# Measuring Power of Locally Testable Languages

Ryoma Sin'ya

Akita University
`ryoma@math.akita-u.ac.jp`

**Abstract.** A language $L$ is said to be $\mathcal{C}$-measurable, where $\mathcal{C}$ is a class of languages, if there is an infinite sequence of languages in $\mathcal{C}$ that converges to $L$. In this paper we investigate the measuring power of LT the class of all locally testable languages. Although each locally testable language only can check some local property (prefix, suffix, and infix of some bounded length), it is shown that many non-locally-testable languages are LT-measurable. In particular, we show that the measuring power of locally testable languages coincides with the measuring power of unambiguous polynomials. We also examine the measuring power of some fragments of unambiguous polynomials.

## 1 Introduction

A language $L$ is called *star-free* if it can be represented as a finite combination of Boolean operations and concatenation of finite languages, and $L$ is called *locally testable* if it is a finite Boolean combination of languages of the form $uA^*$, $A^*v$ and $A^*wA^*$. After the celebrated Schützenberger's theorem giving an algebraic characterisation [18] and McNaughton–Papert theorem giving a logical characterisation [10] of star-free languages, both algebraic and logical counterparts of many fragments of star-free languages are deeply well-investigated: see a survey [6] or [11] for example. In particular, McNaughton [9], Zalcstein [24], and Brzozowski–Simon [4] showed that it is decidable whether a given regular language is locally testable by giving an algebraic counterpart. Although the definition of locally testable languages is quite simple, this result is non-trivial and a proof relies on a deep algebraic decomposition theory.

In this paper, we shed new light on the fragments of star-free languages by using *measurability* which is a measure theoretic notion on formal languages. $\mathcal{C}$-measurability for a class $\mathcal{C}$ of languages is introduced by [21] and it was used for classifying non-regular languages by using regular languages. A language $L$ is said to be $\mathcal{C}$-measurable if there is an infinite sequence of languages in $\mathcal{C}$ that converges to $L$. Roughly speaking, $L$ is $\mathcal{C}$-measurable means that it can be approximated by a language in $\mathcal{C}$ with *arbitrary high precision*: the notion of "precision" is formally defined by the density of formal languages. Hence that a language $L$ is not REG-measurable, where REG is the class of all regular languages, means that $L$ has a complex shape so that it can not be approximated by regular languages. While the membership problem for a given language $L$ and a class $\mathcal{C}$ asks the existence of single language $K \in \mathcal{C}$ such that $L = K$, the $\mathcal{C}$-measurability asks the existence of an infinite sequence of languages in $\mathcal{C}$ that converges to $L$. In this sense, measurability is much more difficult than the membership problem and its analysis is a challenging task. For example, the author [22] showed that, for the class SF of all star-free languages, the class of all SF-measurable regular languages strictly contains SF but does not contain some regular languages. However, the decidability of SF-measurability is still unknown.

Instead of the class of all regular languages or star-free languages, in this paper we consider LT-measurability where LT is the class of all locally testable languages and also consider measuring power of three other fragments of star-free languages: the class UPol of all *unambiguous polynomials*, the class PT of all *piecewise testable* languages and the class AT of all *alphabet testable* languages. The main results of this paper are briefly summarised as follows.

(1) LT-measurability and UPol-measurability are equivalent (Theorem 6 and Theorem 7).
(2) AT- and PT-measurability are strictly weaker than LT-measurability and decidable for regular languages (Theorem 8, Theorem 9–11).

The result (1) is the first example of two *incomparable* subclasses of regular languages with the *same measuring power*. The result (2) (PT-measurability, in particular) is the first non-trivial examples of subclasses of regular languages with *decidable measurability*. Historically, locally testable languages [10] and unambiguous polynomials [17] are originally introduced with two different motivations: **"locality"** versus **"unambiguity"**. But interestingly, they have the same measuring power.

The structure of this paper is as follows. Section 2 provides preliminaries including density, measurability and definitions of fragments of star-free languages. The measuring power of $\mathrm{LT}, \mathrm{UPol}$ and $\mathrm{AT}, \mathrm{PT}$ are investigated in Section 3 and Section 4, respectively. A summary of all results and future work are described in Section 5.

## 2 Preliminaries

This section provides the precise definitions of density, measurability and local varieties of regular languages. $\mathrm{REG}_A$ denotes the family of all regular languages over an alphabet $A$. We assume that the reader has a standard knowledge of automata theory including the concept of syntactic monoids (*cf.* [8]).

### 2.1 Density of formal languages

For a set $X$, we denote by $\#(X)$ the cardinality of $X$. We denote by $\mathbb{N}$ and $\mathbb{Z}$ the set of natural numbers including 0 and the set of integers, respectively. For an alphabet $A$, we denote the set of all words (all non-empty words, respectively) over $A$ by $A^*$ ($A^+$, respectively). We write $|w|$ for the length of $w$ and $A^{\leq n}$ for the set of all words of length less than or equal to $n$. For a word $w \in A^*$ and a letter $a \in A$, $|w|_a$ denotes the number of occurrences of $a$ in $w$. We denote by $\mathrm{alph}(w) = \{a \mid |w|_a > 0\}$ the set of all letters contained in $w$. A word $v$ is said to be a subword of a word $w$ if $w = xvy$ for some $x, y \in A^*$. For a language $L \subseteq A^*$, we denote by $\overline{L} = A^* \setminus L$ the complement of $L$. A language $L$ is said to be *dense* if $L \cap A^*wA^* \neq \emptyset$ holds for any $w \in A^*$. $L$ is not dense means $L \cap A^*wA^* = \emptyset$ for some word $w$ by definition, and such word $w$ is called a forbidden word of $L$.

**Definition 1** (*cf.* [2])**.** The *density* $\delta_A(L)$ of $L \subseteq A^*$ is defined as

$$\delta_A(L) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \frac{\#(L \cap A^k)}{\#(A^k)}$$

if its exists, otherwise we write $\delta_A(L) = \perp$. $L$ is called *null* if $\delta_A(L) = 0$, and conversely $L$ is called *co-null* if $\delta_A(L) = 1$.

*Example 1.* It is known that every regular language has a rational density (*cf.* [16]) and it is computable. Here we explain two examples of (co-)null languages.

(1) For each word $w$, the language $A^*wA^*$, the set of all words that contain $w$ as a subword, is of density 1 (co-null). This fact is sometimes called *infinite monkey theorem.* A language $L$ having a forbidden word $w$ is always null: this means $A^*wA^* \subseteq \overline{L}$ and $\delta_A(A^*wA^*) \leq \delta_A(\overline{L})$ which implies $\delta_A(\overline{L}) = 1$ by infinite monkey theorem.

(2) The semi-Dyck language $D = \{\varepsilon, ab, aabb, abab, aaabbb, \ldots\}$ over $A = \{a, b\}$ is dense but null. This follows from the fact that $\#\big(D \cap A^{2n}\big)$ equals the $n$-th Catalan number whose asymptotic formula is $\Theta(4^n/n^{3/2})$.

As explained above, “dense” does not imply “not null”. But these two notions are equivalent for regular languages as the following theorem says. We denote by $\mathrm{ZO}_A$ the family of all null or co-null regular languages over $A$ (ZO stands for “zero-one”).

