grammars for L, cardinality of $D \leq b$, such that **M**, fed T, outputs, after a finite number of incorrect guesses, only grammars from the set D. This notion was studied by Osherson and Weinstein [OW82a] and by Case [Cas88]. A machine **M** is said to be $\mathbf{TxtFex}_b$-prudent just in case any grammar output by **M** in response to any linguistic input is for a language which **M** can $\mathbf{TxtFex}_b$-identify.

Fulk [Ful85] had conjectured that prudence was not likely to restrict **TxtBc**-identification. Kurtz and Royer [KR88] showed that this conjecture was indeed true, as they showed that for any learning machine **M**, there exists a machine $\mathbf{M}'$ such that $\mathbf{M}'$ is **TxtBc**-prudent and $\mathbf{M}'$ **TxtBc**-identifies every language which **M** **TxtBc**-identifies. However, they left the problem open for $\mathbf{TxtFex}_b$-identification. In the present paper, we settle this question by showing that prudence does not restrict $\mathbf{TxtFex}_b$-identification.

A related topic in the context of function identification has been studied under the title 'class preserving strategies' (see Jantke and Beick [JB81]).

We now proceed formally. Section 2 states the notation and preliminary concepts from formal language learning theory. The main result of the paper is contained in Section 3.

2 Preliminaries

2.1 Notation

Any unexplained recursion theoretic notation is from [Rog67]. The symbol N denotes the set of natural numbers, $\{0, 1, 2, 3, \ldots\}$. The symbol N^+ denotes the set of positive natural numbers, $\{1, 2, 3, \ldots\}$. Unless otherwise specified, i, j, m, n, s, t, x, y, with or without decorations[2], range over N. Symbols $\emptyset$, $\subseteq$, $\subset$, $\supseteq$, and $\supset$ denote empty set, subset, proper subset, superset, and proper superset, respectively. We use the symbol $\Rightarrow$ to denote logical implication. We use symbols $\equiv$ and $\Leftrightarrow$ to denote equivalence.

Symbols P and S, with or without decorations, range over finite sets. Cardinality of a set S is denoted by $\text{card}(S)$. We say that $\text{card}(A) \leq *$ to mean that $\text{card}(A)$ is finite.

[2] Decorations are subscripts, superscripts and the like.

Intuitively, the symbol, $*$, denotes 'finite without any prespecified bound.' We let b range over $N^+ \cup \{*\}$. D_x denotes the finite set with canonical index x [Rog67]. 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) and when we want to display the argument simply as the set itself.

We use the symbol $\uparrow$ to denote 'undefined.' The maximum and minimum of a set are denoted by $\max(\cdot), \min(\cdot)$, respectively, where $\max(\emptyset) = 0$ and $\min(\emptyset) =\uparrow$.

Letters f, g, and h, with or without decorations, range over *total* functions with arguments and values from N.

A pair $\langle i, j \rangle$ 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 multiple tuples of natural numbers onto N.

By φ we denote a fixed *acceptable* programming system for the partial computable functions: $N \rightarrow N$ [Rog58, Rog67, MY78]. By φ_i we denote the partial computable function computed by program i in the φ-system. By Φ we denote an arbitrary fixed Blum complexity measure [Blu67, HU79] for the φ-system.

By W_i we denote domain(φ_i). W_i is, then, the r.e. set/language ($\subseteq N$) accepted (or equivalently, generated) by the φ-program i. Symbol $\mathcal{E}$ will denote the set of all r.e. languages. Symbol L, with or without decorations, ranges over $\mathcal{E}$. Symbol $\mathcal{L}$, with or without decorations, ranges over subsets of $\mathcal{E}$. We denote by $W_{i,s}$ the set $\{x \leq s : \Phi_i(x) \leq s\}$. We use $\mathcal{FIN}$ to denote the set $\{L : \text{card}(L) < \infty\}$.

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

The quantifiers '$\overset{\infty}{\forall}$' and '$\overset{\infty}{\exists}$' mean 'for all but finitely many' and 'there exist infinitely many', respectively.

2.2 Language Learning Machines and Texts

We now consider language learning machines. Definition 1 below introduces a notion that facilitates discussion about elements of a language being fed to a 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 content(σ), 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. For $n \leq |\sigma|$, $\sigma[n]$ denotes the finite initial sequence of σ with length n. The result of concatenating τ onto the end of σ is denoted by $\sigma \diamond \tau$. We say that $\sigma \subseteq \tau$ just in case σ is an initial segment of τ, that is, $|\sigma| \leq |\tau|$ and $\sigma = \tau[|\sigma|]$. 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. This coding assigns a canonical index to each member of SEQ. We will abuse the notation somewhat, as a reference to σ will mean both the sequence and its canonical index. Hence, a reference to a *least* σ really refers to a σ with least canonical index.

