Computer Science Journal of Moldova, vol.**32**, no.3(96), 2024

Insertion Systems Controlled by Ideals and Codes

Jürgen Dassow   Bianca Truthe

Abstract

We study the generative power of controlled insertion systems where the control languages are special codes or ideals instead of arbitrary regular languages.

Keywords: insertion systems, controlled insertions, subregular languages, ideals, codes.

MSC 2020: 68R45, 68Q09.

1 Introduction

Controlled insertion systems, as considered here, were introduced in [1] as extensions of prefixal systems as well as insertion systems with left context only. In [2], the investigation of controlled insertion systems was continued. It is required that the insertion of a word x into a word uv after the subword u is only allowed if u belongs to a regular set R_x which is associated with x.

A survey about similar systems (insertion-deletion systems, semi-Thue systems, prefixal systems) together with references can be found in [1] and [2]. In the latter paper, controlled insertion systems were investigated where the control languages are all taken from a certain subfamily of the family of the regular languages. The families considered were the sets of all finite, nilpotent, definite, regular non-counting, monoidal, combinational, regular commutative, regular circular, regular suffix-closed, and union-free languages. In the present paper, this research is continued by investigating the impact of control languages which are ideals or codes of a certain type.

doi:10.56415/csjm.v32.19

2 Preliminaries

We assume that the reader is familiar with the basic concepts of formal language theory (see, e. g., [3]). We only recall here some notations used in the paper.

Let V be an alphabet. By V^*, we denote the set of all words (strings) over the alphabet V (including the empty word λ).

By FIN and REG, we denote the families of finite and regular languages, respectively.

2.1 Ideals and Codes

In the sequel, let V be an alphabet. We now introduce the notion of an ideal in V^* from the theory of rings and semigroups.

A non-empty language $L \subseteq V^*$ is called a *right (left) ideal* if and only if, for any words $v \in V^*$ and $u \in L$, we have $uv \in L$ ($vu \in L$, respectively). It is easy to see that the language L is a right (left) ideal if and only if there is a language L' such that $L = L'V^*$ ($L = V^*L'$, respectively).

We now present some notions from coding theory, especially some special codes. For details, we refer to [4] and [5].

A language $L \subseteq V^*$ is called

- a *code* if and only if, for any numbers $n \geq 1$, $m \geq 1$, and words
$$x_1, x_2, \ldots, x_n, y_1, y_2, \ldots y_m \in L$$
such that
$$x_1 x_2 \ldots x_n = y_1 y_2 \ldots y_m,$$
we have the equalities $n = m$ and $x_i = y_i$ for $1 \leq i \leq n$ (i. e., a word of L^* has a unique decomposition into code words.

- *uniform* if and only if $L \subseteq V^n$ for some $n \geq 1$ (all words have the same length);

- *prefix* if and only if, for any words $u \in L$ and $v \in V^*$ such that $uv \in L$, we have $v = \lambda$ (i. e., any proper prefix of a word in L is not in L);

- *suffix* if and only if, for any words $u \in L$ and $v \in V^*$ such that $vu \in L$, we have $v = \lambda$ (i. e., any proper suffix of a word in L is not in L);

- *bifix* if and only if it is prefix as well as suffix;

- *infix* if and only if, for any $u \in L$, and $v, v' \in V^*$ such that $vuv' \in L$, we have $v = v' = \lambda$ (i. e., any proper subword of a word in L is not in L).

Note that uniform, prefix, suffix, bifix, and infix languages are codes.

A code $L \subseteq V^*$ is called

- *outfix* if and only if, for any words $u \in V^*$ and $v, v' \in V^*$ such that $vv' \in L$ and $vuv' \in L$, we have $u = \lambda$;

- *reflective* if and only if, for any words $u, v \in V^*$ such that $uv \in L$, we have $vu \in L$.

By *rId*, *lId*, *C*, *PfC*, *SfC*, *BfC*, *IfC*, *OfC*, *RC*, and *UC*, we denote the families of regular right ideals, regular left ideals, regular codes, regular prefix codes, regular suffix codes, regular bifix codes, regular infix codes, regular outfix codes, regular reflective codes, and uniform codes, respectively.

Let

$$\mathcal{G} = \{FIN, REG, rId, lId, C, PfC, SfC, BfC, IfC, OfC, RC, UC\}.$$

In [6], it was proved that any uniform code, any regular outfix code, and any regular reflective code is finite. Further relations, especially those depicted in Figure 1, are proved in [6], [4], and [5].

