Theoretical Computer Science 136 (1994) 471–485
Elsevier

Congruences, infix and cohesive prefix codes*

M. Ito
Faculty of Science, Kyoto Sangyo University, Kamigamo-Motoyama, Kita-ku, Kyoto, 603, Japan

G. Thierrin
Department of Mathematics, The University of Western Ontario, London, Ont., Canada

Communicated by M. Nivat
Received October 1993
Revised December 1993

Abstract

Ito, M. and G. Thierrin, Congruences, infix and cohesive prefix codes, Theoretical Computer Science 136 (1994) 471–485.

A language $L \subseteq X^*$ is called a *cohesive prefix code* if $xLy \cap L \neq \emptyset$ implies that $y=1$ and $xL \subseteq L$ for any $x, y \in X^*$. An example of cohesive prefix codes is an infix code. We determine first the structure of cohesive prefix codes and then we study several relationships between maximal infix codes and maximal cohesive prefix codes. Finally, we determine the structure of a cohesive prefix code that is a right semaphore code.

1. Introduction

Congruences occur naturally in the theory of languages and automata. With every language L over an alphabet X is associated its syntactic congruence P_L and the corresponding syntactic monoid $Syn(L)$. Several important classes of languages can be characterized by the properties of their syntactic congruences and the connected syntactic monoids. Also many interesting properties of codes are related to the different types of congruences that can be associated with them. For example, infix

Correspondence to: M. Ito, Faculty of Science, Kyoto Sangyo University, Kamigamo-Motoyama, Kita-ku, Kyoto, 603, Japan. Email: ito@ipnksuvx.bitnet.
* This work was supported by Grant-in-Aid for Scientific Research 04044150, Ministry of Education, Science and Culture of Japan and the Natural Science and Engineering Research Council of Canada Grant A7877.

SSDI 0304-3975(94)00018-E

codes and prefix codes are, respectively, classes of their syntactic and right syntactic congruences.

In this paper, we consider properties of languages that are classes of congruences or right congruences, especially in relation with codes. After establishing, in Section 2, several general properties of languages that are classes of congruences, we introduce, in Section 3, an important family of such languages, the family of cohesive prefix codes which is a subfamily of the prefix codes. In particular, we give a characterization of the syntactic monoid of these languages. Section 4 is devoted to the study of closure properties of cohesive prefix codes as well as the properties of maximal cohesive prefix codes. In the last section, we consider cohesive prefix codes that are also right semaphore codes and give some representations of them.

2. Congruences

Let X be a finite alphabet and let X^* be the free monoid generated by X. An equivalence relation ρ on X^* is said to be *right* (*left*) *compatible* if $u \equiv v(\rho)$ implies $ux \equiv vx(\rho)$ $(xu \equiv xv(\rho))$ for every $x \in X^*$. If ρ is right and left compatible, then ρ is said to be *compatible*. Remark that an equivalence relation ρ is compatible if and only if $r \equiv s(\rho)$ and $u \equiv v(\rho)$ implies $ru \equiv sv(\rho)$. Using this property, the product of classes of ρ can be defined naturally to obtain a new monoid X^*/ρ called the *quotient monoid* of X^* modulo ρ. Right (left) compatible and compatible equivalence relations are also called, respectively, *right* (*left*) *congruences* and *congruences*. If L is a language over X and if $u \in X^*$, then let $L.\cdot u = \{x \mid x \in X^*, ux \in L\}$, $L\cdot.u = \{x \mid x \in X^*, xu \in L\}$ and $L..u = \{(x, y) \mid x, y \in X^*, xuy \in L\}$

The relation R_L defined by $u \equiv v(R_L)$ if and only if $L.\cdot u = L.\cdot v$ is a right congruence called the *principal right congruence* or the *right syntactic congruence* defined by L. The *principal left congruence* $_LR$ is defined symmetrically. The relation P_L defined by $u \equiv v(P_L)$ if and only if $L..u = L..v$ is a congruence called the *principal congruence* or *the syntactic congruence* defined by L. The quotient monoid $Syn(L) = X^*/P_L$ is called the *syntactic monoid of* L. If not empty, the set $W_L = \{u \in X^* \mid L.\cdot u = \emptyset\}$ $({}_LW = \{u \in X^* \mid L..u = \emptyset\})$, called the *right* (*left*) *residue* of L, and the set $W(L) = \{u \in X^* \mid L..u = \emptyset\}$, called the *residue* of L, are classes of respectively R_L $({}_LR)$ and P_L. If the (right, left) residue of a language L is empty, then L is said to be (*right, left*) *dense*. Remark that $P_L \subseteq R_L$ and that L is a union of classes of R_L and P_L.

A language $L \subseteq X^*$, $L \neq \emptyset$, is called *rc-simple* if L is a class of a right congruence and *c-simple* if L is a class of a congruence. The following result is well known and easy to prove.

Fact 2.1. *For a language* L, *the following conditions are equivalent*:

(1) L *is rc-simple* (*c-simple*);

(2) L *is a class of* R_L (*a class of* P_L);

(3) $Lx \cap L \neq \emptyset$ $(xLy \cap L \neq \emptyset)$ *implies* $Lx \subseteq L(xLy \subseteq L)$.

Many examples of *rc*-simple or *c*-simple languages can be found in several families of codes. Recall that a *code* C over X is a nonempty language $C \subseteq X^+$ such that $c_1 c_2 \cdots c_m = d_1 d_2 \cdots d_n$, $c_i, d_j \in C$, implies that $m = n$ and $c_1 = d_1, \ldots, c_m = d_m$. Let $L \subseteq X^+$ be a nonempty language over X. If $u, ux \in L$ ($u, xu \in L$) implies $x = 1$, then L is a code called a *prefix (suffix) code*. If $u, xuy \in L$ implies $x = y = 1$, then L is a code called an *infix code*. It is immediate that every prefix code is an *rc*-simple language and that every infix code is a *c*-simple language.

Decompositions of regular *rc*-simple languages in connection with prefix codes have been given in [6] where *rc*-simple languages are called right simple languages. These decompositions can easily be extended to the general case in the following way.

Proposition 2.2. *Let L be a nonempty rc-simple language. Then*

(1) *if $1 \in L$, then $L = \{1\}$ or $L = P^*$ where P is a prefix code;*

(2) *if $1 \notin L$, then either L is a prefix code or $L = PQ^*$ where P and Q are prefix codes.*

Proof. (1) Clearly, L is a submonoid of X^*. If $u, ux \in L$, then from $1 \equiv u(R_L)$ follows $x \equiv ux(R_L)$ and $x \in L$. Therefore, L is right unitary and the conclusion follows.

