レター　Letter

Some families of codes under the pure codes and the bifix codes

Tetsuo Moriya* and Itaru Kataoka*

Abstract: In this paper, we investigate the inclusion relation among familes of codes under the pure codes and the bifix codes. The family of pure codes and the family of bifix codes are incomparable, and both families properly include the d-codes, the infix codes, the intercodes, and the comma-free codes.

1. Preliminaries

Let Σ be an alphabet. Σ^*denotes the free moniod generated by Σ, that is, the set of all finite words over , Σ including the empty word ε, and $+ = \Sigma^* - \varepsilon$. For w in Σ^*, $|w|$ denotes the length of w. Any subset of Σ^* is called a *language* over Σ. A language L is a *code* if X freely generates the submonoid L^* of Σ^*. ([1], [2])

A nonempty word u is called a *primitive word* if $u = f^n$, for some $f \in \Sigma^+$, and some $n \geq 1$ always implies that $n = 1$. Let Q be the set of all primitive words over Σ. For a word $w \in \Sigma^+$, there exist a unique primitive word f and a uique integer $i \geq 1$ such that $w = f^i$. Let $f = \sqrt{w}$ and call f the root of w.

For a given word $x \in \Sigma^+$, we define the following sets.

$p(x) = \{y \in \Sigma^+ | x \in y\Sigma^*\}$, $s(x) = \{y \in \Sigma^+ | x \in \Sigma^* y\}$.

For a language $L \subseteq \Sigma^*$, $P(L) = \cup_{x \in L}\, p(x)$, and $S(L) = \cup_{x \in L}\, s(x)$.

A language L is a pure code if it is a code such that, for any $x \in L^*, \sqrt{x} \in L^*$. A nonempty word u is a *non-overlapping word* if $u = vx = yv$ for some $x, y \in \Sigma^+$ always implies that $v = \varepsilon$. Let $D(1)$ be the set of all non-overlapping words over Σ. A word in $D(1)$ is also called a *d-primitive word*.

A language $L \subseteq \Sigma^+$ is a prefix code (suffix code) if the condition $L \cap L\Sigma^+ = \phi$ ($L \cap \Sigma^+ L = \phi$) is true. L is a bifix code if L is both a prefix code and also a suffix code. L is an infix code if, for all $x, y, u \in \Sigma^*$, $u \in L$ and $xuy \in L$ together imply $x = y = \varepsilon$. L is an intercode if $L^{m+1} \cap \Sigma^+ L^m \Sigma^+ = \phi$ for some $m \geq 1$. The integer m is called the index of L. An intercode of index 1 is called a comma-free code. L is a d-code if L is a bifix code and $P(L) \cap S(L) = L$.

2. Inclusion relations

Proposition 1 *([3]) Let $u \in \Sigma^+$. An intercode of index greater than or equal to 2 is a pure code.*

The family of pure codes properly includes the family of intercodes. Let $L_2 = \{a(ab)^n b | n \geq 0\}$. The language L_2 is a pure code but not an intercode.

Proposition 2 *([4]) For any $m \geq 1$, every intercode of index m is an intercode of index m+1.*

Corollary 3 *Every comma-free code is a pure code.*

The family of pure codes properly includes the family of comma-free codes. Let $L_4 = \{aaab, bbba, aabb\}$. The language L_4 is a pure code but not a comma-free code.

Proposition 4 *([5]) The family of d-codes is contained in the family of pure codes.*

Proposition 5 *([5]) Every pure code is contained in Q.*

Proposition 6 *Every d-code is contained in $D(1)$.*

Proof. Let L be a d-code. Suppose that L is not contained in $D(1)$. There exists a word $w = xyx$ such that $x \in \Sigma^+$ and $y \in \Sigma^*$. Then x is not in L since L is a bifix code. Thus $x \in P(L) \cap S(L)$, which contradicts that $L = P(L) \cap S(L)$. Hence L is contained in $D(1)$. □

Let $L_8 = \{a, aba\}$. We shall show that L_8 is a pure code.