Lemma 1. *The hierarchy of the classes in $\mathcal{G}$ is presented in Figure 1.* □

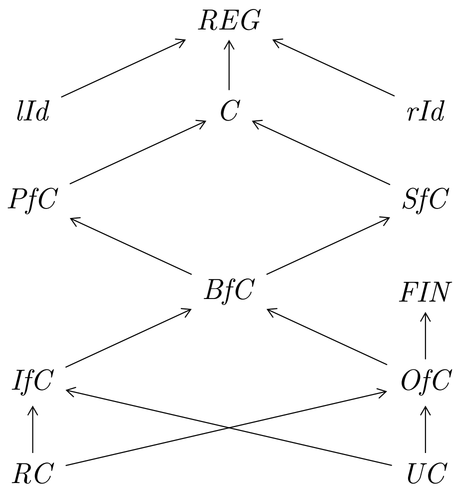

Figure 1. Hierarchy of subregular languages families of $\mathcal{G}$ (an arrow from X to Y denotes $X \subset Y$, and if two families are not connected by a directed path, then they are incomparable)

2.2 Controlled Insertion Systems

We now give the definition of the central concept of this paper.

Definition 1. *A controlled insertion system is an $(n+2)$-tuple*

$$G = (V, (R_1, I_1), (R_2, I_2), \ldots, (R_n, I_n), A), \tag{1}$$

where $n \geq 1$ is a natural number, V is an alphabet, A is a finite non-empty subset of V^*, and, for $1 \leq j \leq n$, $R_j \subseteq V^*$ is a regular set and $I_j \subset V^*$ is a finite set of non-empty words.

We say that a word x generates or derives the word y according to G, written as $x \Longrightarrow_G y$, if there are words $x_1 \in V^*$ and $x_2 \in V^*$, an integer j, $1 \leq j \leq n$, and a word $w \in I_j$, such that $x = x_1x_2$, $x_1 \in R_j$, and $y = x_1wx_2$. By $\Longrightarrow_G^*$, we denote the reflexive and transitive closure of $\Longrightarrow_G$.

The language $L(G)$ generated by G consists of all words z such that $a \Longrightarrow_G^* z$ for some $a \in A$.

For $1 \leq j \leq n$, I_j is called an insertion set and R_j is called the control set of I_j. The elements of A are called axioms.

If we want to specify which pair (R_j, I_j) is used in a derivation step, then we write $x \Longrightarrow_{(R_j,I_j)} y$. Moreover, if G and/or the pair is clear from the context, then we simply write $x \Longrightarrow y$.

We set

$$m(G) = \max\{\ |w| \mid w \in A\ \} + \max\{\ |p| \mid p \in I_j, 1 \leq j \leq n\ \} + 1.$$

If $x \in I_j$ for some j, $1 \leq j \leq n$, then $|x| \leq m(G) - 1$.

We note some properties of words in $L(G)$. If $z \in L(G)$ and $z \notin A$, then there is a derivation

$$a = w_0 \Longrightarrow_G w_1 \Longrightarrow_G w_2 \Longrightarrow_G \cdots \Longrightarrow_G w_{k-1} \Longrightarrow_G w_k = z$$

for some $k \geq 1$ and words $a \in A$, $w_1, w_2, \ldots, w_{k-1} \in V^*$. Obviously, by definition, $w_i \in L(G)$ for $0 \leq i \leq k$. Thus, for any $z \in L(G)$, $z \notin A$, there is a word $z' \in L(G)$ such that $z' \Longrightarrow_G z$ and $|z'| > |z| - m(G)$.

On the other hand, if $z \in L(G)$ and $z = uv$ with $u \in R_j$ for some index j with $1 \leq j \leq n$, then $uxv \in L(G)$ for all $x \in I_j$.

Let us give an example.

Example 2. *We consider the controlled insertion system*

$$G_1 = (\{a, b, c\},\ (\{bca, ca, a, \lambda\}, \{a\}),\ \{bca, bcb\})$$

with only one pair of control set and insertion set. The only word that can be inserted is a. Moreover, since λ is in the control set, if z is in $L(G_1)$, then az is in $L(G_1)$, too, because $z = \lambda z$ allows an insertion of a at the beginning of the word. Furthermore, if az is in $L(G_1)$, then we also have $aaz \in L(G_1)$, since a is in the control set. Because bca is in the control set, we can insert an a after bca in any word $bcaw$. Since ca is not the beginning of an axiom and no word starting with ca can be produced by insertions of a, the control word ca has no influence on the derivations in G_1. Combining these considerations, we obtain

$$L(G_1) = \{\ a^p bca^q \mid p \geq 0, q \geq 1\ \} \cup \{\ a^r bcb \mid r \geq 0\ \},$$

where the two given sets are generated from bca and bcb, respectively.

