ELSEVIER

Theoretical Computer Science 183 (1997) 3–19

Theoretical Computer Science

Insertion and deletion closure of languages[1]

Masami Ito[a], Lila Kari[b,*], Gabriel Thierrin[b]

[a]Faculty of Science, Kyoto Sangyo University, Kyoto 603, Japan
[b]Department of Computer Science, University of Western Ontario, London, Ont., Canada N6A 5B7

Abstract

To a given language L, we associate the sets $ins(L)$ (resp. $del(L)$) consisting of words with the following property: their insertion into (deletion from) any word of L yields words which also belong to L. Properties of these sets and of languages which are insertion (deletion) closed are obtained. Of special interest is the case when the language is ins-closed (del-closed) and finitely generated. Then the minimal set of generators turns out to be a maximal prefix and suffix code, which is regular if L is regular. In addition, we study the insertion-base of a language and languages which have the property that both they and their complements are ins-closed.

1. Introduction

The insertion and deletion are word (language) operations that have been extensively studied, for example, in [5–8]. They are natural generalizations of the catenation, respectively left/right quotient: instead of adding (erasing) a word to the right (from the left/right) extremity of another, we insert (delete) it into (from) an arbitrary position. The result is usually a set of cardinality greater than two, which contains the catenation (left/right quotient) of the words as one of its elements.

A natural question which arises is to consider sets of words with the property that, when inserted (deleted) into (from) any word of a given language L, produce words which remain in L. These sets, denoted in the sequel by $ins(L)$ (resp. $del(L)$) are defined and investigated in Sections 2 and 3. In particular, a method of constructing them from the language L by using the dipolar deletion, is obtained. Moreover, a procedure of constructing the insertion (deletion) closure of a language is given. Results concerning similar concepts in relation with codes can be found in [4].

* Corresponding author.

[1] This research was supported by Grant-in-Aid for Science Research 04044150, Ministry of Education, Science and Culture of Japan and Grant OGP0007877 of the Natural Sciences and Engineering Research Council of Canada.

PII S0304-3975(96)00307-6

When a language equals its insertion (deletion) closure, it is called ins-closed (del-closed). Section 4 deals with ins-closed (del-closed) languages that are finitely generated. Namely, properties of such languages and of their minimal sets of generators are obtained. For example, if a regular language is ins-closed and del-closed, its minimal set of generators is a regular maximal bifix code.

If a language L is ins-closed, its words can either be obtained from other words of L by insertion, or can be "minimal" in this sense. The insertion base of L consists of all words which belong to the second category, that is, cannot be obtained from other words of L by insertion. In Section 5 it is shown that if an ins-closed language is regular, its ins-base is also regular. If, in addition, the language is del-closed, its ins-base is finite.

Finally, we consider the special case of languages L with the property that *any* word belongs to $ins(L)$. This amounts to the fact that the insertion of any word into a word of L is a subset of L. Such languages are called fully ins-closed, and their properties are investigated in Section 6.

In the sequel, for a set S, $card(S)$ is the cardinality of S and S^c the complement of S. X denotes a finite alphabet and X^* the free monoid generated by X under the catenation operation. 1 is the empty word and, for a word $w \in X^*$ and a letter $a \in X$, $|w|$ denotes the length of w and $|w|_a$ the number of occurrences of the letter a in w. For a language $L \subseteq X^*$, $alph(L)$ is the set $\{a \in X \mid \exists x, y \in X^*, xay \in L\}$.

For further undefined notions and notations in formal language theory and theory of codes the reader is referred to [9] (resp. [10]).

2. Insertion closure

Let $L \subseteq X^*$. To the language L one can associate the set $ins(L)$ consisting of all words with the following property: their insertion into *any* word of L yields a word belonging to L. Formally, $ins(L)$ is defined by:

$$ins(L) = \{x \in X^* \mid \forall u \in L, u = u_1u_2 \Rightarrow u_1xu_2 \in L\}.$$

Example. Let $X = \{a, b\}$. Then,
- $ins(X^*) = X^*$;
- $ins(L_{ab}) = L_{ab}$, where $L_{ab} = \{x \in X^* \mid |x|_a = |x|_b\}$;
- if $L = \{a^nb^n \mid n \geqslant 0\}$ then $ins(L) = \{1\}$;
- if $L_1 = (a^2)^*$, $L_2 = aL_1$ then $ins(L_1) = L_1$ and $ins(L_2) = L_1$;
- if $L = b^*ab^*$ then $ins(L) = b^* = ins^2(L)$;
- if $L = aX^*b$ then $ins(L) = L$.

A language L is called *dense* (*right dense*, *left dense*) if for any $w \in X^*$ there exist $x, y \in X^*$ (resp. $x \in X^*$) such that $xwy \in L$ (resp. $wx \in L$, $xw \in L$). If L is not dense (left dense, right dense), it is called *thin* (*left thin*, *right thin*). Note that if $ins(L) = X^*$ then the language L is dense, but the converse is not true.

A word $w \in X^+$ is *primitive* if $w = y^n$ for some $y \in X^+$ implies $n = 1$ and $w = y$. Let Q be the set of the primitive words over X. The language Q is dense, but $ins(Q) \neq X^*$ because $ab \in Q$ but $abab = (ab)^2 \notin Q$.

A language L is *commutative* if for any $w \in L$, L contains all the words obtained from w by arbitrarily permuting its letters.

Proposition 2.1. *$ins(L)$ is a submonoid of X^*. Moreover, if L is a commutative language, then $ins(L)$ is also a commutative language.*

Proof. Let $x, y \in ins(L)$ and $u = u_1u_2 \in L$. Then $u_1xu_2 \in L$, $u_1xyu_2 \in L$, hence $xy \in ins(L)$. Since $1 \in ins(L)$, $ins(L)$ is not empty.

For the second claim, it is sufficient to show that $xuvy \in ins(L)$ implies $xvuy \in ins(L)$. If $w \in L$, $w = w_1w_2$, then $w_1xuvyw_2 \in L$, hence $w_1xvuyw_2 \in L$. Therefore $xvuy \in ins(L)$. $\square$

In the following we give some properties and characterize $ins(L)$ for a given language L. We begin by noticing the connection between $ins(L)$ and the insertion operation, which has been studied in [5].

Let L_1, L_2 be two languages over X. The *sequential insertion* (in short *insertion*) of L_2 into L_1 is defined as

$$L_1 \longleftarrow L_2 = \{u_1vu_2 \mid u_1u_2 \in L_1, v \in L_2\}.$$

The insertion is a generalization of catenation: given $u, v \in X^*$, instead of adding v to the right extremity of u, the insertion places v in an arbitrary position in u. The result of the insertion of two words is thus in general a set of words with cardinality greater than 1.

The iterated insertion can then be defined as

$$L_1 \longleftarrow^* L_2 = \bigcup_{n=0}^{\infty} (L_1 \longleftarrow^n L_2),$$

where $L_1 \longleftarrow^0 L_2 = L_1$ and $L_1 \longleftarrow^{i+1} L_2 = (L_1 \longleftarrow^i L_2) \longleftarrow L_2$, for all $i \geqslant 0$.