**Theorem 1** (*cf.* [16])**.** *A regular language $L$ is not null if and only if $L$ is dense.*

### 2.2 Measurability of formal languages

The notion of “measurability” on formal languages is defined by a standard measure theoretic approach as follows.

**Definition 2** ([21])**.** Let $\mathcal{C}_A$ be a family of languages over $A$. For a language $L \subseteq A^*$, we define its $\mathcal{C}_A$*-inner-density* $\underline{\mu}_{\mathcal{C}_A}(L)$ and $\mathcal{C}_A$*-outer-density* $\overline{\mu}_{\mathcal{C}_A}(L)$ over $A$ as

$$\begin{aligned} \underline{\mu}_{\mathcal{C}_A}(L) &= \sup\{\delta_A(K) \mid K \subseteq L, K \in \mathcal{C}_A, \delta_A(K) \neq \perp\} \text{ and} \\ \overline{\mu}_{\mathcal{C}_A}(L) &= \ \inf\{\delta_A(K) \mid L \subseteq K, K \in \mathcal{C}_A, \delta_A(K) \neq \perp\}, \text{ respectively.} \end{aligned}$$

A language $L$ is said to be $\mathcal{C}_A$*-measurable* if $\underline{\mu}_{\mathcal{C}_A}(L) = \overline{\mu}_{\mathcal{C}_A}(L)$ holds. We say that an infinite sequence $(L_n)_n$ of languages over $A$ *converges to $L$ from inner (from outer,* respectively*)* if $L_n \subseteq L$ ($L_n \supseteq L$, respectively) for each $n$ and $\lim_{n\to\infty} \delta_A(L_n) = \delta_A(L)$.

*Example 2 ([21]).* The semi-Dyck language $D = \{\varepsilon, ab, aabb, abab, aaabbb, \ldots\}$ over $A = \{a, b\}$ is REG-measurable. We notice that there is no regular language $L$ such that $\delta_A(L) = 0$ and $D \subseteq L$, since any null regular language has a forbidden word but $D$ has no forbidden word. Hence we should construct an infinite sequence $(L_k)_k$ of different regular languages that converges to $D$ from outer. This can be done by letting $L_k = \{w \in A^* \mid |w|_a = |w|_b \mod k\}$. Clearly, $D \subseteq L_k$ holds and it can be shown that $\delta_A(L_k) = 1/k$ holds. Hence $\delta_A(L_k)$ tends to zero if $k$ tends to infinity. We will see this type of languages $L_k$ again in the next section.

For a family $\mathcal{C}_A$ of languages over $A$, we define its *Carathéodory extension* and *regular extension* as $\mathrm{Ext}_A(\mathcal{C}_A) = \{L \subseteq A^* \mid L \text{ is } \mathcal{C}_A\text{-measurable}\}$ and $\mathrm{RExt}_A(\mathcal{C}_A) = \mathrm{Ext}_A(\mathcal{C}_A) \cap \mathrm{REG}_A$, respectively. We say that "*$\mathcal{C}_A$ has a stronger measuring power than $\mathcal{D}_A$*" if $\mathrm{Ext}_A(\mathcal{C}_A) \supseteq \mathrm{Ext}_A(\mathcal{D}_A)$ holds.

**Theorem 2 ([22]).** *Let $\mathcal{C}_A \subseteq \mathrm{REG}_A$ be a family of regular languages over $A$. Then $L \in \mathrm{REG}_A$ is $\mathcal{C}_A$-measurable if and only if $L$ satisfies the following Carathéodory's condition:*

$$\forall X \subseteq A^* \qquad \overline{\mu}_{\mathcal{C}_A}(X) = \overline{\mu}_{\mathcal{C}_A}(X \cap L) + \overline{\mu}_{\mathcal{C}_A}(X \cap \overline{L}).$$

*Moreover, this is equivalent to $\overline{\mu}_{\mathcal{C}_A}(L) + \overline{\mu}_{\mathcal{C}_A}(\overline{L}) = 1$ (the case $X = A^*$ in the above condition).*

### 2.3 Fragments of Star-Free Languages

In this paper we examine measuring power of several subclasses of star-free languages equipping rich closure properties. For a family $\mathcal{C}_A$ of languages over $A$, we denote by $\mathscr{B}\mathcal{C}_A$ the Boolean closure of $\mathcal{C}_A$. Then the class $\mathrm{LT}_A$ of all locally testable languages can be defined as $\mathrm{LT}_A = \mathscr{B}\{wA^*, A^*w, A^*wA^* \mid w \in A^*\}$. A family of regular languages over $A$ is called *local variety* [1] over $A$ if it is closed under Boolean operations and left-and-right quotients. The reason why we focus on this type of families is that, the notion of measurability is well-behaved on Boolean operations and quotients as the following theorem says.

**Theorem 3 ([22]).** $\mathrm{Ext}_A$ *is a closure operator, i.e., it satisfies the following three properties for each $\mathcal{C} \subseteq \mathcal{D} \subseteq 2^{A^*}$: (extensive) $\mathcal{C} \subseteq \mathrm{Ext}_A(\mathcal{C})$, (monotone) $\mathrm{Ext}_A(\mathcal{C}) \subseteq \mathrm{Ext}_A(\mathcal{D})$, and (idempotent) $\mathrm{Ext}_A(\mathrm{Ext}_A(\mathcal{C})) = \mathrm{Ext}_A(\mathcal{C})$. Moreover,* $\mathrm{RExt}_A$ *is a closure operator over the class of all local varieties of regular languages over $A$, i.e., $\mathcal{C}_A$-measurability is preserved under Boolean operations and quotients for any local variety $\mathcal{C}_A$.*

*Example 3.* By Theorem 1, for any regular language $L$ in $\mathrm{ZO}_A$, $L$ or its complement has a forbidden word, which implies $\emptyset \subseteq L \subseteq \overline{A^*wA^*}$ or $A^*wA^* \subseteq L \subseteq A^*$. This fact and infinite monkey theorem implies that $\mathrm{ZO}_A \subseteq \mathrm{RExt}_A(\mathscr{B}\{A^*wA^* \mid w \in A^*\})$ holds. On the other hand, $\mathscr{B}\{A^*wA^* \mid w \in A^*\} \subseteq \mathrm{ZO}_A$ holds because $\mathrm{ZO}_A$ forms a local variety (*cf.* [20]). Moreover, it was shown that $\mathrm{RExt}_A(\mathrm{ZO}_A) =$

$\mathrm{ZO}_A$ in [22]. By combining these facts with Theorem 3 we have the following chain of inclusion: $\mathrm{ZO}_A \subseteq \mathrm{RExt}_A(\mathscr{B}\{A^*wA^* \mid w \in A^*\}) \subseteq \mathrm{RExt}_A(\mathrm{ZO}_A) = \mathrm{ZO}_A$ where the second inclusion $\subseteq$ follows from the monotonicity of $\mathrm{RExt}_A$.

