$K-$insertion and $k-$deletion closure of languages

Lila Kari and Gabriel Thierrin
Department of Mathematics
University of Western Ontario
N6A 5B7 London, Ontario
Canada
email: lkari@julian.uwo.ca

Abstract

The operation of $k-$insertion ($k-$deletion) is a generalization of the catenation (quotient) of words and languages. The $k-$insertion of w into $u = u_1u_2$ consists of all words u_1wu_2 where the length of u_2 is at most k ($k-$deletion is performed in a analogous fashion). To a language L we associate the set $k-ins(L)$ ($k-del(L)$) consisting of words with the following property: their $k-$insertion into ($k-$deletion from) any word of L yields words which also belong to L. We study properties of these languages and of languages which are $k-$insertion ($k-$deletion) closed.

[1]Mathematics Subject Classification. Primary 68Q45; Secondary 20M35.
[2]This research was supported by Grant OGP0007877 of the Natural Sciences and Engineering Research Council of Canada
[3]All correspondence to Lila Kari.

1 Introduction

The insertion and deletion operations have been introduced in [4] as natural generalizations of catenation, respectively right/left quotient: instead of adding (erasing) a word to (from) the right extremity of another, we insert (delete) it into (from) an arbitrary position.

Even though insertion generalizes catenation, catenation cannot be obtained as a particular case of it, as we cannot force the insertion to take place at the end of the word. The $k-$insertion (introduced in [3] under the name of $k-$catenation) provides the control needed to overcome this drawback. The $k-$insertion is thus more nondeterministic than catenation, but more restrictive than insertion. When $k-$inserting w into $u = u_1u_2$ we obtain all words u_1wu_2 where the length of u_2 is at most k. Remark that now $0-$insertion is exactly the classical catenation. The operation of $k-$deletion (introduced in [3] under the name of $k-$quotient) is analogously defined: the deletion takes place in at most $k+1$ positions. The $0-$deletion amounts thus to the right quotient.

Among other notions connected with insertion and deletion operations, in [1] the insertion and deletion closure of a language L have been defined and studied. This paper introduces similar notions related to the operations of $k-$insertion and $k-$deletion. Namely, to a language L we associate the set $k-ins(L)$ (respectively $k-del(L)$) consisting of the words with the property that, when $k-$inserted into ($k-$deleted from) any word of L, produce words still belonging to L.

Sections 2, 3 focus on these notions. Using the dual operation of dipolar $k-$deletion, both $k-ins(L)$ and $k-del(L)$ can be constructed. Moreover, procedures of constructing the $k-$insertion closure and $k-$deletion closure of a language are given.

When a language equals its $k-$insertion ($k-$deletion) closure, it is called $k-$ins$-$closed (resp. $k-$del$-$closed). If a language L is $k-$ins$-$closed, its words can either be obtained from other words of L by $k-$insertion, or can be "minimal" in this sense. The $k-$insertion base of L consists of all words which belong to the second category. We show that, if a language is regular, then its $k-$ins$-$base is also regular.

Section 4 deals with right $k-$unitary subsemigroups, that is subsemigroups S with the property that $u = u_1u_2 \in S$, $u_1xu_2 \in S$, $|u_2| \leq k$ implies $x \in S$. A method for obtaining the right $k-$unitary closure of a language is given.

Section 5 addresses the issue of minimal $k-$ins$-$closed languages. In general there is no minimal $k-$ins$-$closed language in X^*, therefore other restrictions on the minimality condition have to be added in order to obtain a positive result. Right $m-$density is such a condition: every right $m-$dense and $k-$ins$-$closed language L contains a minimal right $m-$dense and $k-$ins$-$closed language. Moreover, we show that every minimal right $m-$dense and $k-$ins$-$closed language contains a maximal prefix code P such that P^* is right $m-$dense.

In the sequel, 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 $u \in X^*$, $|u|$ denotes the length of w. A language L is called commutative if for every word $w \in L$, the language L contains all the words obtained from w by arbitrarily permuting its letters. For further undefined notions and notations the reader is referred to [6] and [7].

2 $K-$insertion closure

Given two words $u, v \in X^*$, the insertion of v into u is defined as $u \leftarrow v = \{u_1 v u_2 |\ u = u_1 u_2\}$. The operation of $k-$insertion restricts the generality of insertion by allowing words to be inserted only in at most $k+1$ positions. More precisely, let $L_1, L_2 \subseteq X^*$ and let k be a non-negative integer. The $k-$*insertion* of L_2 into L_1 is the language $L_1 \leftarrow^k L_2 = \bigcup_{u \in L_1, v \in L_2} (u \leftarrow^k v)$ where

$$u \leftarrow^k v = \{u_1 v u_2 \mid u = u_1 u_2,\ |u_2| \leq k\}$$

The $0-$insertion of L_2 into L_1 is the catenation $L_1 L_2$.