Lemma 2.1. *Let $L \subseteq X^*$ and let $u, v \in ins(L)$. Then $(v \longleftarrow^* u) \subseteq ins(L)$.*

Proof. Let $w \in (v \longleftarrow^* u)$. There exists $k \geqslant 0$ such that $w \in (v \longleftarrow^k u)$.

We will show, by induction on k, that $w \in ins(L)$. If $k = 0$, then $w = v \in ins(L)$. Assume the assertion true for k and take $w \in (v \longleftarrow^{k+1} u)$ and $z = z_1z_2 \in L$. Then, $w = w_1uw_2$ where $w_1w_2 \in (v \longleftarrow^k u) \subseteq ins(L)$. Consequently, $z_1w_1w_2z_2 \in L$. This, together with the fact that $u \in ins(L)$ imply that $z_1w_1uw_2z_2 \in L$. As $z = z_1z_2$ was an arbitrary word in L, we deduce that $w \in ins(L)$. $\square$

Proposition 2.2. *Let $L \subseteq X^*$. Then $ins^2(L) = ins(ins(L)) = ins(L)$.*

Proof. Assume $u \in ins(ins(L))$. As $1 \in ins(L)$, we have $u = 1u \in ins(L)$, i.e. $ins(ins(L)) \subseteq ins(L)$. Assume now that $u \in ins(L)$. Let $v = v_1v_2 \in ins(L)$. Consider $v_1uv_2 \in X^*$.

Obviously, $v_1uv_2 \in (v \longleftarrow^* u)$. By Lemma 2.1, $v_1uv_2 \in ins(L)$, hence $u \in ins(ins(L))$, i.e. $ins(L) \subseteq ins(ins(L))$. □

For u, v words over X, the *dipolar deletion* $u \rightleftharpoons v$ is defined by (see [5]) $u \rightleftharpoons v = \{x \in X^* \mid u = v_1xv_2, v = v_1v_2\}$. In other words, the dipolar deletion erases from u a prefix and a suffix whose catenation equals v. The operation can be extended to languages in the natural fashion.

We are now ready to construct the set $ins(L)$ for a given language L.

Proposition 2.3. $ins(L) = (L^c \rightleftharpoons L)^c$.

Proof. Take $x \in ins(L)$. Assume, for the sake of contradiction, that $x \notin (L^c \rightleftharpoons L)^c$. Then, $x \in (L^c \rightleftharpoons L)$, that is, there exist $u_1xu_2 \in L^c$, $u_1u_2 \in L$ such that $x \in u_1xu_2 \rightleftharpoons u_1u_2$. We arrived at a contradiction, as $x \in ins(L)$ and $u_1u_2 \in L$ but the insertion of x into u_1u_2 belongs to L^c.

Consider now a word $x \in (L^c \rightleftharpoons L)^c$. If $x \notin ins(L)$, there exists $u_1u_2 \in L$ such that $u_1xu_2 \notin L$. This further implies $u_1xu_2 \in L^c$ and $x \in L^c \rightleftharpoons L$ – a contradiction with the original assumptions about x. □

Corollary 2.1. *If a language L is regular, then $ins(L)$ is regular and can be effectively constructed.*

Proof. It follows as the family of regular languages is closed under dipolar deletion (see [5]) and complementation. □

A language L such that $L \subseteq ins(L)$ is called *ins-closed*. A language L is ins-closed iff $u = u_1u_2 \in L$ and $v \in L$ imply $u_1vu_2 \in L$. As a consequence, note that every ins-closed language is a subsemigroup of X^*.

In general, submonoids of X^* are not ins-closed. For example, let $X = \{a, b, c\}$ and let $L = (a(bc)^*)^*$. Then L is a submonoid that is not ins-closed, because $a, abc \in L$, but $abac \notin L$.

If nonempty, the intersection of ins-closed languages is also an ins-closed language. Let L be a nonempty language and let I_L be the family of all the ins-closed languages containing L. This family is nonempty because $X^* \in I_L$. The intersection

$$I(L) = \bigcap_{L_i \in I_L} L_i$$

of the languages of the family I_L is clearly an ins-closed language containing L and it is called the *ins-closure* of L. The ins-closure of a language L is the smallest ins-closed language containing L.

Notice that a language L is ins-closed iff $L \longleftarrow L \subseteq L$. Indeed, if $x \in L$, $u_1u_2 \in L$ then, as $x \in L \subseteq ins(L)$ we have that $u_1xu_2 \in L$. For the other implication, take $x \in L$ and $u_1u_2 \in L$. As $L \longleftarrow L \subseteq L$ we have that $u_1xu_2 \in L$ which shows that $x \in ins(L)$.

Proposition 2.4. *The insertion closure of a language L is $I(L) = L \longleftarrow^* L$.*

Proof. "$I(L) \subseteq L \longleftarrow^* L$": Obvious, as $L \longleftarrow^* L$ is ins-closed and L is included in $L \longleftarrow^* L$.

"$L \longleftarrow^* L \subseteq I(L)$": We show by induction on k that $L \longleftarrow^k L \subseteq I(L)$. For $k=0$ the assertion holds, as $L \subseteq I(L)$.

Assume that $L \longleftarrow^k L \subseteq I(L)$ and consider a word $u \in L \longleftarrow^{k+1} L = (L \longleftarrow^k L) \longleftarrow L$. Then $u = u_1vu_2$ where $u_1u_2 \in L \longleftarrow^k L$ and $v \in L$. As both $L \longleftarrow^k L$ and L are included in $I(L)$ and $I(L)$ is ins-closed, we deduce that $u \in I(L)$.

The induction step, and therefore the requested equality are proved. □

Proposition 2.5. (i) *If L is a context-free or context-sensitive language, then $I(L)$ is a context-free or a context-sensitive language.*

(ii) *If L is a regular language, then $I(L)$ is not in general a regular language.*

Proof. (i) If L is context-free or context-sensitive, then so is also $L \cup \{1\}$. Since by [5], the families of context-free and context-sensitive languages are closed under iterated insertion, it follows that $I(L)$ is context-free or context-sensitive.

(ii) By [5], the family of regular languages is not closed under iterated insertion. For example, let $X = \{(,)\}$ and let $L = \{1, (\ )\}$. The iterated insertion of L into L is the Dyck language of order one. Therefore $I(L)$ is the Dyck language which is not a regular language. □

Note that if L is ins-closed then $L \longleftarrow^* L = L$. Indeed, as L is ins-closed, we have that $L = I(L)$. On the other hand, according to Proposition 2.4, $I(L) = L \longleftarrow^* L$.

3. Deletion closure

Let $L \subseteq X^*$ and let $Sub(L) = \{u \in X^* \mid xuy \in L\}$, that is $Sub(L)$ is the set of the subwords of the words in L. To the language L one can associate the set $del(L)$ consisting of all words x with the following property: x is subword of at least one word of L, and the deletion of x from any word of L containing x as subword yields words belonging to L. Formally, $del(L)$ is defined by

$$del(L) = \{x \in Sub(L) \mid \forall u \in L, u = u_1xu_2 \ \Rightarrow\ u_1u_2 \in L\}.$$

The condition that $x \in Sub(L)$ has been added because otherwise $del(L)$ would contain irrelevant elements: words which are not subwords of any word of L and thus yield $\emptyset$ as a result of the deletion from L.