We now define the families of sets generated by controlled insertion systems with special control sets.

Definition 2. *Let $F \in \mathcal{G}$. We say that a controlled insertion system G as in* (1) *is of type F if $R_i \in F$ holds for all i, $1 \leq i \leq n$.*

We define

$$\mathcal{I}(F) = \{ \, L(G) \mid G \textit{ is a controlled insertion system of type } F \, \}$$

as the family of all languages generated by controlled insertion systems of type F.

If $F_1 \subseteq F_2$ for two families F_1 and F_2 of $\mathcal{F}$, then it is obvious that any controlled insertion system of type F_1 is a system of type F_2, too. Thus, we immediately get the following lemma.

Lemma 3. *For any two families F_1 and F_2 of $\mathcal{F}$, it holds that if $F_1 \subseteq F_2$, then $\mathcal{I}(F_1) \subseteq \mathcal{I}(F_2)$.* □

3 Control by Codes

We start with a general result on families where every language can be represented as a union of languages of another family.

Lemma 4. *Let X and Y be two language families such that $Y \subseteq X$ and every language in X has a representation as a finite union of languages in Y. Then we have $\mathcal{I}(X) = \mathcal{I}(Y)$.*

Proof. By Lemma 3, we have $\mathcal{I}(Y) \subseteq \mathcal{I}(X)$.

Let L be a language in $\mathcal{I}(X)$. Then $L = L(H)$ for some controlled insertion system $H = (V, (R_1, I_1), (R_2, I_2), \ldots, (R_n, I_n), A)$ with control sets $R_i \in X$ and insertion sets $I_i \subset V^*$ for $1 \leq i \leq n$. By supposition, for $1 \leq i \leq n$, each R_i can be represented as

$$R_i = C_{i,1} \cup C_{i,2} \cup \cdots \cup C_{i,m_i}$$

for some integers $m_i \geq 1$ and languages $C_{i,j} \in Y$, $1 \leq j \leq m_i$. We consider the controlled insertion system

$$\begin{aligned} H' \;\; = \;\; & (V, (C_{1,1}, I_1), \ldots, (C_{1,m_1}, I_1), (C_{2,1}, I_2), \ldots, (C_{2,m_2}, I_2), \\ & \ldots, (C_{n,1}, I_n), \ldots, (C_{n,m_n}, I_n), A). \end{aligned}$$

If $z \Longrightarrow_{(R_i,I_i)} z'$ is a derivation step in H, then there is a decomposition $z = z_1z_2$ with $z_1 \in R_i$ and $z' = z_1vz_2$ with $v \in I_i$. Moreover, $z_1 \in C_{i,j}$ holds for some j, $1 \leq j \leq m_i$. Therefore, $z \Longrightarrow_{(C_{i,j},I_i)} z'$ is a derivation step in H'. Conversely, if $u \Longrightarrow_{(C_{i,j},I_i)} u'$ is a derivation step in H' for some i and j, $1 \leq i \leq n$ and $1 \leq j \leq m_i$, then $u = u_1u_2$ with $u_1 \in C_{i,j}$ and $u' = u_1v'u_2$ with $v' \in I_i$. Because $C_{i,j} \subseteq R_i$, we obtain that $u \Longrightarrow_{(R_i,I_i)} u'$ is a derivation in H. Thus, a derivation step can be performed in H if and only if it can be performed in H'. This implies $L(H) = L(H')$. Since all control sets of H' are in Y, we obtain $L(H) = L(H') \in \mathcal{I}(Y)$ and $\mathcal{I}(X) \subseteq \mathcal{I}(Y)$.

Now the equality follows. □

Corollary 5. *The following equalities hold:*

$$\mathcal{I}(FIN) = \mathcal{I}(OfC) = \mathcal{I}(UC) \text{ and } \mathcal{I}(BfC) = \mathcal{I}(IfC).$$

Proof. Obviously, every finite language L is a finite union of uniform codes (each uniform part of the union consists of all words of L with the same length). By Lemma 4, we obtain that $\mathcal{I}(FIN) = \mathcal{I}(UC)$ follows. This further leads, by the relations in Figure 1 and Lemma 3, to

$$\mathcal{I}(FIN) = \mathcal{I}(UC) \subseteq \mathcal{I}(OfC) \subseteq \mathcal{I}(FIN).$$

So, both equalities hold.