Clearly $L_1 \leftarrow^k L_2 \subseteq\ L_1 \leftarrow^{k+i} L_2$ and $L_1 \leftarrow L_2 = \bigcup_{k \geq 0} L_1 \leftarrow^k L_2$ where $L_1 \leftarrow L_2$ is the sequential insertion of L_2 into L_1.

Examples. Let $X = \{a, b\}$.

(i) Let $L_1 = a^*$, $L_2 = \{b\}$. Then, $L_1 \leftarrow^k L_2 = a^*b \cup a^*ba \cup \ldots \cup a^*ba^k$. Note that $L_1 \leftarrow^k L_2 \subset\ L_1 \leftarrow^{k+i} L_2$, i.e. the sequence of $k-$insertions is infinite (and strict).

(ii) Let $L_1 = a^*b \cup a^*b^2 \cup a^*b^3$ and $L_2 = \{a^+\}$. Note that $L_1 \leftarrow^3 L_2 = L_1 \leftarrow^{3+i} L_2 = \{a^*b^ia^+|\ i = 1,2,3\} \cup \{a^+b^i|\ i = 1,2,3\} \cup a^*ba^+b \cup a^*b^2a^+b \cup a^*ba^+b^2$. The sequence of $k-$insertions can be finite even with infinite languages.

Proposition 2.1 *([3]) The families of regular, context-free and context-sensitive languages are closed under $k-$insertion.*

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

$$k - ins(L) = \{x \in X^* | \forall u \in L, u = u_1 u_2, |u_2| \leq k\ \Rightarrow\ u_1 x u_2 \in L\}.$$

Examples. Let $X = \{a, b\}$. Then,

– $k-ins(X^*) = X^*$ while $k-ins(L_{ab}) = L_{ab}$, where $L_{ab} = \{w \in X^* |\ w$ has the same number of a's and b's$\}$;

– if $L = \{a^n b^n | n \geq 0\}$ then $k-ins(L) = \{1\}$;

– if $L_1 = (a^2)^*$, $L_2 = aL_1$ then $k-ins(L_1) = L_1$ and $k-ins(L_2) = L_1$;

– if $L = b^*ab^*$ then $k-ins(L) = b^*$;

– if $L = aX^*b$ then $0-ins(L) = 1-ins(L) = X^*b$ and $k-ins(L) = aX^*b$ for $k \geq 2$.

Proposition 2.2 *$k-ins(L)$ is a submonoid of X^*. Moreover, if L is a commutative language, then $k-ins(L)$ is also a commutative language.*

Proof. Let $x, y \in k-ins(L)$ and $u = u_1u_2 \in L$, $|u_2| \le k$. Then $u_1xu_2 \in L$, $u_1xyu_2 \in L$, hence $xy \in k-ins(L)$. Since $1 \in k-ins(L)$, $k-ins(L)$ is not empty and hence a submonoid of X^*.

For the second claim, it is sufficient to show that $xuvy \in k-ins(L)$ implies $xvuy \in k-ins(L)$. If $w \in L$, $w = w_1w_2$, $|w_2| \le k$, then $w_1xuvyw_2 \in L$, hence $w_1xvuyw_2 \in L$. Therefore $xvuy \in k-ins(L)$. □

In [1], in order to construct the language $ins(L)$ from L, the dipolar deletion operation was used: $u \rightleftharpoons v = \{x \in X^* | \ u = v_1xv_2, v = v_1v_2\}$. In the case of $k-$insertion, for the same purpose, we will make use of a similar operation, the dipolar $k-$deletion. For u, v words over X, the *dipolar $k-$deletion* $u \rightleftharpoons^k v$ is defined by $u \rightleftharpoons^k v = \{x \in X^* | \ u = v_1xv_2, v = v_1v_2, |v_2| \le k\}$. (The operation has been introduced in [3] under the name of $k-$deletion.) In other words, the dipolar $k-$deletion erases from u a prefix v_1 of any length and a suffix v_2 of length $\le k$ whose catenation v_1v_2 equals v. The operation can be extended to languages in the natural fashion. If L_1 and L_2 are two languages, then the *dipolar $k-$deletion* of L_2 from L_1 is the language

$$L_1 \rightleftharpoons^k L_2 = \bigcup_{u \in L_1, v \in L_2} u \rightleftharpoons^k v$$

Note that the dipolar deletion of L_2 from L_1, satisfies

$$L_1 \rightleftharpoons L_2 = \bigcup\nolimits_{k \ge 0} L_1 \rightleftharpoons^k L_2 \ .$$

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

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

Proof. Take $x \in k-ins(L)$. Assume, for the sake of contradiction, that $x \notin (L^c \rightleftharpoons^k L)^c$. Then $x \in (L^c \rightleftharpoons^k L)$, that is, there exist $u_1xu_2 \in L^c$, $u_1u_2 \in L$, $|u_2| \le k$, such that $x \in u_1xu_2 \rightleftharpoons^k u_1u_2$. We arrived at a contradiction, as $x \in k-ins(L)$ and $u_1u_2 \in L$, $|u_2| \le k$, but the $k-$insertion of x into u_1u_2 belongs to L^c.