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

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

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 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 decorations, range over texts. $T[n]$ denotes the finite initial sequence of T with length n. The reader should note that $T[n]$ does not contain $T(n)$, the n^{th} element of T. Hence, domain$(T[n]) = \{x : x < n\}$. We say that $\sigma \subset T$ just in case σ is an initial segment of T, that is, $\sigma = T[|\sigma|]$.

We next present three criteria for successful learning of languages by learning machines.

2.3 Language Identification Criteria

2.3.1 Explanatory Learning (TxtEx-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*) $\iff$ $(\exists i)(\overset{\infty}{\forall} n)$ $[\mathbf{M}(T[n]) = i]$. If $\mathbf{M}(T)\downarrow$, then $\mathbf{M}(T)$ is defined as the unique i such that $(\overset{\infty}{\forall} n)[\mathbf{M}(T[n]) = i]$; otherwise, we say that $\mathbf{M}(T)$ *diverges* (written: $\mathbf{M}(T)\uparrow$).

The next definition describes the first criteria of success and is essentially Gold's paradigm of identification in the limit.

Definition 5 [Gol67]

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

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

The notation in the above definition is from [CL82].

2.3.2 Vacillatory Learning (TxtFex-identification)

In Definition 6 below we spell out what it means for a learning machine on a text to converge in the limit to a finite set of grammars.

Definition 6 [Cas88] $\mathbf{M}(T)$ *finitely-converges* (written: $\mathbf{M}(T)\Downarrow$) $\iff$ $\{\mathbf{M}(\sigma) : \sigma \subset T\}$ is finite, otherwise we say that $\mathbf{M}(T)$ *finitely-diverges* (written: $\mathbf{M}(T)\Uparrow$). If $\mathbf{M}(T)\Downarrow$, then $\mathbf{M}(T)$ is defined as P, where $P = \{i : (\stackrel{\infty}{\exists} \sigma \subset T)[\mathbf{M}(\sigma) = i]\}$.

Definition 7 [Cas88] Let $b \in N^+ \cup \{*\}$.

(a) $\mathbf{M}$ $\mathbf{TxtFex}_b$-*identifies* L (written: $L \in \mathbf{TxtFex}_b(\mathbf{M})$) $\iff$ ($\forall$ texts T for L)$(\exists P : \mathrm{card}(P) \leq b \wedge (\forall i \in P)[W_i = L])[\mathbf{M}(T)\Downarrow \wedge \mathbf{M}(T) = P]$.

(b) $\mathbf{TxtFex}_b = \{\mathcal{L} : (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{TxtFex}_b(\mathbf{M})]\}$.

The $b = *$ case in the above was first studied by Osherson and Weinstein [OW82a].

2.3.3 Behaviorally Correct Learning (TxtBc-identification)

Definition 8 [CL82, OW82b, OW82a]

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

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

The following theorem summarizes the relationship between the criteria defined above.

Theorem 1 [OW82a, CL82, Cas88]
$\mathbf{TxtEx} = \mathbf{TxtFex}_1 \subset \mathbf{TxtFex}_2 \subset \ldots \subset \mathbf{TxtFex}_i \subset \mathbf{TxtFex}_{i+1} \subset \ldots \subset \mathbf{TxtFex}_* \subset \mathbf{TxtBc}$.

3 Prudence and Language Learning

The following definition describes the notion of prudence for each of the three criteria described in the previous section. The notion of prudence was first introduced by Osherson, Stob, and Weinstein [OSW82].

Definition 9 Let $b \in N^+ \cup \{*\}$.

(a) A machine $\mathbf{M}$ is $\mathbf{TxtEx}$-*prudent* just in case $\{W_{\mathbf{M}(\sigma)} : \sigma \in \mathrm{SEQ}\} = \mathbf{TxtEx}(\mathbf{M})$.

(b) A machine $\mathbf{M}$ is $\mathbf{TxtBc}$-*prudent* just in case $\{W_{\mathbf{M}(\sigma)} : \sigma \in \mathrm{SEQ}\} = \mathbf{TxtBc}(\mathbf{M})$.

(c) A machine $\mathbf{M}$ is $\mathbf{TxtFex}_b$-*prudent* just in case $\{W_{\mathbf{M}(\sigma)} : \sigma \in \mathrm{SEQ}\} = \mathbf{TxtFex}_b(\mathbf{M})$.