The corresponding notion of a family of finite monoids is called *local pseudovariety* [1], and there is a natural one-to-one correspondence between the class of all local varieties and the class of all local pseudovarieties [7]. The class $\mathrm{SF}_A$ of all star-free languages over $A$ forms a local variety and its corresponding local pseudovariety is the class of all aperiodic monoids [18]. Thanks to Theorem 3, the regular extension $\mathrm{RExt}_A(\mathrm{SF}_A)$ of star-free languages is also a local variety. The following theorem says that $\mathrm{RExt}_A$ extends $\mathrm{SF}_A$ non-trivially, while it does not for $\mathrm{ZO}_A$.

**Theorem 4 ([22]).** $\mathrm{SF}_A \subsetneq \mathrm{RExt}_A(\mathrm{SF}_A) \subsetneq \mathrm{REG}_A$ *if* $\#(A) \geq 2$.

The class $\mathrm{LT}_A$ of all locally testable languages over $A$ is also a local variety. We use this algebraic characterisation of $\mathrm{LT}_A$ in the next section, hence we give a precise definition here. An element $e$ of a monoid $M$ is called idempotent if $e^2 = e$ holds. For each idempotent $e \in M$, $eMe$ is a submonoid of $M$ with the identity $e$ and it is called *local monoid in* $M$ (*cf.* [8]). A monoid $M$ is said to be locally idempotent and commutative if, for each idempotent $e \in M$, the local monoid $eMe$ only contains idempotents and the multiplication on $eMe$ is commutative ($x, y \in eMe \Rightarrow x^2 = x$ and $xy = yx$). The characterisation given in [9,24,4] says that $L$ is locally testable if and only if *its syntactic semigroup is locally idempotent and commutative* (see the full version [23] for more details).

We end this section by giving precise definitions of three additional subclasses of star-free languages. We denote by $\mathrm{AT}_A$ the Boolean combination of languages of the form $B^*$ where $B \subseteq A$ (AT stands for "alphabet testable", *cf.* [15]). This class also can be represented as $\mathrm{AT}_A = \mathscr{B}\{A^*aA^* \mid a \in A\}$ and hence $\mathrm{AT}_A \subsetneq \mathrm{LT}_A$. $\mathrm{AT}_A$ forms a (finite) local variety, and its corresponding local pseudovariety is idempotent and commutative monoids (*cf.* [6]). Clearly, the density of every language in $\mathrm{AT}_A$ is either zero or one, thus we have $\mathrm{AT}_A \subseteq \mathrm{ZO}_A$. A language $L$ is called *monomial* if it is of the form $A_0^*a_1A_1^*a_2A_2^* \cdots A_{n-1}^*a_nA_n^*$ where each $a_i \in A, A_i \subseteq A$ and $n \geq 0$. A monomial defined above is said to be *simple* if $A_i = A$ for each $i$. For $w = a_1a_2 \cdots a_n$ we denote by $L_w$ the simple monomial $A^*a_1A^*a_2A^* \cdots A^*a_nA^*$. A language is called *piecewise testable* if it can be represented as a finite Boolean combination of simple monomials. The class $\mathrm{PT}_A$ of all piecewise testable languages over $A$ forms a local variety. The corresponding local pseudovariety of $\mathrm{PT}_A$ is the class of all $\mathscr{J}$-trivial monoids [19]. A monomial $L = A_0^*a_1A_1^* \cdots a_nA_n^*$ is *unambiguous* if for all $w \in L$ there exists exactly one factorisation $w = w_0a_1w_1 \cdots a_nw_n$ where each $i$ satisfies $w_i \in A_i^*$. A language is an *unambiguous polynomial* if it is a finite disjoint union of unambiguous monomials. The family $\mathrm{UPol}_A$ of all unambiguous polynomials over $A$ forms a local variety [17]. In particular, the complement of an unambiguous polynomial is also an unambiguous polynomial. This fact plays a key role in the next section. By definition we have the following chain of inclusion $\mathrm{AT}_A \subsetneq \mathrm{PT}_A \subsetneq \mathrm{UPol}_A \subsetneq \mathrm{SF}_A$ and every inclusion is strict. We also notice that $\mathrm{PT}_A$ ($\mathrm{UPol}_A$, respectively) and $\mathrm{LT}_A$ are incomparable. For example, $A^*abaA^* \in \mathrm{LT}_A \setminus \mathrm{PT}_A$ (because the syntactic monoid of $A^*abaA^*$ is not $\mathcal{J}$-trivial) and $L_{aba} = A^*aA^*bA^*aA^* \in \mathrm{PT}_A \setminus \mathrm{LT}_A$ (because the syntactic monoid of $L_{aba}$ is not locally idempotent). Every $\mathcal{J}$-trivial finite monoid has a zero element, and a language whose syntactic monoid has a zero is of density zero or one (*cf.* [20]), thus we have $\mathrm{PT}_A \subsetneq \mathrm{ZO}_A$.

## 3 Measuring Power of Locally Testable Languages

In this section we examine the measuring power of locally testable languages: what kind of languages are $\mathrm{LT}_A$-measurable and what are not? First we show there are "many" $\mathrm{LT}_A$-measurable languages.

**Proposition 1.** *For any language $L \subseteq A^*$, $A^*LA^*$ is $\mathrm{LT}_A$-measurable.*

*Proof.* If $L = \emptyset$ then $A^*LA^* = \emptyset$ is in $\mathrm{LT}_A$. If $L \neq \emptyset$, we can choose $w \in L$ and the ideal language $A^*wA^* \subseteq A^*LA^*$ is co-null by infinite monkey theorem. Hence $\underline{\mu}_{\mathrm{LT}_A}(A^*LA^*) = 1$ *i.e.*, $A^*LA^* \in \mathrm{Ext}_A(\mathrm{LT}_A)$. □

If $A$ contains two distinct letters $a$ and $b$, then the subword relation $x \sqsubseteq y$ ($\Leftrightarrow$ "$x$ is a subword of $y$") has an infinite antichain in $A^*$, *e.g.*, $\{ab^na \mid n \geq 0\}$. Two different subsets $L_1$ and $L_2$ of this infinite antichain produce two different languages $A^*L_1A^*$ and $A^*L_2A^*$. Hence the above theorem implies there are uncountably many $\mathrm{LT}_A$-measurable languages. In fact, in [22], a stronger statement was shown as follows[1].

**Theorem 5 ([22]).** *For any real number $\alpha \in [0, 1]$ there is a $\mathrm{LT}_A$-measurable language with density $\alpha$ if $\#(A) \geq 2$.*

Next we show that languages with modulo counting, which were used for the convergent sequence to the semi-Dyck language in Example 2, are $\mathrm{LT}_A$-*im*measurable.

**Proposition 2.** *The language $L_k = \{w \in A^* \mid |w|_a = |w|_b \mod k\}$ over $A = \{a, b\}$ is $\mathrm{LT}_A$-immeasurable for any $k \geq 2$.*

*Proof.* We show that any non-null locally testable language contains some words in $L_k$ and $\overline{L_k}$. Suppose $L \in \mathrm{LT}_A$ is non-null ($\delta_A(L) > 0$) and let $M_L$ and $\eta : A^* \to M_L$ be its syntactic monoid and morphism, and let $S = \eta(L)$.