(2) Suppose that L is not a prefix code and let $P = \{u \in L \mid v \in L,\ x \in X^*,\ u = vx \Rightarrow x = 1\}$. Then P is a prefix code and $P \neq L$. Let $T = \{x \mid x \in X^*, Lx \subseteq L\}$. Clearly, T is a nonempty submonoid of X^* and, since L is not a prefix code, $T \neq \{1\}$. Let $v \in L$. Then $v = uy$ for some $u \in P$, $y \in X^*$. From $u \in L$, $uy \in L$ follows $Ly \subseteq L$ and hence $y \in T$. Therefore $L = PT$. Let $t, tz \in T$. Then $Ltz \subseteq L$ with $Lt \subseteq L$. Hence, $Lz \cap L \neq \emptyset$, $Lz \subseteq L$ and $z \in T$. The submonoid T is right unitary and therefore generated by a prefix code Q, i.e. $T = Q^*$. □

Fact 2.3. *Let L be a nonempty language with $L \subseteq X^+$. Then*

(1) *L is a prefix code if and only if every finite nonempty subset of L is rc-simple;*

(2) *L is an infix code if and only if every finite nonempty subset of L is c-simple.*

Proof. (1) If L is a prefix code, then every nonempty subset of L is also a prefix code and hence *rc*-simple. Conversely, suppose that $u, ux \in L$. Then $\{u, ux\}$ is *rc*-simple and hence a class of a right congruence. This implies that $\{u, ux, ux^2, \ldots\}$ is contained in the same class as u. This is possible only if $x = 1$. Therefore L is a prefix code.

(2) The proof is similar by replacing $\{u, ux\}$ by $\{u, xuy\}$. □

If L is a prefix code, then, since L is a union of classes of P_L and every nonempty subset of a prefix code is also a prefix code, L is a union of classes that are prefix codes. If ${}_LW$ is the left residue of L, then $W(L)$ is contained in ${}_LW$ and, if not empty, ${}_LW$ is a union of classes of P_L.

Proposition 2.4. *Let L be a prefix code. Then*

(1) *the P_L-class of 1 is $\{1\}$;*

(2) *if A is a class of P_L with $A \neq \{1\}$ and if A is not contained in ${}_LW$, then A is a prefix code.*

Proof. (1) If $u \equiv 1(P_L)$ and if $v \in L$, then $vu \equiv v(P_L)$. Hence v, $vu \in L$ and $u = 1$.

(2) Let u, $v \in A$ with $v = uz$. Then there exists $x \in X^*$ such that $xu \in L$. From $u \equiv v(P_L)$ it follows that $xu \equiv xv(P_L)$. Since $xu \in L$, $xv \in L$. Hence, $xu \in L$ and $xv = xuz \in L$. Since L is a prefix code, $z = 1$, i.e. A is a prefix code. □

Corollary 2.5. *If L is a prefix code that is left dense, then every class $\neq \{1\}$ of P_L is a prefix code and L is not a regular language.*

Corollary 2.6. *If L is a prefix code that is left dense, then the syntactic monoid $M = Syn(L)$ of L has the following properties:*

(a) *For every $u \in M$, $ux = u$ implies $x = 1$.*

(b) *M is infinite.*

(c) *No element $u \neq 1$ is periodic,*

Proof. This follows from the fact that every class of the syntactic congruence not containing the identity is a prefix code and that L is not regular. □

3. Cohesive prefix codes

A nonempty language $L \subseteq X^+$ such that $xLy \cap L \neq \emptyset$ implies $y = 1$ and $xL \subseteq L$ is a prefix code and a class of its syntactic congruence. Such a prefix code is called a *cohesive prefix code*.

Clearly, every infix code is a cohesive prefix code, but the converse is not true in general. For example, $P = a^*b$ over $X = \{a, b\}$ is a cohesive prefix code, but not an infix code. This example is a special case of the following general class of prefix codes that are cohesive prefix codes. Let $X = Y \cup Z$ with $Y \cap Z = \emptyset$ and Y, Z not empty. Then $P = Y^*Q$, where Q is an infix code over the subalphabet Z, is a cohesive prefix code in X^*. Remark that these codes are not suffix and hence not infix.

Recall [3] that a nonempty language $L \subseteq X^+$ such that $xLy \cap L \neq \emptyset$ implies $y = 1$ is a prefix code called a *p-infix code*. Clearly, every cohesive prefix code is p-infix, but the converse is not true. For example, $\{a, ba\}$ over $X = \{a, b\}$ is a p-infix code, but not a cohesive prefix code. A prefix code that is a class of its syntactic congruence is not necessarily a cohesive prefix code. For example, take $\{a^n b^n \mid n \geqslant 1\}$ over $X = \{a, b\}$.

Fact 3.1. *Let L be a finite language. Then L is a cohesive prefix code if and only if L is an infix code.*

Proof. ($\Leftarrow$) Obvious.

($\Rightarrow$) Suppose that L is not an infix code. Then there exist x, $y \in X^*$ and $u \in L$ such that $|xy| \geqslant 1$ and $xuy \in L$. By definition, $y = 1$ and $xL \subseteq L$. Remark that $|x| \geqslant 1$. For any $n \geqslant 1$, $x^n L \subseteq L$. Hence, L is infinite, a contradiction. □

A consequence of the above proposition is that every nonempty subset of a finite cohesive prefix code is also a cohesive prefix code. However, this is no more the case in general for infinite cohesive prefix codes. Take, for example, the cohesive prefix code $P=b^*a$ over $X=\{a, b\}$. The subset $b^*a\setminus\{ba\}$ is clearly not a cohesive prefix code.

Proposition 3.2. *A nonempty language $L\subseteq X^+$ is an infix code if and only if every nonempty finite subset of L is a cohesive prefix code.*

Proof. ($\Rightarrow$) Obvious.

($\Leftarrow$) Suppose that u, $xuy=v\in L$ with $x, y\in X^*$. Since the subset $A=\{u, v\}$ is a finite subset of L, A is a cohesive prefix code. Hence, $y=1$ and $xA=\{xu, x^2u\}\subseteq A$. This implies $x=1$. $\square$

Fact 3.3. *If L is a cohesive prefix code, then its residue $W(L)$ is not empty.*

Proof. Suppose that $W(L)$ is empty. Let $a\in X$, $u\in L$ and let $xuay\in L$ for some $x, y\in X^*$. Then $ay=1$, a contradiction. $\square$

If L is a cohesive prefix code, then, since L is a class of P_L, L is a disjunctive element of $Syn(L)$. Since the residue of L is not empty, $Syn(L)$ has a zero element and hence a core.

Remark that the residue of a cohesive prefix code can be strictly contained in the left residue. This is the case for $P=b^*a$ over $X=\{a, b\}$.

Proposition 3.4. *A monoid M is isomorphic to the syntactic monoid $Syn(L)$ of a cohesive prefix code L if and only if the following conditions are satisfied:*

(1) *If e is the identity element of M, then $M\setminus\{e\}$ is a subsemigroup of M.*

(2) *M has a zero element 0.*

(3) *M has a disjunctive element c such that $c\notin\{e, 0\}$ and $c=xcy$ implies $y=e$.*

Proof. ($\Rightarrow$) Since L is a cohesive prefix code, L is a P_L-class and the P_L-class of 1 is trivial. Hence, (1) holds for $Syn(L)$ and thus also for M. By Fact 3.3, the residue $W(L)$ is not empty. Hence $Syn(L)$ has a zero and (2) holds. Let c be the image of L is $Syn(L)$. Since $L\subseteq X^+$, by hypothesis, we have $c\neq e$, and since $L\cap W(L)=\emptyset$, we also have $c\neq 0$. It follows easily that c is disjunctive. For any $w\in X^*$, let $[w]$ be its image in $Syn(L)$, and let $u\in L$. Assume that for $x, y\in X^*$ we have $c=[x]c[y]$. Then u, $xuy\in L$ implies that $y=1$ whence $[y]=e$. Therefore, M satisfies (3) as well.

