Learning Recursive Languages with Bounded Mind Changes

Steffen Lange*
TH Leipzig
FB Mathematik und Informatik
PF 66
O–7030 Leipzig
steffen@informatik.th-leipzig.de

Thomas Zeugmann
TH Darmstadt
Institut für Theoretische Informatik
Alexanderstr. 10
W–6100 Darmstadt
zeugmann@iti.informatik.th-darmstadt.de

September 1992
Key words: Learning Theory, Automata and Formal Languages

Abstract

In the present paper we study the learnability of enumerable families $\mathcal{L}$ of uniformly recursive languages in dependence on the number of allowed mind changes, i.e., with respect to a well–studied measure of efficiency.

We distinguish between *exact* learnability ($\mathcal{L}$ has to be inferred w.r.t. $\mathcal{L}$) and *class preserving* learning ($\mathcal{L}$ has to be inferred w.r.t. some suitable chosen enumeration of all the languages from $\mathcal{L}$) as well as between learning from *positive* and from both, *positive and negative* data.

The measure of efficiency is applied to prove the superiority of class preserving learning algorithms over exact learning. In particular, we considerably improve results obtained previously and establish two infinite hierarchies. Furthermore, we separate exact and class preserving learning from positive data that avoids *overgeneralization*. Finally, language learning with a bounded number of mind changes is completely characterized in terms of recursively generable finite sets. These characterizations offer a new method to handle overgeneralizations and resolve an open question of Mukouchi (1992).

1. Introduction

Inductive inference is the process of hypothesizing a general rule from eventually incomplete data. It has its historical origins in the philosophy of science. However, within the last three decades it received much attention from computer scientists. Nowadays inductive inference can be considered as a form of machine learning with potential applications to artificial intelligence (cf. e.g. Angluin and Smith (1983, 1987), Osherson, Stob and Weinstein (1986)).

The present paper deals with inductive inference of formal languages, a field in which many interesting and sometimes surprising results have been obtained (cf. e.g. Case and Lynes (1982), Case (1988), Fulk (1990)). Looking at potential applications it seemed reasonable to restrict ourselves to study language learning of families of uniformly recursive languages. Recently, this topic has attracted much attention (cf. e.g. Shinohara (1990), Kapur and Bilardi (1992),

*This research has been supported by the German Ministry for Research and Technology (BMFT) under grant no. 01 IW 101.

Lange and Zeugmann (1992), Mukouchi (1992)). The general situation investigated in language learning can be described as follows: Given more and more eventually incomplete information concerning the language to be learnt, the inference device has to produce, ¿from time to time, a hypothesis about the phenomenon to be inferred. The set of all admissible hypotheses is called space of hypotheses. Furthermore, the information given may contain only *positive examples*, i.e., exactly all the strings contained in the language to be recognized, as well as both *positive and negative examples*, i.e., the learner is fed with arbitrary strings over the underlying alphabet which are classified with respect to their containment to the unknown language. The sequence of hypotheses has to converge to a hypothesis correctly describing the object to be learnt. Consequently, the inference process is an ongoing one. If $d_1, d_2, ...$ denotes the sequence of data the inference machine M is successively fed with, then we use $h_1, h_2, ...$ to denote the corresponding hypotheses produced by M. We say that M changes its mind, or synonymously, M performs a mind change, iff $h_i \neq h_{i+1}$. The number of mind changes is a measure of efficiency and has been introduced by Barzdin and Freivalds (1972). Subsequently, this measure has been studied intensively. Barzdin and Freivalds (1974) proved the following remarkable result concerning inductive inference of enumerable classes of recursive functions. Gold's (1967) *identification by enumeration* technique yields successful inference *within the enumeration* but $n-1$ mind changes may be necessary to learn the nth function. On the other hand, there are a learning algorithm and a space of hypotheses such that the nth function in enumeration can be learnt with at most $O(\log n) + o(\log n)$ mind changes. This bound is optimal. Their result shows impressively that a careful choice of the space of hypotheses may considerably influence the efficiency of learning. Moreover, Case and Smith (1983) established a hierarchy in terms of mind changes and anomalies. Wiehagen, Freivalds and Kinber (1984) used the number of mind changes to prove advantages of probabilistic learning algorithms over deterministic ones. Gasarch and Velauthapillai (1992) in particular studied *active learning* in dependence on the number of mind changes.

Consequently, it is only natural to ask whether or not this measure of efficiency is of equal importance in language learning. Answering this question is by no means trivial, since, in general, at least inductive inference from positive data may behave totally different than inductive inference of recursive functions does (cf. e.g. Case (1988), Fulk (1990)). This is already caused by the fact that Gold's (1967) identification by enumeration technique does not necessarily succeed. The main new problem consists in detecting or avoiding overgeneralizations, i.e., hypotheses describing proper supersets of the target language. Mukouchi (1992) studied the power of mind changes for learning algorithms that infer indexed families of recursive languages *within the given enumeration*. Moreover, he characterized a special case of language learning with a bounded number of mind changes, i.e., in case that equality of languages within the given enumeration is decidable.

What we like to present in the sequel is an almost complete investigation of the power of mind changes. For the sake of presentation we introduce some notations. An indexed family $\mathcal{L}$ is said to be *exactly* learnable if there is a learning algorithm inferring $\mathcal{L}$ with respect to $\mathcal{L}$ itself. Furthermore, $\mathcal{L}$ is learnable by a *class preserving* learning algorithm M, if there is a space $\mathcal{G} = (G_j)_{j \in \mathbb{N}^+}$ of hypotheses such that any G_j describes a language from $\mathcal{L}$, and M infers $\mathcal{L}$ with respect to $\mathcal{G}$. In other words, any produced hypothesis is required to describe a language contained in $\mathcal{L}$ but we have the freedom to use a possibly *different enumeration* of $\mathcal{L}$ and possibly *different descriptions* of any $L \in \mathcal{L}$.

We compare exact and class preserving language learning in dependence on the allowed number of mind changes as well as in dependence on the choice of the space of hypotheses and on information presentation. In particular, we establish the strongest possible separation, i.e.,

we prove that there are indexed families $\mathcal{L}$ which are exactly learnable from positive data with at most $k+1$ mind changes but that are not class preservingly learnable from *positive and negative* data with at most k mind changes. This result sheds considerably more light on the power of one additional mind change than Mukouchi's (1992) hierarchy of exact learning in terms of mind changes (cf. Theorem 3). Furthermore, we compare exact and class preserving language learning avoiding overgeneralization and separate them (cf. Corollary 9). Applying the proof technique developed we show that exact language learning from positive data with a bounded number of mind changes is always less powerful than class preserving inference restricted to the same number of mind changes (cf. Theorem 10). Finally, we completely characterize class preserving language learning in terms of recursively generable finite sets (cf. Theorem 11 and 12). In particular, we offer a different possibility to handle overgeneralization than Angluin (1980) did.

The paper is structured as follows. Section 2 presents preliminaries. Separations are established in Section 3. The announced characterizations are presented in Section 4. Finally, we discuss the results obtained and present open problems.

2. Preliminaries

By $\mathbb{N} = \{0, 1, 2, 3, ...\}$ we denote the set of all natural numbers. Moreover, we set $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. In the sequel we assume familiarity with formal language theory (cf. e.g. Bucher and Maurer (1984)). By Σ we denote any fixed finite alphabet of symbols. Let Σ^* be the free monoid over Σ. The length of a string $s \in \Sigma^*$ is denoted by $|s|$. Any subset $L \subseteq \Sigma^*$ is called a language. By $co - L$ we denote the complement of L, i.e., $co - L = \Sigma^* \setminus L$. Let L be a language and $t = s_1, s_2, s_3, ...$ an infinite sequence of strings from Σ^* such that $range(t) = \{s_k \mid k \in \mathbb{N}^+\} = L$. Then t is said to be a *text* for L or, synonymously, a *positive presentation*. Furthermore, let $i = (s_1, b_1), (s_2, b_2), ...$ be a sequence of elements of $\Sigma^* \times \{+, -\}$ such that $range(i) = \{s_k \mid k \in \mathbb{N}^+\} = \Sigma^*$, $i^+ = \{s_k \mid (s_k, b_k) = (s_k, +), k \in \mathbb{N}^+\} = L$ and $i^- = \{s_k \mid (s_k, b_k) = (s_k, -), k \in \mathbb{N}^+\} = co - L$. Then we refer to i as an *informant*. If L is classified via an informant then we also say that L is represented by *positive and negative data*. Moreover, let t, i be a text and an informant, respectively, and let x be a number. Then t_x, i_x denote the initial segment of t and i of length x, respectively, e.g., $i_3 = (s_1, b_1), (s_2, b_2), (s_3, b_3)$. Let t be a text and let $x \in \mathbb{N}^+$. We write t_x^+ as an abbreviation for $range^+(t_x) := \{s_k \mid k \leq x\}$. Furthermore, by i_x^+ and i_x^- we denote the sets $range^+(i_x) := \{s_k \mid (s_k, +) \in i, k \leq x\}$ and $range^-(i_x) := \{s_k \mid (s_k, -) \in i, k \leq x\}$, respectively. Finally, we write $i_x \sqsubseteq i_y$, iff i_x is a prefix of i_y.

Following Angluin (1980) we restrict ourselves to deal exclusively with indexed families of recursive languages defined as follows:
A sequence $L_1, L_2, L_3, ...$ is said to be an *indexed family* $\mathcal{L}$ of recursive languages provided all L_j are non–empty and there is a recursive function f such that for all numbers $j \in \mathbb{N}^+$ and all strings $s \in \Sigma^*$ we have

$$f(j, s) = \begin{cases} 1 & , \quad if \quad s \in L_j \\ 0 & , \quad otherwise. \end{cases}$$

In the sequel we often denote an indexed family and its range by the same symbol $\mathcal{L}$. What is meant will be clear from the context.

As in Gold (1967) we define an *inductive inference machine* (abbr. IIM) to be an algorithmic device which works as follows: The IIM takes as its input larger and larger initial segments of a text t (an informant i) and it either requires the next input string, or it first outputs a hypothesis, i.e., a number encoding a certain computer program, and then it requires the next

input string (cf. e.g. Angluin (1980)).