Let $K \subseteq M_L$ be the minimal ideal of $M_L$. We can easily obtain $\delta_A(\eta^{-1}(K)) = 1$ as a corollary of the infinite monkey theorem. Hence the assumption $\delta_A(L) > 0$ implies that there is some $t \in K \cap S$.

Let $e$ be an idempotent in $M_L$ (since $M_L$ is finite, there is at least one idempotent) and $w_e \in \eta^{-1}(e)$ be a word of $e$. Without loss of generality, we

---

[1] [22] considered $\mathrm{REG}_A$-measurability instead of $\mathrm{LT}_A$-measurability, but the convergent sequence constructed in the proof of Theorem 5 is actually a sequence of locally testable languages

can assume $|w_e|_a \geq |w_e|_b$. Let $n = |w_e|_a - |w_e|_b \geq 0$, $u = w_e b^{2n} a w_e$ and $v = w_e b^{2n+1} w_e$. By construction, $|u|_a - |u|_b = 1$ and $|v|_b - |v|_a = 1$ holds. By the minimality of $K$, there exist $x$ and $y$ such that $\eta(xuvy) = t$. Because $M_L$ is locally idempotent, $\eta(u), \eta(v) \in eM_Le$ are both idempotent. This fact implies that, for any $i, j \geq 1$, $xu^iv^jy$ is in $L$ because

$$\eta(xu^iv^jy) = \eta(x)\eta(u)^i\eta(v)^j\eta(y) = \eta(x)\eta(u)\eta(v)\eta(y) = \eta(xuvy) = t$$

holds. Thus for any $m \in \mathbb{Z}$, we can choose $i, j \geq 1$ such that $|xu^iv^jy|_a - |xu^iv^jy|_b = m$. Hence $L$ contains words in $L_k$ and $\overline{L_k}$ simultaneously, which implies $\underline{\mu}_{\mathrm{LT}_A}(L_k) = 0$ and $\overline{\mu}_{\mathrm{LT}_A}(L_k) = 1$, *i.e.*, $L_k$ is $\mathrm{LT}_A$-immeasurable. $\square$

The next theorem says that $\mathrm{LT}_A$ has a stronger measuring power than $\mathrm{UPol}_A$.

**Theorem 6.** $\mathrm{Ext}_A(\mathrm{LT}_A) \supseteq \mathrm{Ext}_A(\mathrm{UPol}_A)$ *for any $A$.*

We use the following simple lemma for proving this theorem.

**Lemma 1.** *The concatenation $LK$ of two null regular languages $L$ and $K$ is also null.*

*Proof.* By Theorem 1, $L$ and $K$ have some forbidden words $u, v \in A^*$, *i.e.*, $L \subseteq \overline{A^*uA^*}$ and $K \subseteq \overline{A^*vA^*}$. Then $uv$ is a forbidden word of $LK$ as follows. For any word $w \in A^*uvA^*$ and any factorisation $w = xy$, either $x$ contains $u$ or $y$ contains $v$ as a subword. This means $x \notin L$ or $y \notin K$, thus $w$ is not in $LK$. $\square$

One might think that the above lemma is also true for non-regular languages, but it is *false*. Consider a language $L_{\mathtt{sq}} = \{w \in A^* \mid |w| = n^2 \text{ for some } n \geq 0\}$. This language $L_{\mathtt{sq}}$ is null, because almost every natural number is not square. However, by Lagrange's four square theorem stating that every natural number can be represented as the sum of four integer squares, we have $L_{\mathtt{sq}}^4 = A^*$ which is clearly co-null.

*Proof (of Theorem 6).* By the monotonicity and idempotency of $\mathrm{Ext}_A$ (Theorem 3), it is enough to show $\mathrm{UPol}_A \subseteq \mathrm{Ext}_A(\mathrm{LT}_A)$: this implies $\mathrm{Ext}_A(\mathrm{UPol}_A) \subseteq \mathrm{Ext}_A(\mathrm{Ext}_A(\mathrm{LT}_A)) = \mathrm{Ext}_A(\mathrm{LT}_A)$. Let $L = \biguplus_{i=1}^{k} M_i$ be an unambiguous polynomial where each $M_i$ is an unambiguous monomial and $\uplus$ represents the disjoint union.

We show that, for each monomial $M_i$, $\underline{\mu}_{\mathrm{LT}_A}(M_i) = \delta_A(M_i)$ holds, *i.e.*, we can construct a convergent sequence $(L_{i,j})_j$ of locally testable languages to $M_i$ *from inner*: $L_{i,j} \subseteq M_i$ for each $j$ and $\lim_{j\to\infty} \delta_A(L_{i,j}) = \delta_A(M_i)$. If $M_i$ is null, then clearly we can take $L_{i,j} = \emptyset$ for each $j$. Hence we assume $M_i$ is not null. In this case, $M_i$ should be of the form $M_i = A_0^* a_1 A_1^* \cdots A_{n-1}^* a_n A_n^*$ and (★) there is a *unique* $\ell$ satisfying $A_\ell = A$. We show (★). Notice that at least one $\ell$ satisfies $A_\ell = A$, because if not every $A_\ell^*$ and every $a_\ell$ is clearly null and hence these concatenation $M_i$ is also null by Lemma 1. Suppose there are two $\ell < \ell'$

with $A_\ell = A_{\ell'} = A$. In this case the word $(a_1 \cdots a_n)^2 \in M_i$ has two different factorisations:

$$(\varepsilon, a_1, \ldots, a_\ell, \underbrace{a_{\ell+1} \cdots a_n a_1 \cdots a_\ell}_{A_\ell^*}, a_{\ell+1}, \ldots, a_{\ell'}, \underbrace{\varepsilon}_{A_{\ell'}^*}, a_{\ell'+1}, \ldots, a_n, \varepsilon)$$

$$(\varepsilon, a_1, \ldots, a_\ell, \underbrace{\varepsilon}_{A_\ell^*}, a_{\ell+1}, \ldots, a_{\ell'}, \underbrace{a_{\ell'+1} \cdots a_n a_1 \cdots a_{\ell'}}_{A_{\ell'}^*}, a_{\ell'+1}, \ldots, a_n, \varepsilon)$$