($\Leftarrow$) Let $X=M\setminus\{e\}$ and define a mapping φ on the free monoid X^* by

$$\varphi: x_1x_2\ldots x_n\to x_1x_2\ldots x_n\in M\setminus\{e\}$$

if $x_1, x_2, \ldots, x_n\in M\setminus\{e\}$ and

$$\varphi: 1\to e.$$

Then φ is a homomorphism of X^* onto M such that $w\varphi = e$ if and only if $w = 1$. Since c is a disjunctive element of M, for $L = c\varphi^{-1}$, we have $\bar{\varphi} = P_L$ where $\bar{\varphi}$ is the equivalence defined on X^* by $u \equiv v(\bar{\varphi})$ if and only if $\varphi(u) = \varphi(b)$. Therefore, $Syn(L) = X^*/P_L$ is isomorphic to M. Suppose that $xLy \cap L \neq \emptyset$, i.e. $xuy = v$ for some $u, v \in L$. Then $\varphi(xuy) = \varphi(x)\varphi(u)\varphi(y) = \varphi(v)$ and, since $\varphi(u) = \varphi(v) = c$, $\varphi(x)c\varphi(y) = c$ and $\varphi(y) = e$. Therefore $y = 1$, $xL \subseteq L$ and L is a cohesive prefix code. $\square$

The following lemma and its corollary will be used several times in the sequel.

Lemma 3.5. *Let $L \subseteq X^+$ be a cohesive prefix code. Then $L = S^*T$ where S is a suffix code over X or $\{1\}$, and T is an infix code over X.*

Proof. If L is an infix code, then take $S = \{1\}$ and $T = L$. Assume now that L is not an infix code. Let $T = \{u \in L \mid v \in L,\ x, y \in X^*,\ u = xvy \Rightarrow x = y = 1\}$. Then obviously T is an infix code. Now let $U = \{x \in X^* \mid xL \subseteq L\}$. Since L is a cohesive prefix code, U is a left unitary submonoid of X^* and hence its root S is a suffix code. Thus $U = S^*$ and $L = S^*T$. $\square$

It can be easily verified that the above representation for a cohesive prefix code L is uniquely determined.

Corollary 3.6. *Let $L \subseteq X^+$ be a cohesive prefix code with $L = S^*T$ where $S \neq \{1\}$, let $s \in S^+$ and let V be a suffix code with $V \subseteq L$. Then s^*V is a cohesive prefix code.*

Proof. Since $s^*V \subseteq L$, $x(s^*V)y \cap s^*V \neq \emptyset$ implies that $x \in S^*$ and $y = 1$ for any $x, y \in X^*$. Let $xs^iv = s^jv'$ where $i, j \geqslant 0$ and $v, v' \in V$. Since V is a suffix code, $v = v'$ and $xs^i = s^j$. Thus $x = s^{j-i} \in s^*$. This completes the proof of the corollary. $\square$

4. Closure properties

By $COH(X)$ we denote the family of all cohesive prefix codes over X.

The property for a language to be a prefix code is preserved under the operation of taking a nonempty subset. This is no more true for cohesive prefix codes.

Fact 4.1. *Let $L \in COH(X)$. If L is infinite and not an infix code, then there exists a subset $L' \subseteq L$ such that $L' \notin COH(X)$.*

Proof. Since L is not an infix code, there exist $u \in L$ and $x, y \in X^*$, $xy \neq 1$, such that $xuy = v \in L$. Since $L \in COH(X)$, $y = 1$ and $x^n u \in L$ for any $n \geqslant 0$. If $L' = \{u, v = xu\}$, then L' is a subset of L that is not a cohesive prefix code. $\square$

Fact 4.2. (i) *$COH(X)$ is not closed under union.*

(ii) *$COH(X)$ is closed under intersection. More precisely, let $\{L_i\}_{i\in I}$ be a family of elements of $COH(X)$. Then, if not empty, $\bigcap_{i\in I} L_i \in COH(X)$.*

(iii) *$COH(X)$ is not closed under catenation.*

(iv) *$COH(X)$ is not closed under $+$. More precisely, for any $L\in COH(X)$, $L^+\notin COH(X)$.*

Proof. (i) Let $X=\{a,b,\dots\}$, $L=\{a\}$ and $L'=\{ab\}$. Then $L, L'\in COH(X)$ but $L\cup L'\notin COH(X)$.

(ii) Let $\{L_i\}_{i\in I}$ where $L_i\in COH(X)$ for any $i\in I$ and consider $\bigcap_{i\in I} L_i$. Assume $x(\bigcap_{i\in I} L_i)y\cap(\bigcap_{i\in I} L_i)\neq\emptyset$. Then there exist $x, y\in X^*$ and $u\in\bigcap_{i\in I} L_i$ such that $xuy\in\bigcap_{i\in I} L_i$. Let $i\in I$. Then $u, xuy\in L_i$. Since $L_i\in COH(X)$, $y=1$ and $xL_i\subseteq L_i$. Therefore

$$\bigcap_{i\in I}(xL_i)=x\left(\bigcap_{i\in I} L_i\right)\subseteq\bigcap_{i\in I} L_i.$$

This means that $\bigcap_{i\in I} L_i\in COH(X)$.

(iii) Let $X=\{a,b,\dots\}$, $L=b^*a$ and let $L'=a^*b$. Then $L, L'\in COH(X)$. Consider $LL'=b^*a^+b$. Suppose $LL'\in COH(X)$. Since $ab, (ba)(ab)\in LL'$, $(ba)^n ab\in LL'$ for any $n\geqslant 0$. However this is a contradiction. Therefore $LL'\notin COH(X)$.

(iv) Immediate. □

Proposition 4.3. *Let $L\in COH(X)$. Then there exists a maximal element $L'\in COH(X)$ such that $L\subseteq L'$.*

Proof. Let $\{L_i\}_{i\in I}$ be an ascending chain in $COH(X)$ such that $L\subseteq L_i$ for any $i\in I$ and let $L'=\bigcup_{i\in I} L_i$. Suppose that $xL'y\cap L'\neq\emptyset$. Then there exist $x, y\in X^*$ and $u\in L'$ such that $xuy\in L'$. Since $\{L_i\}_{i\in I}$ is an ascending chain, there exists an element L_k of this chain such that $u, xuy\in L_k$ and hence $xL_ky\cap L_k\neq\emptyset$. Since $L_k\in COH(X)$, $y=1$ and $xL_k\subseteq L_k$.