At this point we have to clarify what space of hypotheses we should choose, thereby also specifying the goal of the learning process. Gold (1967) and Wiehagen (1977) pointed out that there is a difference in what can be inferred in dependence on whether we want to synthesize in the limit grammars (i.e., procedures generating languages) or decision procedures, i.e., programs of characteristic functions. Case and Lynes (1982) investigated this phenomenon in detail. As it turns out, IIMs synthesizing grammars can be more powerful than those ones which are requested to output decision procedures. However, in the context of identification of indexed families both concepts are of equal power. Nevertheless, we decided to require the IIMs to output grammars, since this learning goal fits better with the intuitive idea of language learning. Furthermore, since we exclusively deal with indexed families $\mathcal{L} = (L_j)_{j \in \mathbb{N}^+}$ of recursive languages we almost always take as space of hypotheses an enumerable family of grammars $G_1, G_2, G_3, ...$ over the terminal alphabet Σ satisfying $\mathcal{L} = \{L(G_j) \mid j \in \mathbb{N}^+\}$. Moreover, we require that membership in $L(G_j)$ is uniformly decidable for all $j \in \mathbb{N}^+$ and all strings $s \in \Sigma^*$. As it turns out, it is sometimes very important to choose the space of hypotheses appropriately in order to achieve the desired learning goal. Then, the IIM outputs numbers j which we interpret as G_j.

A sequence $(j_x)_{x \in \mathbb{N}^+}$ of numbers is said to be convergent in the limit if and only if there is a number j such that $j_x = j$ for almost all numbers x.

Definition 1. (Gold, 1967) *Let $\mathcal{L}$ be an indexed family of languages, $L \in \mathcal{L}$, and let $\mathcal{G} = (G_j)_{j \in \mathbb{N}^+}$ be a space of hypotheses. An IIM M $LIM-TXT$ $(LIM-INF)$-identifies L on a text t (an informant i) with respect to $\mathcal{G}$ iff it almost always outputs a hypothesis and the sequence $(M(t_x))_{x \in \mathbb{N}^+}$ $((M(i_x))_{x \in \mathbb{N}^+})$ converges in the limit to a number j such that $L = L(G_j)$. Moreover, M $LIM-TXT$ $(LIM-INF)$-identifies L, iff M $LIM-TXT$ $(LIM-INF)$-identifies L on every text (informant) for L. We set:*
$LIM-TXT(M) = \{L \in \mathcal{L} \mid M\ LIM-TXT-identifies\ L\}$ and define $LIM-INF(M)$ analogously.
Finally, let $LIM-TXT$ $(LIM-INF)$ denote the collection of all families $\mathcal{L}$ of indexed families of recursive languages for which there is an IIM M such that $\mathcal{L} \subseteq LIM-TXT(M)$ $(\mathcal{L} \subseteq LIM-INF(M))$.

Definition 1 could be easily generalized to arbitrary families of recursively enumerable languages (cf. Osherson et al. (1986)). Nevertheless, we exclusively consider the restricted case defined above, since our motivating examples are all indexed families of recursive languages. Moreover, it may be well conceivable that the weakening of $\mathcal{L} = \{L(G_j) \mid j \in \mathbb{N}^+\}$ to $\mathcal{L} \subseteq \{L(G_j) \mid j \in \mathbb{N}^+\}$ may increase the collection of inferable indexed families. However, it does not, as the following proposition shows.

Proposition 1. *Let $\mathcal{L}$ be an indexed family and let $\mathcal{G} = (G_j)_{j \in \mathbb{N}^+}$ be any space of hypotheses such that $\mathcal{L} \subseteq \{L(G_j) \mid j \in \mathbb{N}^+\}$ and membership in $L(G_j)$ is uniformly decidable. Then we have: If there is an IIM M inferring $\mathcal{L}$ on text (informant) with respect to $\mathcal{G}$, then there is also an IIM $\hat{M}$ that learns $\mathcal{L}$ on text (informant) with respect to $\mathcal{L}$.*

Proof. Note that the following proof uses an idea of Gold (1967) (cf. Theorem 10.1., pp. 464). Let $\mathcal{L} \in LIM-TXT(M)$ w.r.t. $\mathcal{G}$. The wanted IIM $\hat{M}$ is defined as follows. Let $L \in \mathcal{L}$, let t be any text of L, and let $x \in \mathbb{N}^+$. We define:

$\hat{M}(t_x) =$ "Simulate $M(t_x)$. In case M on t_x does request its next input, output nothing and request the next input.
Otherwise, let $j_x = M(t_x)$. Generate the initial segment i_x of length x of the lexicographically ordered informant i for $L(G_{j_x})$. Search for the least $k \leq x$ satisfying $i_x = i_x^{L_k}$,

where $i_x^{L_k}$ denotes the initial segment of the lexicographically ordered informant i^{L_k} of the language $L_k \in \mathcal{L}$ of length x. In case it is found, output k. Otherwise do not output a guess and request the next input."

It remains to show that $\mathcal{L} \in LIM - TXT(\hat{M})$ w.r.t. $\mathcal{L}$. First, since membership in $L(G_j)$ is uniformly decidable and since $\mathcal{L}$ is an indexed family, we immediately see that $\hat{M}$ is indeed an IIM. Now let $L \in \mathcal{L}$ be arbitrarily fixed and let t be any text for L. By assumption M converges on t to an index j such that $L = L(G_j)$. Hence, there is an x_1 such that $M(t_x) = j$ for all $x \geq x_1$. Moreover, since $L \in \mathcal{L}$ there exists an x_2 with $x_2 = \mu z[L = L_z]$. Therefore, for all $x \geq max\{x_1, x_2\}$ the IIM $\hat{M}$ outputs a hypothesis. Finally, since $x_2 = \mu z[L = L_z]$, the IIM $\hat{M}$ eventually rejects any guess $k < x_2$, and verifies $i_x = i_x^{L_{x_2}}$ afterwards. Hence, $\hat{M}$ converges in the limit to x_2.

q.e.d.

However, the proof of Proposition 1 does not preserve the number of mind changes. As we shall see later, the efficiency of learning may be well influenced by the choice of the space of hypotheses.

Note that, in general, it is not decidable whether or not M has already inferred L. Within the next definition, we consider the special case that it has to be decidable whether or not an IIM has successfully finished the learning task.

Definition 2, (Trakhtenbrot and Barzdin, 1970) *Let $\mathcal{L}$ be an indexed family of languages, $L \in \mathcal{L}$, and let $\mathcal{G} = (G_j)_{j \in \mathbb{N}^+}$ be a space of hypotheses. An IIM M $FIN-TXT$ $(FIN-INF)$–identifies L on a text t (an informant i) with respect to $\mathcal{G}$ iff it outputs only a single and correct hypothesis j, i.e., $L = L(G_j)$, and stops.*
Moreover, M $FIN-TXT$ $(FIN-INF)$–identifies L, iff M $FIN-TXT$ $(FIN-INF)$–identifies L on every text (informant) for L. We set:
$FIN-TXT(M) = \{L \in \mathcal{L} \mid M\ FIN-TXT-identifies\ L\}$ and define $FIN-INF(M)$ analogously.

The resulting identification type is denoted by $FIN-TXT$ $(FIN-INF)$.

Note that finite identification can always be achieved with respect to the target indexed family $\mathcal{L}$, too (cf. Lange and Zeugmann (1992)).

The next definition shows a natural way of weakening the requirement of finite identification. Here, the number of mind changes which an IIM M may perform when inferring a target language is bounded by a number fixed a priori. When dealing with mind changes it is technically much more convenient to require the IIMs to behave as follows. Let t be any text (i be any informant), and $x \in \mathbb{N}^+$. If M on t_x (i_x) outputs for the first time a guess, then it has to output at any subsequent step a hypothesis. It is easy to see that any IIM M may be straightforwardly converted into an IIM $\hat{M}$ behaving as required such that both machines produce the same sequence of mind changes.

Definition 3. (Barzdin and Freivalds, 1974) *Let $\mathcal{L}$ be an indexed family of languages, $\mathcal{G} = (G_j)_{j \in \mathbb{N}^+}$ a space of hypotheses, $k \in \mathbb{N} \cup \{*\}$, and $L \in \mathcal{L}$. An IIM M LIM_k-TXT (LIM_k-INF)–identifies L on text t (informant i) with respect to $\mathcal{G}$ iff for every text t (informant i) the following conditions are fulfilled:*

(1) $L \in LIM-TXT(M)$ $\quad (L \in LIM-INF(M))$

(2) For any $L \in \mathcal{L}$ and any text t (informant i) of L the IIM M performs, when fed with t (i), at most k ($k = *$ means at most finitely many) mind changes, i.e., $card(\{x \mid M(t_x) \neq$

$M(t_{x+1})\}) \leq k$ *(*$card(\{x \mid M(i_x) \neq M(i_{x+1})\}) \leq k$*).*

$LIM_k - TXT(M)$, $LIM_k - INF(M)$ as well as $LIM_k - TXT$ and $LIM_k - INF$ are defined in the same way as above.

Obviously, $FIN - TXT = LIM_0 - TXT$ and $FIN - INF = LIM_0 - INF$. Moreover, it is easy to see that $LIM_* - TXT = LIM - TXT$ as well as $LIM_* - INF = LIM - INF$.

Next to we sharpen Definition 1 in additionally requiring that any mind change has to be caused by a "provable misclassification" of the hypothesis to be rejected.

Definition 4. (Angluin, 1980) *Let $\mathcal{L}$ be an indexed family, $L \in \mathcal{L}$, and let $\mathcal{G} = (G_j)_{j \in \mathbb{N}^+}$ be a space of hypotheses. An IIM M CONSERVATIVE–TXT (CONSERVATIVE–INF)–identifies L on text t (on informant i) with respect to $\mathcal{G}$ iff for every text t (informant i) the following conditions are satisfied:*

(1) $L \in LIM - TXT(M) \quad (L \in LIM - INF(M))$

(2) If M on input t_x makes the guess j_x and then makes the guess $j_{x+k} \neq j_x$ at some subsequent step, then $L(G_{j_x})$ must fail to contain some string from t_{x+k} ($L(G_{j_x})$ must fail either to contain some string $s \in i^+_{x+k}$ or it generates some string $s \in i^-_{x+k}$).

CONSERVATIVE–TXT(M) and *CONSERVATIVE–INF(M)* as well as the collections of sets *CONSERVATIVE–TXT* and *CONSERVATIVE–INF* are defined in an analogous manner as above.

Intuitively, a conservatively working IIM performs *exclusively* justified mind changes.

For any mode of inference defined above we use the prefix E to denote exact learning, i.e., the fact that $\mathcal{L}$ has to be inferred with respect to $\mathcal{L}$ itself. For example, $ELIM_k - TXT$ denotes exact learnability with at most k mind changes from text. Despite the fact that $LIM - TXT = ELIM - TXT$, $FIN - TXT = EFIN - TXT$ as well as $LIM - INF = ELIM - INF$, $FIN - INF = EFIN - INF$ none of the analogous statements is true for the other modes of inference defined above, as we shall see.

3. Separations

The aim of the present chapter is to relate the different types of language learning defined above one to the other.

