Relations between language classes in terms of insertion and locality

Kaoru Fujioka *

1 Introduction

Insertion systems use only insertion operations of the form $(u, x, v)$ and produce a string $\alpha uv\beta$ for a given string $\alpha u v \beta$ by inserting the string $x$ between $u$ and $v$. From the definition of insertion operations, using only insertion operations, we generate only context-sensitive languages.

Using insertion systems together with some morphisms, characterizing recursively enumerable languages is obtained in [8], [6]. Furthermore, similarly to the Chomsky–Schützenberger representation theorem [1], each recursively enumerable language can be expressed using an insertion system and a Dyck language in [7], and each context-free language can be expressed using an insertion system and a star language in [5].

In [2] and [3], within the framework of the Chomsky–Schützenberger representation theorem, some characterizations and representation theorems of languages in the Chomsky hierarchy have been provided by insertion system $\gamma$, strictly locally testable language $R$, and morphism $h$ such as $h(L(\gamma) \cap R)$.

The purpose of this paper is to clarify the relation between the classes of languages $h(L(\gamma) \cap R)$ using insertion systems of weight $(i, 0)$ for $i \geq 1$ and those using insertion systems of weight $(i, 1)$ for $i \geq 1$.

2 Preliminaries

For a string $x \in V^*$ with an alphabet $V$, $|z|$ is the length of $z$. For $0 \leq k \leq |x|$, let $Pre_k(x)$ and $Suf_k(x)$ respectively denote the prefix and the suffix of $x$ with length $k$. For $0 \leq k \leq |x|$, let $Int_k(x)$ be the set of intermediate substrings of $x$ with length $k$.

For a positive integer $k$, a language $L$ over $T$ is strictly $k$-testable if a triplet $S_k = (A, B, C)$ exists with $A, B, C \subseteq T^k$ such that, for any $w$ with $|w| \geq k$, $w$ is in $L$ iff $Pre_k(w) \in A$, $Suf_k(w) \in B$, $Int_k(w) \subseteq C$. A language $L$ is strictly locally testable iff there exists an integer $k \geq 1$ such that $L$ is strictly $k$-testable.

Note that, for an alphabet $T$, a language $T^*$ is a strictly 1-testable language.

Let $LOC(k)$ be the class of strictly $k$-testable languages. There is the following result.

**Theorem 1** [4] $LOC(1) \subset LOC(2) \subset \cdots \subset LOC(k) \subset \cdots \subset REG$.

We define an insertion system $\gamma = (T, P, A)$, where $T$ is an alphabet, $P$ is a finite set of insertion rules of the form $(u, x, v)$ with $u, x, v \in T^*$, and $A$ is a finite set of strings over $T$ called axioms.

We write $\alpha \Rightarrow^r_\gamma \beta$ if $\alpha = \alpha_1ux\alpha_2$ and $\beta = \alpha_1uxv\alpha_2$ for some insertion rule $r : (u, x, v) \in P$ with $\alpha_1, \alpha_2 \in T^*$. We write $\alpha \implies \beta$ if no confusion exists. The reflexive and transitive closure of $\implies$ is defined as $\implies^*$.
A language generated by $\gamma$ is defined as

$L(\gamma) = \{w \in T^* \mid s \Rightarrow^*_\gamma w, \text{ for some } s \in A\}$.

An insertion system $\gamma = (T, P, A)$ is said to be of weight $(m, n)$ if

$$m = \max\{ |x| \mid (u, x, v) \in P \},$$

$$n = \max\{ |u| \mid (u, x, v) \in P \text{ or } (v, x, u) \in P \}.$$  

For $m, n \geq 0$, let $INS^m_n$ be the class of all languages generated by insertion systems of weight $(m', n')$ with $m' \leq m$ and $n' \leq n$. We use $*$ instead of $m$ or $n$ if the parameter is not bounded.

**Theorem 2** [8]

1. $INS^i_0 \subseteq INS^{i'}_{i'}$ (0 \leq i \leq i', 0 \leq j \leq j').$

2. $INS^1_0 \subset CF$.

A mapping $h : V^* \to T^*$ is called morphism if $h(\lambda) = \lambda$ and $h(xy) = h(x)h(y)$ hold for all $x, y \in V^*$. For any $a$ in $T$, if $h(a) = a$ holds, then $h$ is an identity morphism.

The following results related to Chomsky-Schützenberger like characterization are obtained using insertion systems of weight $(i, 0)$ or $(i, 1)$ for $i \geq 1$ and strictly $k$-testable languages $(k \geq 1)$.

**Theorem 3** [2]

1. $H(INS^0_0 \cap LOC(1)) \subset REG$.

2. $H(INS^0_0 \cap LOC(k)) = REG$ $(k \geq 2)$.

3. $H(INS^0_0 \cap LOC(1))$ and $REG$ are incomparable $(i \geq 2)$.

4. $H(INS^0_0 \cap LOC(1)) \subset CF$ $(i \geq 2)$.