Example. Let $X = \{a, b\}$. Then,
- $del(X^*) = X^*$;
- $del(L_{ab}) = L_{ab}$;

– if $L = \{a^n b^n \mid n \geqslant 0\}$ then $del(L) = L$;
– if $L = b^* ab^*$ then $del(L) = b^*$.

Proposition 3.1. *Let $L \subseteq X^*$.*
(i) *If $x, y \in del(L)$ and $xy \in Sub(L)$, then $xy \in del(L)$.*
(ii) *If $Sub(L)$ is a submonoid of X^*, then $del(L)$ is a submonoid of X^*.*
(iii) *If L is a commutative language, then $del(L)$ is also commutative.*

Proof. (i) Let $x, y \in del(L)$ with $xy \in Sub(L)$. If $u = u_1 xyu_2 \in L$, then $u_1 yu_2 \in L$ and consequently $u_1 u_2 \in L$. Therefore $xy \in del(L)$.
(ii) Immediate.
(iii) It is sufficient to show that $xuvy \in del(L)$ implies $xvuy \in del(L)$. Since L is commutative, $u_1 xuvyu_2 \in L$ if and only if $u_1 xvuyu_2 \in L$. As $xuvy \in del(L)$ we have that $u_1 u_2 \in L$. This further implies that $xvuy \in del(L)$. $\square$

In the following we show how, for a given language L, the set $del(L)$ can be constructed. The construction is similar to the one for $ins(L)$ and involves the same operation, the dipolar deletion.

Proposition 3.2. $del(L) = (L \rightleftharpoons L^c)^c \cap Sub(L)$.

Proof. Let $x \in del(L)$. From the definition of $del(L)$ it follows that $x \in Sub(L)$. Assume that $x \notin (L \rightleftharpoons L^c)^c$. This means there exist $u_1 xu_2 \in L$ and $u_1 u_2 \in L^c$ such that $x \in u_1 xu_2 \rightleftharpoons u_1 u_2$. We arrived at a contradiction as $x \in del(L)$ but $u_1 xu_2 \in L$ and $u_1 u_2 \notin L$.

For the other inclusion, let $x \in (L \rightleftharpoons L^c)^c \cap Sub(L)$. As $x \in Sub(L)$, if $x \notin del(L)$ there exist $u_1 xu_2 \in L$ such that $u_1 u_2 \notin L$. This further implies that $u_1 u_2 \in L^c$, that is, $x \in L \rightleftharpoons L^c$ – a contradiction with the initial assumption about x. $\square$

A language L is called *del-closed* if $v \in L$ and $u_1 vu_2 \in L$ imply $u_1 u_2 \in L$.

For example, X^* and L_{ab} are del-closed languages that are also ins-closed. Furthermore, they are both submonoids of X^*.

The notion of a del-closed language is strongly connected with the operation of deletion, defined in [5]. Related issues have recently been investigated in [7, 8].

Let L_1, L_2 be two languages over the alphabet X. The *sequential deletion* (in short *deletion*) of L_2 from L_1 is defined as

$$L_1 \longrightarrow L_2 = \{u_1 u_2 \in X^* \mid u_1 wu_2 \in L_1, w \in L_2\}.$$

The deletion generalizes the left/right quotient of words and languages. Given words $u, v \in X^*$, instead of erasing v from the left/right extremity of u, the deletion erases it from any place in u. If v does not occur as subword of u, the result of the deletion is the empty set. The result of deletion can also be a set of cardinality greater than 1.

Notice that a language $L \subseteq X^*$ is del-closed iff $L \longrightarrow L \subseteq L$.

Proposition 3.3. *Let $L \subseteq X^*$ be an ins-closed language. Then L is del-closed if and only if $L = (L \longrightarrow L)$.*

Proof. Since L is del-closed, $L \longrightarrow L \subseteq L$. Now let $u \in L$. Since L is ins-closed, $uu \in L$. Therefore $u \in (L \longrightarrow L)$, i.e. $L \subseteq (L \longrightarrow L)$. We can conclude that $L = (L \longrightarrow L)$. The other implication is obvious. $\square$

If L is a nonempty language and if D_L is the family of all the del-closed languages L_i containing L, then the intersection

$$\bigcap_{L_i \in D_L} L_i$$

of all the del-closed languages containing L is a del-closed language called the *del-closure* of L. The del-closure of L is the smallest del-closed language containing L.

We will now define a sequence of languages whose union is the del-closure of a given language L. Let

$$D_0(L) = L,$$

$$D_1(L) = D_0(L) \longrightarrow (D_0(L) \cup \{1\}),$$

$$D_2(L) = D_1(L) \longrightarrow (D_1(L) \cup \{1\}),$$

$$\ldots$$

$$D_{k+1}(L) = D_k(L) \longrightarrow (D_k(L) \cup \{1\}).$$

$$\ldots$$

Clearly $D_k(L) \subseteq D_{k+1}(L)$. Let

$$D(L) = \bigcup_{k \geqslant 0} D_k(L).$$

Proposition 3.4. *$D(L)$ is the del-closure of the language L.*

Proof. Clearly $L \subseteq D(L)$. Let now $v \in D(L)$ and $u_1vu_2 \in D(L)$. Then $v \in D_i(L)$ and $u_1vu_2 \in D_j(L)$ for some integers $i, j \geqslant 0$. If $k = \max\{i, j\}$, then $v \in D_k(L)$ and $u_1vu_2 \in D_k(L)$. This implies $u_1u_2 \in D_{k+1}(L) \subseteq D(L)$. Therefore, $D(L)$ is a del-closed language containing L.

Let T be a del-closed language such that $L = D_0(L) \subseteq T$. Since T is del-closed, if $D_k(L) \subseteq T$ then $D_{k+1}(L) \subseteq T$. By an induction argument, it follows that $D(L) \subseteq T$. $\square$

Since, by [5], the family of regular languages is closed under deletion, it follows that if L is regular, then the languages $D_k(L)$, $k \geqslant 0$, are also regular. However, it is an open question whether $D(L)$ is regular for any regular language $L \subseteq X^*$.

Recall that, for a language L, the *principal congruence* P_L is defined by:

$$u \equiv v(P_L) \text{ iff } \forall x, y \in X^* \text{ we have } xuy \in L \Longleftrightarrow xvy \in L.$$

When the principal congruence of L has a finite index (finite number of classes) the language L is regular.

If L is commutative, we have the following result.