Theorem 1. (Mukouchi, 1992)

$$ELIM_0 - TXT \subset ELIM_1 - TXT \subset ELIM_2 - TXT \subset ... \subset ELIM_* - TXT$$

First, we want to strengthen the theorem above in two directions. But before doing this we introduce the notion of canonical text defined as follows.

Let $\mathcal{L}$ be an indexed family and let L any of its languages. Moreover, let $s_1, s_2, ...$ be the lexicographically ordered text of Σ^*. The canonical text t of L is obtained as follows. Test sequentially whether $s_z \in L$ for $z = 1, 2, 3, ...$ until the first z is found such that $s_z \in L$. Since $L \neq \emptyset$ there must be at least one z fulfilling the test. Set $t_1 = s_z$. We proceed inductively:

$$t_{x+1} = \begin{cases} t_x s_{z+x+1} & \text{if} \quad s_{z+x+1} \in L \\ t_x s & \text{otherwise, where } s \text{ is the last string in } t_x \end{cases}$$

This notion is very useful in proving several theorems below, since its helps considerably to overcome the technical difficulty to deal simultaneously with finite and infinite languages.

Now we ready to present the first announced sharpening of Theorem 1.

Theorem 2. $\bigcup_{k\in\mathbb{N}} LIM_k - TXT \subset LIM - TXT$

Proof. Obviously, from Definition 3 it follows $\bigcup_{k\in\mathbb{N}} LIM_k - TXT \subseteq LIM - TXT$. It remains to show that $LIM - TXT \setminus \bigcup_{k\in\mathbb{N}} LIM_k - TXT \neq \emptyset$. Consider the indexed family $\mathcal{L}$ with $L_k = \{a\}^* \setminus \{a^k\}$. In Lange and Zeugmann (1991) it was shown that $\mathcal{L} \in LIM - TXT$. It is easy to recognize that there does not exist any $k \in \mathbb{N}$ with $\mathcal{L} \in LIM_k - TXT$. Suppose the converse, i.e. there is a k such that $\mathcal{L} \in LIM_k - TXT$. Let M be any IIM witnessing $\mathcal{L} \in LIM_k - TXT$. Let us consider M's behavior on the following text t_{fool}. We feed $a^2, a^3, ...$ to M until it outputs a guess j_{z_1}, say on $a^2, a^3, ..., a^{z_1}$, such that $L(G_{j_{z_1}}) = L_1$. In case it does not, we are done, since M does not identify L_1 on its canonical text. Then we extend this initial segment of text with $a, a^{z_1+2}, a^{z_1+3}, ...$ until M produces a correct guess j_{z_2} for L_{z_1+1}. Again there has to be an z_2 such that M on $a^2, a^3, ..., a^{z_1}, a, a^{z_1+2}, a^{z_1+3}, ..., a^{z_2}$ outputs j_{z_2}, since otherwise M never infers L_{z_2+1}. Obviously, the same idea can be iterated until we obtain a text t_{fool} for a language $L \in \mathcal{L}$ on which M changes at least $k+1$ times its mind. This yields the desired contradiction.

q.e.d.

Our next theorem shows that, in general, one additional mind change can neither be traded versus information presentation nor versus an appropriate choice of the space of hypotheses.

Theorem 3. *For all* $k \geq 0$: $ELIM_{k+1} - TXT \setminus LIM_k - INF \neq \emptyset$

Proof. Let $k \geq 0$. By $\mathcal{L}_{LIM_{k+1}}$ we denote the set of all finite languages with at most $k+2$ elements over the alphabet $\Sigma = \{a\}$. Assume any appropriate enumeration $(L_j)_{j\in\mathbb{N}^+}$ of $\mathcal{L}_{LIM_{k+1}}$. In order to construct an IIM M exactly inferring $\mathcal{L}_{LIM_{k+1}}$, we define a family of non–empty, finite and recursive sets $(T_j)_{j\in\mathbb{N}^+}$ as follows: For any $j \geq 1$ set $T_j = L_j$. Then M may be defined mutatis mutandis as in Angluin's (1980) proof of her Theorem 1. It is easy to verify that M exactly infers any $L \in \mathcal{L}_{LIM_{k+1}}$ with at most $k+1$ mind changes on any text.

We proceed in showing $\mathcal{L}_{LIM_{k+1}} \notin LIM_k - INF$. Suppose the converse, i.e., there is an IIM M which infers $\mathcal{L}_{LIM_{k+1}}$ on informant with at most k mind changes. Let i denote the lexicographically ordered informant of $L = \{a\}$. Since M infers L on i, there has to be some x such that $M(i_x) = j_x$ with $L = L(G_{j_x})$. Obviously, i_x may be extended to become the lexicographically ordered informant of $\tilde{L} = \{a, a^{x+1}\}$. Hence, M has to perform at least one mind change when inferring $\tilde{L}$ on its lexicographically informant. (* This directly yields the contradiction for the subcase $k = 0$. *) By iterating the same argument, one can easily show that there is a language $\hat{L} \in \mathcal{L}_{LIM_{k+1}}$ with exactly $k+2$ elements such that M performs $k+1$ mind changes when inferring $\hat{L}$ on its lexicographically ordered informant. This yields a contradiction.

q.e.d.

The following hierarchy is an immediate consequence of the latter theorems.

$$LIM_0 - TXT \subset LIM_1 - TXT \subset ... \subset LIM_k - TXT \subset ... \subset \bigcup_{k\in\mathbb{N}} LIM_k - TXT \subset LIM - TXT$$

In Lange and Zeugmann (1991) we have shown that $FIN - INF \subset CONSERVATIVE\text{–}TXT$. Surprisingly, it makes a real difference, if an IIM is allowed to change its mind at most one time.

Theorem 4. $ELIM_1 - INF \setminus LIM - TXT \neq \emptyset$

Proof. We define $\mathcal{L} = L_1, L_2, ...$ as follows: $L_1 = \{a\}^*$, $L_k = \{a^n \mid 1 \leq n \leq k\}$ with $k \geq 2$. It

is folklore that $\mathcal{L} \notin LIM - TXT$. An IIM M which identifies $\mathcal{L}$ with at most one mind change on informant may work as follows: As long as some $(a^n, -)$ does not appear in the informant, the machine outputs a grammar for L_1. If a pair $(a^n, -)$ is presented, M repeats the current hypothesis until $(a^k, +)$ and $(a^{k+1}, -)$ have been seen. In this case, M outputs a grammar for L_k. This hypothesis will be output in any subsequent step. Obviously, M infers $\mathcal{L}$ with at most one mind change on informant.

q.e.d.

Moreover, $LIM_k - TXT \subseteq LIM_k - INF$. Since $FIN - TXT \subset FIN - INF$ (cf. Lange and Zeugmann (1991)), Theorem 2 allows the following observation.

Corollary 5.

(1) *For all* $k \geq 0$, $LIM_k - TXT \subset LIM_k - INF$

(2) $LIM_0 - INF \subset LIM_1 - INF \subset LIM_2 - INF \subset ... \subset LIM_* - INF$

As above, this result can be sharpened, too.

Lemma 6. $\bigcup_{k \in \mathbb{N}} LIM_k - INF \subset LIM - INF$

Proof. From Definition 3 it directly follows that $\bigcup_{k \in \mathbb{N}} LIM_k - INF \subseteq LIM - INF$. It remains to show that $LIM - INF \setminus \bigcup_{k \in \mathbb{N}} LIM_k - INF \neq \emptyset$. By $\mathcal{L}_{fin}$ we denote any appropriate enumeration of all finite sets over the alphabet $\Sigma = \{a\}$. Obviously, $\mathcal{L}_{fin} \in LIM - INF$. It is easy to recognize that there does not exist any $k \in \mathbb{N}$ with $\mathcal{L} \in LIM_k - INF$.

q.e.d.

The latter result has the following consequence.

Corollary 7. $LIM - TXT \# \bigcup_{k \in \mathbb{N}} LIM_k - INF$.

Proof. Obviously, $\mathcal{L}_{fin} \in LIM - TXT$. Hence, we conclude $\mathcal{L}_{fin} \in LIM - TXT \setminus \bigcup_{k \in \mathbb{N}} LIM_k - INF$. Applying Theorem 3 we immediately get $\bigcup_{k \in \mathbb{N}} LIM_k - INF \setminus LIM - TXT \neq \emptyset$.

q.e.d.

Summarizing the results above we obtain the following hierarchy:

$$LIM_0 - INF \subset LIM_1 - INF \subset ... \subset LIM_k - INF \subset ... \subset \bigcup_{k \in \mathbb{N}} LIM_k - INF \subset LIM - INF$$

Finally, we want to compare exact and class preserving language learning with an a priori bounded number of mind changes. Moreover, we compare conservatively working IIMs with those ones performing a bounded number of mind changes. The next theorems and a corollary thereof relate these different modes of inference one to the other. In particular, we show class preserving learning with at most one mind change has to be performed by conservatively working IIMs. Moreover, one mind change is already sufficient to beat exact conservative learning.

Theorem 8.

(1) $LIM_1 - TXT \subset$ *CONSERVATIVE-TXT*

(2) $LIM_1 - TXT \setminus$ *ECONSERVATIVE-TXT* $\neq \emptyset$

Proof. It is easy to see that $LIM_1 - TXT \subseteq$ *CONSERVATIVE-TXT*. This is forced by the fact that an IIM which is allowed to perform at most one mind change may never produce an overgeneralized hypothesis.

Assertion (2) is much harder to prove. In order to define the wanted indexed family we fix any Gödel numbering φ_1, φ_2, φ_3, ... of the partial recursive functions over $\mathbb{N}^+$. As usual, we use the following abbreviations. Let $k \in \mathbb{N}^+$. By $\varphi_k(k)\downarrow$ we denote that $\varphi_k(k)$ is defined. If $\varphi_k(k)$ is not defined, we write $\varphi_k(k)\uparrow$. Furthermore, we use $\varphi_k^{(j)}(k)\downarrow$, if $\varphi_k(k)$ can be computed within at most j steps. Otherwise, we write $\varphi_k^{(j)}(k)\uparrow$. By $c : \mathbb{N}^+ \times \mathbb{N}^+ \to \mathbb{N}^+$ we denotes Cantor's pairing function.