From [7], we know that every regular bifix code can be represented as a union of finitely many regular infix codes. Therefore, Lemma 4 implies $\mathcal{I}(BfC) = \mathcal{I}(IfC)$. □

We now show that reflective codes as control languages have a smaller power than uniform codes.

Lemma 6. *We have $\mathcal{I}(RC) \subset \mathcal{I}(UC)$.*

Proof. Since every regular reflective code is a finite code, we get the relation $\mathcal{I}(RC) \subseteq \mathcal{I}(FIN) = \mathcal{I}(UC)$ by Lemma 3 and Corollary 5.

We now present a controlled insertion system with control sets which are all uniform codes whose language cannot be generated by a controlled insertion system with reflective codes as control sets. We consider

$$H = (\{a,b\},\ (\{ab\},\{ab\}),(\{aab\},\{ab\}),(\{ba\},\{ba\}),\{aab,ab,ba\}).$$

We can only insert ab after ab or aab and ba after ba. Hence, we get

$$L(H) = \{aab,ab\}\{ab\}^* \cup \{ba\}^+.$$

Now assume that there is a controlled insertion system

$$H' = (V,(R_1,I_1),(R_2,I_2),\ldots,(R_n,I_n),A)$$

with $L(H') = L(H)$ where all control sets R_i, $1 \le i \le n$, are regular reflective codes.

Let m be the maximal length of words in A. We consider a word $(ab)^r$ with $r \ge 2m$. Then there are a word w and an i, $1 \le i \le n$, such that $w \Longrightarrow_{(R_i,I_i)} (ab)^r$. By the structure of the words in $L(H)$, we have the following possibilities for w, R_i, and I_i:

- $w = aab(ab)^s$ with $s \le r-2$. Then we obtain $a \in R_i$, $b(ab)^{r-2-s} \in I_i$. But then we can insert $b(ab)^{r-2-s}$ also after a in the word $ab \in L(H)$ which gives $ab(ab)^{r-2-s}bab \notin L(H)$.

- $w = (ab)^s$. Then we have $(ab)^t a \in R_i$, $t \ge 1$, and $(ba)^u \in I_i$ or $(ab)^{t'} \in R_i$, $t' \ge 1$, and $(ab)^{u'} \in I_i$. If $t = 0$ in the former case, we can insert $(ba)^u$ after the first letter of $aab \in L(H)$ and obtain $a(ba)^u ab \notin L(H)$. If $t \ge 1$ in the former case, we have $a(ab)^t \in R_i$ because R_i is reflective. Then we can insert $(ba)^u$ after $a(ab)^t \in L(H)$ and get $a(ab)^t(ba)^u \notin L(H)$. In the latter case, we have $(ba)^{t'}$ since R_i is reflective. Inserting $(ab)^{u'}$ after $(ba)^{t'} \in L(H)$, we obtain $(ba)^{t'}(ab)^{u'} \notin L(H)$.

Therefore, in all cases, we can generate a word not in $L(H)$ which contradicts the assumption $L(H) = L(H')$. Consequently, our assumption is false and $L(H) \notin \mathcal{I}(RC)$. □

Now we present a partial incomparability result for suffix and prefix codes.

Lemma 7. *Let L be the language generated by the controlled insertion system*

$$G = (\{a,b\}, (\{ba^n \mid n \geq 0\}, \{ab\}), \{bab\}).$$

Then we have $L \in \mathcal{I}(SfC)$ and $L \notin \mathcal{I}(PfC)$.

Proof. We first prove $L \in \mathcal{I}(SfC)$. Since the only control set of G is a regular suffix code, we have $L \in \mathcal{I}(SfC)$ by definition.

We now prove $L \notin \mathcal{I}(PfC)$. We do not determine the language L completely. We only mention a property of the words in L. By induction on the length of the derivation of words in L, it is easy to show that any word in L has the form $ba^{n_1}b^{m_1}a^{n_2}b^{m_2}\dots a^{n_k}b^{m_k}$ with $k \geq 1$, and $n_i > 0$, $m_i > 0$ for $1 \leq i \leq k$, as well as $n_1 + n_2 + \cdots + n_k = m_1 + m_2 + \cdots + m_k$.

Moreover, if we insert ab in all steps after the last a, we get

$$\begin{aligned} bab &\Longrightarrow baabb = ba^2b^2 \Longrightarrow ba^2abb^2 = ba^3b^3 \\ &\Longrightarrow ba^4b^4 \Longrightarrow \cdots \Longrightarrow ba^nb^n, \end{aligned}$$

which proves that $ba^nb^n \in L$ for all $n \geq 1$.