Now we prove that $xL'\subseteq L'$. Remark that if $k\leqslant i$, then $xL_i\subseteq L_i$ because $u, xuy\in L_k\subseteq L_i$. Furthermore,

$$\bigcup_{i\in I}(xL_i)=x\left(\bigcup_{i\in I} L_i\right)\subseteq\bigcup_{i\in I} L_i,\text{ i.e. } xL'\subseteq L'.$$

Therefore $\{L_i\}_{i\in I}\in COH(X)$. Using the Zorn's lemma, it follows then that L is contained in a maximal cohesive prefix code. □

Fact 4.4. *If L is a finite maximal infix code, then L is a maximal cohesive prefix code.*

Proof. Suppose that L is not a maximal cohesive prefix code. Then there exists a maximal cohesive prefix code $\tilde{L}$ such that $L\subset\tilde{L}$. Since L is a maximal infix code, there exist $u, xuy\in\tilde{L}$ where $x, y\in X^*$ and $xy\neq 1$. Since $\tilde{L}$ is a cohesive prefix code, $y=1$,

$x \neq 1$ and $x^n u \in \tilde{L}$ for any $n \geq 0$. Let $m = \max\{|v| \mid v \in L\}$ and let $n > m$. Since $|x^n| > m$ and L is a maximal infix code, there exist r, $s \in X^*$ and $v \in L$ such that $rs \neq 1$ and $rvs = x^n$. Consequently, $x^n u = rv(su) \in \tilde{L}$. This contradicts the fact that $\tilde{L}$ is a cohesive prefix code because $v, rv(su) \in \tilde{L}$ and $su \neq 1$. Hence, L must be a maximal cohesive prefix code. □

Corollary 4.5. *Let $L \subseteq X^*$ be a finite language. Then L is maximal cohesive prefix code if and only if L is a maximal infix code.*

The following example shows the existence of an infinite maximal infix code that is a maximal cohesive prefix code.

Example 4.6. Let $X = \{a, b, \ldots\}$ and $L = ab^+a \cup bab \cup a^2 \cup (X \setminus \{a, b\})$. Then it is easy to see that L is a maximal infix code. Now suppose L is not a maximal cohesive prefix code. Let $\tilde{L}$ be a maximal cohesive prefix code with $L \subset \tilde{L}$. Since $L \neq \tilde{L}$, there exist $u, xu \in X^+$ such that $u \in L$, $xu \in \tilde{L}$, $x^* \tilde{L} \subseteq \tilde{L}$ and $x \in X^+$. If $x = x'a$, then $xa^2 = x'a^3 = x'(a^2)a \in \tilde{L}$, a contradiction. If $x = x'b$, then $xaba = x'(bab)a \in \tilde{L}$, a contradiction. If $x = x'c$ where $c \in X \setminus \{a, b\}$, then $xa^2 = x'(c)a^2 \in \tilde{L}$, a contradiction. Therefore, L must be a maximal cohesive prefix code.

In the above example, L is a regular language. This suggests the following result that is a generalization of Fact 4.4.

Proposition 4.7. *Let $|X| \geq 2$ and let $L \subseteq X^*$ be a regular language. If L is a maximal infix code, then L is a maximal cohesive prefix code.*

Proof. Suppose that L is not a maximal cohesive prefix code. Then, by Corollary 3.6, there exists $s \in X^+$ such that s^*L is a cohesive prefix code.

First we prove the existence of $\alpha \in X^+$ satisfying the following condition: For any $i, i \geq 1$ there exists $\beta_i \in X^+$ such that $\alpha s^i \beta_i \in L$.

Let $k \geq 1$. Since L is a maximal infix code, we have $u \leq_i s^{k+2}$ or $s^{k+2} \leq_i u$ for some $u \in L$ where $u \leq_i v$ means that $v = xuy$ for some $x, y \in X^*$.

If $u \leq_i s^{k+2}$, then we have a contradiction with the assumption that s^*L is a cohesive prefix code. If $s^{k+2} \leq_i u$, then there exist $\alpha'_k, \beta'_k \in X^*$ such that $\alpha'_k s^{k+2} \beta'_k \in L$. Now let $\alpha_k = \alpha'_k s$ and $\beta_k = s\beta'_k$. Then $\alpha_k s^k \beta_k \in L$ and $\alpha_k, \beta_k \in X^+$. Since L is a regular language, we can assume that $|\alpha_k| \leq N$ for some positive integer N (for instance, we can take for N the number of states of an automaton accepting L), without loss of generality. Remark that we can take infinitely many numbers as k. Moreover, because of the restriction of the length of α_k, we can see the following: There exist $\alpha \in X^+$ with $|\alpha| \leq N$, an infinite sequence of positive integers $k_1 < k_2 < \cdots < k_r < \cdots$ and $\beta_{k_r} \in X^+$, $r \geq 1$ such that $\alpha s^{k_r} \beta_{k_r} \in L$.

Now let $i, i \geq 1$. Take any k_r with $k_r > i$. Then $\alpha s^{k_r} \beta_{k_r} = \alpha s^i (s^{k_r - i} \beta_{k_r}) \in L$. Put $\beta_i = s^{k_r - i} \beta_{k_r}$. Hence we have $\alpha s^i \beta_i \in L$.

Let p be a positive integer. Assume now that, for n, $1\leqslant n\leqslant p$, we have a sequence $\alpha_1, \alpha_2, \ldots, \alpha_n$ of elements in X^+ satisfying the following condition:

$$\forall i,\ i\geqslant 1,\ \exists\beta_{i_n}\in X^+ \text{ such that } \alpha_n\alpha_{n-1}\ldots\alpha_2\alpha_1 s^i\beta_{i_n}\in L.$$

Let $k\geqslant 1$. Consider $s\alpha_p\alpha_{p-1}\cdots\alpha_2\alpha_1 s^k\in X^+$. Then there exists $u_k\in L$ such that

$$u_k \leqslant_i s\alpha_p\alpha_{p-1}\cdots\alpha_2\alpha_1 s^k \quad\text{or}\quad s\alpha_p\alpha_{p-1}\cdots\alpha_2\alpha_1 s^k \leqslant_i u_k.$$

If $u_k \leqslant_i s\alpha_p\alpha_{p-1}\cdots\alpha_2\alpha_1 s^k$, then $\alpha u_k\beta = s\alpha_p\alpha_{p-1}\cdots\alpha_2\alpha_1 s^k$ for some $\alpha, \beta\in X^*$. Remark that

$$(\alpha u_k\beta)\beta_{kp} = s(\alpha_p\alpha_{p-1}\cdots\alpha_2\alpha_1 s^k\beta_{kp})\in sL.$$

Since $|\beta_{kp}|\geqslant 1$, this contradicts the assumption that s^*L is a cohesive prefix code. Therefore $s\alpha_p\alpha_{p-1}\cdots\alpha_2\alpha_1 s^k \leqslant_i u_k$. In this case, $\alpha'_k(s\alpha_p\alpha_{p-1}\cdots\alpha_2\alpha_1 s^k)\beta'_k = u_k\in L$ for some $\alpha'_k, \beta'_k\in X^*$. For any n, $1\leqslant n\leqslant p$, we can assume that $|\alpha_n|\leqslant N$, without loss of generality. By the above remark, there exist $\alpha_{p+1}\in X^+$ with $|\alpha_{p+1}|\leqslant N$, an infinite sequence of positive integers $k_1<k_2<\cdots<k_r<\cdots$ and $\beta'_{k_r}\in X^*$, $r\geqslant 1$ such that $\alpha_{p+1}\alpha_p\cdots\alpha_2\alpha_1 s^{k_r}\beta'_{k_r}\in L$. Let $i\geqslant 1$ and $k_r>i$. Then $\alpha_{p+1}\alpha_p\cdots\alpha_2\alpha_1 s^i(s^{k_r-i}\beta'_{k_r})\in L$. Put $\beta_{ip+1}=s^{k_r-i}\beta'_{k_r}$. Then $\beta_{ip+1}\in X^+$ and $\alpha_{p+1}\alpha_p\cdots\alpha_2\alpha_1 s^i\beta_{ip+1}\in L$. By induction, we have the following result.