Proposition 3.5. *Let $L \subseteq X^*$ be a regular language. If L is commutative, then $D(L)$ is commutative and regular.*

Proof. Let us prove first that $D(L)$ is commutative. To this end, it is sufficient to show that $D_{k+1}(L)$ is commutative if $D_k(L)$ is commutative. Let $xuvy \in D_{k+1}(L)$. If $xuvy \in D_k(L)$, then we are done. Otherwise, by the definition of $D_{k+1}(L)$, there exist $w, z \in D_k(L)$ such that $w \in (xuvy \longleftarrow z)$. Since $D_k(L)$ is commutative, $xuvyz \in D_k(L)$ and $xvuyz \in D_k(L)$. From the fact that $z, xvuyz \in D_k(L)$ and the definition of $D_{k+1}(L)$, it follows that $xvuy \in D_{k+1}(L)$, i.e. $D_{k+1}(L)$ is commutative.

We will show next that $D(L)$ is regular. To this aim, we show that if $u \equiv v(P_{D_k(L)})$ then $u \equiv v(P_{D_{k+1}(L)})$. Let $u \equiv v(P_{D_k(L)})$ and let $xuy \in D_{k+1}(L)$. By the definition of $D_{k+1}(L)$, there exist $w, z \in D_k(L)$ such that $w \in (xuy \longleftarrow z)$. Since $D_k(L)$ is commutative, $xuyz \in D_k(L)$. Hence $xvyz \in D_k(L)$. From the fact that $z \in D_k(L)$ and by the definition of $D_{k+1}(L)$, it follows that $xvy \in D_{k+1}(L)$. In the same way, $xvy \in D_{k+1}(L)$ implies $xuy \in D_{k+1}(L)$. Consequently, $u \equiv v(P_{D_{k+1}(L)})$ holds. This means that the number of congruence classes of $P_{D_{k+1}(L)}$ is smaller or equal to that of $P_{D_k(L)}$. Remark that

$$D_0(L) \subseteq D_1(L) \subseteq \cdots \subseteq D_n(L) \subseteq D_{n+1}(L) \ldots$$

It can be shown (see [3]) that $D_t(L) = D_{t+1}(L)$ for some t, $t \geqslant 1$. Thus, $D(L) = D_t(L)$ which implies that $D(L)$ is regular. $\square$

4. Generators of insertion-closed and deletion-closed languages

This section is focused on ins-closed and del-closed languages that are finitely generated. Namely, properties of such languages and of their minimal sets of generators are obtained. One of the main results of the section states that, if L is regular, ins-closed and del-closed, then its minimal set of generators is a regular maximal bifix code, where the notion of bifix code is defined in the following.

A nonempty language $L \subseteq X^+$ is called a *prefix* (*suffix*) *code* if $x, xy \in L$ ($x, yx \in L$) implies $y = 1$. It is called a *bifix code* if it is both a prefix and a suffix code. L is called an *infix code* if $u \in L$, $xuy \in L$ imply $x = y = 1$. L is an *outfix code* if $xy \in L$, $xuy \in L$ imply $u = 1$.

Lemma 4.1. *Let $L \subseteq X^*$ be an ins-closed language that is finitely generated. Then for any $a \in alph(L)$, there exists a positive integer i_a such that $a^{i_a} \in L$.*

Proof. We begin by showing that for any $a \in alph(L)$ there exists $w \in X^*$ such that $aw \in L$ or $wa \in L$. Suppose this is not the case. Then there exists $a \in alph(L)$ such that $uav \in L$ for some $u, v \in X^*$ with $|u|, |v| \geqslant 1$. Let $n = \min\{|x|, |y| \mid xay \in L, x, y \in X^+\}$. Now let $uav \in L$ with $\min\{|u|, |v|\} = n$. Moreover, let $m = \max\{|z| \mid z \in K\}$, where K is the minimal set of generators of L.

Case 1: $|u| \leqslant |v|$. Consider $(ua)^m v^m \in L$. Since $m|ua| > m$, $u'av' \in K$ or $v'au'' \in K$, where $v' \in alph(L)^+$ and u' is a suffix of u or u'' is a prefix of u with $|u'|, |u''| < |u|$. From the assumption that $(aX^* \cup X^*a) \cap K = \emptyset$, it follows that $|u'|, |u''| \geqslant 1$ and $u', u'' \neq u$. However, this contradicts the minimality of $|u|$.

Case 2: $|v| < |u|$. Considering $u^m(av)^m \in L$, we can prove in a similar way as above that we reach a contradiction.

As both cases lead to contradictions, our assumption was false and $aw \in L$ or $wa \in L$ for some $w \in X^*$.

Now consider $a^m w^m \in L$ or $w^m a^m \in L$. In this case, it is easy to verify that $a^{i_a} \in L$ for some positive integer i_a, by taking $m = \max\{|z| \mid z \in K\}$. □

Proposition 4.1. *Let $L \subseteq X^*$ be a finitely generated ins-closed language and K be its minimal set of generators. Then*:

(i) *K contains a finite maximal prefix (suffix) code over $alph(L)$*;

(ii) *If K is a code over $alph(L)$ then $K = alph(L)^n$ for some $n \geqslant 1$*;

(iii) *If L is del-closed then $K = alph(L)^n$ for some $n \geqslant 1$.*

Proof. (i) Let $P = \{u \in L \mid u \neq 1,\ u = vx,\ v \in L \setminus \{1\},\ x \in X^* \Longrightarrow x = 1\}$. Then obviously P is a prefix code over $alph(L)$ and $P \subseteq K$. Let $x \in alph(L)^+$ and let $x = a_1 a_2 \ldots a_n$ where $a_i \in alph(L)$, $1 \leqslant i \leqslant n$. By Lemma 4.1, for any a_i, $1 \leqslant i \leqslant n$, there exists a positive integer t_i such that $a_i^{t_i} \in L$. Therefore, $a_1 a_1^{t_1 - 1} \in L$. Then $a_1 a_2 a_2^{t_2 - 1} a_1^{t_1 - 1} \in L$.

Assume $a_1 a_2 \ldots a_i w \in L$. Then $a_1 a_2 \ldots a_i a_{i+1} a_{i+1}^{t_{i+1} - 1} w \in L$. By induction, there exists $y \in alph(L)^*$ such that $xy \in L$. Hence $xy \in P alph(L)^*$. This means that P is a maximal prefix code.

The proof that L contains a finite maximal suffix code can be carried out symmetrically.

(ii) By (i), the code K contains a finite maximal prefix code P. Since P is finite, P is thin. This implies $K = P$, because every maximal prefix code that is thin is also a maximal code.

Claim. For any $a \in alph(L)$, there exists $v \in K$, $|v| = \max\{|y| \mid y \in K\}$ such that $v \in X^* a$.

Indeed, let $v = v'b$ be a word of maximal length in K. Consider the word $v'a$. Recall that for two words $x, y \in X^*$, $x \leqslant_p y$ iff x is a prefix of y. As K is a maximal prefix code, there exists $w \in K$ such that $v'a \leqslant_p w$ or $w \leqslant_p v'a$. (Otherwise $v'a$ can be added to K – a contradiction with the fact that K is maximal.)

If $v'a \leqslant_p w$, as $|v'a| = |v'b|$ it follows that $w = v'a$, which implies $v'a \in K$. We have found therefore a word in K which ends in a.

Assume now that $w \leqslant_p v'a$. If $w \neq v'a$ then $w \leqslant_p v' <_p v'b$ – a contradiction with the fact that K is a prefix code.