Assume that $L \in \mathcal{I}(PfC)$. Then there is a controlled insertion system

$$G' = (V, (R'_1, I'_1), (R'_2, I'_2), \dots, (R'_n, I'_n), A')$$

such that all control sets R'_j, $1 \leq j \leq n$, are regular prefix codes and $L(G') = L$.

Let

$$r \geq \max\{|w| \mid w \in A'\} + n \cdot \max\{|v| \mid v \in \bigcup_{j=1}^{n} I'_j\} + 1.$$

We consider the word $ba^rb^r \in L$. By the facts mentioned above, there is a derivation

$$\begin{aligned} ba^{t_0}b^{t_0} &\Longrightarrow_{(R'_{j_1}, I'_{j_1})} ba^{t_1}b^{t_1} \Longrightarrow_{(R'_{j_2}, I'_{j_2})} ba^{t_2}b^{t_2} \\ &\Longrightarrow \cdots \Longrightarrow_{(R'_{j_p}, I'_{j_p})} ba^{t_p}b^{t_p} = ba^rb^r, \end{aligned}$$

where $ba^{t_{i-1}} \in R'_{j_i}$ and $a^{t_i - t_{i-1}} b^{t_i - t_{i-1}} \in I'_{j_i}$ for $1 \leq i \leq p$. By the choice of r, we have $p > n$ and, therefore, there are numbers k and l with $1 \leq k < l \leq p$ such that $(R'_{j_k}, I'_{j_k}) = (R'_{j_l}, I'_{j_l})$. Thus, we get that not only $ba^{t_{k-1}} \in R_{j_k}$ but also $ba^{t_{l-1}} \in R_{j_k}$. But $t_{k-1} < t_{l-1}$ which contradicts the prefix-freeness of R_{j_k}. □

We conjecture that the converse does not hold.

Lemma 8. *Let $G = (V, (R, I), A)$ be a control insertion system, where R is in PfC. Then there are words $u_1, u_2, \ldots, u_n, v_1, v_2, \ldots v_n$ and a finite set A' such that*

$$L(G) = A' \cup \bigcup_{i=1}^{n} \{u_i\} I^* \{v_i\}.$$

Proof. Let w be a word of A. If w has no prefix in R, then no word can be obtained from w in G by insertions. Then we define A' as the set of all words of A which have no prefix in R.

Now assume that w has a prefix in R, that is, $w = uv$ with $u \in R$. We note that there is no other decomposition of w of this form. If $w = u'v'$ with $u' \in R$ also holds, then we have that u is a prefix of u' or u' is a prefix of u. Since R is a prefix code, the only possibility is $u = u'$. Therefore, by an insertion of a word z in w, we can only generate the word $w' = uzv$. As above, we obtain that u is the only prefix of w' in R. Thus, again, we can only insert after u and obtain $w'' = uz'zv$ with $z' \in I$. We can iterate this process and get that $\{u\} I^* \{v\}$ is the set of all words which can be generated from w.

Now the result follows immediately. □

From Lemma 8, we immediately obtain that all languages generated by controlled insertion systems with only one prefix control language are regular. Moreover, such languages belong to $\mathcal{I}(FIN)$ because $L(G') = L(G)$ holds for the controlled insertion system $G' = (V, (\{u_1, u_2, \ldots, u_n\}, I), A)$ which has a finite control set.

4 Results Concerning Ideals

We present some results which imply that $\mathcal{I}(X) \setminus \mathcal{I}(Y)$ is not empty. Such statements are the basis for incomparabilities of some families. We start with the situation that Y is the family of right ideals.

Lemma 9. *Let $L = \{ab\}^+$. Then we have $L \in \mathcal{I}(RC)$, $L \in \mathcal{I}(lId)$, and $L \notin \mathcal{I}(rId)$.*

Proof. The language L is generated by the controlled insertion system

$$G = (\{a,b\}, (\{ab, ba\}, \{ab\}), \{ab\}).$$

Starting with the axiom ab, only ab in the control set can be used to insert another word ab; thus, the system G exactly generates the language L. The control language of G is a reflective code. Hence, we have $L \in \mathcal{I}(RC)$.

Furthermore, we consider the controlled insertion system

$$H = (\{a,b\}, (\{a,b\}^*\{ab\}, \{ab\}), \{ab\}).$$

We can insert the word ab after any prefix ending with ab. Thus, starting with ab, we produce successively all words $(ab)^n$ with $n \geq 1$. Other words cannot be obtained. Hence, $L(H) = L$. Because $\{a,b\}^*\{ab\}$ is a left ideal, we get $L \in \mathcal{I}(lId)$.