This contradicts with the unambiguity of $M_i$. Hence ($\bigstar$) is true and we can write $M_i = PA^*S$ where $P = A_0^* a_1 A_1^* \cdots A_{\ell-1}^* a_\ell$ and $S = a_{\ell+1} A_{\ell+1}^* \cdots A_{n-1}^* a_n A_n^*$. Because $M_i$ is unambiguous, for each word $w \in M_i$, there is a unique factorisation $w = xyz$ where $x \in P$, $y \in A^*$ and $z \in S$, respectively. Hence, for any $n \geq 0$, we have

$$\begin{aligned}\frac{\#(M_i \cap A^n)}{\#(A^n)} &= \frac{\#(\{(x,y,z) \in P \times A^* \times S \mid |xyz| = n\})}{\#(A^n)} = \frac{\#\left(\biguplus_{(x,z)\in U_n} xA^*z \cap A^n\right)}{\#(A^n)} \\ &= \frac{\sum_{(x,z)\in U_n} \#(xA^*z \cap A^n)}{\#(A^n)} = \sum_{(x,z)\in U_n} \#(A)^{-|xz|} \end{aligned} \tag{1}$$

holds where $U_n = \bigl\{(x,z) \in P \times S \mid |x| + |z| \leq n\bigr\}$. Because the sequence $(\#(M_i \cap A^n) / \#(A^n))_n$ is bounded above by 1 and non-decreasing, the limit of (1) exists, say $\lim_{n\to\infty} (1) = \alpha$. In general, if a sequence converges to some value, then its average also converges to the same value. Hence we have $\delta_A(M_i) = \alpha$. For each $j \in \mathbb{N}$, the language $L_{i,j} = \bigcup_{(x,z)\in U_j} xA^*z$ is locally testable, because (i) for each $x, z \in A^*$, $xA^*z = (xA^* \cap A^*z) \setminus \{w \in A^* \mid |w| < |x| + |z|\}$ is locally testable, and (ii) $U_j$ is finite. Moreover, $L_{i,j} \subseteq M_i$ for each $j$ and $\delta_A(L_{i,j}) = \sum_{(x,z)\in U_j} \#(A)^{-|xz|}$. Hence $\lim_{j\to\infty} \delta_A(L_{i,j}) = \alpha = \delta_A(M_i)$, *i.e.*, $\underline{\mu}_{\mathrm{LT}_A}(M_i) = \delta_A(M_i)$. This fact implies that $\underline{\mu}_{\mathrm{LT}_A}(L) = \delta_A(L)$ because we have the following equality:

$$\underline{\mu}_{\mathrm{LT}_A}(L) = \underline{\mu}_{\mathrm{LT}_A}\left(\biguplus_{i=1}^{k} M_i\right) = \sum_{i=1}^{k} \underline{\mu}_{\mathrm{LT}_A}(M_i) = \sum_{i=1}^{k} \delta_A(M_i) = \delta_A(L).$$

Next we show $\underline{\mu}_{\mathrm{LT}_A}(\overline{L}) = \delta_A(\overline{L})$. Notice that the complement of $L$ is also an unambiguous polynomial since $\mathrm{UPol}_A$ is a local variety. Thus $\overline{L} = \biguplus_{i=1}^{k'} M_i'$ holds for some unambiguous monomials $M_i'$. Hence we can conclude that $\underline{\mu}_{\mathrm{LT}_A}(\overline{L}) = \delta_A(\overline{L}) = 1 - \delta_A(L)$ which implies $\underline{\mu}_{\mathrm{LT}_A}(L) + \underline{\mu}_{\mathrm{LT}_A}(\overline{L}) = 1$. Because $\mathrm{LT}_A$ is closed under complementation, we have $\underline{\mu}_{\mathrm{LT}_A}(K) = 1 - \overline{\mu}_{\mathrm{LT}_A}(\overline{K})$ for any $K$. Thus $\overline{\mu}_{\mathrm{LT}_A}(L) + \overline{\mu}_{\mathrm{LT}_A}(\overline{L}) = 1$, *i.e.*, $L$ is $\mathrm{LT}_A$-measurable by Theorem 2. ☐

Next we show the reverse inclusion of Theorem 6. This direction is more easy.

**Theorem 7.** $\mathrm{Ext}_A(\mathrm{UPol}_A) \supseteq \mathrm{Ext}_A(\mathrm{LT}_A)$ *for any* $A$.

*Proof.* By the monotonicity and idempotency of $\mathrm{Ext}_A$ (Theorem 3), this is equivalent to $\mathrm{LT}_A \subseteq \mathrm{Ext}_A(\mathrm{UPol}_A)$. Moreover, $\mathrm{UPol}_A$-measurability is preserved under Boolean operations by Theorem 3, we only have to show that $wA^*$, $A^*w$ and $A^*wA^*$ are all $\mathrm{UPol}_A$-measurable for each $w \in A^*$. Let $w = a_1 \cdots a_n$ where each $a_i \in A$.

First we show $wA^* \in \mathrm{Ext}_A(\mathrm{UPol}_A)$. This is easy because the language $wA^* = \emptyset^* a_1 \emptyset^* a_2 \emptyset^* \cdots \emptyset^* a_n A^*$ itself is actually an unambiguous polynomial. Similarly, we also have $A^*w \in \mathrm{UPol}_A$.

Next we show $A^*wA^* \in \mathrm{Ext}_A(\mathrm{UPol}_A)$. This language is not in $\mathrm{UPol}_A$ in general. For example, $A^*abA^*$ is not an unambiguous polynomial if $A = \{a, b, c\}$ (*cf.* [6]). Since the case $w = \varepsilon$ is trivial, we assume $w = a_1 \cdots a_n$ where $a_i \in A$ and $n \geq 1$. Define $W_k = (A^k \backslash K_k)wA^*$ where $K_k = \{u \in A^k \mid ua_1 \cdots a_{n-1} \in A^*wA^*\}$ for each $k \geq 0$. Intuitively, $W_k$ is the set of all words in which $w$ *firstly* appears at the position $k+1$ as a subword. $W_k$ is in $\mathrm{UPol}_A$ for each $k$, because it can be written as $W_k = \biguplus_{v \in (A^k \backslash K_k)} vwA^*$, where each $vwA^*$ is an unambiguous polynomial as shown above, which means that this disjoint finite union $W_k$ is also an unambiguous polynomial. Clearly, $W_i \cap W_j = \emptyset$ and $\delta_A(W_i) > 0$ for each $i \neq j$, thus we have $\biguplus_{k \geq 0} W_k = A^*wA^*$ and hence $\lim_{n \to \infty} \delta_A \left( \biguplus_{k \geq 0}^{n} W_k \right) = 1$ *i.e.*, $\underline{\mu}_{\mathrm{UPol}_A}(A^*wA^*) = 1$. Thus $A^*wA^* \in \mathrm{Ext}_A(\mathrm{UPol}_A)$. □

Combining Theorem 6 and Theorem 7, we have the following equivalence.

**Corollary 1.** $\mathrm{Ext}_A(\mathrm{LT}_A) = \mathrm{Ext}_A(\mathrm{UPol}_A)$ *for each* $A$.

We showed that $\mathrm{LT}_A$ has a certain measuring power, but yet we do not know whether $\mathrm{LT}_A$-measurability on $\mathrm{REG}_A$ is decidable or not. We only know that $\mathrm{RExt}_A(\mathrm{LT}_A)$ forms a local variety thanks to Theorem 3.

## 4 Measuring Power of Alphabet and Piecewise Testable Languages

For any alphabet $A$, $\mathrm{AT}_A$ is a finite family of regular languages, hence we can decide, for a given regular language $L \subseteq A^*$, whether $L$ is $\mathrm{AT}_A$-measurable or not: enumerate every pair $(L_1, L_2)$ of languages in $\mathrm{AT}_A$ and check $L_1 \subseteq L \subseteq L_2$ and $\delta_A(L_1) = \delta_A(L_2) = \delta_A(L)$ holds. But the next theorem gives us a more simpler way to check $\mathrm{AT}_A$-measurability than this naïve approach.