The desired indexed family is defined as follows. For all $k, j \in \mathbb{N}^+$:

$L_{c(k,1)} = \{a^k b^n \mid n \in \mathbb{N}^+\}$, and for $j > 1$ let

$$L_{c(k,j)} = \begin{cases} L_{c(k,1)} & if\ \varphi_k^{(j)}(k)\uparrow \\ \{a^k b^m \mid m \leq z\} & if\ z = \mu n\{n \mid n \leq j\ ,\ \varphi_k^{(n)}(k)\downarrow\} \end{cases}$$

Obviously, $\mathcal{L}$ is an indexed family of recursive languages. The definition above has the following consequences. First, for any $k,\ j \in \mathbb{N}^+$, if $a^k b^{j+1} \notin L_{c(k,j)}$, then $L_{c(k,j)}$ is a finite language. Hence, it is easy to decide whether or not $L_{c(k,j)}$ is finite. This will help to prove Claim A. Second, for any $k,\ j \in \mathbb{N}^+$, $a^k b^j \in L_{c(k,j)}$ if and only if $a^k b^j \in L_{c(k,j+r)}$, for any $r \in \mathbb{N}^+$. This will be used when proving Claim B.

Claim A. $\mathcal{L} \notin$ *ECONSERVATIVE–TXT*

Suppose the converse, i.e., $\mathcal{L} \in$ *ECONSERVATIVE–TXT*. Without loss of generality, we may assume that there is an IIM M which conservatively and consistently infers $\mathcal{L}$ w.r.t. the space of hypotheses $\mathcal{L}$ (cf. Lange and Zeugmann (1991)). We shall show that M decides the halting problem for φ, i.e., M may be used for deciding whether or not $\varphi_k(k)$ is defined. Let $k \in \mathbb{N}^+$, and let t be the canonical text for $L_{c(k,1)}$. Since M in particular infers L on t, there is an minimal z such that M on input t_z computes a hypothesis. Moreover, because M works consistently, we obtain $M(t_z) = c(k,j)$ for some $j \in \mathbb{N}^+$. First of all, we have to test whether or not $a^k b^{j+1} \in L_{c(k,j)}$. This test can be effectively performed.

Case 1. $a^k b^{j+1} \notin L_{c(k,j)}$
Then $L_{c(k,j)}$ is finite. Due to our construction, we obtain $\varphi_k(k)\downarrow$.

Case 2. $a^k b^{j+1} \in L_{c(k,j)}$
By definition of $\mathcal{L}$ we obtain that $L_{c(k,j)}$ is infinite, i.e., $L_{c(k,j)} = L_{c(k,1)}$. For deciding whether or not $\varphi_k(k)$ is defined, we have to compare j and z.

Subcase 2.1. $j \geq z$
We claim that $\varphi_k(k)\uparrow$. Suppose the converse. Then due to our construction, there has to be an $n > j$ such that $L_{c(k,n)}$ is finite. Since $n > j \geq z$, t_z is an initial segment of the canonical text for $L_{c(k,n)}$. Hence, M has produced an overgeneralized hypothesis when inferring $L_{c(k,n)}$ on its canonical text. This contradicts our assumption that M infers $\mathcal{L}$ exactly and conservatively.

Subcase 2.2. $j < z$
Now, one has to test whether or not $a^k b^{z+1} \in L_{c(k,z)}$. If not, $L_{c(k,z)}$ is finite. Hence, $\varphi_k(k)\downarrow$. If $a^k b^{z+1} \in L_{c(k,z)}$, then $L_{c(k,z)} = L_{c(k,1)}$. Now, one may easily conclude that $\varphi_k(k)\uparrow$. Assume the opposite. Then there has to be an $n > z$ such that $L_{c(k,n)}$ is finite. Since $n > z$, t_z is again an initial segment of the canonical text for $L_{c(k,n)}$ which yields an contradiction as in Subcase 2.1.

Hence, there is indeed no IIM at all inferring $\mathcal{L}$ exactly and conservatively.

Claim B. $\mathcal{L} \in LIM_1 - TXT$
We define an indexed family $\hat{\mathcal{L}} = (\hat{L}_j)_{j \in \mathbb{N}^+}$ such that $range(\hat{\mathcal{L}}) = range(\mathcal{L})$. Afterwards, we

show that there is an IIM M which weak-monotonically infers $\mathcal{L}$ w.r.t. the space of hypotheses $\hat{\mathcal{L}}$. Let $k \in \mathbb{N}^+$.

$$\hat{L}_{2k-1} = L_{c(k,1)}$$

$$\hat{L}_{2k} = \bigcap_{j \in \mathbb{N}^+} L_{c(k,j)}$$

First of all, we show that $\hat{\mathcal{L}}$ is an indexed family satisfying the announced properties. Due to our construction, we may immediately conclude that $range(\hat{\mathcal{L}}) = range(\mathcal{L})$. Moreover, it is easy to verify that, for any number k and any string s, we have $s \in \hat{L}_{2k}$ if and only if there is an j such that $s = a^k b^j$ as well as $s \in L_{c(k,j)}$. Since $\mathcal{L}$ is an indexed family, it is not hard to prove that $\hat{\mathcal{L}}$ is an indexed family, too.

It remains to show that there is an IIM M which performs at most one mind change when inferring $\mathcal{L}$ w.r.t. the space of hypotheses $\hat{\mathcal{L}}$. Assume any $L \in \mathcal{L}$, any text t for L, and any $x \in \mathbb{N}^+$. We set:

$$M(t_x) = \begin{cases} 2k & if\ t_x^+ \subseteq \hat{L}_{2k} \\ 2k-1 & otherwise \end{cases}$$

Since $\hat{\mathcal{L}}$ is an indexed family, M is indeed an IIM. It is easy to recognize that M performs at most one mind change. If L is a finite language, then M outputs in every step a correct hypothesis. If L is an infinite language, M may compute at the beginning a wrong hypothesis j. But in this case, $\hat{L}_j$ is finite and, therefore, there has to be a y such that $t_y^+ \not\subseteq \hat{L}_j$. Due to our definition, M changes its mind to a correct hypothesis for L which will be repeated in every subsequent step. Hence, M infers L on any text with at most one mind change.

q.e.d.

The following corollary is an immediate consequence of the latter theorem.

Corollary 9.

(1) $ECONSERVATIVE\text{-}TXT \subset CONSERVATIVE\text{-}TXT$

(2) $ELIM_1 - TXT \subset LIM_1 - TXT$

Finally, we modify the proof technique introduced above and show the desired separation of exact and class preserving language learning with a bounded number of mind changes.

Theorem 10. For all $k \geq 1$, $ELIM_k - TXT \subset LIM_k - TXT$

Proof. First we separate $ELIM_2 - TXT$ and $LIM_2 - TXT$ thereby explaining the basic idea. Then we describe how the construction can be generalized. Let $\mathcal{L}$ be the indexed family introduced in the proof of assertion (2) of Theorem 8. For any $L_m \in \mathcal{L}$, $m \in \mathbb{N}^+$ we define $\tilde{L}_m = L_m \cup \{b\}$. Moreover, let $\mathcal{L}_{sep2}$ be any canonical enumeration of all the languages L_m and $\tilde{L}_m$ (cf. proof of Theorem 8). As above, we define mutatis mutandis a space $\hat{\mathcal{L}}$ of hypotheses such that $range(\hat{\mathcal{L}}) = range(\mathcal{L}_{sep2})$. Now it is straightforward to show in the same way as above that $\mathcal{L}_{sep2} \in LIM_2 - TXT$ w.r.t. $\hat{\mathcal{L}}$. We omit the details. It remains to prove that $\mathcal{L}_{sep2} \notin ELIM_2 - TXT$. This is indirectly done via the following claim.

Claim. If $\mathcal{L}_{sep2} \in ELIM_2 - TXT$, then $\mathcal{L} \in ELIM_1 - TXT$.

Let M be any IIM witnessing $\mathcal{L}_{sep2} \in ELIM_2 - TXT$. Furthermore, let $\tilde{L}_m = L_m \cup \{b\}$ be arbitrarily fixed, and let t be any text for L_m. Since M has to infer L_m too, there must be a

z such that $M(t_x) = j$ for all $x \geq z$ and $L_m = L_j$. But now we may feed the following text $\tilde{t}$ for $\tilde{L}_m$ to M. First, we input the initial segment t_z, then the string b, and subsequently all the remaining strings of t. By construction, $M(t_z) = j$ but $\tilde{L}_m \neq L_j$. Consequently, M has to perform one more mind change. That means, on t the machine has performed at most one mind change, since otherwise $\mathcal{L}_{sep2} \notin ELIM_2 - TXT(M)$. Hence, $\mathcal{L} \in ELIM_1 - TXT(M)$. This proves the claim.

On the other hand, as we have already seen, there is no IIM inferring $\mathcal{L}$ with at most one mind change. Thus, by contraposition of the claim made above we directly obtain that $\mathcal{L}_{sep2} \notin ELIM_2 - TXT$. This proves the theorem in case $k = 2$.

The general case is handled by induction. Suppose, we have already separated $ELIM_k - TXT$ and $LIM_k - TXT$ in using the indexed family $\mathcal{L}_{sepk}$. For any $m \in \mathbb{N}^+$ we set the language $\tilde{L}_m = L_m \cup \{b, b^2, ..., b^k\}$ and define $\mathcal{L}_{sepk+1}$ to be any canonical enumeration of all the languages from $\mathcal{L}_{sepk}$ and of all $\tilde{L}_m$. Again, one straightforwardly defines analogously as above a space $\hat{\mathcal{L}}$ of hypotheses and shows that $\mathcal{L}_{sepk+1} \in LIM_{k+1} - TXT$ w.r.t. $\hat{\mathcal{L}}$. Moreover, an easy modification of the demonstration above directly yields the following claim.

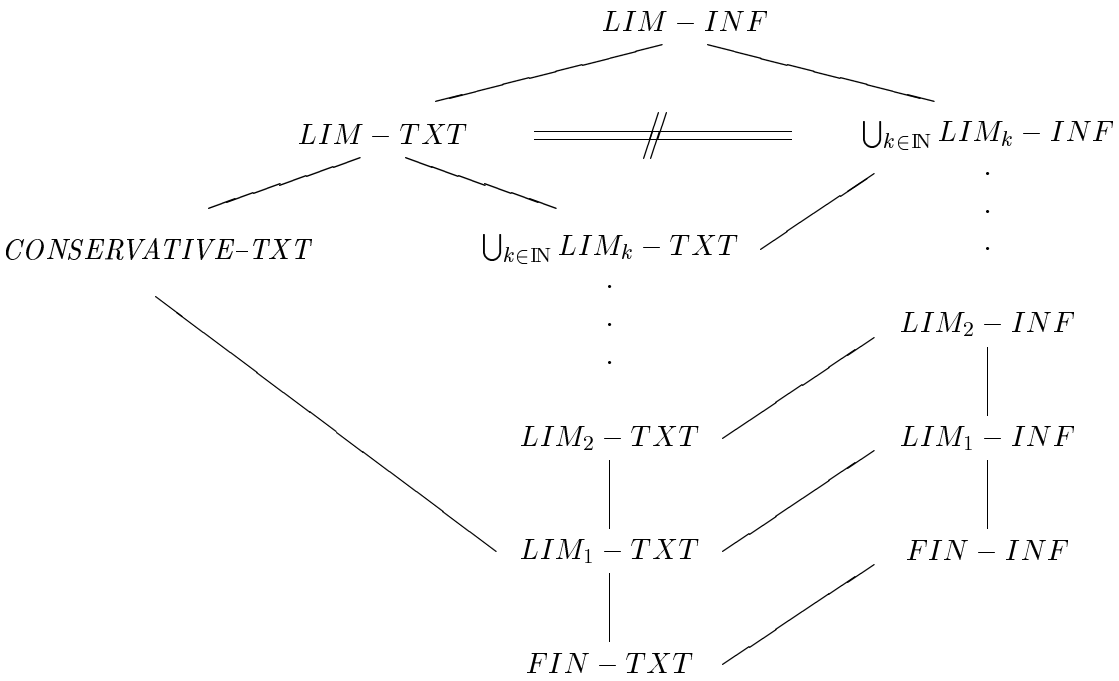
Claim. If $\mathcal{L}_{sepk+1} \in ELIM_{k+1} - TXT$, then $\mathcal{L}_{sepk} \in ELIM_k - TXT$.