Consider now a word $x \in (L^c \rightleftharpoons^k L)^c$. If $x \notin k-ins(L)$, there exists $u_1u_2 \in L$, $|u_2| \le k$ such that $u_1xu_2 \notin L$. This further implies $u_1xu_2 \in L^c$ and $x \in L^c \rightleftharpoons^k L$ – a contradiction with the original assumptions about x. □

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

Proof. It has been proved in [3] that if L is regular, then $L \rightleftharpoons^k R$ is regular and moreover, the proof is constructive. Since L is regular, L^c is regular. This implies that $L^c \rightleftharpoons^k L$ is regular, hence $k - ins(L) = (L^c \rightleftharpoons^k L)^c$ is regular. □

A nonempty subset $S \subseteq X^*$ such that $u \in S$, $v \in S$, imply $u \leftarrow^k v \subseteq S$ is called a $k-$*subsemigroup* (see [3]). Clearly S is a subsemigroup of X^*. If S contains the identity, it is called a $k-$submonoid. A language L such that $L \subseteq k-ins(L)$ is called $k-$ins$-$closed. It is easy to see that a language L is $k-$ins$-$closed iff it is a $k-$subsemigroup.

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

$$KI(L) = \bigcap_{L_i \in KI_L} L_i$$

of the languages belonging to the family KI_L is clearly a $k-$ins$-$closed language containing L and it is called the $k-$*ins*$-$*closure* of L. The $k-$ins$-$closure of a language L is the smallest $k-$ins$-$closed language containing L.

Notice that a language L is $k-$ins$-$closed iff $L \leftarrow^k L \subseteq L$. Indeed, if $x \in L$, $u_1u_2 \in L$, $|u_2| \leq k$, then, as $x \in L \subseteq k-ins(L)$ we have that $u_1xu_2 \in L$. For the other implication, take $x \in L$ and $u_1u_2 \in L$, $|u_2| \leq k$. As $L \leftarrow^k L \subseteq L$ we have that $u_1xu_2 \in L$ which shows that $x \in k-ins(L)$.

The $k-$insertion of order n of L_2 into L_1 is inductively defined by the equations:

$$L_1 \leftarrow^{k(0)} L_2 = L_1$$

$$\ldots$$

$$L_1 \leftarrow^{k(i+1)} L_2 = (L_1 \leftarrow^{k(i)} L_2) \leftarrow^k L_2, \ i \geq 0.$$

The iterated sequential $k-$insertion of L_2 into L_1 is defined by:

$$L_1 \leftarrow^{k*} L_2 = \bigcup_{n \geq 0} (L_1 \leftarrow^{k(n)} L_2).$$

Examples. (i) Let $L_1 = a^+ba^k$ and $L_2 = \{a\}$. Then: $L_1 \leftarrow^{k*} L_2 = a^+b(a^*)a^k$.

(ii) Let $L_1 = ba^+ba^k$ and $L_2 = \{a\}$. Then $L_1 \leftarrow^{k*} L_2 = ba^+b(a^*)a^k$.

Proposition 2.4 *The $k-$insertion closure of a language L is $KI(L) = L \leftarrow^{k*} L$.*

Proof. "$KI(L) \subseteq L \leftarrow^{k*} L$". Obvious, as $L \leftarrow^{k*} L$ is $k-$ins$-$closed and L is included in $L \leftarrow^{k*} L$.

"$L \leftarrow^{k*} L \subseteq KI(L)$" We show by induction on n that $L \leftarrow^{k(n)} L \subseteq KI(L)$. For $n = 0$ the assertion holds, as $L \subseteq KI(L)$. Assume that $L \leftarrow^{k(n)} L \subseteq KI(L)$ and consider a word $u \in L \leftarrow^{k(n+1)} L = (L \leftarrow^{k(n)} L) \leftarrow^k L$. Then $u = u_1vu_2$, $|u_2| \leq k$, where $u_1u_2 \in L \leftarrow^{k(n)} L$ and $v \in L$. As both $L \leftarrow^{k(n)} L$ and L are included in $KI(L)$ and $KI(L)$ is $k-$ins$-$closed, we deduce that $u \in KI(L)$.

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

Let $L \subseteq X^*$ be a $k-$ins$-$closed language. As the result of the $k-$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 $k-$insertions of other words of L, and words that cannot be obtained in this fashion.