**Theorem 8.** *A co-null language* $L \subseteq A^*$ *is* $\mathrm{AT}_A$*-measurable if and only if* $L$ *contains* $\bigcap_{a \in A} A^*aA^*$.

*Proof.* Clearly, $\bigcap_{a \in A} A^*aA^* \in \mathrm{AT}_A$ and $\delta_A(\bigcap_{a \in A} A^*aA^*) = 1$ holds. Thus any language $L \supseteq \bigcap_{a \in A} A^*aA^*$ is $\mathrm{AT}_A$-measurable. If $L \not\supseteq \bigcap_{a \in A} A^*aA^*$, then any subset of $L$ in $\mathrm{AT}_A$ is null, because every language in $\mathrm{AT}_A$ not containing $\bigcap_{a \in A} A^*aA^*$ is a subset of $\bigcup_{B \subsetneq A} B^*$ and hence it is clearly null. □

We notice that the above theorem also gives a characterisation of null $\mathrm{AT}_A$-measurable languages: because $\mathrm{AT}_A$ is closed under complementation, $L$ is $\mathrm{AT}_A$-measurable if and only if $\overline{L}$ is $\mathrm{AT}_A$-measurable by Theorem 2. Hence a null language $L \subsetneq A^*$ is $\mathrm{AT}_A$-measurable if and only if $\overline{L}$ contains $\bigcap_{a \in A} A^*aA^*$. The latter condition is equivalent to the following: $\mathrm{alph}(w) \neq A$ for any $w \in L$.

Next we give a simple different characterisation of PT-measurability. The following lemma can be considered as a specialised version of Theorem 1 (a regular language is co-null if and only if it contains an ideal language $A^*wA^*$) to piecewise testable languages. Notice that $A^*wA^* \subseteq L_w$ always holds hence $L_w$ is more "larger" than $A^*wA^*$.

**Lemma 2.** *A piecewise testable language $L \in \mathrm{PT}_A$ is co-null if and only if it contains a simple monomial.*

*Proof.* ($\Leftarrow$:) this is trivial: every simple monomial $L_w$ is co-null by infinite monkey theorem.
($\Rightarrow$:) Let $L \in \mathrm{PT}_A$ be a co-null piecewise testable language. By definition of $\mathrm{PT}_A$, $L$ can be written as a finite Boolean combination of simple monomials, hence it can be written as a disjunctive normal form $L = I_1 \cup \cdots \cup I_n$ where $n \geq 1$ and each $I_i$ is the intersection of some simple monomials or complements of simple monomials. $\delta_A(L) = 1$ implies that, at least one $I_i$ is the intersection of some simple monomials (otherwise $\delta_A(L) = 0$), say $I_i = L_{w_1} \cap \cdots \cap L_{w_k}$. Hence we can conclude that $L$ contains a simple monomial $L_{w_1 \cdots w_k} \subseteq I_i \subseteq L$. $\square$

**Theorem 9.** *A co-null language $L \subseteq A^*$ is $\mathrm{PT}_A$-measurable if and only if $L_w \subseteq L$ holds for some $w \in A^*$.*

*Proof.* ($\Leftarrow$): trivial.
($\Rightarrow$): $L$ is $\mathrm{PT}_A$-measurable means there is a convergent sequence $(L_k)_k$ of piecewise testable languages to $L$ from inner. This means that, for some $i \geq 0$, $\delta_A(L_j) = 1$ holds for any $j \geq i$ because the density of each $L_k$ is either zero or one. By Lemma 2, $L_j$ contains a simple monomial $L_{w_j}$ for each $j \geq i$. Hence $L_{w_i} \subseteq L_i \subseteq L$, in particular. $\square$

We notice that the above theorem also gives a characterisation of null $\mathrm{PT}_A$-measurable languages: because $\mathrm{PT}_A$ is closed under complementation, $L$ is $\mathrm{PT}_A$-measurable if and only if $\overline{L}$ is $\mathrm{PT}_A$-measurable by Theorem 2. By using Lemma 2, we can also show that the measuring power of $\mathrm{PT}_A$ is strictly weaker than $\mathrm{ZO}_A$ as follows.

**Theorem 10.** $\mathrm{PT}_A \subsetneq \mathrm{RExt}_A(\mathrm{PT}_A) \subsetneq \mathrm{ZO}_A$ *if* $\#(A) \geq 2$.

*Proof.* $\mathrm{PT}_A \subsetneq \mathrm{RExt}_A(\mathrm{PT}_A)$ follows from the fact that any regular language $L \subseteq B^*$ is in $\mathrm{RExt}_A(\mathrm{PT}_A)$ for $B \subsetneq A$. Also, we have $\mathrm{RExt}_A(\mathrm{PT}_A) \subseteq \mathrm{RExt}_A(\mathrm{ZO}_A) = \mathrm{ZO}_A$ because $\mathrm{PT}_A \subseteq \mathrm{ZO}_A$ holds. Hence it is enough to show $\mathrm{RExt}_A(\mathrm{PT}_A) \neq \mathrm{ZO}_A$.

We show $A^*wA^* \notin \mathrm{Ext}_A(\mathrm{PT}_A)$ for any $w \in A^*$ with $|w| \geq 3$. Let $L \in \mathrm{PT}_A$ be a co-null piecewise testable language. By Lemma 2, there exists some word $u$

such that $L_u \subseteq L$. We now show that there exits $v \in L_u$ such that $v \notin A^*wA^*$ which implies $L \not\subseteq A^*wA^*$. Let $u = a_1 \cdots a_\ell$ where $a_i \in A$ for each $i$, and let $w = w'b_1b_2b_3$ where $w' \in A^*, b_j \in A$ for each $j$. If $\ell \leq 2$, it is clear that $L_u \ni a_1 \cdots a_\ell \notin A^*wA^*$ because $|w| \geq 3$. Hence we consider the case $\ell \geq 3$. We perform case analysis of $\#(A)$.

**(Case $\#(A) \geq 3$):** Let $v_1 = a_1$. We choose $v_i$ in order from $i = 2$ to $\ell$ as follows: (1) if $|v_1v_2 \cdots v_{i-1}| \geq |w|$ and the suffix of $v_1v_2 \cdots v_{i-1}a_i$ of length $|w|$ equals $w$ ($a_i = b_3$ in this case), then put $v_i = aa_i$ where $a \in (A \setminus \{b_2, b_3\})$. (2) otherwise, put $v_i = a_i$.

**(Case $\#(A) = 2$):** This case is a bit more involved. Let $A = \{a, b\}$ and let $v_1 = ab$. We choose $v_i$ in order from $i = 2$ to $\ell$ as follows: (1) if $b_1b_2b_3 \in \{aaa, aab, abb, baa, bba, bbb\}$, then put $v_i = ab$. (2) if $b_1b_2b_3 \in \{aba, bab\}$, then put $v_i = baab$. Observe that the each suffix of $v_i$ of length 2 is $ab$, hence $v_1 \cdots v_\ell \in L_{u'}$ for any $u' \in A^\ell$. Also, one can easily observe that no subword of $v_iv_{i+1}$ equals to $b_1b_2b_3$ for each $1 \leq i \leq \ell - 1$.