Lemma 7 *For any $w \in F(L^*_8)$, if $w \in a\Sigma^* a$, then $w \in L^*_8$.*

Proof. Let $w \in a\Sigma^* a$. By induction on $n = \sharp_b(w)$.

For $n = 0$, obviously $w \in L^*_8$. Suppose that the result holds for $k \geq 0$. Let $\sharp_b(w) = k+1$. There exist an integer $i \geq 1$ and $z \in \Sigma^*$ such that $w = a^i baz$.

(Case 1) $z = \varepsilon$.

We have that $w = a^i ba \in L^*_8$.

(Case 2) $z \neq \varepsilon$.

(Case 2.1) $z = a$. Obviously, $w \in L^*_8$.

(Case 2.2) $z \neq a$ We can write $z = az'a$ for some $z' \in \Sigma^*$. By inductive hypothesis, $az'a \in L^*_8$. Thus $w \in L^*_8$. □

Proposition 8 *L_8 is a pure code.*

Proof. It is obvious that L_8 is a code. We show that L_8 is pure. Let $w = p^n \in L^*_8$, where p is in Q and $n \geq 2$ (It is trivial for the case $n = 1$). If $p = a$, then $p \in L^*_8$. Assume that $p \neq a$. It follows that $p \in a\Sigma^* a$. By the previous lemma, $p. \in L^*_8$

* School of Science and Engineering Kokushikan University

Thus L_8 is pure. □

Proposition 9 *The family of pure codes and the family of infix codes are incomparable.*

Proof. The language $L_9 = \{aba, bab\}$ is an infix code but not pure. The language $L_8 = \{a, aba\}$ is a pure code but not an infix code. □

Proposition 10 *The family of d-codes and the family of inter codes are incomparable.*

Proof. The language $L_2 = \{a(ab)^n b | n \geq 0\}$ is a d-code, but not an intercode. The language $L_6 = \{bbababb, a\}$ is an intercode of index 2([2]) but not a d-code.

□

Proposition 11 *The family of d-codes and the family of comma-free codes are incomparable.*
Proof. The language $L_7 = \{aabb, ab\}$ is a d-code, but not a comma-free code. The language $L_3 = \{aaab, bbba\}$ is a comma-free code but not a d-code. □

Proposition 12 *The family of d-codes and the family of infix codes are incomparable.*
Proof. The language $L_7 = \{aabb, ab\}$ is a d-code, but not an infix code. The language $L_5 = \{ab, ba\}$ is an infix code but not a d-code. □

Let $L_{11} = \{aabaaa, aaabaa, aaaaab\}$.