Fulk showed the following result which says that prudence is not a restriction on **TxtEx**-identification.

Theorem 2 [Ful85, Ful90] *For each machine* $\mathbf{M}$ *there exists a* $\mathbf{TxtEx}$*-prudent machine* $\mathbf{M}'$*, such that* $\mathbf{TxtEx}(\mathbf{M}) \subseteq \mathbf{TxtEx}(\mathbf{M}')$*.*

Fulk left the question of prudence open for **TxtBc**-identification and $\mathbf{TxtFex}_b$-identification. But, he conjectured that a counterpart of the above theorem was likely to hold for **TxtBc**-identification. Kurtz and Royer [KR88] showed that Fulk's conjecture was indeed true, as they established the following result.

Theorem 3 [KR88] *For each machine* $\mathbf{M}$ *there exists a* $\mathbf{TxtBc}$*-prudent machine* $\mathbf{M}'$*, such that* $\mathbf{TxtBc}(\mathbf{M}) \subseteq \mathbf{TxtBc}(\mathbf{M}')$*.*

However, prudence for vacillatory identification remained open. Theorem 4 in the present paper settles this question. Our proof of this result builds on notions of stabilizing and locking sequences. To facilitate discussion of these technical concepts, we first extend the definition of pairing function. Towards this end, fix a canonical index for members of $\mathcal{FIN}$. (See Rogers [Rog67]). Recall from Section 2.2 that there is a canonical index for members of SEQ. Now, let i be the canonical index for $\sigma \in$ SEQ and j be the canonical index for $D \in \mathcal{FIN}$. Then, define $\langle \sigma, D \rangle = \langle i, j \rangle$. Clearly, the pairing function, $\langle ., . \rangle$, defines a total order on SEQ $\times\, \mathcal{FIN}$, and we have a canonical index for members of SEQ $\times\, \mathcal{FIN}$. Hence, a reference to *least* $\langle \sigma, D \rangle$ in the sequel is well defined.

Definition 10 (Based on [Cas88, Ful85]) Let $b \in N^+ \cup \{*\}$. Then $\langle \sigma, D \rangle$ is a $\mathbf{TxtFex}_b$-*stabilizing sequence* for $\mathbf{M}$ on L just in case the following hold:

(a) $\text{content}(\sigma) \subseteq L$,

(b) $\text{card}(D) \leq b$, and

(c) $(\forall \tau : \sigma \subseteq \tau \wedge \text{content}(\tau) \subseteq L)[\mathbf{M}(\tau) \in D]$.

Definition 11 (Based on [BB75, Cas88, Ful85]) Let $b \in N^+ \cup \{*\}$. Then $\langle \sigma, D \rangle$ is a $\mathbf{TxtFex}_b$-*locking sequence* for $\mathbf{M}$ on L just in case the following hold:

(a) $\langle \sigma, D \rangle$ is a $\mathbf{TxtFex}_b$-stabilizing sequence for $\mathbf{M}$ on L, and

(b) $(\forall j \in D)[W_j = L]$.

Lemma 1 (Based on [BB75, Cas88, Ful85]) *Suppose* $b \in N^+ \cup \{*\}$ *and* $\mathbf{M}$ $\mathbf{TxtFex}_b$*-identifies* L*. Then there exists a* $\mathbf{TxtFex}_b$*-locking sequence for* $\mathbf{M}$ *on* L*.*

Proof. For $\sigma \in$ SEQ, let $\text{Last}_b(\sigma)$ denote the set,

$$\{\mathbf{M}(\sigma[l]) : l \leq |\sigma| \wedge \text{card}(\{\mathbf{M}(\sigma[l']) : |\sigma| \geq l' \geq l\}) \leq b\}.$$

Roughly, $\text{Last}_b(\sigma)$ denotes the set of b most recent conjectures of $\mathbf{M}$ on σ.

Suppose by way of contradiction that no σ is a $\mathbf{TxtFex}_b$-locking sequence for $\mathbf{M}$ on L. Then we have,

$$(\forall \sigma : \text{content}(\sigma) \subseteq L)(\exists \sigma' : \sigma \subseteq \sigma' \wedge \text{content}(\sigma') \subseteq L)[W_{\mathbf{M}(\sigma')} \neq L \vee \mathbf{M}(\sigma') \notin \text{Last}_b(\sigma)] \tag{1}$$

Let T^0 be a text for L. We now define σ_i, τ_i inductively. It will be the case that for all i, $\sigma_i \subseteq \tau_i \subseteq \sigma_{i+1}$ and $\text{content}(\sigma_{i+1}) \subseteq L$.