Consider the set

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

$$L\backslash((L\backslash\{1\}) \longleftarrow^{k+} (L\backslash\{1\})),$$

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

The following result shows that if L is regular, its $k-$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 [6]) g such that $g(L)$ is the set of words in L that can be obtained as results of $k-$insertions.

Proposition 2.5 *If L is a regular $k-$ins$-$closed language, then its $k-$ins$-$base $KB(L)$ is a regular language.*

Proof. Let L be a regular $k-$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 $s_i a \longrightarrow s_j$, $s_i, s_j \in S$, $a \in X$.

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

Notice first that, as L is $k-$ins$-$closed, $L\backslash KB(L) = \{u \in L \mid\ u = v_1 w v_2, w \in L, v_1 v_2 \in L, |v_2| \leq k\}$.

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

$$\begin{array}{rcll}
S' & = & S \cup \{s_j^{(i)} \mid\ 0 \leq j \leq n, 0 \leq i \leq n\} \cup \{s_{i,j} \mid\ s_i \in F, 0 \leq j \leq k\} & \\
F' & = & \{s_{i,j} \mid\ s_i \in F, 1 \leq j \leq k\} & \\
P' & = & \{s_i a \longrightarrow a s_m \mid\ s_i a \longrightarrow s_m \in P\} & (1) \\
 & \cup & \{s_i a \longrightarrow a s_j^{(i)} \mid\ s_0 a \longrightarrow s_j \in P\} & (2) \\
 & \cup & \{s_j^{(i)} a \longrightarrow a s_m^{(i)} \mid\ s_j a \longrightarrow s_m \in P, 0 \leq i \leq n\} & (3) \\
 & \cup & \{s_j^{(i)} a \longrightarrow a s_{i,0} \mid\ s_j a \longrightarrow s_l \in P, s_l \in F\} & (4) \\
 & \cup & \{s_{i,j} a \longrightarrow a s_{m,j+1} \mid\ s_i a \longrightarrow s_m \in P, 1 \leq j \leq k-1\} & (5)
\end{array}$$

The idea of the proof is the following. We have constructed *card*(S) indexed copies of the automaton A, $A^{(i)} = (X, S^{(i)}, s_0^{(i)}, F^{(i)}, P^{(i)})$, $1 \leq i \leq n$. Given a word $v_1 w v_2 \in L$, 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. 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, the states $s_{i,j}$, $0 \leq j \leq k$ are the states s_i in disguise; the second index j makes sure that the length of v_2 is at most k and that at least one word w has been encountered.)

From the above explanations it follows that g reaches a final state iff the input word u is of the form v_1wv_2, $v_1v_2 \in L$, $|v_2| \leq k$, $w \in L$. Consequently, $g(L) = \{v_1wv_2|\ v_1v_2 \in L, |v_2| \leq k, w \in L\}$. □

3 $K-$deletion closure

Given words $u, v \in X^*$, the deletion of v from u is

$$u \rightarrow v = \{u_1u_2|\ u = u_1vu_2\}.$$

The $k-$deletion operation puts some restrictions on the positions where the deletion can take place, being thus more deterministic than deletion, but more nondeterministic than the right quotient. More precisely, let $L_1, L_2 \subseteq X^*$ and let k be a non-negative integer. The $k-$*deletion* of L_2 from L_1 is the language $L_1 \rightarrow^k L_2 = \cup_{u \in L_1, v \in L_2}(u \rightarrow^k v)$ where

$$u \rightarrow^k v = \{u_1u_2|\ u = u_1vu_2, |u_2| \leq k\}.$$

If $k = 0$ we obtain the usual right quotient.

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

$$k - del(L) = \{x \in k - Sub(L)|\ \forall u \in L, u = u_1xu_2, |u_2| \leq k \ \Rightarrow\ u_1u_2 \in L\}.$$

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

Examples.

– $k-del(X^*) = X^*$, $k-del(L_{ab}) = L_{ab}$;

– $k-del(\{a^nb^n|\ n \geq 0\} = \{a^nb^n|\ n \geq 0\}$;

– $k-del(ba^*b^m) = \emptyset$ if $k < m$ and $k-del(ba^*b^m) = a^*$ for $k \geq m$.

Proposition 3.1 *Let $L \subseteq X^*$.*

(i) If $x, y \in k{-}del(L)$ and $xy \in k{-}Sub(L)$, then $xy \in k{-}\ del(L)$.

(ii) If $k{-}Sub(L)$ is a submonoid of X^*, then $k{-}del(L)$ is a submonoid of X^*.

(iii) If L is a commutative language, then $k{-}del(L)$ is also commutative.

Proof. *(i)* Let $x, y \in k{-}del(L)$ with $xy \in k - Sub(L)$. If $u = u_1xyu_2 \in L, |u_2| \leq k$, then $u_1yu_2 \in L$ and consequently $u_1u_2 \in L$. Therefore $xy \in k{-}del(L)$.