There exists an infinite sequence $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ of elements in X^+ such that $\alpha_n\alpha_{n-1}\cdots\alpha_2\alpha_1 s^i\beta_{in}\in L$ for any i, j, $i, j\geqslant 1$ where $\beta_{in}\in X^+$. As a special case, we consider the case $i=1$. Then we have

$$\alpha_1 s\beta_{11}\in L,\ \alpha_2\alpha_1 s\beta_{12}\in L, \ldots, \alpha_n\alpha_{n-1}\cdots\alpha_2\alpha_1 s\beta_{1n}\in L, \ldots$$

Since we can assume that each β_{1n} satisfies the condition $|\beta_{1n}|\leqslant N$, there exist $p, q\geqslant 1$, $p\neq q$, such that $\beta_{1p}=\beta_{1q}=\beta\in X^+$. In this case, $\alpha_p\alpha_{p-1}\cdots\alpha_2\alpha_1 s\beta\in L$ and $\alpha_q\alpha_{q-1}\cdots\alpha_2\alpha_1 s\beta\in L$. This contradicts the assumption that L is an infix code.

Therefore, L must be a maximal cohesive prefix code. □

Unlike the case of finite maximal infix codes, the converse of the above proposition does not hold true. Let $X=\{a, b\}$ and let $L=a^*b$. Then L is a regular language that is a maximal cohesive prefix code, but it is not an infix code. □

Proposition 4.8. *Let $L\subseteq X^*$ be a finite infix code. Then there exists a finite maximal infix code $\tilde{L}$ such that $L\subseteq\tilde{L}$.*

Proof. Let $n=\max\{|u|\mid u\in L\}$ and let $X_n=\{v\in X^*\mid |v|\leqslant n\}$. Since X_n is finite, there exists an infix code $\tilde{L}$ with $L\subseteq\tilde{L}\subseteq X_n$ such that, for any $u\in X_n$, $u\leqslant_i\tilde{u}$ or $\tilde{u}\leqslant_i u$ for some $\tilde{u}\in\tilde{L}$.

We show that $\tilde{L}$ is a maximal infix code. Let $w\in X^*$. If $|w|\leqslant n$, then by the definition of $\tilde{L}$ there exists $\tilde{w}\in\tilde{L}$ such that $w\leqslant_i w$ or $\tilde{w}\leqslant_i w$. Now let $|w|>n$. Then $w=w'w''$ where $|w'|=n$ and $w''\in X^+$. Since $w'\in X_n$, there exists $\tilde{w}'\in\tilde{L}$ such that $w'\leqslant_i\tilde{w}'$ or $\tilde{w}'\leqslant_i w'$. However, by $|w'|=n$, $\tilde{w}'\leqslant_i w'$. Hence $\tilde{w}'\leqslant_i w'\leqslant_i w'w''=w$. Both cases indicate that $w \leqslant_i \tilde{w}$ or $\tilde{w} \leqslant_i w$ for some $\tilde{w} \in \tilde{L}$. This means that $\tilde{L}$ is a finite maximal infix code such that $L \subseteq \tilde{L}$. $\square$

Corollary 4.9. *If L is a finite cohesive prefix code, then there exists a finite maximal cohesive prefix code $\tilde{L}$ such that $L \subseteq \tilde{L}$.*

Remark. If L is a finite prefix code over an alphabet X with $|X| \geqslant 2$ and if L is not a maximal prefix code, then there exist both a finite and an infinite maximal prefix code containing L. However, this is no more the case for infix codes and hence for cohesive prefix codes. For example, let $X = \{a_1, a_2, \ldots, a_r\}$ and let $L = \{a_i a_j \mid i \neq j, 1 \leqslant i, j \leqslant n\}$. Then L is an infix code. Moreover, if $\tilde{L}$ is a maximal infix code such that $L \subseteq \tilde{L}$, then $\tilde{L}$ is represented as $\tilde{L} = L \cup \{a_i^{n_i} \mid n_i \geqslant 2, 1 \leqslant i \leqslant r\}$. Hence $|\tilde{L}| = |L| + r$.

From the preceding results, one may guess that every maximal infix code is a maximal cohesive prefix code. However, this is not the case in general.

Proposition 4.10. *Let $|X| \geqslant 2$. Then there exists a maximal infix code $L \subseteq X^*$ such that L is not a maximal cohesive prefix code.*

Proof. First we consider the case $|X| = 2$. Let $X^+ = \{w_1, w_2, \ldots\}$ and let $L' = \{\bar{a}_i^{|w_i|+1} a_i^{|w_i|+1} \bar{a}_i w_i \bar{b}_i b_i^{|w_i|+1} \bar{b}_i^{|w_i|+1} \mid i \geqslant 1\}$ where $a_i, b_i \in X$, $w_i \in a_i X^* \cap X^* b_i$, $\{\bar{a}_i\} = X \setminus \{a_i\}$, $\{\bar{b}_i\} = X \setminus \{b_i\}$ for $i \geqslant 1$. We prove that L' contains a maximal infix code.

Let $\{L_j\}_{j \in J}$ be an ascending chain of infix codes such that $L_j \subseteq L'$ for every $j \in J$. Let $\bar{L} = \bigcup_{j \in J} L_j$. Then $\bar{L}$ is an infix code and $\bar{L} \subseteq L'$. Therefore, by the Zorn's lemma, there exists an infix code $L \subseteq L'$ satisfying the following condition:

($*$) For any $u \in L' \setminus L$, $L \cup \{u\}$ is not an infix code.

Now we show that L is a maximal infix code. Let $w \in X^+$. Then there exists $p \geqslant 1$ such that $w = w_p$. Consider the word $\alpha_p = \bar{a}_p^{|w_p|+1} \bar{a}_p^{|w_p|+1} \bar{a}_p w_p \bar{b}_p \bar{b}_p^{|w_p|+1} \bar{b}_p^{|w_p|+1} \in L'$. If $\alpha_p \in L$, then $w = w_p \leqslant_i \alpha_p \in L$. Let $\alpha_p \notin L$. There are two cases.

Case 1: $\alpha_p \leqslant_i \alpha_q$ where $\alpha_q = a_q^{|w_q|+1} a_q^{|w_q|+1} \bar{a}_q w_q \bar{b}_q b_q^{|w_q|+1} \bar{b}_q^{|w_q|+1} \in L$. Then obviously $w = w_p \leqslant_i \alpha_q \in L$.