Let σ_0 be the empty sequence. Let $\tau_i = \sigma_i \diamond T^0(i)$. Let σ_{i+1} be an (arbitrary) extension of τ_i such that $\text{content}(\sigma_{i+1}) \subseteq L$ and $[W_{\mathbf{M}(\sigma_{i+1})} \neq L \vee \mathbf{M}(\sigma_{i+1}) \notin \text{Last}_b(\tau_i)]$. By (1) there exists such an extension. Now let $T = \bigcup_{i \in N} \sigma_i$. Clearly, $\text{content}(T) = L$ (since, the construction of T guarantees that only elements of L and all the elements of L appear in T). Now, for all i, either $[W_{\mathbf{M}(\sigma_{i+1})} \neq L$ or $\mathbf{M}(\sigma_{i+1}) \notin \text{Last}_b(\tau_i)]$. This implies that $\mathbf{M}$ does not $\mathbf{TxtFex}_b$-identify L. A contradiction. ∎ (Lemma 1)

The following corollary to Lemma 1 is evident.

Corollary 1 *Suppose $b \in N^+ \cup \{*\}$ and* $\mathbf{M}$ $\mathbf{TxtFex}_b$*-identifies L. Then there exists a* $\mathbf{TxtFex}_b$*-stabilizing sequence for* $\mathbf{M}$ *on L.*

Similarly, using the technique in the proof of Lemma 1, we can establish the following lemma. The reader is referred to [Cas92] for a number of related results about $\mathbf{TxtFex}_b$-identification.

Lemma 2 *If* $\mathbf{M}$ $\mathbf{TxtFex}_b$*-identifies L and $\langle \sigma, D \rangle$ is a* $\mathbf{TxtFex}_b$*-stabilizing sequence for* $\mathbf{M}$ *on L, then $(\exists i \in D)[W_i = L]$.*

We now present our main result.

Theorem 4 *Let $b \in N^+ \cup \{*\}$. For each* $\mathbf{M}$ *there exists a* $\mathbf{TxtFex}_b$*-prudent machine* $\mathbf{M}''$*, such that* $\mathbf{TxtFex}_b(\mathbf{M}) \subseteq \mathbf{TxtFex}_b(\mathbf{M}'')$.

Proof of Theorem 4. Let $\mathbf{M}$ and b be as given in the hypothesis of the theorem. We will show that there exist machines $\mathbf{M}'$ and $\mathbf{M}''$ such that $\mathbf{TxtFex}_b(\mathbf{M}) \subseteq \mathbf{TxtFex}_b(\mathbf{M}') \subseteq \mathbf{TxtFex}_b(\mathbf{M}'')$ and $\mathbf{M}''$ is $\mathbf{TxtFex}_b$-prudent.

Our proof may be thought of as a nontrivial extension of Fulk's technique [Ful90]. Like Fulk's proof, our proof is also nonconstructive. For a given $\mathbf{M}$, $\mathbf{M}'$ will depend on whether $\mathbf{M}$ $\mathbf{TxtFex}_b$-identifies N or not. Since this is not known in advance, this technique does not allow for an effective construction of $\mathbf{M}''$ from $\mathbf{M}$. It is open at this stage whether our proof can be made constructive.

Given $\mathbf{M}$ and b, we define a predicate MidentN as follows: MidentN is true iff $\mathbf{M}$ $\mathbf{TxtFex}_b$-identifies N.

Let g be a recursive function such that, for all n, $W_{g(n)} = \{x : x < n\}$. Define $\mathbf{M}'$ as follows.

$$\mathbf{M}'(\sigma) = \begin{cases} g(0), & \text{if content}(\sigma) = \emptyset; \\ g(n), & \text{if } \neg\text{MidentN} \wedge \text{content}(\sigma) = \{x : x < n\}; \\ \mathbf{M}(\sigma), & \text{otherwise.} \end{cases}$$

Let i_N be such that $W_{i_N} = N$. Let $\mathcal{C}_N = \{N\}$, if MidentN holds, and $\mathcal{C}_N = \{L : (\exists n)[L = \{x : x < n\}]\}$, otherwise.

It is easy to see that $\mathbf{TxtFex}_b(\mathbf{M}) \cup \{\emptyset\} \cup \mathcal{C}_N \subseteq \mathbf{TxtFex}_b(\mathbf{M}')$. Intuitively, $\mathbf{M}'$ behaves just like $\mathbf{M}$ except with minor changes to allow it to $\mathbf{TxtFex}_b$-identify $\emptyset$ and each element of $\mathcal{C}_N$.