The proof of the Claim is thus complete.

Let us return to the proof of the proposition. Let $u \in K$ such that $|u| = \min\{|x| \mid x \in K\}$ and let $a \in alph(L)$. Assume that $u = bu'$ for some $b \in alph(L)$. According to the Claim, there exists $v \in K$ of maximal length such that $v = v'a$, $v' \in X^*$. Consider the word $v'ua = v'bu'a \in L$. From the facts that K is a prefix code, $|v| = \max\{|y| \mid y \in K\}$ and $|u| = \min\{|x| \mid x \in K\}$ we deduce that $u'a \in K$. Thus, $u'alph(L) \subseteq K$.

If $|u'| \neq 0$, we continue the same procedure. Namely, take $u'a$ in the role of u. Let $a' \in alph(L)$. According to the Claim, there exists $v \in K$, of maximal length, such that $v = v'a'$. As L is ins-closed, the word $v'u'aa' \in L$. If $u' \neq 1$ then $u' = cu''$ and $v'cu''aa' \in L$. From the fact that K is a prefix code and that $v'c$ is of maximal length, $u''aa'$ of minimal length, we deduce that $v'c, u''aa' \in K$. This means $u''alph(L)^2 \subseteq L$. Continuing this procedure, we can get $alph(L)^{|u|} \subseteq K$. Since K is a code, $K = alph(L)^{|u|}$.

(iii) Assume $u, ux \in K$ for $x \in alph(L)^+$. Since L is del-closed, $x \in L$. If $x \neq 1$, then $ux \in K^2K^*$, a contradiction. Therefore, $x = 1$. This means that K is a finite prefix code, i.e. a code. By (ii), $L = (alph(L)^n)^*$ for some n, $n \geqslant 1$. Notice that $1 \in L$. □

Finitely generated ins-closed languages are not always of the form $(alph(L)^n)^*$. For instance, if $X = \{a,b\}$ and $L = (\{a\} \cup X^2 \cup X^3)^+ = \{a, ab, ba, bb, bab, bbb\}^+$, then L is finitely generated and ins-closed but does not equal $\{a,b\}^*$.

A language L is called *strongly ins-closed* if both L and its complement L^c are ins-closed.

Example. Let $X = \{a_1, a_2, \ldots, a_n\}$, let $Y \subset X$ be a nonempty subalphabet of X and let $Z = X \backslash Y$.

(1) If $L = Y^*$, then $L^c = X^* \backslash Y^*$. Both L and L^c are ins-closed, hence both are strongly ins-closed languages and L^c is an ideal of X^*. Recall that a language L is a *right* (*left*) *ideal* of X^* if $u \in L$ and $x \in X^*$ implies $ux \in L$ ($xu \in L$). L is an *ideal* of X^* if it is both left and right ideal.

(2) If $L = YX^*$, then $L^c = ZX^* \cup \{1\}$. Both L and L^c are ins-closed, hence both are strongly ins-closed languages. Remark that L and $L^c \backslash \{1\}$ are both right ideals.

If L is strongly ins-closed, then $uv \in L$ implies either $u \in L$ or $v \in L$. If $u \in L$ and if $ux \in L^c$, then $x \in L^c$.

Proposition 4.2. *Assume* $1 \notin L$. *Let* $L \subseteq X^*$ *be a strongly ins-closed language such that* $L \neq \emptyset, X^+$. *Then the following assertions are equivalent*:

(1) *L is an ideal of* X^*,

(2) *L^c is finitely generated*,

(3) *There exist* $Y, Z \subseteq X$ *such that* $Y, Z \neq \emptyset$, $Y \cap Z = \emptyset$, $X = Y \cup Z$, $L^c = Z^*$ *and* $L = X^* \backslash Z^*$.

Proof. (1) $\Rightarrow$ (3): Let $Y = L \cap X$ and $Z = X \backslash Y$. Since $L \cup L^c = X^*$ and L, L^c are nontrivial subsemigroups, Y and Z are nonempty. From $Y \subseteq L$ and L a subsemigroup follows $Y^+ \subseteq L$. Similarly, it can be shown that $Z^+ \subseteq L^c$. Note that $1 \in L^c$. Indeed, as L is an ideal, $1 \in L$ would imply $L = X^*$ – a contradiction. Consequently, we have that $Z^* \subseteq L^c$.

Let now $u \in L^c$. If $u \notin Z^*$, then $u = xay$ with $a \in Y \subseteq L$. Since L is an ideal, $u \in L$, a contradiction. Hence $u \in Z^*$, which means $L^c \subseteq Z^*$.

The proofs of (3) $\Rightarrow$ (2) and (3) $\Rightarrow$ (1) are obvious.

We show (2) $\Rightarrow$ (3). Let $Y = X \cap L$ and let $Z = X \cap L^c$. We have that $alph(L^c) \cap Y = \emptyset$. Indeed, suppose $c \in alph(L^c) \cap Y$. Since L^c is finitely generated, by Lemma 4.1, $c^n \in L^c$ for some n, $n \geqslant 1$. On the other hand, since $c \in Y \subseteq L$ and L is ins-closed, $c^n \in L$, a contradiction.

Therefore, $alph(L^c) = Z$, $Y \cap Z = \emptyset$, $L^c = Z^*$ and $L = X^* \backslash Z^*$. □

Let $L \subseteq X^*$. Then $ins(L) = M$ is ins-closed. On the other hand, let M be an ins-closed monoid. Then $ins(M) = M$.

Proposition 4.3. *Let $L \subseteq X^*$ be a regular language. If L is ins-closed and del-closed, then the minimal set of generators K of L is a regular maximal bifix code over $alph(L)$. In fact, K is a maximal prefix code and a maximal suffix code over $alph(L)$.*

Proof. Since $K = (L \backslash \{1\}) \backslash (L \backslash \{1\})^2$, K is regular. Moreover, since L is del-closed, K is a bifix code. By the same procedure as in the proof of Proposition 4.1(i), it can be shown that K is a maximal prefix (suffix) code over $alph(L)$.

Indeed, let $a \in alph(L)$. Then $uav \in L$ for some $u, v \in alph(L)^*$. Therefore, $u^n(av)^n \in L$ for any n, $n \geqslant 1$. Let $A = (X, S, s_0, F, P)$ be a finite deterministic automaton accepting L and let $m \geqslant card(S)$. Then there exist s, t, $1 \leqslant s \leqslant t \leqslant card(S) + 1$ such that $s_0 u^s \Longrightarrow s'$ and $s_0 u^t \Longrightarrow s'$. Let $i = t - s$. Then $s_0 u^m \Longrightarrow s''$ and $s_0 u^{m+i} \Longrightarrow s''$. Since $u^m(av)^m \in L$, $u^{m+i}(av)^m \in L$. Hence $u^i \in L$, because L is del-closed. On the other hand, $u^i(av)^i \in L$. Therefore, $(av)^i \in L$, i.e. $aw \in L$ for some $w \in alph(L)^*$. By the same automaton argument, there exists a positive integer i_a such that $a^{i_a} \in L$. (Since $aw \in L$, we have $a^n w^n \in L$ for any n, $n \geqslant 1$. Notice that $s_0 a^i \Longrightarrow s'$, $s_0 a^j \Longrightarrow s'$ for some $s' \in S$ and i, j, $1 \leqslant i < j \leqslant card(S) + 1$. Let $i_a = j - i$. Then $s_0 a^{n+i_a} w^n \Longrightarrow s''$, $s_0 a^n w^n \Longrightarrow s''$ for some $s'' \in S$, if $n > card(S)$. Hence $a^{n+i_a} w^n$, $a^n w^n \in L$. Since L is del-closed, $a^{i_a} \in L$.)