Finally, the latter claim immediately yields a contradiction to the induction hypothesis that $\mathcal{L}_{sepk} \notin ELIM_k - TXT$. Hence, by contraposition of the claim one obtains $\mathcal{L}_{sepk+1} \notin ELIM_{k+1} - TXT$. We omit the details.

q.e.d.

However, some problems remained open. We shall discuss them in Section 5.

The following picture illustrates the results presented above.



All lines between identification types indicate inclusions in the sense that the upper type always properly contains the lower one. For the sake of readability, the ascending line from

$FIN - INF$ to *CONSERVATIVE–TXT* is not drawn in the figure above.

4. Characterization Theorems

Characterizations play an important role in that they lead to a deeper insight into the problem how algorithms performing the inference process may work (cf. e.g. Blum and Blum (1975), Wiehagen (1977), Angluin (1980), Zeugmann (1983), Jain and Sharma (1989)). Moreover, characterizations may help gain a better understanding of the properties objects should have in order to be inferable in the desired sense. A very illustrative example is Angluin's (1980) characterization of those indexed families for which learning in the limit from positive data is possible. In particular, this theorem provides insight into the problem how to deal with overgeneralizations. Our next theorem offers an alternative way to resolve this question. We characterize $LIM_k - TXT$ in terms of recursively generable finite tell–tales. A family of finite sets $(T_j)_{j\in\mathbb{N}^+}$ is said to be recursively generable, iff there is a total effective procedure g which, on input j, generates all elements of T_j and stops. If the computation of $g(j)$ stops and there is no output, then T_j is considered to be empty. Finally, for notational convenience we use $L(\mathcal{G})$ to denote $\{L(G_j) \mid j \in \mathbb{N}^+\}$ for any space $\mathcal{G} = (G_j)_{j\in\mathbb{N}^+}$ of hypotheses.

Theorem 11. *Let $\mathcal{L}$ be an indexed family of recursive languages, and $k \in \mathbb{N}$. Then: $\mathcal{L} \in LIM_k - TXT$ if and only if there is a space of hypotheses $\hat{\mathcal{G}} = (\hat{G}_j)_{j\in\mathbb{N}^+}$, a computable relation $\prec$ over $\mathbb{N}^+$, and a recursively generable family $(\hat{T}_j)_{j\in\mathbb{N}^+}$ of finite and non–empty tell-tale sets such that*

(1) $range(\mathcal{L}) = L(\hat{\mathcal{G}})$.

(2) For all $z \in \mathbb{N}^+$, $\hat{T}_z \subseteq L(\hat{G}_z)$.

(3) For all $L \in \mathcal{L}$, any $z \in \mathbb{N}^+$, and all finite $A \subseteq L$, if $\hat{T}_z \subseteq L$, $L(\hat{G}_z) \neq L$, then there is a j such that $z \prec j$, $A \subseteq \hat{T}_j$ and $L(\hat{G}_j) = L$.

(4) For all $L \in \mathcal{L}$, there is no sequence $(h_j)_{j\in\mathbb{N}^+}$ of length greater than $k+1$ such that $h_j \prec h_{j+1}$ as well as $\hat{T}_{h_j} \subseteq \hat{T}_{h_{j+1}} \subseteq L$.

Proof. First of all we prove the following technical lemma.

Lemma A. *Let $\mathcal{L}$ be an indexed family, $\mathcal{G} = (G_j)_{j\in\mathbb{N}^+}$ a space of hypotheses, and let $k \in \mathbb{N}$. If there is an IIM M such that $\mathcal{L} \in LIM_k - TXT(M)$ w.r.t. $\mathcal{G}$, then there is an IIM $\hat{M}$ such that*

(1) $\mathcal{L} \in LIM_k - TXT(\hat{M})$

(2) Whenever $\hat{M}$ on t_x produces a hypothesis j_x, then $t_x^+ \subseteq L(G_{j_x})$.

Proof. Let t be any text for some language $L \in \mathcal{L}$ and let $x \in \mathbb{N}^+$. When fed t_x the machine $\hat{M}$ simulates M's computation on t_y for all $y \leq x$. In case that M does not produce any hypothesis, $\hat{M}$ requests the next input and outputs nothing. Otherwise, let j_z be M's last guess. The machine $\hat{M}$ tests whether or not $t_x^+ \subseteq L(G_{j_z})$. In case it is, $\hat{M}$ outputs j_z and requests the next input. Otherwise, the IIM $\hat{M}$ outputs nothing and requests the next input. It is easy to see that $\hat{M}$ performs at most as many mind changes than M does. Finally, since M converges on t to a correct hypothesis j, at least the test $t_x^+ \subseteq L(G_j)$ is always satisfied. Hence $\hat{M}$ converges to j, too. This proves Lemma A.

If an IIM M fulfils property (2) of the lemma above, then we refer to M as a consistently working IIM. Now we are ready to prove the theorem.

Necessity: We start by defining the relation $\prec$. For this purpose some additional notation is needed. Let $\mathbb{N}^{+*}$ be the set of all finite sequences over $\mathbb{N}^+$, and for $\alpha \in \mathbb{N}^{+*}$ let $|\alpha|$ denote the length of α. Whenever appropriate, we interpret a number h as a bijective encoding of a 4-tupel $(n, \alpha, \beta, \gamma)$, where $n \in \mathbb{N}^+$, α, β, $\gamma \in \mathbb{N}^{+*}$. Let h, $j \in \mathbb{N}^+$. Then $h \preceq j$ iff $h = (n, \alpha, \beta, \gamma)$, $j = (m, \alpha\delta, \beta n\tau, \gamma\kappa)$, where $|\alpha| = |\beta| + 1 = |\gamma|$ as well as $|\delta| = |n\tau| = |\kappa|$. Moreover, $h \approx j$ iff $h = (n, \alpha, \beta, \gamma)$, $j = (m, \alpha\delta, \beta n\tau, , \gamma\kappa)$, where $|\alpha| = |\beta| + 1 = |\gamma|$ as well as $|\delta| = |n\tau| = |\kappa|$ and $range(\tau) = \{n\}$. Finally, $h \prec j$ iff $h \preceq j$ and not $h \approx j$. Note that $\prec$ is a transitive relation.

Let M be an IIM such that $\mathcal{L} \in LIM_k - TXT(M)$ with respect to some space $\mathcal{G} = (G_j)_{j \in \mathbb{N}^+}$ of hypotheses. Without loss of generality we may assume M to work consistently too (cf. Lemma A.)

For all $n \in \mathbb{N}^+$, $\alpha, \beta, \gamma \in \mathbb{N}^{+*}$ we define $\tilde{G}_{(n,\alpha,\beta,\gamma)} = G_n$. Moreover, we define $\tilde{T}_{(n,\alpha,\beta,\gamma)}$ as follows:

(i) If not $(|\alpha| = |\beta| + 1 = |\gamma|)$, then $\tilde{T}_{(n,\alpha,\beta,\gamma)} = \emptyset$.

(ii) If $(|\alpha| = |\beta| + 1 = |\gamma|)$, $\beta = \hat{\beta} n \rho$, $n \notin range(\hat{\beta})$, and $range(\rho) \neq \{n\}$, then $\tilde{T}_{(n,\alpha,\beta,\gamma)} = \emptyset$.

(iii) Otherwise let $\alpha = y_1, ..., y_r$, $\beta = n_1, ..., n_{r-1}$, $\gamma = z_1, ..., z_r$ and do the following:

Generate the canonical text $\hat{\sigma}_{n_1}$ of $L(G_{n_1})$ of length y_1 and compute $visible(\hat{\sigma}_{n_1}) = \{\tau \mid |\tau| \leq |\hat{\sigma}_{n_1}|, range(\tau) \subseteq range(\sigma_{n_1})\}$. If $|visible(\hat{\sigma}_{n_1})| < z_1$, then $\tilde{T}_{(n,\alpha,\beta,\gamma)} = \emptyset$. Else test whether the z_1st element σ_{n_1} (with respect to the lexicographical ordering) of $visible(\hat{\sigma}_{n_1})$, when fed successively to M, exactly yields the sequence n_1 of hypotheses. If not, then $\tilde{T}_{(n,\alpha,\beta,\gamma)} = \emptyset$, and stop.
In case it does, generate the canonical text $\hat{\sigma}_{n_2}$ of $L(G_{n_2})$ of length $y_1 + y_2$. Compute $visible(\hat{\sigma}_{n_2})$, and test whether $|visible(\hat{\sigma}_{n_2})| < z_2$. In case it is, set $\tilde{T}_{(n,\alpha,\beta,\gamma)} = \emptyset$ and stop. For if not, test whether the z_2nd element of σ_{n_2} (with respect to the lexicographical ordering) of $visible(\hat{\sigma}_{n_2})$ does fulfil the following properties:

(a) σ_{n_1} is a prefix of σ_{n_2} , and
(b) When fed successively with σ_{n_2}, the machine M exactly produces the sequence n_1, n_2 of guesses.