For constructing $\mathbf{M}''$ as claimed in the theorem, we will define, using s-m-n theorem, a recursive function h, from $N \times \text{SEQ} \times \mathcal{FIN} \times \mathcal{FIN}$ to N. But before defining h, we introduce a few predicates.

Let $\text{Good}(j, \sigma, P, S)$ be a conjunction of the following three conditions:

(a) $j \in P$,
(b) $\langle \sigma, P \rangle$ is the least $\mathbf{TxtFex}_b$-stabilizing sequence for $\mathbf{M}'$ on W_j, and
(c) $S = W_j \cap \bigcup_{\langle \sigma', P' \rangle \leq \langle \sigma, P \rangle} \text{content}(\sigma')$.

Intuitively, h will be such that, if $\text{Good}(j, \sigma, P, S)$ holds, then $W_{h(j,\sigma,P,S)} = W_j$; otherwise $W_{h(j,\sigma,P,S)}$ is a member of $\mathcal{C}_N \cup \{\emptyset\}$. Parameter S helps in the verification of part (b) in the definition of Good, that is, S contains enough information to verify that no $\langle \sigma', P' \rangle < \langle \sigma, P \rangle$ is a $\mathbf{TxtFex}_b$-stabilizing sequence for $\mathbf{M}'$ on W_j. $\mathbf{M}''$ will use this h to achieve its goal as claimed in the theorem. The idea is that $\mathbf{M}''$ upon seeing the initial segment of a text for some language L, will attempt to find a candidate for the least $\mathbf{TxtFex}_b$-stabilizing sequence, $\langle \sigma, P \rangle$, for $\mathbf{M}'$ on L. $\mathbf{M}''$ will then pick from P a seemingly best grammar, j_0, for L. $\mathbf{M}''$ will then output $h(j_0, \sigma, P, S)$ (where S is chosen according to part (c) in the definition of Good above). We will discuss details of $\mathbf{M}''$ later.

We now define two predicates, "plausible" and "impossible." Intuitively, $\text{plausible}(j, \sigma, P, S, t)$ is true just in case it can be verified, in at most t steps, that condition (a) and parts of (b) and (c) of $\text{Good}(j, \sigma, P, S)$ hold. And, $\text{impossible}(j, \sigma, P, S, t)$ is true just in case it can be verified, in at most t steps, that $\neg\text{Good}(j, \sigma, P, S)$ holds.

Let

$$\begin{aligned} &\text{plausible}(j, \sigma, P, S, t) \equiv \\ &\quad [t > 0] \wedge [\text{card}(P) \leq b] \wedge [j \in P] \wedge \\ &\quad [S \subseteq \textstyle\bigcup_{\langle \sigma', P' \rangle \leq \langle \sigma, P \rangle} \text{content}(\sigma')] \wedge [\text{content}(\sigma) \subseteq S \subseteq W_{j,t}] \wedge \\ &\quad (\forall \langle \sigma', P' \rangle < \langle \sigma, P \rangle)(\exists \tau : |\tau| \leq t)[[\text{content}(\sigma') \not\subseteq S] \vee [\text{card}(P') > b] \vee \\ &\qquad [[\text{content}(\sigma') \subseteq \text{content}(\tau) \subseteq W_{j,t}] \wedge [\mathbf{M}'(\tau) \notin P']]]. \end{aligned}$$

Let

impossible$(j,\sigma,P,S,t) \equiv$

$[\overline{S} \cap W_{j,t} \cap \bigcup_{\langle\sigma',P'\rangle \leq \langle\sigma,P\rangle} \text{content}(\sigma') \neq \emptyset] \vee$

$(\exists\tau \supseteq \sigma : \text{content}(\tau) \subseteq W_{j,t} \wedge |\tau| \leq t)[\mathbf{M}'(\tau) \notin P]$.

We now let h be a recursive function such that $W_{h(j,\sigma,P,S)} = \bigcup_{t\in N} W_{h'(j,\sigma,P,S,t)}$, where

$$W_{h'(j,\sigma,P,S,t)} = \begin{cases} \emptyset, & \text{if } \neg\text{plausible}(j,\sigma,P,S,t); \\ W_{j,t}, & \text{if plausible}(j,\sigma,P,S,t) \\ & \ \wedge \neg\text{impossible}(j,\sigma,P,S,t); \\ \{x : x \leq t\}, & \text{if MidentN} \wedge \text{plausible}(j,\sigma,P,S,t) \\ & \ \wedge \text{impossible}(j,\sigma,P,S,t); \\ \{x : x \leq \max(W_{h'(j,\sigma,P,S,t-1)})\}, & \text{if } \neg\text{MidentN} \wedge \text{plausible}(j,\sigma,P,S,t) \\ & \ \wedge \text{impossible}(j,\sigma,P,S,t). \end{cases}$$