Let now $x \in X^+$. Then $x = a_1 a_2 \dots a_r$, where $a_i \in X$, $1 \leqslant i \leqslant r$. Let $H = \{a_1^{i_1}, a_2^{i_2}, \dots, a_r^{i_r}\} \subseteq L$. Obviously, $xy = a_1 a_2 \dots a_r y \in (H \longleftarrow^* H) \subseteq L$ for some $y \in X^*$. This means that K is a maximal prefix code. □

In the above proof, the condition of regularity is necessary.

Example 1. Let $X = \{a, b\}$ and let $L = (ab \longleftarrow^* ab) \cup \{1\}$. Then L is ins-closed and del-closed. Moreover, K is a bifix code. But K is not a maximal bifix code over $\{a, b\}$ since $\{ba\} \cup K$ is a bifix code where K is the minimal set of generators of L.

The converse of Proposition 4.3 does not hold.

Example 2. Let $X=\{a,b\}$ and let $C=\{a^2\}\cup\{b^2\}\cup ab^+a\cup ba^+b$. Then C is a regular maximal bifix code over $\{a,b\}$. But C^* is not ins-closed since $ab(aba)a=ababaa\notin L$ though $aba\in L$.

A language L is *reflective* if for all $x,y\in X^*$ we have $xy\in L\Longleftrightarrow yx\in L$.

Proposition 4.4. *Let $L\subseteq X^*$ be a reflective ins-closed language. If L is del-closed, then the minimal set of generators K of L is a maximal bifix code over $alph(L)$.*

Proof. Let $a\in alph(L)$. Then there exist $u,v\in alph(L)^*$ such that $uav\in L$. Since L is reflective, $avu\in L$ (resp. $vua\in L$). Let $x\in alph(L)^+$. Then $x=a_1a_2\ldots a_r$, $a_i\in alph(L)$, $1\leqslant i\leqslant r$. Let $a_i(v_iu_i)\in L$ (resp. $(v_iu_i)a_i\in L$). As L is ins-closed, by inserting into $a_1(v_1u_1)$ the words $a_2(v_2u_2)$, $\ldots a_r(v_ru_r)$ (resp. by inserting into $(v_ru_r)a_r$ the words $(v_{r-1}u_{r-1})a_{r-1}$, $\ldots(v_1u_1)a_1$) we obtain that the word $a_1a_2\ldots a_r(v_ru_r)(v_{r-1}u_{r-1})\ldots(v_2u_2)(v_1u_1)$ belongs to L (resp. $(v_ru_r)\ldots(v_2u_2)(v_1u_1)a_1a_2\ldots a_r\in L$). Consequently, $xw\in L$ (resp. $wx\in L$) for some $w\in alph(L)^*$. This means L is left (resp. right) dense in $alph(L)$ which implies $K=(L\backslash\{1\})\backslash(L\backslash\{1\})^2$ is a maximal prefix (resp. suffix) code. □

Example 3. Let $X=\{a,b\}$ and let $L=\{u\in X^*\mid |u|_a=|u|_b\}$. Then L is a reflective language that is ins-closed and del-closed. $K=\{v\in X^*\mid v=a_1a_2\ldots a_r,\ a_i\in X,\ 1\leqslant i\leqslant r,$ $|v|_a=|v|_b$ and $|a_1a_2\ldots a_t|_a\neq|a_1a_2\ldots a_t|_b$ for any t, $1\leqslant t<r\}$ is a maximal bifix code. Notice that $aabb\in K$, but $baab\notin K$. This means that K is not reflective.

Example 4. Let $X=\{a,b\}$ and let $L=(a^*ba^*ba^*)^*\cup a^*$. Then L is a regular language that is ins-closed and del-closed. The minimal set of generators K of L, i.e. $K=(L\backslash\{1\})\backslash(L\backslash\{1\})^2$ is the set $\{a\}\cup ba^*b$. Moreover, $L=(\{a,bb\}\longleftarrow^*\{a,bb\})\cup\{1\}$.

Proposition 4.5. *If a language L is a del-closed submonoid of X^*, then L is generated by a bifix code P, i.e. $L=P^*$.*

Proof. Let $v,vx\in L$. Since L is del-closed, $x\in L$, i.e. L is a right unitary submonoid. Similarly, it can be shown that L is also a left unitary submonoid. By a well known result from the theory of codes (see, for example, [1, 10]), it follows that L is generated by a bifix code. □

5. Properties of insertion-closed and deletion-closed languages

Let $L\subseteq X^*$ be an ins-closed language. As the result of the insertion of two words in L always belongs to L, we can divide the words of L into two categories: words that can be obtained as the result of insertions of other words of L, and words that cannot be obtained in this fashion.

Consider the set

$$J=\{u\in L\mid u\neq 1, u\notin ((L\backslash\{1\})\longleftarrow (L\backslash\{1\}))\}=L\backslash((L\backslash\{1\})\longleftarrow^{+} (L\backslash\{1\})),$$

i.e., J consists of the words of L that are not the result of insertions of any words of L. Then J is uniquely determined and $L\backslash\{1\}=(J\longleftarrow^{*} J)$. J is called the *ins-base* of L.

The following result shows that if L is regular, its ins-base is also regular. The proof is based on the fact that one can construct a generalized sequential machine (for the definition see, for example, [9]) g such that $g(L)$ is the set of words in L that can be obtained as a result of insertions.

Proposition 5.1. *If L is a regular ins-closed language, then its ins-base J is a regular language.*

Proof. Let L be a regular ins-closed language. We can assume, without loss of generality, that L is 1-free. Let $A=(X,S,s_0,F,P)$ be a finite deterministic automaton accepting L, where $S=\{s_0,s_1,\dots,s_n\}$ and the rules of P are of the form $s_ia\longrightarrow s_j$, $s_i,s_j\in S$, $a\in X$.

We will show that there exists a generalized sequential machine g such that $g(L)=L\backslash J$. As the family of regular languages is closed under gsm mappings and set difference, it will follow that J is regular.

Notice first that, as L is ins-closed, $L\backslash J=\{u\in L\mid u=v_1wv_2, v_1v_2\in L, w\in L\}$.