(ii) Immediate.

(iii) It is sufficient to show that $xuvy \in k{-}del(L)$ implies $xvuy \in k{-}del(L)$. Since L is commutative, $u_1xuvyu_2 \in L$ $|u_2| \leq k$ if and only if $u_1xvuyu_2 \in L$, $|u_2| \leq k$. If $u = u_1xuvyu_2$ we have that $u_1u_2 \in L$ and $u_1xvuyu_2 \in L$. This implies $xvuy \in k{-}del(L)$. □

Proposition 3.2 $k{-}del(L) = (L \rightleftharpoons^k L^c)^c \cap k - Sub(L)$.

Proof. Let $x \in k{-}del(L)$. From the definition of $k{-}del(L)$ it follows that $x \in k - Sub(L)$. Assume that $x \notin (L \rightleftharpoons^k L^c)^c$. This means there exist $u_1xu_2 \in L$, $|u_2| \leq k$ and $u_1u_2 \in L^c$ such that $x \in u_1xu_2 \rightleftharpoons^k u_1u_2$. We arrived at a contradiction as $x \in k{-}del(L)$, but $u_1xu_2 \in L$, $|u_2| \leq k$ and $u_1u_2 \notin L$.

For the other inclusion, let $x \in (L \rightleftharpoons^k L^c)^c \cap k - Sub(L)$. As $x \in k - Sub(L)$, if $x \notin k{-}del(L)$ there exist $u_1xu_2 \in L$, $|u_2| \leq k$ such that $u_1u_2 \notin L$. This further implies that $u_1u_2 \in L^c$, that is, $x \in L \rightleftharpoons^k L^c$ – a contradiction with the initial assumption about x. □

A language L is called *$k{-}del{-}closed$* if $v \in L$ and $u_1vu_2 \in L$, $|u_2| \leq k$, implies $u_1u_2 \in L$. Remark that every k−del−closed language contains the identity 1.

Proposition 3.3 *Let $L \subseteq X^*$ be a $k{-}ins{-}closed$ language. L is $k{-}del{-}closed$ if and only if $L = (L \rightarrow^k L)$.*

Proof. ($\Rightarrow$) Let $u \in (L \rightarrow^k L)$. Then there exists $u_1, u_2 \in X^*$, $|u_2| \leq k$, and $v \in L$ such that $u = u_1u_2$ and $u_1vu_2 \in L$. Since L is k−del−closed, $u = u_1u_2 \in L$. This means that $(L \rightarrow^k L) \subseteq L$.

Now let $u \in L$. Since L is k−ins−closed, $uu \in L$. Therefore $u \in (L \rightarrow^k L)$, i.e. $L \subseteq (L \rightarrow^k L)$. We can conclude that $L = (L \rightarrow^k L)$.

($\Leftarrow$) Let $u_1vu_2 \in L$ for some $u_1, u_2 \in X^*$, $|u_2| \leq k$ and $v \in L$. Consider $u_1u_2 \in X^*$. Since $u_1u_2 \in (L \rightarrow^k L)$ and $(L \rightarrow^k L) = L$, $u_1u_2 \in L$. This means that L is k−del−closed. □

If L is a nonempty language and if KD_L is the family of all the k−del−closed languages L_i containing L, then the intersection $\bigcap_{L_i \in KD_L} L_i$ of all the k−del−closed languages containing L is also a k−del−closed language called the k−del−closure of L. The k−del−closure of L is the smallest k−del−closed language containing L.

We will now define a sequences of languages whose union is the $k-$del$-$closure of a given language L. Let:

$$KD_0(L) = L$$
$$KD_1(L) = KD_0(L) \to^k (KD_0(L) \cup \{1\})$$
$$KD_2(L) = KD_1(L) \to^k (KD_1(L) \cup \{1\})$$
$$\dots$$
$$KD_{j+1}(L) = KD_j(L) \to^k (KD_j(L) \cup \{1\})$$
$$\dots$$

Clearly $KD_j(L) \subseteq KD_{j+1}(L)$. Let

$$KD(L) = \bigcup_{j \geq 0} KD_j(L)$$

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

Proof. Clearly $L \subseteq KD(L)$.

Let $v \in KD(L)$ and $u_1vu_2 \in KD(L)$, $|u_2| \leq k$. Then $v \in KD_i(L)$ and $u_1vu_2 \in KD_j(L)$ for some integers $i, j \geq 0$. If $l = max\{i, j\}$, then $v \in KD_l(L)$ and $u_1vu_2 \in KD_l(L)$. This implies $u_1u_2 \in KD_{l+1}(L) \subseteq KD(L)$. Therefore $KD(L)$ is a $k-$del$-$closed language containing L.