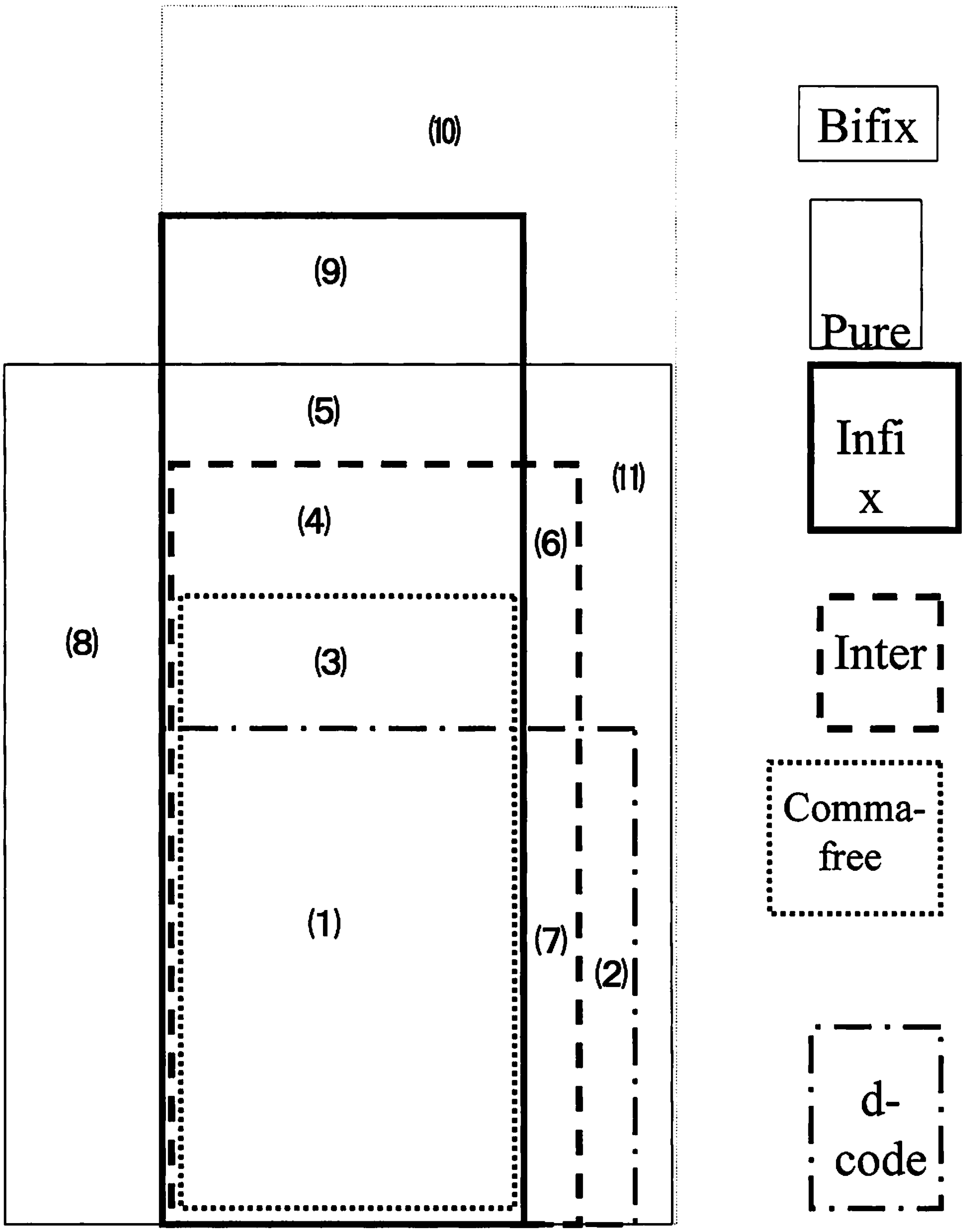
Proposition 13 L_{11} *is a pure code, and also is a bifix code. However, it is neither an infix code, an intercode, nor a d-code.*
□

The inclusion relation among these families mentioned above is shown in figure, with the following languages.

(1) $L_1 = \{aabbb, ababbb\}$
(2) $L_2 = \{a(ab)^n b | n \geq 0\}$
(3) $L_3 = \{aaab, bbba\}$
(4) $L_4 = \{aaab, bbba, aabb\}$
(5) $L_5 = \{ab, ba\}$
(6) $L_6 = \{bbababb, a\}$
(7) $L_7 = \{aabb, ab\}$
(8) $L_8 = \{aba, a\}$
(9) $L_9 = \{aba, bab\}$
(10) $L_{10} = \{aba, b\}$
(11) $L_{11} = \{aabaaa, aaabaa, aaaaab\}$

References

[1] J.Berstel and D. Perrin, Theory of codes, Academic Press, New York, 1985.
[2] H.J.Shyr, "Free Monoids and Languages", Hon Min Book, 2001.
[3] Chen-Ming Fan and Chen-Chih Huang, "Properties of pure codes", International Journal of Computer Mathematics, vol. 85, no.1, pp.9-17, 2008.
[4] H.J.Shyr and S.S.Yu, Intercodes and Related Properties, Soochow J. of Mathematics, vol.16, no.1, pp.95-107, 1990.
[5] Y.Y. Lin, Properties of words and related homomorphisms. MS Thesis, Institute of Applied Mathematics, Chun-Yuan University.


(10)
(9)
(5)
(11)
(4)
(6)
(8)
(3)
(1)
(7)
(2)
Bifix
Pure
Infi
x
Inter
Comma-
free
d-
code