Case 2: $\alpha_q \in L$ and $\alpha_q \leqslant_i \alpha_p$. We will show that $\alpha_q \leqslant_i w_p = w$. Since $\alpha_q \leqslant_i \alpha_p$, $\alpha_p = x \alpha_p y$ for some $x, y \in X^*$. Suppose that $|y| < 2|w_p| + 3$. If $|y| = 0$, then $p = q$. This contradicts the assumption that $\alpha_p \notin L$. Therefore $|y| \neq 0$.

Case (i): $1 \leqslant |y| < |w_p| + 1$. In this case

$$\bar{a}_p^{|w_p|+1} a_p^{|w_p|+1} \bar{a}_p w_p \bar{b}_p b_p^{|w_p|+1} \bar{b}_p^{|w_p|+1-|y|} = \cdots \bar{b}_q b_q^{|w_q|+1} \bar{b}_q^{|w_q|+1},$$

where $b_p = b_q$, $\bar{b}_p = \bar{b}_q$ and $|w_q| < |w_p|$, a contradiction.

Case (ii): $|w_p| + 1 \leqslant |y| < 2|w_p| + 2$. In this case

$$\bar{a}_p^{|w_p|+1} a_p^{|w_p|+1} \bar{a}_p w_p \bar{b}_p b_p^{2|w_p|+2-|y|} = \cdots \bar{b}_q b_q^{|w_q|+1} \bar{b}_q^{|w_q|+1},$$

where $b_p = \bar{b}_q$, $\bar{b}_p = b_q$, $w_p \in X^* b_p$ and $|w_q| + 1 > 1$, a contradiction.

Case (iii): $|y|=2|w_p|+2$. In this case

$$\bar{a}_p^{|w_p|+1}a_p^{|w_p|+1}\bar{a}_p w_p \bar{b}_p = \cdots \bar{b}_q b_q^{|w_q|+1}\bar{b}_q^{|w_q|+1},$$

where $b_q=b_q$, $\bar{b}_p=\bar{b}_q$, $w_p\in X^*b_p$ and $|w_q|+1>1$, a contradiction.

Consequently $|y|\geqslant 2|w_p|+3$. By symmetry, $|x|\geqslant 2|w_p|+3$. This means that $\alpha_q \leqslant_i w_p = w$ with $\alpha_q\in L$.

Hence, all cases indicate that for any $u\in X^+$ there exists $v\in L$ such that $u\leqslant_i v$ or $v\leqslant_i u$. This means that L is a maximal infix code.

Let $\tilde{L}=(ab)^*L$ where $X=\{a,b\}$. Obviously $L\subset\tilde{L}$. We show that $\tilde{L}$ is a cohesive prefix code. If not, then there exist $x\in X^*$, $y\in X^+$, $\alpha_p,\alpha_q\in L$ such that $(ab)^r\alpha_p=x\alpha_q y$ where $\alpha_p=\bar{a}_p^{|w_p|+1}a_p^{|w_p|+1}\bar{a}_p w_p\bar{b}_p b_p^{|w_p|+1}\bar{b}_p^{|w_p|+1}$ and $\alpha_q=\bar{a}_q^{|w_q|+1}a_q^{|w_q|+1}\bar{a}_q w_q\bar{b}_q b_q^{|w_q|+1}\bar{b}_q^{|w_q|+1}$. Suppose $y\neq 1$. By the same considerations as above, we have $\alpha_q\leqslant_i w_p\leqslant_i \alpha_p$. However, this contradicts the fact that L is an infix code. Hence

$$x\tilde{L}y\cap\tilde{L}\neq\emptyset \Leftrightarrow x\in(ab)^*,\quad y=1$$

and $\tilde{L}$ is a cohesive prefix code. This completes the proof of the proposition for the case $|X|=2$.

Now let $|X|>2$. Let $X=Y\cup Z$ where $|Y|=2$ and $Y\cap Z=\emptyset$. Let $L_Y\subseteq Y^*$ be a maximal infix code over Y that is not a maximal cohesive prefix code over Y. Moreover, let $\tilde{L}_Y$ be a maximal cohesive prefix code over Y such that $L_Y\subset\tilde{L}_Y$. Let $L=L_Y\cup Z$. It is easy to see that L is a maximal infix code over X. Let $\tilde{L}=\tilde{L}_Y\cup\{x\in Y^+\mid x\tilde{L}_Y\subseteq\tilde{L}_Y\}Z$. Then $\tilde{L}$ is a cohesive prefix code, in fact, a maximal cohesive prefix code over X. Obviously $\tilde{L}\supset L$. This completes the proof of the proposition. □

In the above proposition, we established the existence of an infinite maximal infix code that is not a maximal cohesive prefix code. As it has already been shown, all regular maximal infix codes are maximal cohesive prefix codes.

Now we want to know whether there exists a nonregular maximal infix code that is a maximal cohesive prefix code. In order to do that, we will need to establish some properties of reflective codes and maximal reflective codes.

Definition. For every $u\in X^*$ and $L\subseteq X^*$, let

$$Ref(u)=\{wv\mid v,w\in X^*,\ u=vw\}\quad\text{and}\quad Ref(L)=\bigcup_{u\in L}Ref(u).$$

A language $L\subseteq X^*$ is said to be *reflective* if $L=Ref(L)$. A code is called *reflective* if it is a reflective language. By [4], every reflective code is an infix code.

Lemma 4.11. *Let $L\subseteq X^*$ be a reflective code. Then there exists a maximal reflective code $\tilde{L}$ with $L\subseteq\tilde{L}$.*

Proof. Immediate by the Zorn's lemma. □

Proposition 4.12. *Let $L \subseteq X^*$ be a maximal reflective code and let L' be a cohesive prefix code with $L \subseteq L'$. Then L' is an infix code.*

Proof. Let $L' = S^*T$ where $S = \{1\}$ or is a suffix code and T is an infix code. Suppose that $S \neq \{1\}$ and let $s \in S$. Since L' is a cohesive prefix code and L is an infix code with $L \subseteq L'$, by Corollary 3.6, s^*L is a cohesive prefix code. First, notice that $Ref(s) \cap L = \emptyset$. It is obvious that $L \cup Ref(s)$ is a reflective language. Since L is a maximal reflective code, $L \cup Ref(s)$ is not an infix code. Hence, for some $u \in L$, $u \leqslant_i s''s'$ or $s''s' \neq u$ and $s''s' \leqslant_i u$ where $s = s's''$, $s', s'' \in X^*$. If $u \leqslant_i s''s'$, then $s^2 = s'(s''s')s''$ contains u as a subword, a contradiction. If $s''s' \neq u$ and $s''s' \leqslant_i u$, then there exist $x, y \in X^*$ such that $xy \in X^+$ and $xs''s'y = u$. Since L is reflective, $s''s'yx \in L$ and $s'yxs'' \in L$. Consider $s(s'yxs'') \in sL$. Then $s(s'yxs'') = s'(s''s'yx)s''$. If $s'' \in X^+$, then $s'Ls'' \cap sL \neq \emptyset$, a contradiction. If $s'' = 1$, then $s''s' = s$ and $xsy = u$. Moreover, $yxs \in L$ and $syx \in L$. Hence, $s(yxs) = (syx)s$, i.e. $sL \cap Ls \neq \emptyset$, a contradiction. Consequently, $S = \{1\}$ and $L' = T$. This completes the proof of the proposition. □