Let T be a $k-$del$-$closed language such that $L = KD_0(L) \subseteq T$. Since T is $k-$del$-$closed, if $KD_j(L) \subseteq T$ then $KD_{j+1}(L) \subseteq T$. Using induction, it follows then that $KD(L) \subseteq T$. □

Since, by [3], the family of regular languages is closed under $k-$deletion, it follows that if L is regular, then the languages $KD_j(L)$, $j \geq 0$, are also regular. However, it is an open question whether $KD(L)$ is regular for any regular language $L \subseteq X^*$. 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 $KD(L)$ is commutative and regular.*

Proof. Let us show first that $KD(L)$ is commutative. To this end, it is sufficient to show that $KD_{j+1}(L)$ is commutative if $KD_j(L)$ is commutative. Let $xuvy \in KD_{j+1}(L)$. By the definition of $KD_{j+1}(L)$, there exist $w \in KD_j(L)$, $z \in KD_j(L) \cup \{1\}$, such that $w \in (xuvy \leftarrow^k z)$. Since $KD_j(L)$ is commutative, $xuvyz \in KD_j(L)$ and $xvuyz \in KD_j(L)$. From the fact that $z, xvuyz \in KD_j(L)$ and the definition of $KD_{j+1}(L)$, it follows that $xvuy \in KD_{j+1}(L)$, i.e. $KD_{j+1}(L)$ is commutative.

We will show next that $KD(L)$ is regular. To this aim, we show that $u \equiv v(P_{KD_j(L)})$ implies $u \equiv v(P_{KD_{j+1}(L)})$. Let $u \equiv v(P_{KD_j(L)})$ and let $xuy \in KD_{j+1}(L)$. By the definition of $KD_{j+1}(L)$, there exists $w \in KD_j(L)$,

$z \in KD_j(L) \cup \{1\}$, such that $w \in (xuy \leftarrow^k z)$. Since $KD_j(L)$ is commutative, $xuyz \in KD_j(L)$. Hence $xvyz \in KD_j(L)$. From the fact that $z \in KD_j(L)$ and by the definition of $KD_{j+1}(L)$, it follows that $xvy \in KD_{j+1}(L)$. In the same way, $xvy \in KD_{j+1}(L)$ implies $xuy \in KD_{j+1}(L)$. Consequently, $u \equiv v(P_{KD_{j+1}(L)})$ holds. This means that the number of congruence classes of $P_{KD_{j+1}(L)}$ is smaller than or equal to that of $P_{KD_j(L)}$. Remark that

$$KD_0(L) \subseteq KD_1(L) \subseteq \ldots \subseteq KD_n(L) \subseteq KD_{n+1}(L) \ldots$$

It can be shown that $KD_t(L) = KD_{t+1}(L)$ for some $t, t \geq 1$. Thus, $KD(L) = KD_t(L)$ which implies that $KD(L)$ is regular. $\square$

4 Right $k-$unitary languages and $k-$prefix codes

Recall (see [3]) that a $k-$prefix code P is a nonempty language, $P \subseteq X^+$, such that $u, u_1xu_2 \in P$ with $u = u_1u_2$ and $|u_2| \leq k$ implies $x = 1$. A code is a prefix code iff it is a $0-$prefix code and an outfix code iff it is a $k-$prefix code for $k \geq 0$.

If $P_k(X)$ is the family of all the $k-$prefix codes over X with $|X| \geq 2$, then we have the following strict hierarchy:

$$\cdots \subset P_{i+1}(X) \subset P_i(X) \subset \cdots \subset P_1(X) \subset P_0(X)$$

It is immediate that $P_{i+1}(X) \subseteq P_i(X)$. However $P_{i+1}(X) \subset P_i(X)$. Suppose that $X = \{a, b, \cdots\}$ and let $T_i = \{a^n b^n | n \geq i+1\}$. Then T_i is a $i-$prefix code, but not a $(i+1)-$prefix code.

The *relation* ρ_k defined on X^* by:

$$u\rho_k v \;\Leftrightarrow\; v = u_1xu_2, u = u_1u_2, |u_2| \leq k,$$

is reflexive, antisymmetric and left compatible. The *transitive closure* $\bar{\rho}_k$ of ρ_k is a left compatible partial order. The language P is a $k-$prefix code iff it is an anti-chain with respect to ρ_k (see [3]). Remark that if $k = 0$, ρ_0 is the usual prefix order.

Let $L \subseteq X^+$ be a nonempty language and let:

$$\mathrm{Prf}_k(L) = \{u \in L | \; u = v_1xv_2, v = v_1v_2 \in L, |v_2| \leq k, \;\Rightarrow\; x = 1\}.$$

It is easy to see that $\mathrm{Prf}_k(L)$ is a $k-$prefix code and that $\mathrm{Prf}_k(L) = \{u \in L | v\rho_k u, v \in L \;\Rightarrow\; v = u\}$, i.e., $\mathrm{Prf}_k(L)$ is the set of words in L that are minimal with respect to the relation ρ_k or $\bar{\rho}_k$ (see [3]).