Let $\mathcal{C} = \{W_j : (\exists\sigma,P)[[\langle\sigma,P\rangle$ is the least $\mathbf{TxtFex}_b$-stabilizing sequence for $\mathbf{M}'$ on $W_j] \wedge j \in P]\}$,

and let $\mathcal{C}' = \{W_{h(j,\sigma,P,S)} : j \in N \wedge \sigma \in \text{SEQ} \wedge P \in \mathcal{FIN} \wedge S \in \mathcal{FIN}\}$.

Claim 1 *$\mathcal{C} = \mathcal{C}'$.*

Proof. It is easy to see that,

$$(\forall j,\sigma,P,S,t)[\text{plausible}(j,\sigma,P,S,t) \Rightarrow \text{plausible}(j,\sigma,P,S,t+1)]$$

and

$$(\forall j,\sigma,P,S,t)[\text{impossible}(j,\sigma,P,S,t) \Rightarrow \text{impossible}(j,\sigma,P,S,t+1)].$$

Thus, for each j,σ and finite sets P,S: $W_{h(j,\sigma,P,S)} \in \{W_j,\emptyset\} \cup \mathcal{C}_N$. Now, there are two cases:

Case 1: $W_{h(j,\sigma,P,S)} \in \{\emptyset\} \cup \mathcal{C}_N$.

In this case, it is easy to see that by the description of $\mathbf{M}'$ from $\mathbf{M}$, $\mathbf{M}'$ $\mathbf{TxtFex}_b$-identifies $W_{h(j,\sigma,P,S)}$. Now, by Corollary 1 and Lemma 2, there exists $\langle\sigma',P'\rangle$ such that $\langle\sigma',P'\rangle$ is the least $\mathbf{TxtFex}_b$-stabilizing sequence of $\mathbf{M}'$ on $W_{h(j,\sigma,P,S)}$, and there exists a $j' \in P'$ such that $W_{j'} = W_{h(j,\sigma,P,S)}$. Hence, $W_{h(j,\sigma,P,S)} \in \mathcal{C}$.

Case 2: $W_{h(j,\sigma,P,S)} \notin \{\emptyset\} \cup \mathcal{C}_N$.

In this case, $(\overset{\infty}{\forall} t)[\text{plausible}(j,\sigma,P,S,t)]$ and $(\forall t)[\neg\text{impossible}(j,\sigma,P,S,t)]$. Thus, $\langle\sigma,P\rangle$ is the least $\mathbf{TxtFex}_b$-stabilizing sequence for $\mathbf{M}'$ on W_j and $j \in P$. Hence, $W_{h(j,\sigma,P,S)} \in \mathcal{C}$.

It follows from the above two cases that $\mathcal{C}' \subseteq \mathcal{C}$.

Now, for each $L \in \mathcal{C}$, let $\langle\sigma,P\rangle$ be the least $\mathbf{TxtFex}_b$-stabilizing sequence for $\mathbf{M}'$ on L. Let $j \in P$ be such that $W_j = L$ (by definition of $\mathcal{C}$ there exists such a j). Let $S = W_j \cap \bigcup_{\langle\sigma',P'\rangle \leq \langle\sigma,P\rangle} \text{content}(\sigma')$. It is easy to see that $(\exists t)[\text{plausible}(j,\sigma,P,S,t)]$ and $(\forall t)[\neg\text{impossible}(j,\sigma,P,S,t)]$. It follows that $W_{h(j,\sigma,P,S)} = W_j \in \mathcal{C}'$. Thus, $\mathcal{C} \subseteq \mathcal{C}'$. ▮ (Claim 1)

Claim 2 *$(\forall j)(\forall\sigma)(\forall P,S \in \mathcal{FIN})[Good(j,\sigma,P,S) \Rightarrow W_{h(j,\sigma,P,S)} = W_j]$.*

PROOF. Suppose $\text{Good}(j, \sigma, P, S)$ holds. Then, we have

(a) $(\overset{\infty}{\forall} t)[\text{plausible}(j, \sigma, P, S, t)]$, and

(b) $(\forall t)[\neg\text{impossible}(j, \sigma, P, S, t)]$.