Corollary 4.13. *Let $L \subseteq X^*$ be a reflective code. Then there exists a maximal infix code $\tilde{L} \subseteq X^*$ with $L \subseteq \tilde{L}$ that is a maximal cohesive prefix code.*

Proof. By Lemma 4.11, there exists a maximal reflective code $\bar{L}$ such that $L \subseteq \bar{L}$. Let $\tilde{L}$ be a maximal cohesive prefix code with $\bar{L} \subseteq \tilde{L}$. By the proposition, $\tilde{L}$ is an infix code. The maximality of $\tilde{L}$ as a cohesive prefix code implies that $\tilde{L}$ is a maximal infix code. This completes the proof of the corollary. □

Proposition 4.14. *Let $|X| \geqslant 2$. Then there exists a maximal infix code that is not a regular language, but a maximal cohesive prefix code.*

Proof. Let $X = \{a, b, \ldots\}$ and let $L = Ref(\{ab^nab^n \mid n \geqslant 1\})$. Then L is an infinite reflective code. By the above corollary, there exists a maximal infix code $\tilde{L}$ such that $L \subseteq \tilde{L}$ and $\tilde{L}$ is a maximal cohesive prefix code. To complete the proof of the proposition, we must show that $\tilde{L}$ is not regular. Suppose that $\tilde{L}$ is regular. Since $ab^nab^n \in \tilde{L}$ for $n \geqslant 1$, by a pumping lemma for regular languages, follows the existence of k, $k \geqslant 1$ such that $ab^nab^{n+ki} \in \tilde{L}$ for any i, $i \geqslant 1$. This contradicts the fact that $\tilde{L}$ is an infix code. Hence, $\tilde{L}$ is not regular. □

5. Relations between right semaphore codes and cohesive prefix codes

Recall [1, 2] that a *right semaphore code* P is a prefix code such that for every $u \in P$, $x \in X^*$ there exist $v \in P$, $y \in X^*$ such that $xu = vy$. Let $|X| \geqslant 2$. By $RSC(X)$ we denote the class of all right semaphore codes over X. In general, there is no inclusion relation between $COH(X)$ and $RSC(X)$.

Example 5.1. Let $X=\{a, b, \dots\}$ and let $L=\{aaa, bbb, aab, bba, ab, ba\}\cup(X\setminus\{a, b\})\cup\{a, b\}\ (X\setminus\{a, b\})$. Then L is a right semaphore code, but not a cohesive prefix code, i.e. $RSC(X)\setminus COH(X)\neq\emptyset$.

Example 5.2. *Let* $X=\{a, b, \dots\}$ and let $L=\{a, bb\}$. Since L is an infix code, L is a cohesive prefix code. However, L is not a right semaphore code, i.e. $COH(X)\setminus RSC(X)\neq\emptyset$.

We are now interested in the class $RSC(X)\cap COH(X)$. First consider the case where L is an infix code.

Fact 5.3. *Let* $|X|\geqslant 2$ *and let* $L\subseteq X^*$ *be an infix code. Then* $L\in RSC(X)\cap COH(X)$ *if and only if* $L=X^n$ *for some* $n\geqslant 1$.

Proof. This follows immediately from the fact that if B is a right semaphore code that is a biprefix code, i.e. prefix and suffix code, then $B=X^n$ for some $n\geqslant 1$ [1]. □

Now let $L\subseteq X^*$ be a language that is not an infix code. If $L\in COH(X)$ then $L=S^*T$ where S is a suffix code and T is an infix code. Let $Y=X\cap S$ and let $Z=X\setminus Y$.

Lemma 5.4. *Let* $L\subseteq X^*$ *be a language such that* $L\in COH(X)\cap RSC(X)$ *and assume that L is not an infix code. Then*

(i) $\emptyset\neq Y\neq X$.
(ii) *There exists* $n\geqslant 1$ *such that for any* $b\in Z$ *we have* $b^n\in T$.
(iii) $S=Y$.
(iv) $T\cap YX^*=\emptyset$.

Proof. (i) If $Y=\emptyset$, then $a\notin S$ for any $a\in X$. Let $\tilde{u}\in L$ with $|\tilde{u}|=\min\{|u|\mid u\in L\}$. Let $a\in X$ and let $\tilde{u}\in\tilde{u}'X$. Since $a\tilde{u}\in aL$ and $L\in RSC(X)$, $a\tilde{u}\in LX^*$. From the minimality of $|\tilde{u}|$, it follows that $a\tilde{u}'\in L$ or $a\tilde{u}\in L$. If $a\tilde{u}\in L$, then $a\in S$, a contradiction. Hence $a\tilde{u}'\in L$. Consequently, $X\tilde{u}'\subseteq L$. Applying the same process to elements of $X\tilde{u}'$ and by induction we have $L=X^{|\tilde{u}|}$. This contradicts the assumption that L is not an infix code. Therefore $\emptyset\neq Y$.

Now suppose $Y=X$. Let $u\in S^*T$ and let $t\in T$. Then $ut\in uT\subseteq S^*T=L$, i.e. $u, ut\in L$. This contradicts the assumption that L is a cohesive prefix code. Thus $Y\neq X$.

(ii) Let $\tilde{u}\in L$ with $|\tilde{u}|=n=\min\{|u|\mid u\in L\}$. Since $b\tilde{u}\in bL\subseteq LX^*$, $b\notin S$ and $L\in RSC(X)$, $b\tilde{u}'\in L$ where $\tilde{u}\in\tilde{u}'X$. Now we apply the same procedure for $b\tilde{u}'\in L$ and get $b^2\tilde{u}''\in L$ where $\tilde{u}'\in\tilde{u}''X$. Continuing this process, we have $b^n\in L$. Moreover, by the minimality of $|\tilde{u}|$, $b^n\in T$.

(iii) Suppose that there exists $s\in S$ such that $|s|\geqslant 2$.

Case 1: $s=s'b$, $b\in Y$. Since $b\in S$ and S is a suffix code, this case does not occur.

Case 2: $s=s'b, b\in Z$. Since $b^n\in T$, $sb^n\in T, sb^n\in ST\subseteq L$. On the other hand, $sb^n=s'b^{n+1}=s'(b^n)b\in L$. Together with $b^n\in L$, this yields a contradiction because L is a cohesive prefix code. Therefore $S=Y$.

(iv) Suppose there exists $y \in Y$ such that $yu \in T$ for some $u \in X^*$. Let $t \in T$. Since $ut \in uT \subseteq uS^*T = uL \subseteq LX^*$, $ut = vt'x$ for some $v \in S^*$, $t' \in T$ and $x \in X^*$. We have the following three cases.

Case 1: $|t'x| < |t|$. Then $t \neq t'$ and $t' \leqslant_i t$. This contradicts the fact that T is an infix code.