In case it does not, set $\tilde{T}_{(n,\alpha,\beta,\gamma)} = \emptyset$ and stop. Otherwise continue analogously.
Finally, if $\tilde{T}_{(n,\alpha,\beta,\gamma)}$ has not been defined til now, generate the canonical text $\hat{\sigma}_n$ of $L(G_n)$ of length $y_1 + ... + y_r$ and compute $visible(\hat{\sigma}_n)$. If $|visible(\hat{\sigma}_n)| < z_r$, then set $\tilde{T}_{(n,\alpha,\beta,\gamma)} = \emptyset$. Else let σ_n be the z_rth element of $visible(\hat{\sigma}_n)$ with respect to the lexicographical ordering. Test whether

(a) $\sigma_{n_{r-1}}$ is a prefix of σ_n
(b) When fed successively with σ_n, the machine M exactly produces the sequence $n_1, n_2, ..., n_{r-1}, n$ of guesses.

In case the test is not completely fulfilled set $\tilde{T}_{(n,\alpha,\beta,\gamma)} = \emptyset$. Otherwise, set $\tilde{T}_{(n,\alpha,\beta,\gamma)} = range(\sigma_n)$.

Obviously, $\tilde{T}_{(n,\alpha,\beta,\gamma)}$ is uniformly recursively generable and finite. The families $(\hat{T}_j)_{j \in \mathbb{N}^+}$ as well as $\hat{\mathcal{G}} = (\hat{G}_j)_{j \in \mathbb{N}^+}$ are obtained by simply striking off all $\tilde{T}_j$ that are empty as well as the corresponding $\tilde{G}_j$.

It remains to show that property (1) to (4) are satisfied. In order to prove (1), let $L \in \mathcal{L}$, and let $n \in \mathbb{N}^+$ be chosen such that M on L's canonical text t^L converges to n. Furthermore, let $y = \mu z[M(t^L_z) = n]$. We proceed in showing that there are $\alpha, \beta, \gamma \in \mathbb{N}^{+*}$ such that $\tilde{T}_{(n,\alpha,\beta,\gamma)} \neq \emptyset$. The latter statement yields (1), since we may conclude that $L(\hat{G}_{(n,\alpha,\beta,\gamma)}) = L(G_n) = L$. We define β to be the sequence $n_1, n_2, ..., n_r$ of M's hypotheses when successively fed with t^L_y where the last element n is deleted.
Let $\sigma_{n_1} \sqsubset \sigma_{n_2} \sqsubset ... \sqsubset \sigma_{n_r} \sqsubset t^L_y$ be the corresponding initial segments of text on which M produces its hypotheses ($* \sqsubset$ denotes proper prefix relation of finite sequences $*$). Since M works consistently, we obtain $\sigma^+_{n_j} \subseteq L(G_{n_j})$ for $j = 1, ..., r$. Compute the canonical text $t^{n_1}_{a_1}$, $t^{n_1}_{a_2}$, ... of $L(G_{n_1})$ of length $a_i = |\sigma_{n_1}| + i$, $i = 0, 1, ...$ until the least i with $\sigma_{n_1} \in visible(t^{n_1}_{a_i})$ has been found. Then let z_1 be the lexicographical number of σ_{n_1} with respect to $visible(t^{n_1}_{a_i})$, and set $y_1 = a_i$. Next generate the canonical text $t^{n_2}_{a_1}$, $t^{n_2}_{a_2}$, ... of $L(G_{n_2})$ of length $a_i = |\sigma_{n_2}| + i$ until $\sigma_{n_2} \in visible(t^{n_2}_{a_i})$ for the first time. Then define z_2 to be the lexicographical number of σ_{n_2} with respect to $visible(t^{n_2}_{a_i})$, and set $y_2 = a_i - y_1$, a.s.o.

Finally, suppose $y_1, ..., y_r$ as well as $z_1, ..., z_r$ are already defined. Then we set $y_{r+1} = max\{y,\ y_1 + ... + y_r\}$. Define z_{r+1} to be lexicographical number of t^L_y with respect to $visible(t^L_{y_{r+1}})$. Then, setting $\alpha = y_1, ..., y_{r+1}$, $\beta = n_1, ..., n_r$ and $\gamma = z_1, ..., z_{r+1}$ directly yields $\tilde{T}_{(n,\alpha,\beta,\gamma)} \neq \emptyset$. This proves property (1).

Assertion (2) directly follows from the construction described above.

The technique applied above to prove (1) applies mutatis mutandis to obtain (3). Hence we describe only the modification that has to be made. Let $A \subseteq L$ and let $\sigma = s_1, ..., s_m$ be the sequence of A's strings written in lexicographical order. Moreover, let $\hat{T}_h \subseteq L$ and $L(\hat{G}_h) \neq L$. Then there are $n \in \mathbb{N}^+$, $\alpha, \beta, \gamma \in \mathbb{N}^{+*}$ with $|\alpha| = |\beta| + 1 = |\gamma|$ such that $h = (n, \alpha, \beta, \gamma)$. Furthermore, let $\alpha = y_1, ..., y_r$, $\beta = n_1, ..., n_{r-1}$ $\gamma = z_1, ..., z_r$, $q = y_1 + ... + y_r$, and let $w_1, ..., w_q$ be the uniquely determined sequence of all elements of $\hat{T}_h$ on which, when fed successively to M, the machine M produces β as its sequence of hypotheses. Since $\hat{T}_h \subseteq L$ but $L \neq L(\hat{G}_h) = L(\hat{G}_{(n,\alpha,\beta,\gamma)})$, we conclude that M has not yet converged on this particular initial segment $w_1, ..., w_q$ of some text for L. Next we consider M's behavior when fed with $w_1, ..., w_q, \sigma$. There are two cases to distinguish, i.e., either the computation of $M(w_1, ..., w_q, \sigma)$ ends in M's request state, or it yields a guess. However, in both cases we extend $w_1, ..., w_q, \sigma$ with a sufficiently long initial segment t^L_y of L's canonical text until M outputs a hypothesis that is correct for L. Finally, j is obtained analogously as in the proof of property (1) where the construction is performed with respect to $w_1, ..., w_q, \sigma, t^L_y$ instead of t^L_y. Obviously, $h \prec j$, $A \subseteq \hat{T}_j$ and $L(\hat{G}_j) = L$. Hence (3) is proved.

We continue in proving property (4). We start with an observation. Recall the definition of the relation $\prec$. Since $h \prec j$, in particular there are $n, m \in \mathbb{N}^+$, $\alpha, \beta, \gamma, \delta, \tau, \kappa \in \mathbb{N}^{+*}$ such that $h = (n, \alpha, \beta, \gamma)$ and $j = (m, \alpha\delta, \beta n \tau, \gamma\kappa)$. Due to the definition of $\hat{T}_j$ we obtain an initial segment σ of text for L on which M, when successively fed with, sometime outputs n, and in some subsequent step m. In accordance with our construction we know that $L(\hat{G}_h) = L(\hat{G}_{(n,\alpha,\beta,\gamma)}) = L(G_n)$ and $L(\hat{G}_j) = L(\hat{G}_{(m,\alpha\delta,\beta n\tau,\gamma\kappa)}) = L(G_m)$. Thus, we directly see that $\hat{T}_h \subseteq \hat{T}_j$. This observation is used in demonstrating (4).

As above, if $h_j \prec h_{j+1}$, then there are $n, m \in \mathbb{N}^+$, $\alpha, \beta, \gamma, \delta, \tau, \kappa \in \mathbb{N}^{+*}$ such that $h_j = (n, \alpha, \beta, \gamma)$ and $h_{j+1} = (m, \alpha\delta, \beta n\tau, \gamma\kappa)$. Moreover, since not $h_j \approx h_{j+1}$, we additionally have $range(\tau) \neq \{n\}$. Now suppose there is a sequence $(h_j)_{j \in \mathbb{N}^+}$ such that $h_j \prec h_{j+1}$ and $\hat{T}_{h_j} \subseteq \hat{T}_{h_{j+1}} \subseteq L$ of length greater than $k+1$. Since $\hat{T}_h \subseteq \hat{T}_j$, we get a segment of a text t for L on which M changes its mind at least $k+1$ times, a contradiction. Hence, (4) is proved.

Sufficiency: The main problem we have to deal with is the detection and handling of over-generalized hypotheses. We proceed in defining an IIM M inferring $\mathcal{L}$ with respect to $\hat{\mathcal{G}}$. Let $L \in \mathcal{L}$, let t be any text for L, and let $x \in \mathbb{N}^+$. The wanted IIM M is defined as follows:

$M(t_x) =$ "If $x = 1$ or M, when fed successively with t_{x-1}, does not produce any guess, then goto (A). Else goto (B).

(A) Search for the least $j \leq x$ for which $\hat{T}_j \subseteq t_x^+$. In case it is found, output j and request next input. Otherwise, request next input.

(B) Let h be the hypothesis produced last by M on input t_{x-1}.

(i) Execute the following tests:

- Check whether $t_x^+ \subseteq L(\hat{G}_h)$. In case it is not, goto (ii). Otherwise, continue as follows.
- Check for all $j \leq x$ satisfying $h \prec j$, $\hat{T}_h \subseteq \hat{T}_j$ and $\hat{T}_j \subseteq t_x^+ \subseteq L(\hat{G}_j)$ whether or not there is a string s with $|s| \leq max\{x, |w| \mid w \in t_x^+\}$ fulfilling $s \in L(\hat{G}_h) \setminus L(\hat{G}_j)$. In case at least one j has been found, output the minimal one and request the next input. Otherwise, output h and request next input.

(ii) Search for the least $j \leq x$ satisfying $h \prec j$, $\hat{T}_h \subseteq \hat{T}_j$ and $\hat{T}_j \subseteq t_x^+ \subseteq L(\hat{G}_j)$. In case it is found, output j and request next input. Otherwise, output h and request next input."

We have to show that M infers L with at most k mind changes.

Claim 1. If M converges, say to z, then $L = L(\hat{G}_z)$.

First observe that $\hat{T}_z \subseteq L$, since otherwise z cannot be any of M's guesses. Suppose, $L \neq L(\hat{G}_z)$. Then we have to distinguish between the following two cases.
Case 1. $L \setminus L(\hat{G}_z) \neq \emptyset$.
Consequently, there is a string $s \in L \setminus L(\hat{G}_z)$. Since t is a text for L, there has to be an x such that $s \in t_x^+$. Therefore, M has to reject z in its consistency test in (i), a contradiction.
Case 2. $L(\hat{G}_z) \setminus L \neq \emptyset$.
Then we may assume $L(\hat{G}_z) \supset L$, since otherwise we are again in Case 1. Let x_z be the corresponding x used to find z. Since $\hat{T}_z \subseteq L$ and $L(\hat{G}_z) \neq L$, we may set $A := t_{x_z}^+$ and apply property (3). Hence, there has to be a j such that $z \prec j$, $A \subset \hat{T}_j \subseteq L$ and $L(\hat{G}_j) = L$. Therefore, there must be an $x \geq j$ such that $t_x^+ \subseteq \hat{T}_j$ and the IIM M eventually finds a string s satisfying $s \in L(\hat{G}_z) \setminus L(\hat{G}_j)$ as well as $|s| \leq max\{x, |w| \mid w \in t_x^+\}$. This event would force M to change its mind from z to j, a contradiction. This proves Claim 1.