Consider now the gsm $g=(X,X,S',s_0,F',P')$ where

$$\begin{aligned}
S' &= S\cup\{s_j^{(i)}\mid 0\leqslant j\leqslant n, 0\leqslant i\leqslant n\}\cup\{s_i'\mid s_i\in F\}\\
F' &= \{s_i'\mid s_i\in F\}\\
P' &= \{s_ia\longrightarrow as_k\mid s_ia\longrightarrow s_k\in P\} && (1)\\
&\quad \cup\{s_ia\longrightarrow as_j^{(i)}\mid s_0a\longrightarrow s_j\in P\} && (2)\\
&\quad \cup\{s_j^{(i)}a\longrightarrow as_k^{(i)}\mid s_ja\longrightarrow s_k\in P, 0\leqslant i\leqslant n\} && (3)\\
&\quad \cup\{s_j^{(i)}a\longrightarrow as_i'\mid s_ja\longrightarrow s_l\in P, s_l\in F\} && (4)\\
&\quad \cup\{s_i'a\longrightarrow as_k'\mid s_ia\longrightarrow s_k\in P\} && (5)
\end{aligned}$$

The idea of the proof is the following. We have constructed $card(S)$ indexed copies of the automaton A, $A^{(i)}=(X,S^{(i)},s_0^{(i)},F^{(i)},P^{(i)})$, $1\leqslant i\leqslant n$. Given a word $v_1wv_2\in L$, the gsm g works as follows.

The rules (1) scan the word v_1, using the corresponding productions of P. Suppose that after scanning v_1, the automaton is in state s_i. Rules (2) switch the derivation to the automaton $A^{(i)}$, starting thus to scan the word w. The word w is parsed by using rules (3) of the automaton $A^{(i)}$. If a final state is reached, that is if $w\in L$, rules (4) switch the derivation back to A. The fact that the index of the automaton was (i) allows us to remember the state s_i where we left the scanning of v_1v_2. Now, rules (5) continue the scanning of v_2. If a final state is reached, this means $v_1v_2 \in L$. (In this second part of the derivation for v_1v_2 primed versions of the states are used, in order to make sure that at least one word w has been encountered in the meantime.)

From the above explanations it follows that g reaches a final state iff the input word u is of the form v_1wv_2, $v_1v_2 \in L$, $w \in L$. Consequently, $g(L) = \{v_1wv_2 \mid v_1v_2 \in L, w \in L\}$. □

As we have seen in Proposition 5.1, if L is regular and ins-closed, its ins-base is regular. The ins-base of L can be an infinite language. For example, $L = ba^*b \longleftarrow^* ba^*b = \{bxb \mid x \in X^*, |x|_b \text{ is even}\}$ is regular and ins-closed but J contains the infinite set ba^*b. If we put the additional constraint that L is del-closed, the ins-base will always be finite, as shown by the following proposition. First, we need the following lemma.

Lemma 5.1. *Let $L \subseteq X^*$ be a regular language that is ins-closed and del-closed. Then there exists a positive integer n such that for every $u \in X^+$ we have that $u^n \in L$.*

Proof. Since the minimal set of generators of L is a maximal prefix code, for any $u \in alph(L)^+$ there exists $y \in X^*$ such that $uy \in L$. Let $p \geqslant 1$ be an integer. Then $u^py^p \in L$. In the same way as in the proof of Proposition 4.3, there exists a positive integer r satisfying the following condition: $\forall p, p \geqslant r, \exists t, 1 \leqslant t \leqslant r$ such that $u^{p+t}y^p \in L$. Since $u^py^p \in L$ and L is del-closed, $u^t \in L$. If we now let $n = r!$ then $u^n \in L$. □

Proposition 5.2. *Let $L \subseteq X^*$ be a regular language that is ins-closed and del-closed. Let K be the minimal set of generators of L and J be the ins-base of L. Then,*

(i) *J is finite.*

(ii) *If $J = K$ then $K = alph(L)^n$ for some $n \geqslant 1$.*

Proof. (i) Suppose J is infinite. Then, by a pumping lemma for a regular language, there exist $u, v, w \in alph(L)^*$ such that $v \in alph(L)^+$, $uvw \in J$ and $uv^*w \in L$. Hence $uw \in L$. Notice that, by Lemma 5.1, there exists n, $n \geqslant 1$ such that $z^n \in L$ for any $z \in alph(L)^*$. Since L is ins-closed, $uv(u^{n-1}w^{n-1})w = uvu^{n-1}w^n \in L$. From the assumption that L is del-closed, it follows that $uvu^{n-1} \in L$. On the other hand, since $u^n \in L$, $u^{n-1}(uvu^{n-1})u = u^nvu^n \in L$. By the assumption that L is del-closed, $v \in L$. However, this contradicts the assumption that $uvw \in J$ because $uw \in L$ and $v \in L$. Therefore, J must be finite.

(ii) By (i), J is finite, hence K is finite. Therefore, by Proposition 4.1(iii), $L = (alph(L)^n)^*$ for some n, $n \geqslant 1$. □

Proposition 5.3. *Let $L \subseteq X^*$ be a reflective language that is ins-closed and del-closed. Then the following assertions are equivalent:*

(1) *K is reflective;*

(2) $K=J$;
(3) $L=(alph(L)^n)^*$ *for some* $n, n\geqslant 1$.

Proof. (3) $\Rightarrow$ (1): Obvious.

(1) $\Rightarrow$ (2): Notice that $J\subseteq K$. Let $u\in K\backslash J$. Then there exist $u'u'', v\in L$ such that $u=u'vu''$. Since K is reflective, $(u''u')v\in K$. On the other hand, since $u'u''\in L$, $u''u'\in L$ and hence $(u''u')v\notin K$, a contradiction. This means that $J=K$.

(2) $\Rightarrow$ (3): Let $u\in K$ with $|u|=\min\{|z|\,|\,z\in L\}$ and let $u=bu'$ where $b\in alph(L)$ and $u'\in alph(L)^*$. Let $a\in alph(L)$. Then there exists $v\in K$ such that $v=v'a$ with $v'\in alph(L)^*$ because L is reflective. Consider $v'bu'a\in L$. Since $J=K$, $v'bu'a\in K^2K^*$. Notice that v' is a prefix of v and $|u'a|=|u|=\min\{|z|\,|\,z\in L\}$. This implies that $u'a\in K$. Consequently, $u'alph(L)\subseteq K$. Let $u'a\in K$ for any $a\in alph(L)$. Let $u'a=bu''a$ where $b\in alph(L)$. Then the above method implies $(u''a)alph(L)\subseteq K$. Since $a\in alph(L)$ was taken arbitrarily, $u''alph(L)alph(L)\subseteq K$. By induction, we have $alph(L)^{|u|}\subseteq K$. The fact that K is a prefix code implies $K=alph(L)^{|u|}$ and $L=(alph(L)^{|u|})^*$. □

6. Fully insertion-closed languages

We have been considering so far languages L with the property that $ins(L)\subseteq L$ (ins-closed languages). Of special interest are the languages with the property that all words of X^* belong to $ins(L)$, that is, any word has the property that its insertion in a word of L still belongs to L.

A language $L\subseteq X^*$ is said to be:
- *fully ins-closed* or simply *fins-closed* if $ins(L)=X^*$;
- *extensible* (see [2]) if $u=u_1u_2\in L$, $x\in X^*$ implies $u_1xu_2=rvs$ for some $v\in L$ and $r,s\in X^*$.