A subsemigroup $S \subseteq X^*$ is called *right* $k-$*unitary* if $u = u_1u_2, u_1xu_2 \in S$, $|u_2| \leq k$, implies $x \in S$. Clearly, $1 \in S$. Hence every right $k-$unitary subsemigroup is a submonoid, called *right* $k-$*unitary submonoid*.

Let $X = \{a, b\}$. Then a^* and L_{ab} are right $k-$unitary for every $k \geq 0$.

Proposition 4.1 *Let $L \subseteq X^*$ be a nonempty language and let $KU(L)$ (respectively $KUK(L)$) be the intersection of all the right $k-$unitary submonoids ($k-$submonoids) containing L. Then $KU(L)$ (respectively $KUK(L)$) is a right $k-$unitary submonoid ($k-$submonoid) of X^*.*

Proof. Immediate. □

Remark that $KU(L)$ (resp. $KUK(L)$) is the smallest $k-$unitary submonoid ($k-$submonoid) of X^* containing L. The submonoid $KU(L)$ is called the *right $k-$unitary closure* of L.

Let $U_k(L) = \{x \in X^* | \exists u = u_1u_2 \in L, |u_2| \leq k \ \text{ with } u_1xu_2 \in L\}$. Note that $U_k(L) = L \rightleftharpoons^k L$. If L is regular then, from a result of [3], it follows that $U_k(L)$ is also regular.

Define the sequence:

$$U_k^0(L) = L \cup \{1\},\ \ U_k^1(L) = U_k(U_k^0(L)), \ldots, U_k^{i+1}(L) = U_k(U_k^i(L)) \ldots$$

Clearly, $U_k^i(L) \subseteq U_k^{i+1}(L)$.

Proposition 4.2 *If $L \subseteq X^*$ is nonempty, then $KU(L) = \bigcup_{i \geq 0} U_k^i(L)$.*

Proof. Let $T = \bigcup_{i \geq 0} U_k^i(L)$. Clearly $L \subseteq T$. Let $u = u_1u_2 \in T$ and $u_1xu_2 \in T$ with $|u_2| \leq k$. Then $u = u_1u_2, u_1xu_2 \in U_k^i(L)$ for some $i \geq 0$. This implies $x \in U_k^{i+1}(L) \subseteq T$. Hence T is right $k-$unitary and $KU(L) \subseteq \bigcup_{i \geq 0} U_k^i(L)$.

We show by induction that $\bigcup_{i \geq 0} U_k^i(L) \subseteq KU(L)$. For $i = 0$, the assertion holds because $L \subseteq KU(L)$ and hence $U_k^0(L) \subseteq KU(L)$. Assume that $U_k^i(L) \subseteq KU(L)$ and let $x \in U_k^{i+1}(L)$. Then there exists $u \in U_k^i(L)$, $u = u_1u_2$ with $|u_2| \leq k$ such that $u_1xu_2 \in U_k^i(L)$. Hence:

$$u = u_1u_2 \in KU(L),\ u_1xu_2 \in KU(L).$$

Since $KU(L)$ is right $k-$unitary, it follows that $x \in KU(L)$, which implies $U_k^{i+1}(L) \subseteq KU(L)$. Consequently, $KU(L) = \bigcup_{i \geq 0} U_k^i(L)$. □

5 Right $m-$dense and $k-$ins$-$closed languages

A $k-$ins$-$closed language L is said to be minimal if $L' \subseteq L$, with L' a $k-$ins$-$closed language, implies $L' = L$. The next result shows that a $k-$ins$-$closed language in X^+ cannot be minimal and hence other restrictions to the minimality are necessary in order to get positive results.

Proposition 5.1 *There is no minimal $k-$ins$-$closed language in X^+.*

Proof. Suppose that $L \subseteq X^+$ is a minimal $k-$ins$-$closed language. Let $w \in L$ with minimal length $m = |w|$ and let $L' = L \backslash \{w\}$. The language L' is not $k-$ins$-$closed, therefore there exist $u = u_1u_2 \in L'$, $v \in L'$, $|u_2| \leq k$, such that $u_1vu_2 \notin L'$. However, since $L' \subseteq L$ and L is $k-$ins$-$closed, we have that $u_1vu_2 \in L$. Therefore $u_1vu_2 = w$, which implies $|w| > |u|$ - a contradiction. □

A language $L \subseteq X^*$ is called *right $m-$dense* if for any $w \in X^*$, there exists $x \in X^*$, $|x| \leq m$, such that $wx \in L$. A right $m-$dense and $k-$ins$-$closed language L is said to be minimal if it does not properly contain any right $m-$dense and $k-$ins$-$closed language. It has been shown in [2] that every right $m-$dense language contains a minimal one.