Claim 2. M converges on t, and $card(\{x \mid M(t_x) \neq M(t_{x+1})\}) \leq k$.

In accordance with (2) one easily shows that M outputs at least once a hypothesis. This guess is repeated as long as in (B) the tests in (i) are fulfilled. Moreover, if M's last guess is rejected, it always outputs a new hypothesis in some subsequent step. This is obvious as long as M does not reject its last guess h by verifying $t_x^+ \not\subseteq L(\hat{G}_h)$. If M detects an inconsistency, then property (3) ensures that M finds via (ii) a new hypothesis. Consequently, the only remaining case we have to deal with is that M performs more than k mind changes. Let $(h_j)_{j \in \mathbb{N}^+}$ be M's sequence of hypotheses produced when successively fed with t. However, by construction we obtain the following: $h_j \prec h_{j+1}$ and $\hat{T}_{h_j} \subseteq \hat{T}_{h_{j+1}} \subseteq L$ for all $j \in \mathbb{N}^+$. Hence, property (4) is contradicted. This proves the Claim 2 as well as the theorem.

q.e.d.

Next we give a characterization of $LIM_k - INF$. For that purpose we define a relation $\prec$ over pairs of sets as follows. Let A, B, C, D be sets. Then $(A, B) \prec (C, D)$ iff $A \subseteq C,\ B \subseteq D$

and $A \subset C$ or $B \subset D$. Note that $\prec$ is computable if A, B, C, D are finitely generable. Now we are ready to state the announced characterization.

Theorem 12. *Let $\mathcal{L}$ be an indexed family of recursive languages and $k \in \mathbb{N}$. Then: $\mathcal{L} \in LIM_k - INF$ if and only if there are a space of hypotheses $\hat{\mathcal{G}} = (\hat{G}_j)_{j \in \mathbb{N}^+}$ and recursively generable families $(\hat{P}_j)_{j \in \mathbb{N}^+}$ and $(\hat{N}_j)_{j \in \mathbb{N}^+}$ of finite sets such that*

(1) $range(\mathcal{L}) = L(\hat{\mathcal{G}})$

(2) For all $j \in \mathbb{N}^+$, $\emptyset \neq \hat{P}_j \subseteq L(\hat{G}_j)$ and $\hat{N}_j \subseteq co - L(\hat{G}_j)$

(3) For all $L \in \mathcal{L}$ and $z \in \mathbb{N}^+$, if $\hat{P}_z \subseteq L \neq L(\hat{G}_z)$ and $\hat{N}_z \subseteq co - L$, then there is a $j \in \mathbb{N}^+$ such that $(\hat{P}_z, \hat{N}_z) \prec (\hat{P}_j, \hat{N}_j)$ as well as $L = L(\hat{G}_j)$.

(4) For all $L \in \mathcal{L}$ there is no sequence $(\hat{P}_{z_j}, \hat{N}_{z_j})_{j=1,\ldots,m}$ with $m > k+1$ such that $(\hat{P}_{z_j}, \hat{N}_{z_j}) \prec (\hat{P}_{z_{j+1}}, \hat{N}_{z_{j+1}}) \prec (L, co - L)$, for all $j < m$.

Proof. Necessity: Let $k \in \mathbb{N}$, and $\mathcal{L} \in LIM_k - INF$. Then there are an IIM M and a space of hypotheses $(G_j)_{j \in \mathbb{N}^+}$ such that M infers any $L \in \mathcal{L}$ with at most k mind changes according to $(G_j)_{j \in \mathbb{N}^+}$. Without loss of generality, we may assume that M works conservatively, i.e., M performs justified mind changes, only. We proceed in showing how to construct $(\hat{G}_j)_{j \in \mathbb{N}^+}$. This will be done in two steps. In the first step, we define a space of hypotheses $(\tilde{G}_j)_{j \in \mathbb{N}^+}$ as well as corresponding recursively generable families $(\tilde{P}_j)_{j \in \mathbb{N}^+}$ and $(\tilde{N}_j)_{j \in \mathbb{N}^+}$ of finite sets, where $\tilde{P}_j$ may be empty for some $j \in \mathbb{N}^+$. Afterwards, we define a procedure which enumerates a certain subset of $\tilde{\mathcal{G}}$.

First step: Let $c : \mathbb{N}^+ \times \mathbb{N}^+ \to \mathbb{N}^+$ be Cantor's pairing function. For all $m,\ x \in \mathbb{N}^+$ we set $\tilde{G}_{c(m,x)} = G_m$. Obviously, it holds $range(\mathcal{L}) = L(\tilde{\mathcal{G}})$. Let i^m be the lexicographically ordered informant for $L(G_m)$, and let $x \in \mathbb{N}^+$.
We define:

$$\tilde{P}_{c(m,x)} = \begin{cases} range^+(i^m_y), & if \quad y = min\{z \mid z \leq x,\ M(i^m_z) = m,\ range^+(i^m_z) \neq \emptyset\} \\ \emptyset, & otherwise \end{cases}$$

If $\tilde{P}_{c(m,x)} = range^+(i^m_y) \neq \emptyset$, then we set $\tilde{N}_{c(m,x)} = range^-(i^m_y)$. Otherwise, we define $\tilde{N}_{c(m,x)} = \emptyset$.

Second step: The space of hypotheses $(\hat{G}_j)_{j \in \mathbb{N}^+}$ will be defined by simply striking off all grammars $\tilde{G}_{c(m,x)}$ with $\tilde{P}_{c(m,x)} = \emptyset$. In order to save readability, we omit the corresponding bijective mapping yielding the enumeration $(\hat{G}_j)_{j \in \mathbb{N}^+}$ from $(\tilde{G}_j)_{j \in \mathbb{N}^+}$. If $\hat{G}_j$ is referring to $\tilde{G}_{c(m,x)}$, we set $\hat{P}_j = \tilde{P}_{c(m,x)}$ as well as $\hat{N}_j = \tilde{N}_{c(m,x)}$.

We have to show that $(\hat{G}_j)_{j \in \mathbb{N}^+}$, $(\hat{N}_j)_{j \in \mathbb{N}^+}$, and $(\hat{P}_j)_{j \in \mathbb{N}^+}$ do fulfil the announced properties. Obviously, $(\hat{P}_j)_{j \in \mathbb{N}^+}$ and $(\hat{N}_j)_{j \in \mathbb{N}^+}$ are recursively generable families of finite sets. Furthermore, it is easy to see that $L(\hat{\mathcal{G}}) \subseteq range(\mathcal{L})$. In order to prove (1), it suffices to show that for every $L \in \mathcal{L}$ there is at least one $j \in \mathbb{N}^+$ with $L = L(\tilde{G}_j)$ and $\tilde{P}_j \neq \emptyset$. Let i^L be L's lexicographically ordered informant. Since M has to infer L on i^L, too, and $L \neq \emptyset$, there are $m,\ x \in \mathbb{N}^+$ such that $M(i^L_x) = m,\ L = L(G_m),\ range^+(i^L_x) \neq \emptyset$ as well as $M(i^L_y) \neq m$ for all $y < x$. From that we immediately conclude that $L = L(\tilde{G}_j)$ and that $\tilde{P}_j \neq \emptyset$ for $j = c(m, x)$. Due to our construction, property (2) is obviously fulfilled.

Next to, we show (3). Let $L \in \mathcal{L}$ and $z \in \mathbb{N}^+$ such that $\hat{P}_z \subseteq L \neq L(\hat{G}_z)$ and $\hat{N}_z \subseteq co - L$. We have to show that there is a $j \in \mathbb{N}^+$ such that $(\hat{P}_z, \hat{N}_z) \prec (\hat{P}_j, \hat{N}_j)$ as well as $L = L(\hat{G}_j)$. In accordance to our construction one can easily observe: There is a uniquely defined initial segment, say i_x^z, of the lexicographically ordered informant for $L(\hat{G}_z)$ such that $range(i_x^z) = \hat{P}_z \cup \hat{N}_z$. Furthermore, $M(i_x^z) = m$ with $L(\hat{G}_z) = L(G_m)$. Additionally, since $\hat{P}_z \subseteq L$ as well as $\hat{N}_z \subseteq co - L$, i_x^z is an initial segment of the lexicographically ordered informant i^L of L. Since M infers L on its lexicographically ordered informant i^L, there exist r, $n \in \mathbb{N}^+$ such that $M(i_{x+r}) = n$ and $L = L(G_n)$ for the first time. (*Recall, that M works conservatively. This implies that M never outputs n on an initial segment of i_x.*) Due to our construction, for $j = c(n, x + r)$, we obtain that $\hat{P}_j = range^+(i_{x+r}^L)$, $\hat{N}_j = range^-(i_{x+r}^L)$, and $L(G_n) = L(\hat{G}_j)$. Hence, $(\hat{P}_z, \hat{N}_z) \prec (\hat{P}_j, \hat{N}_j)$ as well as $L = L(\hat{G}_j)$.

It remains to show (4). Let $L \in \mathcal{L}$. Assume that there is a sequence $(\hat{P}_{z_j}, \hat{N}_{z_j})_{j=1,...,m}$ with $m > k + 1$ such that, for all $j < m$, $(\hat{P}_{z_j}, \hat{N}_{z_j}) \prec (\hat{P}_{z_{j+1}}, \hat{N}_{z_{j+1}}) \prec (L, co - L)$. Hence, there is a uniquely defined initial segment, say $i_{x_m}^L$, of L's lexicographically ordered informant i^L such that $range(i_{x_m}^L) = \bigcup_{j \leq m} (\hat{P}_{z_j} \cup \hat{N}_{z_j})$. Due to our construction, M performs on $i_{x_m}^L$ at least $m - 1$ mind changes when inferring L. In detail, when fed with the initial segment $i_{x_1}^L$ of $i_{x_m}^L$ with $range(i_{x_1}^L) = \hat{P}_{z_1} \cup \hat{N}_{z_1}$, M outputs a hypothesis n_1 with $L(G_{n_1}) = L(\hat{G}_{z_1})$. Afterwards, when processing the initial segment $i_{x_2}^L$ of $i_{x_m}^L$ with $range(i_{x_2}^L) = \bigcup_{j \leq 2} \hat{P}_{z_j} \cup \hat{N}_{z_j}$, M outputs n_2 with $L(G_{n_2}) = L(\hat{G}_{z_2})$, a.s.o. Since $m > k + 1$, this contradicts our assumption that M infers any $L \in \mathcal{L}$ with at most k mind changes on i^L.