Let $L\subseteq X^*$ and let $inf(L)=\{u\in L\,|\,u\neq 1,\ u=xvy,\ v\in L\Rightarrow u=v\}$. Hence $inf(L)$ is the set of all the words in L that are not empty and are minimal in L relatively to the infix order. Clearly if $L\neq\{1\}$, then $inf(L)$ is an infix code. For every ideal L of X^*, $inf(L)$ is the unique infix code with the property $L=X^*inf(L)X^*$.

A language L is fins-closed iff $u=u_1u_2\in L$ implies $u_1xu_2\in L$ for all $x\in X^*$. Such a language is ins-closed and it is an ideal of X^* that has also been called μ-ideal in [11]. Every fins-closed language is extensible and is a regular language [11].

Proposition 6.1. *A language L is extensible if and only if $T=X^*LX^*$ is a fins-closed language.*

Proof. ($\Rightarrow$) Let $u\in T$ and $x\in X^*$. Let $u=x_1x_1'v_1v_2y_1y_1'$ with $v=v_1v_2\in L$ and x_1x_1', $y_1y_1'\in X^*$. Then $x_1xx_1'v_1v_2y_1y_1'\in T$ and $x_1x_1'v_1v_2y_1xy_1'\in T$. Since L is extensible, $v_1xv_2=rwt$ with $w\in L$ and $r,t\in X^*$, therefore $x_1x_1'v_1xv_2y_1y_1'=x_1x_1'rwty_1y_1'\in T$. Hence T is fins-closed.

($\Leftarrow$) Let $u=u_1u_2\in L$ and let $x\in X^*$. Since T is fins-closed, $u_1xu_2\in T=X^*LX^*$ and $u_1xu_2=rvt$ for some $v\in L$ and $r,t\in X^*$. Hence L is an extensible language. □

Corollary 6.1. *An ideal L is fins-closed if and only if $inf(L)$ is an extensible infix code.*

Proof. ($\Rightarrow$) Clearly $L \neq \{1\}$, hence $inf(L)$ is an infix code. Let $u = u_1u_2 \in inf(L) \subseteq L$ and $x \in X^*$. Since L is fins-closed, $u_1xu_2 \in L \subseteq X^*inf(L)X^*$. Hence $u_1xu_2 = rvs$ with $v \in inf(L)$. This shows that $inf(L)$ is extensible.

($\Leftarrow$) Let $u = u_1u_2 \in L = X^*inf(L)X^*$ and let $x \in X^*$. Then $u_1xu_2 = rvs$ with $v \in inf(L)$. Since L is an ideal, then $rvs \in L$ which implies $u_1xu_2 \in L$. Therefore $X^* = ins(L)$, i.e., L is fins-closed. □

Corollary 6.2. *Every extensible infix code U is regular.*

Proof. If $T = X^*UX^*$, then T is fins-closed and hence regular [11]. This implies that $inf(T)$ is regular. Clearly $U = inf(T)$ and hence U is regular. □

Proposition 6.2. *Let $L \subseteq X^*$. Then L is extensible if and only if $inf(L)$ is extensible.*

Proof. ($\Rightarrow$) Let $u \in inf(L)$ and let $x \in X^*$. Since L is extensible, $u'xu'' \in X^*LX^*$ with $u = u'u''$. Hence, there exist $v \in inf(L)$ and $r, s \in X^*$ such that $u'xu'' = rvs$, i.e. $u'xu'' \in X^*inf(L)X^*$. This means that $inf(L)$ is extensible.

($\Leftarrow$) Let $u \in L$ and let $x \in X^*$. Consider $u'xu''$ where $u = u'u''$. Since $u \in L$, there exist $v \in inf(L)$ and $r, s \in X^*$ such that $u = rvs$. If $|u'| \leqslant |r|$ or $|u'| \geqslant rv$, then obviously $u'xu'' \in X^*vX^* \subseteq X^*inf(L)X^*$. If $|u'| > |r|$ and $|u'| < |rv|$, then $v = v'v''$, $u' = rv'$ and $u'xu'' = rv'xv''s$. Since $inf(L)$ is extensible, $v'xv'' \in X^*inf(L)X^*$ and hence $u'xu'' \in X^*inf(L)X^*$. In all the cases, we have $u'xu'' \in X^*LX^*$. This means that L is extensible. □

Proposition 6.3. *Let S and T be two ideals of X^*. Then the catenation ST of S and T is a fins-closed language if and only if both S and T are fins-closed languages.*

Proof. ($\Rightarrow$) Let $u \in S$, let $x \in X^*$ and let $v \in T$ with $|v| = \min\{|w| \mid w \in T\}$. Let $u = u'u''$ with $u', u'' \in X^*$. Since ST is fins-closed, $u'u''v \in ST$ implies $u'xu''v \in ST$. The minimality of $|v|$ implies $u'xu'' \in SX^* \subseteq S$. This means that S is fins-closed. By symmetry, we can show that T is fins-closed.

($\Leftarrow$) Let S and T be two fins-closed languages. Let $u \in ST$ where $u = s_1s_2t_1t_2$ with $s_1s_2 \in S$ and $t_1t_2 \in T$. If $x \in X^*$, then $s_1xs_2t_1t_2 \in ST$, $s_1s_2xt_1t_2 \in ST$ because S is fins-closed and $s_1s_2t_1xt_2 \in ST$ because T is fins-closed. Hence ST is fins-closed. □

References

[1] J. Berstel and D. Perrin, *Theory of Codes* (Academic Press, New York, 1985).

[2] Y.Q. Guo, G. Thierrin and S.H. Zhang, Semaphore codes and ideals, *J. Inform. Optim. Sci.* **9** (1988) 73–83.

[3] M. Ito, C.M. Reis and G. Thierrin, Adherence in finitely generated free monoids, *Congressus Numer.* **95** (1993) 37–45.

[4] H. Jürgensen, H.J. Shyr and G. Thierrin, Monoids with disjunctive identity and their codes, *Acta Math. Hungarica* **47** (1986) 299–312.
[5] L. Kari, On insertion and deletion in formal languages, Ph.D. Thesis, University of Turku, 1991.
[6] L. Kari, Insertion and deletion of words: determinism and reversibility, Lecture Notes in Computer Science, Vol. 629 (Springer, Berlin, 1992) 315–327.
[7] L. Kari, A. Mateescu, Gh. Paun and A. Salomaa, Deletion sets, *Fundam. Inform.* 18 **1** (1993) 355–370.
[8] L. Kari, A. Mateescu, Gh. Paun and A. Salomaa, On parallel deletions applied to a word, RAIRO, *Theoret. Inform. Appl.* **29** (1995) 129–144.
[9] A. Salomaa, *Formal Languages* (Academic Press, New York, 1973).
[10] H.J. Shyr, *Free Monoids and Languages*, Lecture Notes, Institute of Applied Mathematics, National Chung-Hsing University, Taichung, Taiwan, 1991.
[11] G. Thierrin, Hypercodes, right convex languages and their syntactic monoids, *Proc. Amer. Math. Soc.* **83** (1981) 255–258.