In both cases, by construction, no subword of $v = v_1 \cdots v_\ell$ equals to $w$ and $v \in L_u$. Thus only a null piecewise testable language can be a subset of $A^*wA^*$, hence $\underline{\mu}_{\mathrm{PT}_A}(A^*wA^*) = 0$, *i.e.*, $A^*wA^* \notin \mathrm{Ext}_A(\mathrm{PT}_A)$. ☐

Finally, we give an algebraic characterisation of $\mathrm{PT}_A$-measurability based on Theorem 9. We notice that the syntactic monoid of every co-null regular language has the zero element 0 (*cf.* [20]). We use Green's $\mathcal{J}$-relation $=_\mathcal{J}$ and $<_\mathcal{J}$ on a monoid $M$ defined by $x =_\mathcal{J} y \Leftrightarrow MxM = MyM$ and $x <_\mathcal{J} y \Leftrightarrow MxM \subsetneq MyM$, respectively (*cf.* [8]).

**Theorem 11.** *A co-null regular language $L \subseteq A^*$ is $\mathrm{PT}_A$-measurable if and only if ($\Diamond$) for every $x \in M \setminus \{0\}$ there is a letter $a \in A$ such that $x'\eta(a) <_\mathcal{J} x'$ for every $x' =_\mathcal{J} x$, where $\eta : A^* \to M$ and $M$ is the syntactic morphism and monoid of $L$, respectively.*

*Proof.* It is clear that 0 is the minimum element of $M$ with respect to $<_\mathcal{J}$. Also, we have $0 \in \eta(L)$ by infinite monkey theorem. We write $[x]$ for the $\mathcal{J}$-class of $x$.
($\Leftarrow$): Assume ($\Diamond$). Let $|M/=_\mathcal{J}| = n$ and let $[x_1], \ldots, [x_n]$ be a sequence of $\mathcal{J}$-classes of $M$ such that (1) for every $x \in M$ there is $i$ such that $x \in [x_i]$, and (2) for every $i < j$ either $x_i >_\mathcal{J} x_j$ or $x_i$ and $x_j$ are incomparable with respect to $<_\mathcal{J}$. By the assumption, for each $\mathcal{J}$-class $[x_i]$ where $i \neq n$ ($x_n = 0$ by definition), there is a letter $a_i \in A$ such that $x'\eta(a_i) <_\mathcal{J} x'$ for every $x' \in [x_i]$. Define $w = a_1 \cdots a_{n-1}$. By construction, it is clear that, for every $w_0, w_1, \ldots, w_{n-1} \in A^*$, we have $\eta(w_0a_1w_1a_2 \cdots w_{n-2}a_{n-1}w_{n-1}) = 0$. Hence $\eta(L_w) = \{0\}$, that is, we obtain $L_w \subseteq L$. This means that $L$ is $\mathrm{PT}_A$-measurable by Theorem 9.
($\Rightarrow$): Assume the contrary of ($\Diamond$). For any $w = a_1 \cdots a_k \in A^*$, we show that $L_w \not\subseteq L$ holds. This implies that $L$ is $\mathrm{PT}_A$-*im*measurable by Theorem 9. By the assumption, there is $y \in M \setminus \{0\}$ such that, for each letter $a_i$, $y_i\eta(a_i) =_\mathcal{J} y_i$ for some $y_i \in [y]$. $y_i\eta(a_i) =_\mathcal{J} y_i =_\mathcal{J} y$ means that there is a pair $(x_i, z_i)$ such that $x_iy_i\eta(a_i)z_i = y$. Also, for each $y_i \in [y]$, there is a pair $(x'_i, z'_i)$ such that $x'_iyz'_i = y_i$. For each $i$, let $u_i \in \eta^{-1}(x_i), v_i \in \eta^{-1}(y_i), w_i \in \eta^{-1}(z_i)$ and $u'_i \in \eta^{-1}(x'_i), w'_i \in \eta^{-1}(z'_i)$. Define $t_1 = u_1v_1a_1w_1$ and $t_i = u_iu'_it_{i-1}w'_ia_iw_i$ for each $2 \leq i \leq k$. By

| **Language** | **Algebra** | **Logic** | **Measurability** |
|---|---|---|---|
| SF | aperiodic | FO | SF $\subsetneq$ $\mathrm{RExt}_A$(SF) $\subsetneq$ REG [22] |
| LT | locally idempotent and commutative | | $\mathrm{Ext}_A$(LT) = $\mathrm{Ext}_A$(UPol) |
| UPol | **DA** | $\mathrm{FO}^2$ | |
| PT | $\mathcal{J}$-trivial | $\mathbb{B}\Sigma_1$ | PT $\subsetneq$ $\mathrm{RExt}_A$(PT) $\subsetneq$ ZO<br>$L$ is PT-measurable iff $L$ or $\overline{L}$ contains a simple monomial |
| AT | idempotent and commutative | $\mathrm{FO}^1$ | AT $\subsetneq$ $\mathrm{RExt}_A$(AT) $\subsetneq$ $\mathrm{RExt}_A$(PT)<br>$L$ is AT-measurable iff $L$ or $\overline{L}$ contains $\bigcap_{a \in A} A^*aA^*$ |

**Table 1.** Correspondence of language-algebra-logic and summary of our results.

straightforward induction, we can show that $\eta(t_i) = y$ holds for every $1 \leq i \leq k$. It also clear that $t_k \in L_w$. Because $y \neq 0$, there is some $x, z \in M$ such that $xyz \notin \eta(L)$ (if not $y = 0$ holds by the definition of the syntactic monoid of $L$). This means that $L_w \ni w_x t_k w_z \notin L$ where $w_x \in \eta^{-1}(x), w_z \in \eta^{-1}(z)$. Hence we obtain $L_w \not\subseteq L$. ☐

Because the syntactic monoid of every regular language is finite, the condition ($\diamondsuit$) is decidable.

**Corollary 2.** $\mathrm{PT}_A$*-measurability is decidable for* $\mathrm{REG}_A$.

## 5 Summary and Future Work

For simplicity, in this section we only consider alphabets with two more letters, and omit the subscript $A$ for denoting local varieties. Table 1 shows algebraic and logical counterparts of local varieties we considered (left) and a summary of our results (right). Here $\mathrm{FO}^n$ stands for first-order logic with $n$-variables and $\mathbb{B}\Sigma_1$ is the Boolean closure of existential first-order logic. The hierarchy of languages is strictly decreasing top down excluding that LT and UPol (PT, respectively) are incomparable. All algebraic and logical counterparts in Table 1 are nicely described in a survey [6], with the sole exception LT [9,24,4].

Our future work are two kinds.

(1) Prove or disprove $\mathrm{Ext}_A(\mathrm{LT}) \subsetneq \mathrm{Ext}_A(\mathrm{SF})$.
(2) Prove or disprove the decidability of LT-measurability.