Sufficiency: Let $k \in \mathbb{N}$. It suffices to prove that there is an IIM M inferring any $L \in \mathcal{L}$ with at most k mind changes with respect to the space of hypotheses $(\hat{G}_j)_{j \in \mathbb{N}^+}$. So let $L \in \mathcal{L}$, let i be any informant for L, and let $x \in \mathbb{N}^+$.

$M(i_x) =$ "If $x = 1$ or M, when successively fed with i_{x-1}, does not output a hypothesis, then goto (A). Otherwise, goto (B).

(A) Generate $\hat{P}_j$ and $\hat{N}_j$ for $j = 1, ..., x$ and test whether
  (A1) $\hat{P}_j \subseteq i_x^+ \subseteq L(\hat{G}_j)$ and
  (A2) $\hat{N}_j \subseteq i_x^- \subseteq co - L(\hat{G}_j)$.
  In case there is at least a j fulfilling the test, output the minimal one and request the next input.
  Otherwise, output nothing and request the next input.

(B) Let j be M's last hypothesis when successively fed with i_{x-1}. Test whether $i_x^+ \subseteq L(\hat{G}_j)$ and $i_x^- \subseteq co - L(\hat{G}_j)$. In case it is, output j and request the next input. Otherwise, goto (C).

(C) Generate $\hat{P}_z$ and $\hat{N}_z$ for $z = 1, ..., x$ and test whether
  (C1) $\hat{P}_z \subseteq i_x^+ \subseteq L(\hat{G}_z)$,
  (C2) $\hat{N}_z \subseteq i_x^- \subseteq co - L(\hat{G}_z)$ and
  (C3) $(\hat{P}_j, \hat{N}_j) \prec (\hat{P}_z, \hat{N}_z)$.
  In case there is at least a z fulfilling the test output the minimal one and request the next input.
  Otherwise, output j and request the next input."

Since all of the $\hat{P}_j$ and $\hat{N}_j$ are uniformly recursively generable and finite we see that M is an IIM. We have to show that it infers L.

Claim A. M converges, and $card(\{x \mid M(t_x) \neq M(t_{x+1})\}) \leq k$.
Because of (2), M generates at least one hypothesis when fed with i. Furthermore, assume for a moment that M performs more than k mind changes when inferring L on i. It is easy to recognize that this assumption would imply that there exists a sequence $(\hat{P}_{z_j}, \hat{N}_{z_j})_{j=1,\ldots,m}$ with $m > k+1$ such that, for all $j < m$, $(\hat{P}_{z_j}, \hat{N}_{z_j}) \prec (\hat{P}_{z_{j+1}} \hat{N}_{z_{j+1}}) \prec (L, co-L)$. This would contradict (4). Since the number of possible mind changes is bounded by k, and, moreover, M outputs after a certain period always a hypothesis, we can conclude that M converges.

Claim B. If M converges, say to j, then $L = L(\hat{G}_j)$.
Assume that M converges to a hypothesis j with $L(\hat{G}_j) \neq L$. Because of (3), M performs an additional mind change when fed with a sufficiently large initial segment i_{x+r} of i. Obviously, if M once outputs a correct hypothesis, then it will never change its mind in any subsequent step.

As already shown, M changes at most k times its mind. Consequently, M infers any $L \in \mathcal{L}$ with at most k mind changes.

q.e.d

5. Conclusions and Open Problems

We have dealt with the learnability of enumerable families $\mathcal{L}$ of uniformly recursive languages in dependence on the number of allowed mind changes. Applying this measure of efficiency we could prove that class preserving learning algorithms are superior to exact learnability. Moreover, in improving Mukouchi's (1992) results we established two new infinite hierarchies. On the other hand, we also proved that even a single additional mind change can neither be compensated by a suitable choice of the space of hypotheses nor by information presentation. Furthermore, we have separated exact and class preserving language learning that avoids overgeneralization. Finally, we presented a complete characterization of language learning in terms of recursively generable finite sets. These theorems resolved the problem that remained open in Mukouchi (1992). Additionally, they offer a new approach to handle overgeneralized hypotheses. However, some problems remained open. It would be very interesting to know how many mind changes are necessary to learn indexed families that cannot be inferred by class preserving conservatively working IIMs. Furthermore, we did not fully succeed in separating $ELIM_k - INF$ and $LIM_k - INF$.

Moreover, it seems to be very challenging to study the question whether or not the results of Barzdin and Freivalds (1972, 1974) as well as of Barzdin, Kinber and Podnieks (1974) may be extended to language learning ¿from positive data.

Acknowledgement The authors gratefully acknowledge valuable comments on the preparation of the paper by Yasuhito Mukouchi.

6. References

Angluin, D. (1980), Inductive inference of formal languages from positive data, *Information and Control* **45**, 117 - 135.

Angluin, D., and Smith, C.H. (1983), Inductive inference: theory and methods, *Computing Surveys* **15**, 237 - 269.

Angluin, D., and Smith, C.H. (1987), Formal inductive inference, *in* "Encyclopedia of Artificial Intelligence" (St.C. Shapiro, Ed.), Vol. 1, pp. 409 - 418, Wiley-Interscience Publication, New York.

Barzdin, Ya.M., Kinber, E.B., and Podnieks, K.M. (1974), Ob uskorenii sinteza i prognosirowanija funkzi, *in* "Teorija Algoritmow i Programm," Vol. 1 (Ya.M. Barzdin, ed.) Latvian State University, Riga, pp. 117 - 128.

Barzdin, Ya.M., and Freivalds, R.V. (1972), On the prediction of general recursive functions, *Sov. Math. Dokl.* **13**, 1224 - 1228.

Barzdin, Ya.M., and Freivalds, R.V. (1974), Prognosirowanie i predelny sintez effektiwno perechislimych klassow funkzi, *in* "Teorija Algoritmow i Programm," Vol. 1 (Ya. M. Barzdin, ed.) Latvian State University, Riga, pp. 101 - 111.

Blum, L., and Blum, M. (1975), Toward a mathematical theory of inductive inference, *Information and Control* **28**, 122 - 155

Bucher, W., and Maurer, H. (1984), "Theoretische Grundlagen der Programmiersprachen, Automaten und Sprachen," Bibliographisches Institut AG, Wissenschaftsverlag, Zürich.

Case, J. (1988), The power of vacillation, *in* "Proceedings 1st Workshop on Computational Learning Theory," (D. Haussler and L. Pitt, Eds.), pp. 196 -205, Morgan Kaufmann Publishers Inc.

Case, J., and Lynes, C. (1982), Machine inductive inference and language identification, *in* "Proceedings Automata, Languages and Programming, Ninth Colloquium, Aarhus, Denmark," (M. Nielsen and E.M. Schmidt, Eds.), Lecture Notes in Computer Science Vol. 140, pp. 107 - 115, Springer-Verlag, Berlin.

Case, J., and Smith, C. (1983), Comparison of identification criteria for machine inductive inference, *Theoretical Computer Science* **25**, 193 - 220.

Fulk, M.(1990), Prudence and other restrictions in formal language learning, *Information and Computation* **85**, 1 - 11.

Gasarch, W.I., and Velauthapillai, M. (1992), Asking questions versus verifiability, to appear *in* "Proceedings 3rd International Workshop on Analogical and Inductive Inference," October 1992, Dagstuhl, Lecture Notes in Artificial Intelligence, Springer-Verlag, Berlin.

Gold, E.M., (1967), Language Identification in the Limit, *Information and Control* **10**, 447 - 474.

Jain, S., and Sharma, A. (1989), Recursion theoretic characterizations of language learning, The University of Rochester, Dept. of Computer Science, TR 281.

Kapur, S., and Bilardi, G. (1992), Language learning without overgeneralization, *in* "Proceedings 9th Annual Symposium on Theoretical Aspects of Computer Science, Cachan, France, February 13 - 15," (A. Finkel and M. Jantzen, Eds.), Lecture Notes in Computer Science Vol. 577, pp. 245 - 256, Springer-Verlag, Berlin.

Lange, S., and Zeugmann, T. (1991), Monotonic versus non–monotonic language learning, *in* "Proceedings 2nd International Workshop on Nonmonotonic and Inductive Logic, December 1991, Reinhardsbrunn," to appear in Lecture Notes in Artificial Intelligence, Springer-Verlag, Berlin.

Lange, S., and Zeugmann, T. (1992), Types of monotonic language learning and their characterization, *in* "Proceedings 5th Annual ACM Workshop on Computational Learning Theory," July 27 - 29, Pittsburgh, pp. 377 - 390, ACM Press.

Lange, S.,Zeugmann, T., and Kapur, S (1992), Class preserving monotonic language learning, submitted to *Theoretical Computer Science*, and
GOSLER–Report 14/92, FB Mathematik und Informatik, TH Leipzig.

Mukouchi, Y. (1992), Inductive Inference with Bounded Mind Changes, to appear *in* Proceedings "Algorithmic Learning Theory," October 1992, Tokyo, Japan, and
Research Institute of Fundamental Information Science, Kyushu University 33, Fukuoka, May 14, 1992, RIFIS-TR-CS-58.

Osherson, D., Stob, M., and Weinstein, S. (1986), "Systems that Learn, An Introduction to Learning Theory for Cognitive and Computer Scientists," MIT-Press, Cambridge, Massachusetts.

Shinohara, T. (1990), Inductive Inference from Positive Data is Powerful, *in* "Proceedings 3rd Annual Workshop on Computational Learning Theory," (M. Fulk and J. Case, Eds.), pp. 97 - 110, Morgan Kaufmann Publishers Inc.

Solomonoff, R. (1964), A formal theory of inductive inference, *Information and Control* **7**, 1 - 22, 234 - 254.

Trakhtenbrot, B.A., and Barzdin, Ya.M. (1970) "Konetschnyje Awtomaty (Powedenie i Sintez)," Nauka, Moskwa, (in Russian)
english translation: "Finite Automata–Behavior and Synthesis, Fundamental Studies in Computer Science 1," North–Holland, Amsterdam.

Wiehagen, R. (1977), Identification of formal languages, *in* "Proceedings Mathematical Foundations of Computer Science, Tatranska Lomnica," (J. Gruska, Ed.), Lecture Notes in Computer Science Vol. 53, pp. 571 - 579, Springer-Verlag, Berlin.

Wiehagen, R., Freivalds, R., and Kinber, B. (1984), On the power of probabilistic strategies in inductive inference, *Theoretical Computer Science* **28**, 111 - 133.

Zeugmann, T. (1983), A–posteriori characterizations in inductive inference of recursive functions, *Journal of Information Processing and Cybernetics (EIK)* **19**, 559 - 594.