Assume that $L \in \mathcal{I}(rId)$. Then there is a controlled insertion system

$$G' = (V, (R'_1, I'_1), (R'_2, I'_2), \ldots, (R'_n, I'_n), A')$$

such that all control sets R'_j, $1 \leq j \leq n$, are right ideals and $L(G') = L$. We consider $(ab)^p \in L$ with

$$p \geq m(G') + \max\{|v| \mid v \in I'_j, 1 \leq j \leq n\}.$$

Then there are a word w and an index j with $1 \leq j \leq n$ such that $w \Longrightarrow_{(R'_j,I'_j)} (ab)^p$. By the structure of the words in L, we have $w = (ab)^m$ for some $m \geq 1$. There are two cases for the derivation $(ab)^m \Longrightarrow_{(R'_j,I'_j)} (ab)^p$.

Case 1. $(ab)^{m'} \in R'_j$, $0 \le m' \le m$, and $(ab)^{p-m} \in I'_j$ with the derivation

$$(ab)^m = (ab)^{m'}(ab)^{m-m'} \Longrightarrow_{(R'_j,I'_j)} (ab)^{m'}(ab)^{p-m}(ab)^{m-m'} = (ab)^p.$$

Since $(ab)^{m'}a \in R'_j$ by the definition of a right ideal, we also have the derivation

$$(ab)^{m'}ab(ab)^{m-m'-1} \Longrightarrow_{(R'_j,I'_j)} (ab)^{m'}a(ab)^{p-m}b(ab)^{m-m'-1} \notin L.$$

Case 2. $(ab)^{m'}a \in R'_j$, $0 \le m' < m$, and $(ba)^{p-m} \in I'_j$ with the derivation

$$\begin{aligned}(ab)^m &= (ab)^{m'}ab(ab)^{m-m'-1}\\ &\Longrightarrow_{(R'_j,I'_j)} (ab)^{m'}a(ba)^{p-m}b(ab)^{m-m'-1} = (ab)^p.\end{aligned}$$

Then also $(ab)^{m'}ab \in R'_j$ since R'_j is a right ideal, which gives the derivation

$$(ab)^{m'}ab(ab)^{m-m'-1} \Longrightarrow_{(R'_j,I'_j)} (ab)^{m'}ab(ba)^{p-m}(ab)^{m-m'-1} \notin L.$$

Since in both cases we get a contradiction to $L = L(G')$, our assumption has to be false. Therefore, $L \notin \mathcal{I}(rId)$. □

The following lemmas show that insertion systems with a control by right and left ideals can generate languages which cannot be obtained by codes as control languages.

Lemma 10. *Let L be the language generated by the controlled insertion system*

$$G = (\{a,b\}, (\{a\}\{a,b\}^*, \{ab\}), \{ab\}).$$

Then we have $L \in \mathcal{I}(rId)$ and $L \notin \mathcal{I}(C)$.

Proof. The only control language of G is a right ideal. Hence, we have the relation $L \in \mathcal{I}(rId)$.

Assume that $L \in \mathcal{I}(C)$. Then there is a controlled insertion system

$$G' = (V, (R'_1, I'_1), (R'_2, I'_2), \ldots, (R'_n, I'_n), A')$$

such that all control sets R'_j, $1 \le j \le n$, are codes and $L(G') = L$.

Similarly to the proof of Lemma 7, one can deduce a contradiction also here. We note that $a^n b^n \in L$ for all $n \ge 1$.

Let

$$r \ge \max\{|w| \mid w \in A'\} + n \cdot \max\{|v| \mid v \in \bigcup_{j=1}^{n} I'_j\} + 1.$$

We consider the word $a^r b^r \in L$. There is a derivation

$$a^{t_0} b^{t_0} \Longrightarrow_{(R'_{j_1}, I'_{j_1})} a^{t_1} b^{t_1} \Longrightarrow_{(R'_{j_2}, I'_{j_2})} a^{t_2} b^{t_2} \Longrightarrow \cdots \Longrightarrow_{(R'_{j_p}, I'_{j_p})} a^{t_p} b^{t_p} = a^r b^r,$$

where $a^{t_{i-1}} \in R'_{j_i}$ and $a^{t_i - t_{i-1}} b^{t_i - t_{i-1}} \in I'_{j_i}$ for $1 \le i \le p$. By the choice of r, we have $p > n$ and, therefore, there are numbers k and l with $1 \le k < l \le p$ such that $(R'_{j_k}, I'_{j_k}) = (R'_{j_l}, I'_{j_l})$. Thus, we get that not only $a^{t_{k-1}} \in R_{j_k}$ but also $a^{t_{l-1}} \in R_{j_k}$. But $t_{k-1} \ne t_{l-1}$ which contradicts that R_{j_k} is a code (for instance, the word $a^{t_{k-1}+t_{l-1}}$ does not have a unique decomposition). □