To show the decidability, perhaps we can use some known techniques related to locally testable languages, for example, the so-called *separation problem* for a language class $\mathcal{C}$: for a given pair of regular languages $(L_1, L_2)$, is there a language $L$ in $\mathcal{C}$ such that $L_1 \subseteq L$ and $L \cap L_2 = \emptyset$ ($L$ "separates" $L_1$ and $L_2$)? It is known that the separation problem for PT, LT, and SF are all decidable [12,13,14]. Theorem 8 and Theorem 9 says that, AT-measurability and PT-measurability does not rely on the existence of an infinite convergent sequence, but relies on the existence of a *single* language $\cap_{a \in A} A^* a A^*$ and $L_w$ as a subset, respectively. But from Theorem 5, we can observe that, the situation of LT-measurability is essentially different: LT-measurability heavily relies on the existence of an *infinite sequence* of different locally testable languages. Because the density of every regular language is rational (*cf.* [16]), for each $\mathrm{LT}_A$-measurable language $L$ with an irrational density, there is no single pair of regular languages $(L_1, L_2)$ such that $L_1 \subseteq L \subseteq L_2$ and $\delta_A(L_1) = \delta_A(L_2) = \delta_A(L)$.

Between SF and LT, there is a fine-grained infinite hierarchy called the *dot-depth hierarchy* originally introduced by Cohen and Brzozowski [5] in 1970. For a family $\mathcal{C}$ of languages, we denote by $\mathscr{M}\mathcal{C} = \{L_1 \cdots L_k \mid k \geq 1, L_1, \ldots, L_k \in \mathcal{C}\} \cup \{\{\varepsilon\}\}$ the monoid closure of $\mathcal{C}$. The dot-depth hierarchy starts with the family $\mathcal{B}_0$ of all finite or co-finite languages, and continues as $\mathcal{B}_{i+1} = \mathscr{B}\mathscr{M}\mathcal{B}_i$ for each $i \geq 0$. Brzozowski and Knast [3] showed that this infinite hierarchy is strict: $\mathcal{B}_i \subsetneq \mathcal{B}_{i+1}$ for each $i \geq 0$. By definition, we have $\mathrm{SF} = \bigcup_{i \geq 0} \mathcal{B}_i$, and actually, we also have $\mathcal{B}_0 \subsetneq \mathrm{LT} \subsetneq \mathcal{B}_1$ because each of $wA^*, A^*w$ and $A^*wA^*$ is obtained by concatenating a finite language $\{w\}$ and a co-finite language $A^*$. Although the dot-depth hierarchy was introduced in a half-century before and much ink has been spent on it, the decidability of the membership problem for $\mathcal{B}_i$ is open for $i \geq 3$ and the research on this topic is still active: see a survey [11] or a recent progress given by Place and Zeitoun [15] that shows the decidability of the separation problem for $\mathcal{B}_2$, which implies the decidability of membership of $\mathcal{B}_2$. The equation $\mathrm{Ext}_A(\mathrm{LT}) = \mathrm{Ext}_A(\mathrm{SF})$ means that the dot-depth hierarchy *collapses* via $\mathrm{Ext}_A$. But if not, it might be interesting to consider the new hierarchy $\mathcal{B}_0 = \mathrm{Ext}_A(\mathcal{B}_0) \subsetneq \mathrm{Ext}_A(\mathcal{B}_1) \subseteq \mathrm{Ext}_A(\mathcal{B}_2) \subseteq \cdots \subseteq \mathrm{Ext}_A(\mathrm{SF})$.

**Acknowledgements:** I am grateful to my colleague Y. Yamaguchi (Osaka University) and Y. Nakamura (Tokyo Tech) for useful discussions. I also thank to anonymous reviewers for many valuable comments. This work was supported by JSPS KAKENHI Grant Number JP19K14582 and JST ACT-X Grant Number JPMJAX210B, Japan.

## References

1. Adámek, J., Milius, S., Myers, R.S.R., Urbat, H.: Generalized eilenberg theorem I: Local varieties of languages. In: Foundations of Software Science and Computation Structures. pp. 366–380 (2014)
2. Berstel, J., Perrin, D.: Theory of codes, Pure and Applied Mathematics, vol. 117. Academic Press Inc. (1985)
3. Brzozowski, J., Knast, R.: The dot-depth hierarchy of star-free languages is infinite. Journal of Computer System Sciences **16**, 37–55 (1978)
4. Brzozowski, J., Simon, I.: Characterizations of locally testable events. vol. 4, pp. 243–271 (1973)
5. Cohen, R.S., Brzozowski, J.: Dot-depth of star-free events. Journal of Computer System Sciences **5**, 1–16 (1971)

6. Diekert, V., Gastin, P., Kufleitner, M.: A survey on small fragments of first-order logic over finite words. Int. J. Found. Comput. Sci. **19**(3), 513–548 (2008)
7. Gehrke, M., Grigorieff, S., Pin, J.: Duality and equational theory of regular languages. In: Automata, Languages and Programming. pp. 246–257 (2008)
8. Lawson, M.V.: Finite Automata. Chapman and Hall/CRC (2004)
9. McNaughton, R.: Algebraic decision procedures for local testability. Mathematical systems theory **8**, 60–76 (1974)
10. McNaughton, R., Papert, S.: Coutner-Free Automata. The MIT Press (1971)
11. Pin, J.: The dot-depth hierarchy, 45 years later. In: The Role of Theory in Computer Science - Essays Dedicated to Janusz Brzozowski. pp. 177–202 (2017)
12. Place, T., van Rooijen, L., Zeitoun, M.: Separating Regular Languages by Locally Testable and Locally Threshold Testable Languages. In: IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science. pp. 363–375 (2013)
13. Place, T., van Rooijen, L., Zeitoun, M.: Separating regular languages by piecewise testable and unambiguous languages. In: Mathematical Foundations of Computer Science 2013. pp. 729–740 (2013)
14. Place, T., Zeitoun, M.: Separating regular languages with first-order logic. Log. Methods Comput. Sci. **12**(1) (2016)
15. Place, T., Zeitoun, M.: Separation for dot-depth two. Logical Methods in Computer Science **17** (Sep 2021)
16. Salomaa, A., Soittola, M.: Automata Theoretic Aspects of Formal Power Series. Springer-Verlag New York, Inc., Secaucus, NJ, USA (1978)
17. Schützenberger, M.: Sur le produit de concaténation non ambigu. Semigroup Forum **13**, 47–75 (1976)
18. Schützenberger, M.P.: On finite monoids having only trivial subgroups. Information and Control **8**(2), 190–194 (1965)
19. Simon, I.: Piecewise testable events. In: Automata Theory and Formal Languages. pp. 214–222 (1975)
20. Sin'ya, R.: An automata theoretic approach to the zero-one law for regular languages. In: Games, Automata, Logics and Formal Verification. pp. 172–185 (2015)
21. Sin'ya, R.: Asymptotic approximation by regular languages. In: Current Trends in Theory and Practice of Computer Science. pp. 74–88 (2021)
22. Sin'ya, R.: Carathéodory extensions of subclasses of regular languages. In: Developments in Language Theory. pp. 355–367 (2021)
23. Sin'ya, R.: Measuring power of locally testable languages (full version) (2022), `http://www.math.akita-u.ac.jp/~ryoma/misc/LTmeasure_full.pdf`
24. Zalcstein, Y.: Locally testable languages. Journal of Computer and System Sciences **6**(2), 151–167 (1972)