Proposition 5.2 *Every right $m-$dense and $k-$ins$-$closed language L contains a minimal right $m-$dense and $k-$ins$-$closed language.*

Proof. Let $D(L) = \{L_\delta | \delta \in \Delta\}$ be the family of the right $m-$dense and $k-$ins$-$closed languages L_δ contained in L and let $I = \{L_\gamma \mid \gamma \in \Gamma\}$ be an infinite descending chain of languages L_γ belonging to the family D(L):

$$L \supseteq \ldots \supseteq L_\alpha \supseteq \ldots \supseteq L_\zeta \supseteq \ldots$$

Let $L_0 = \cap_{\gamma\in\Gamma} L_\gamma$ and let X_m be the set of words of length $\leq m$.

Suppose first that L_0 is not right $m-$dense. Then there exists $u \in X^*$ such that $ux_j \notin L_0$ for all $x_j \in X_m$, that is, for each x_j there exists a $L_{\gamma_j} \in I$ such that $ux_j \notin L_{\gamma_j}$. If $L_\phi = \cap_{1\leq j\leq n} L_{\gamma_j}$ where $n = |X_m|$, then $ux_j \notin L_\phi$ for every $x_j \in X_m$. Since I is a descending chain, $L_\phi \in I$ and L_ϕ is right $m-$dense, a contradiction. Hence L_0 is right $m-$dense and therefore $L_0 \neq \emptyset$. It is immediate that, if not empty, the intersection of $k-$ins$-$closed languages is a $k-$ins$-$closed language. Hence L_0 is also $k-$ins$-$closed.

As every infinite descending chain I in $D(L)$ has a lower bound L_0, the family $D(L)$ is inductive. Consequently, according to Zorn's lemma, $D(L)$ has at least a minimal element which is a minimal right $m-$dense and $k-$ins$-$closed language contained in L. □

Corollary 5.1 *Let L be a regular right dense $k-$ins$-$closed language. Then L contains a minimal right $m-$dense and $k-$ins$-$closed language, m being a positive integer depending on L.*

Proof. It follows from a result of [2] stating that every right dense regular language is $m-$dense for some positive integer. □

Proposition 5.3 *Let L be a minimal right $m-$dense and $k-$ins$-$closed language. Then L contains a maximal prefix code P such that P^* is right $m-$dense.*

Proof. Since L is $k-$ins$-$closed, L is a subsemigroup that is right $m-$dense. By a result of [2] this implies that L contains a maximal prefix code P with P^* right $m-$dense. □

Proposition 5.4 *Let $L \subseteq X^*$. Then for every $k \geq 0$ and $m \geq 1$, there exists a right $m-$dense and $k-$ins$-$closed language L_μ such that:*

(i) $L \subseteq L_\mu$.

(ii) If L' is right $m-$dense and $k-$ins$-$closed with $L \subseteq L' \subseteq L_\mu$, then $L' = L_\mu$.

Proof. Let $I = \{L_\gamma \mid \gamma \in \Gamma\}$ be an infinite descending chain of right $m-$dense and $k-$ins$-$closed languages L_γ containing L:

$$\ldots \supseteq L_\alpha \supseteq \ldots \supseteq L_\delta \supseteq \ldots \supseteq L.$$

This chain is not empty because it contains X^*. Let $L_0 = \cap_{\gamma \in \Gamma} L_\gamma$ and let X_m be the set of words of length $\leq m$.

As in the proof of Proposition 5.2, it can be shown that L_0 is right $m-$dense and $k-$ins$-$closed. If $F(L)$ denotes the family of the right $m-$dense and $k-$ins$-$closed languages then, by applying the Zorn's lemma, we can deduce that $F(L)$ has at least a minimal element L_μ satisfying the conditions (i) and (ii). $\square$

References

[1] M. Ito, L. Kari and G. Thierrin, Insertion and deletion closure of languages, To appear in *Theoretical Computer Science*.

[2] M. Ito, G. Thierrin and S.S. Yu, Right $k-$dense languages, *Semigroup Forum* 48 (1994), 313-325.

[3] L. Kari and G. Thierrin, K-catenation and applications: k-prefix codes, *Journal of Information & Optimization Sciences*, 16 (1995).

[4] L. Kari, On insertion and deletion in formal languages, *Ph.D. Thesis*, University of Turku, 1991.

[5] L.Kari, Language equations with invertible operations, *Theoretical Computer Science*, 132 (1994), 129-150.

[6] A.Salomaa, *Formal Languages*, Academic Press, New York, 1973.

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