Lemma 11. *Let L be the language generated by the controlled insertion system*

$$G = (\{a, b\}, (\{a, b\}^* \{a\}^+, \{ab\}), \{ab\}).$$

Then we have $L \in \mathcal{I}(lId)$ and $L \notin \mathcal{I}(C)$.

Proof. The only control language of G is a left ideal. Hence, we have the relation $L \in \mathcal{I}(lId)$.

Assume that $L \in \mathcal{I}(C)$. Then there is a controlled insertion system

$$G' = (V, (R'_1, I'_1), (R'_2, I'_2), \ldots, (R'_n, I'_n), A')$$

such that all control sets R'_j, $1 \le j \le n$, are codes and $L(G') = L$.

In the same way as in the proof of Lemma 10, one can deduce a contradiction also here. We note that $a^n b^n \in L$ for all $n \ge 1$. Then there is a control language R'_j with $1 \le j \le n$ which contains two different words a^x and a^y. This contradicts that R_j is a code (for instance, the word a^{x+y} does not have a unique decomposition). □

Finally, we consider languages which cannot be generated by insertion systems with left ideals as control languages.

Lemma 12. *Let L be the language generated by the controlled insertion system*

$$H = (\{a, b, c\}, (\{cb\}, \{a\}), (\{\ ba^n \mid n \geq 0\ \}, \{ab\}), \{cb, bab\}).$$

Then we have $L \in \mathcal{I}(SfC)$ and $L \notin \mathcal{I}(lId)$.

Proof. We have $L \in \mathcal{I}(SfC)$ since both control languages of H are suffix-free codes.

We note that we can generate $\{cb\}\{a\}^*$ using the first pair of control and insertion set and $L(G)$ from Lemma 7 using the second pair and there is no derivation where both pairs are applied. Thus, $L = \{cb\}\{a\}^* \cup L(G)$.

Now assume that $L \in \mathcal{I}(lId)$. Then there is a controlled insertion system

$$G' = (V, (R'_1, I'_1), (R'_2, I'_2), \ldots, (R'_n, I'_n), A')$$

such that all control sets R'_j, $1 \leq j \leq n$, are left ideals and $L(G') = L$. For the word ba^pb^p with

$$p \geq m(G') + \max\{|v| \mid v \in I'_j, 1 \leq j \leq n\},$$

there are a word w and a j, $1 \leq j \leq n$, such that $w \Longrightarrow_{(R'_j, I'_j)} ba^pb^p$. By the structure of words in L, we get $w = ba^mb^m$ for some $m \geq 1$, $ba^m \in R'_j$, and $a^{p-m}b^{p-m} \in I'_j$ and the derivation

$$ba^mb^m \Longrightarrow_{(R'_j, I'_j)} ba^ma^{p-m}b^{p-m}b^m = ba^pb^p.$$

Because $cba^m \in R'_j$ holds by the definition of left ideal and $cba^m \in L$, we also have the derivation

$$cba^m \Longrightarrow_{(R'_j, I'_j)} cba^ma^{p-m}b^{p-m} \notin L.$$

This contradiction to $L(G') = L$ proves that $L \notin \mathcal{I}(lId)$. □

Lemma 13. *Let L be the language generated by the controlled insertion system*

$$H' = (\{a,b,c\}, (\{cb\}\{a,b,c\}^*, \{a\}), (\{b\}\{a,b,c\}^*, \{ab\}), \{cb, bab\}).$$

Then we have $L \in \mathcal{I}(rId)$ and $L \notin \mathcal{I}(lId)$.

Proof. Since the two control languages of H' are right ideals, we get $L = L(H') \in \mathcal{I}(rId)$.

The proof for $L \notin \mathcal{I}(lId)$ can be given as in the preceding proof. All words of L which start with the letter c belong to the language $\{cb\}\{a\}^*$ and every word cba^m with $m \geq 0$ belongs to L. Further, all derivations from the axiom bab in the system H from Lemma 12 are also possible in the system H'. With left ideals as control sets, again some word would be derived which starts with the letter c and which contains two letters b. This is a contradiction. □

5 Conclusion

Summarizing the above lemmas, we get the following theorem.