5. $H(INS^0_0 \cap LOC(k)) = CF$ $(i, k \geq 2)$.

**Theorem 4** [3]

1. $H(INS^1_1 \cap LOC(k)) = CF$ $(i \geq 1, k \geq 2)$.

2. $H(INS^1_1 \cap LOC(1)) \subset CF$ $(i \geq 1)$.

In the present paper, we specifically examine the relation between language classes $H(INS^0_0 \cap LOC(k_0))$ and $H(INS^1_1 \cap LOC(k_1))$ for $i_0, k_0, i_1, k_1 \geq 1$.

### 3 Main Results

For context-free languages, from Theorem 3 and Theorem 4, we obtain

$$CF = H(IS^0_0 \cap LOC(k_0))$$

$$= H(IS^1_1 \cap LOC(k_1))$$

with $i_0, k_0, k_1 \geq 2, i_1 \geq 1$.

We next examine the language class $H(IS^0_0 \cap LOC(1))$. From Theorem 3, $H(IS^0_0 \cap LOC(1))$ and $REG$ are known to be incomparable.

**Theorem 5** $H(IS^0_0 \cap LOC(1))$ and $H(IS^1_1 \cap LOC(1))$ are incomparable.

**Proof** Consider an insertion system $\gamma_1 = (T, \{(\lambda, ab, \lambda), \{\lambda\})$ of weight $(2, 0)$ with $T = \{a, b\}$, a strictly 1-testable language $R = T^+$, and an identity morphism $h : T^* \to T^*$. The above definition indicates directly that $L(\gamma) = h(L(\gamma) \cap R)$.

We can show that $L(\gamma_1)$ is not in $H(IS^0_0 \cap LOC(1))$ by contradiction. We omit the proof here.

Now we consider an insertion system $\gamma_2 = (T, \{(a, a, a), \{b, b,\lambda\}, \{a, b\}) of weight $(1, 1)$ with $T = \{a, b\}$, a strictly 1-testable language $R = T^+$, and an identity morphism $h : T^* \to T^*$. From the definition, we have $L(\gamma_2) = h(L(\gamma_2) \cap R) = \{a^i \mid i \geq 1\} \cup \{b^i \mid i \geq 1\}$.

From [2], $L(\gamma_2)$ is not in $H(IS^0_0 \cap LOC(1))$. □

Theorem 5 implies the following Corollaries.

**Corollary 1** $H(IS^0_0 \cap LOC(1))$ and $H(IS^1_1 \cap LOC(1)) \cap H(IS^2_0 \cap LOC(2))$ are incomparable.
Corollary 2 $H(INS_{0}^{i} \cap LOC(1)) \subset H(INS_{1}^{i} \cap LOC(1)) (i \geq 2)$.

For the class of languages $H(INS_{0}^{i} \cap LOC(1))$, from the size of parameters, we have the inclusions $H(INS_{0}^{i} \cap LOC(1)) \subseteq H(INS_{1}^{i} \cap LOC(1))$ and $H(INS_{1}^{i} \cap LOC(1)) \subseteq H(INS_{2}^{i} \cap LOC(1))$. Next we present the following proper inclusion.

Theorem 6 $H(INS_{0}^{i} \cap LOC(1)) \subset H(INS_{1}^{i} \cap LOC(1)) \cap H(INS_{2}^{i} \cap LOC(2))$.

Proof To show the proper inclusion, we consider an insertion system $\gamma_{2} = (T, \{(a,a,\lambda), (b,b,\lambda)\}, \{a,b\})$ of weight $(1,1)$ with $T = \{a,b\}$, a strictly 1-testable language $R = T^{*}$, and an identity morphism $h : T^{*} \to T^{*}$.

In a similar way to Theorem 5, we can show that $L(\gamma_{2})$ is not in $H(INS_{0}^{i} \cap LOC(1))$. $\square$

Corollary 3 $H(INS_{0}^{i} \cap LOC(1)) \subseteq H(INS_{1}^{i} \cap LOC(1)) \cap H(INS_{2}^{i} \cap LOC(2)) \cap H(INS_{3}^{i} \cap LOC(1))$.

4 Concluding Remarks

In the present paper, we specifically examined the language classes $H(INS_{0}^{i} \cap LOC(k_{0}))$ and $H(INS_{1}^{i} \cap LOC(k_{1}))$ for $i_{0}, i_{1}, k_{0}, k_{1} \geq 1$ and considered the relations of those language classes.

The following remain as open problems:

- $H(INS_{0}^{i} \cap LOC(1)) \cap H(INS_{1}^{i} \cap LOC(1)) = H(INS_{0}^{i} \cap LOC(1))$ holds?

- $H(INS_{0}^{i} \cap LOC(1)) \cap H(INS_{1}^{i} \cap LOC(1)) \cap H(INS_{2}^{i} \cap LOC(1)) \cap H(INS_{3}^{i} \cap LOC(1))$ holds?

- $CF = H(INS_{m}^{i} \cap LOC(k))$ holds for some $m, k \geq 1$?

References