Case 2: $|x| \leqslant |t| \leqslant |t'x|$. Let $u = vv'$, $t' = v'\tilde{t}$ and $t = \tilde{t}x$. Then $yu\tilde{t} = (yv)(v'\tilde{t}) = (yv)t' \in S^*T = L$. On the other hand, $yu\tilde{t} \in T\tilde{t} \subseteq L\tilde{t}$. Since L is a cohesive prefix code, $\tilde{t} = 1$. Therefore, $yu \in T$ and $u = vt' \in vT$. This contradicts the fact that T is an infix code.

Case 3: $|x| > |t|$. In this case, $vt' \leqslant_p u$ and $yut' \leqslant_p yu$, i.e. $vt'z = u$ and $yvt'z = yu \in T$ for some $z \in X^*$. However, this case yield a contradiction because T is an infix code.

Therefore $T \cap YX^* = \emptyset$ □

Proposition 5.5. *Let $|X| \geqslant 2$ and let $L \subseteq X^*$ be a language that is not an infix code. Then $L \in COH(X) \cap RSC(X)$ if and only if L can be represented in the following way:*

$$L = Y^*T \quad \textit{where } Y, Z \subseteq X,\ Y, Z \neq \emptyset,\ Y \cap Z = \emptyset,\ X = Y \cup Z,$$

with T an infix code such that $T = \bigcup_{z \in Z} zT_z$ and T_z either $\{1\}$ or a maximal prefix code.

Proof. ($\Rightarrow$) Let Y, Z, T be the sets defined in the previous lemma. From $S = Y$, it follows then that $L = Y^*T$ with T an infix code. Since $T \cap YX^* = \emptyset$, we can express T as $T = \bigcup_{z \in Z} zT_z$. We prove now that T_z is either $\{1\}$ or a maximal prefix code. First suppose that T_z is neither a prefix code nor $\{1\}$. Then there exist u $ux \in T_z$ with $x \neq 1$. In this case, zu, $zux, \in zT_z \subseteq T$. This contradicts the fact that T is an infix code. Now we show that T_z is a maximal prefix code if $T_z \neq \{1\}$. Let $w \in X^*$. Then $zwL \subseteq zT_zX^*$ and $wL \subseteq T_zX^*$. Hence, $wX^* \cap T_zX^* \neq \emptyset$ and T_z is a maximal prefix code.

($\Leftarrow$) Let $L = Y^*T$. First we prove that L is a cohesive prefix code. Assume $y_1y_2 \cdots y_mt = \alpha y_1'y_2' \cdots y_n't'\beta$ for some $\alpha, \beta \in X^*$, $y_i, y_j' \in Y$, $0 \leqslant i \leqslant m$, $0 \leqslant j \leqslant n$ and $t, t' \in T$. If $\alpha \in Y^*$, then $t = t'\beta$. Since T is an infix code, $\beta = 1$. If $\alpha = yzr$ with $y \in Y^*$, $z \in Z$ and $r \in X^*$, then in this case, $t = zry_1'y_2' \cdots y_n't'\beta$ and again $\beta = 1$. Now it is obvious that $\alpha \in Y^*$ and that $\alpha L \subseteq L$ for any $\alpha \in Y^*$. Therefore, L is a cohesive prefix code.

Now we prove that L is a right semaphore code. Let $a \in X$. If $a \in Y$, then $aL = aY^*T \subseteq Y^*T = L$. If $a = z \in Z$, then for any $y \in Y^*$ and $t \in T$ we have $t \in z'T_{z'}$ with $z' \in Z$ and $ayt = zyz't'$ where $t' \in T_{z'}$ and $z't' \in T$. Since T_z is a maximal prefix code, there exists $\tilde{y} \in T_z$ such that $\tilde{y} \leqslant_p yz't'$ or $yz't' \leqslant_p \tilde{y}$. However, $yz't' \leqslant_p \tilde{y}$ cannot occur because in this case $zyz't' \leqslant_p z\tilde{y} \in T$ and $T \ni z't' \leqslant_i zy(z't') \leqslant_i z\tilde{y} \in T$, a contradiction. Consequently, we have $\tilde{y} \leqslant_p yz't'$. In this case, $z\tilde{y} \leqslant_p zyz't'$ and $ayt = zyz't' \in z\tilde{y}X^* \subseteq TX^* \subseteq LX^*$. In any case, we have proved that $LX \subseteq LX^*$. This means that L is a right semaphore code.

Hence $L \in COH(X) \cap RSC(X)$. □

Corollary 5.6. *Let $|X| = 2$.* If *$L \subseteq X^*$ is not an infix code, then $L \in COH(X) \cap RSC(X)$ if and only if $L = a^*bT'$ where $a, b \in X$, $a \neq b$, T' is a maximal prefix code or $T' = \{1\}$ and bT' is an infix code.*

Example 5.7. Let $X=\{a, b, c\}$. Then $L=a^*\{b, c\}\{a, b, c\}^2 \in COH(X) \cap RSC(X)$. For, in this case, $Y=\{a\}$, $Z=\{b, c\}$, $T=\{b, c\}X^2$ and $T_b=T_c=X^2$ satisfy the conditions of Proposition 5.5.

Example 5.8. Let $X=\{a, b, c\}$ and let $L=a^*(b, c\}(bX^2 \cup cX^3)$. Then $L \notin COH(X) \cap RSC(X)$ because $T=\{b, c\}(bX^2 \cup cX^3)$ is not an infix code.

Example 5.9. Let $X=\{a, b\}$ and let $L=a^*ba^*b$. Then $L \in COH(X) \cap RSC(X)$ because $T'=a^*b$ is a maximal prefix code and $bT'=ba^*b$ is an infix code.

By $MPC(X)$ we denote the class of all maximal prefix codes over X.

Proposition 5.10. $RSC(X) \cap COH(X)=MPC(X) \cap COH(X)$.

Proof. Since $RSC(X) \subseteq MPC(X)$,

$$RSC(X) \cap COH(X) \subseteq MPC(X) \cap COH(X).$$

Let $L \in MPC(X) \cap COH(X)$, let $x \in X^*$ and let $u \in L$. Since $L \in MPC(X)$, there exists $v \in L$ such that $xu \leqslant_p v$ or $v \leqslant_p xu$. If $xu \leqslant_p v$, then there exists $y \in X^*$ such that $xuy=v$. This means that $xLy \cap L \neq \emptyset$. Hence $y=1$. Therefore $xu \in L$. On the other hand, if $v \leqslant_p xu$, then obviously $xu \in LX^*$. In any case, $xu \in LX^*$, i.e. $xL \subseteq LX^*$. This means that $L \in RSC(X)$, i.e. $MPC(X) \cap COH(X) \subseteq RSC(X) \cap COH(X)$. This completes the proof of the proposition. $\square$

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, H. Jürgensen, H.J. Shyr and G. Thierrin, Outfix and infix codes and related classes of languages, *J. Comput. System Sci.* **43** (1991) 484–508.

[4] C.M. Reis and G. Thierrin, Reflective star languages and codes, *Inform. and Control* **42** (1979) 1–9.

[5] H.J. Shyr, *Free Monoids and Languages*, Leture Notes (Hon Min Book Company, Taichung, 1991).

[6] G. Thierrin, *Décomposition des Langages Réguliers*, Revue d'Informatique et de Recherche Opérationnelle, No. 3 (1969) 45–50.