Theorem 14.

1. *Figure 2 shows the inclusions for the classes $\mathcal{I}(F)$ with $F \in \mathcal{G}$.*
2. *The family $\mathcal{I}(rId)$ is incomparable to all the families $\mathcal{I}(X)$ with $X \in \mathcal{G} \setminus \{REG, rId\}$.* □

It remains to study whether the inclusions $\mathcal{I}(SfC) \subseteq \mathcal{I}(C)$, $\mathcal{I}(BfC) \subseteq \mathcal{I}(PfC)$, and $\mathcal{I}(UC) \subseteq \mathcal{I}(BfC)$ are proper. Moreover, one has to look whether $\mathcal{I}(lId)$ and $\mathcal{I}(X)$ with $X \in \{PfC, BfC, UC, RC\}$ and $\mathcal{I}(PfC)$ and $\mathcal{I}(SfC)$ are incomparable.

In order to solve some of these problems, it would be helpful to have a language in $\mathcal{I}(PfC) \setminus \mathcal{I}(SfC)$. However, we believe that this is not an easy task to find such a language, since it seems that the languages in $\mathcal{I}(PfC)$ have a very special form. Lemma 8 gives a hint that this conjecture holds for simple insertion systems controlled by prefix codes, but we have no general result on it.

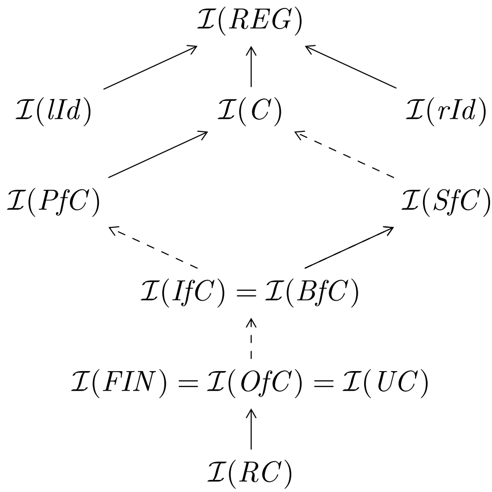

Figure 2. Hierarchy of languages families generated by controlled insertion systems (an arrow from X to Y denotes $X \subset Y$, a dashed arrow from X to Y denotes $X \subseteq Y$; if two families are not connected by a directed path, then they are not necessarily incomparable).

In [2], the special case of controlled insertion systems with only one pair of control and insertion languages and only singletons as insertion languages was considered in continuation of [1]. It remains to find the hierarchy of such families with control by codes and ideals and to compare it with the families of this paper.

References

[1] M. Latteux and Y. Roos, "On rationally controlled one-rule insertion systems," *RAIRO – Theoretical Informatics and Applications*, vol. 56, p. 8, 2022.

[2] J. Dassow and B. Truthe, "Subregularly controlled insertion systems," *Journal of Automata, Languages and Combinatorics*, 2024, accepted.

[3] G. Rozenberg and A. Salomaa, Eds., *Handbook of Formal Languages.* Springer-Verlag, Berlin, 1997.

[4] H. Jürgensen and S. Konstantinidis, "Codes," in *[3], Vol. 1*, 1997, pp. 511–607.

[5] H.-J. Shyr, *Free Monoids and Languages.* Hon Min Book Company, Taichung, Taiwan, 1991.

[6] J. Dassow, "Grammars with control by ideals and codes," *Journal of Automata, Languages and Combinatorics*, vol. 23, no. 1–3, pp. 143–164, 2018.

[7] S. Konstantinidis, "Every regular bifix code is a finite union of regular infix codes," *Journal of Automata, Languages and Combinatorics*, vol. 25, no. 1, pp. 55–59, 2020.

Jürgen Dassow, Bianca Truthe

Received October 9, 2024
Revised October 22, 2024
Accepted October 23, 2024

Jürgen Dassow
ORCID: https://orcid.org/0000-0002-4735-4696
Otto-von-Guericke-Universität Magdeburg
Fakultät für Informatik
PSF 4120
39016 Magdeburg
Germany
E–mail: dassow@iws.cs.uni-magdeburg.de

Bianca Truthe
ORCID: https://orcid.org/0000-0003-0031-5275
Justus-Liebig-Universität Gießen
Institut für Informatik
Arndtstr. 2
35392 Gießen
Germany
E–mail: bianca.truthe@informatik.uni-giessen.de