Thus, by definition of $W_{h(j,\sigma,P,S)}$, $W_{h(j,\sigma,P,S)} = W_j$. ▮ (Claim 2)

Also, clearly, $\mathbf{TxtFex}_b(\mathbf{M}') \subseteq \mathcal{C}$. This is because for each $L \in \mathbf{TxtFex}_b(\mathbf{M}')$, Corollary 1 and Lemma 2 imply that there exists $\langle \sigma, P \rangle$ such that $\langle \sigma, P \rangle$ is the least $\mathbf{TxtFex}_b$-stabilizing sequence for $\mathbf{M}'$ on L and there exists a $j \in P$ such that $W_j = L$. Hence, $L \in \mathcal{C}$.

We now construct $\mathbf{M}''$ such that $\mathbf{M}''$ is $\mathbf{TxtFex}_b$-prudent and $\mathbf{TxtFex}_b(\mathbf{M}') \subseteq \mathbf{TxtFex}_b(\mathbf{M}'')$. Intuitively, $\mathbf{M}''$ tries to find the least $\mathbf{TxtFex}_b$-stabilizing sequence $\langle \sigma, P \rangle$ for $\mathbf{M}'$ on the input language L, and then outputs $h(j_0, \sigma, P, S)$, where j_0 is a "seemingly best" candidate grammar for L in P, and $S = L \cap \bigcup_{\langle \sigma', P' \rangle \leq \langle \sigma, P \rangle} \text{content}(\sigma')$. We now say a few words on how $\mathbf{M}''$ choses a "seemingly best" j_0 from the finite set P. This is achieved using the computable function *match* defined below.

Let T be a text. Define $\text{match}(j, T[n]) = \max(\{s \leq n : \text{content}(T[s]) \subseteq W_{j,n} \wedge W_{j,s} \subseteq \text{content}(T[n])\})$.

The reader should observe the following about $\text{match}(j, T[n])$:

1. If $W_j = \text{content}(T)$, then $\lim_{n\to\infty} \text{match}(j, T[n]) = \infty$.
2. If $W_j \neq \text{content}(T)$, then $\lim_{n\to\infty} \text{match}(j, T[n]) < \infty$.

From the above, it is clear that for each $j \in P$ such that $W_j \neq \text{content}(T)$, $\text{match}(j, T[n])$ will eventually become fixed whereas for each $j' \in P$ such that $W_{j'} = \text{content}(T)$, $\text{match}(j', T[n])$ will keep on increasing. Thus, $\mathbf{M}''$ on $T[n]$ uses as j_0, a j in P for which $\text{match}(j, T[n])$ is maximized (in case there are more than one such j, $\mathbf{M}''$ choses the minimum such j). More formally:

Begin $\mathbf{M}''(T[n])$:

(1) Let $\langle \sigma, P \rangle$ be the least pair such that the following three conditions are satisfied.

(a) $\text{content}(\sigma) \subseteq \text{content}(T[n])$,

(b) $\text{card}(P) \leq b$,

(c) $(\forall \tau : \sigma \subseteq \tau \wedge |\tau| \leq n \wedge \text{content}(\tau) \subseteq \text{content}(T[n]))[\mathbf{M}'(\tau) \in P]$.

(2) Let $m = \max(\{\text{match}(j, T[n]) : j \in P\})$.

(3) Let $j_0 = \min(\{j \in P : \text{match}(j, T[n]) = m\})$.

(4) Output $h(j_0, \sigma, P, \text{content}(T[n]) \cap \bigcup_{\langle \sigma', P' \rangle \leq \langle \sigma, P \rangle} \text{content}(\sigma'))$.

End $\mathbf{M}''(T[n])$

Clearly, $\mathbf{M}''$ outputs grammars only for languages in $\mathcal{C}'$. We now establish that $\mathbf{M}''$ is $\mathbf{TxtFex}_b$-prudent by showing that $\mathbf{M}''$ $\mathbf{TxtFex}_b$-identifies each language in $\mathcal{C}'$. Let $L \in \mathcal{C}'$. Let σ and P, be such that $\langle \sigma, P \rangle$ is the least $\mathbf{TxtFex}_b$-stabilizing sequence for $\mathbf{M}'$ on L (such a $\langle \sigma, P \rangle$ exists for each $L \in \mathcal{C}$ and by Claim 1, $\mathcal{C}' = \mathcal{C}$). Let $P_L = \{j \in P : W_j = L\}$. By definition of $\mathcal{C}$, $P_L \neq \emptyset$. Let $S = L \cap \bigcup_{\langle \sigma', P' \rangle \leq \langle \sigma, P \rangle} \text{content}(\sigma')$. Let T be a text for L. Let n_0 be large enough so that the following conditions are satisfied:

1. $\text{content}(\sigma) \subseteq \text{content}(T[n_0])$;
2. $S \subseteq \text{content}(T[n_0])$;
3. $(\forall n \geq n_0)(\forall p \in P_L)(\forall p' \in P - P_L)\ [\text{match}(p, T[n]) > \text{match}(p', T[n])]$;
4. $(\forall \langle \sigma', P' \rangle < \langle \sigma, P \rangle)[$

   $[\text{content}(\sigma') \not\subseteq L] \vee$
   $[\text{card}(P') > b] \vee$
   $[(\exists \tau : \sigma' \subseteq \tau \wedge \text{content}(\tau) \subseteq \text{content}(T[n_0]) \wedge |\tau| \leq n_0)[\mathbf{M}'(\tau) \notin P']]$

   $]$.

Clearly, such an n_0 exists. Then, it is easy to verify that, for all $n \geq n_0$, $\mathbf{M}''(T[n]) \in \{h(j, \sigma, P, S) : j \in P_L\}$. But, for each $j \in P_L$, $\text{Good}(j, \sigma, P, S)$ holds, and thus by Claim 2, $W_{h(j,\sigma,P,S)} = W_j = L$. Therefore, we have that $\mathbf{M}''$ $\mathbf{TxtFex}_b$-identifies L.

Hence, $\mathbf{M}''$ is $\mathbf{TxtFex}_b$-prudent and $\mathbf{TxtFex}_b(\mathbf{M}) \subseteq \mathbf{TxtFex}_b(\mathbf{M}') \subseteq \mathbf{TxtFex}_b(\mathbf{M}'')$.

▮ (Theorem 4)

4 Conclusion

The problem of prudence for successful learning of languages from positive data was described. It was shown that requiring vacillatory language learners to be prudent does not result in any loss of learning power. This result, together with previous results of Fulk and of Kurtz and Royer, settles the question of prudence for the three popularly investigated criteria of successful language acquisition in formal language learning theory. We would like to note that Kurtz and Royer [KR88] reported that they can make Fulk's proof of **TxtEx**-prudence constructive; it is an interesting open question if their techniques can be adapted to make our proof constructive.

5 Acknowledgements

We thank the referees for many valuable suggestions for improvement in the presentation of the paper. We would also like to thank R. Wiehagen for bringing the reference [JB81] to our attention.

References

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

[BH70] R. Brown and C. Hanlon. Derivational complexity and the order of acquisition in child speech. In J. R. Hayes, editor, *Cognition and the Development of Language*. Wiley, 1970.

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

[Cas88] 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 Publishers, Inc., 1988. Expanded in [Cas92].

[Cas92] John Case. The power of vacillation in language learning. Technical Report 93-08, University of Delaware, 1992. Expands on [Cas88]; journal article under review.

[CL82] J. Case and C. Lynes. Machine inductive inference and language identification. *Lecture Notes in Computer Science*, 140:107–115, 1982.

[DPS86] M. Demetras, K. Post, and C. Snow. Feedback to first language learners: The role of repetitions and clarification questions. *Journal of Child Language*, 13:275–292, 1986.

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

[Ful90] M. Fulk. Prudence and other conditions on formal language learning. *Information and Computation*, 85:1–11, 1990.

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

[HPTS84] K. Hirsh-Pasek, R. Treiman, and M. Schneiderman. Brown and hanlon revisited: Mothers' sensitivity to ungrammatical forms. *Journal of Child Language*, 11:81–88, 1984.

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

[JB81] K. Jantke and H. Beick. Combining postulates of naturalness in inductive inference. *Electronische Informationverarbeitung und Kybernetik*, 17:465–484, 1981.

[KR88] S.A. Kurtz and J.S. Royer. Prudence in language learning. In D. Haussler and L. Pitt, editors, *Proceedings of the Workshop on Computational Learning Theory*, pages 143–156. Morgan Kaufmann Publishers, Inc., 1988.

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

[OSW82] D. Osherson, M. Stob, and S. Weinstein. Learning strategies. *Information and Control*, 53:32–51, 1982.

[OSW84] D. Osherson, M. Stob, and S. Weinstein. Learning theory and natural language. *Cognition*, 17:1–28, 1984.

[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.

[OW92] D. Osherson and S. Weinstein. On the study of first language acquisition. Manuscript, 1992.

[Pen87] S. Penner. Parental responses to grammatical and ungrammatical child utterances. *Child Development*, 58:376–384, 1987.

[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, New York, 1967. Reprinted, MIT Press 1987.

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

[Wex82] K. Wexler. On extensional learnability. *Cognition*, 11:89–95, 1982.