Closure under Composition and Syntactic Monoids of some Codes*

Tetsuo MORIYA†

Department of Electrical Engineering, Faculty of Engineering
Kokushikan University
Setagaya 4-28-1, Setagaya-ku, Tokyo 154-8515, Japan

*This paper was mainly written during my visit at the Institut für Informatik, Martin-Luther-Universität Halle-Wittenberg, 1999.

†Electronic Mail: `moriya@kokushikan.ac.jp`

Contents

Abstract

We consider closure property of some classes of codes under composition and syntactic congruences of some codes. A code X is a strongly infix (outfix) code if X is an infix (outfix) code and any catenation of two words in X has no proper (infix) outfix in X, which is neither a left factor nor a right factor. We show that the class of strongly outfix codes is closed under composition, and as the dual result, that the property to be strongly outfix is inherited by a component of a decomposition. Finally, we show stronger properties for syntactic congruences of strongly infix codes than for those of infix codes.

Keywords: prefix code, suffix code, infix code, composition of codes, syntactic congruence

1 Introduction

The theory of codes has been studied in algebraic direction in connection to automata theory, combinatorics on words, formal languages, and semigroup theory. Many classes of codes have been defined and studied ([1, 2]). Among those classes, the class of infix codes and the class of outfix codes have many remarkable algebraic properties, like that of prefix (suffix, bifix) codes ([2, 3, 4]).

On the other hand, composition on codes ([1]) is very important as a binary operation by which more complicated codes can be constructed from simpler ones. So we are interested in whether the properties of codes are preserved under composition or not.

As though the classes of prefix codes, suffix codes, and bifix codes are closed under composition ([1]), the class of neither infix codes nor outfix codes is not [5]. Recently a strongly infix code has been defined and it has been proved that the class of those codes is closed under composition ([1]).

In section 2 some basic definitions and results are presented.

In section 3 we first show that if a code $X \subseteq \Sigma^+$ is a strongly outfix code, then X^* is mid-unitary. Then we show that the class of strongly outfix codes is closed under composition. Moreover we consider whether properties of codes are inherited by a component of a decomposition or not. We show that for the composition X of two codes Y and Z, if X is strongly outfix, then also is Y.

In section 4 we consider some properties of syntactic congruences of codes. Several properties of syntactic congruences for infix codes and outfix codes have been presented in [2] and [3]. We show stronger properties for syntactic congruences of strongly infix code than for those of infix codes.

2 Preliminaries

Let Σ be an alphabet. Σ^* denotes the free monoid generated by Σ, that is, the set of all finite words over Σ, including the empty word 1, and $\Sigma^+ = \Sigma^* - 1$. For w in Σ^* , $|w|$ denotes the length of w.

A word $x \in \Sigma^*$ is a *factor* or an *infix* of a word $w \in \Sigma^*$ if there exists $u, v \in \Sigma^*$ such that $w = uxv$. A factor x of w is *proper* if $w \neq x$. A catenation xy of two words x and y is an *outfix* of a word $w \in \Sigma^*$ if there exists $u \in \Sigma^*$ such that $w = xuy$. A word $u \in \Sigma^*$ is a *left factor* of a word $w \in \Sigma^*$ if there exists $x \in \Sigma^*$ such that $w = ux$. A left factor u of w is called *proper* if $u \neq w$. A right factor is defined symmetrically. An outfix xy of w is *proper* if $xy \neq w$.

A *language* over Σ is a set $X \subseteq \Sigma^*$. A language $X \subseteq \Sigma^*$ is a *code* if X freely generates the submonoid X^* of Σ^* (See [1] about the definition.). A language $X \subseteq \Sigma^+$ is a *prefix code* (resp. *suffix code*) if no word in X has a proper left factor (a proper right factor) in X. A language $X \subseteq \Sigma^+$ is a *bifix code* if X is both prefix and suffix. A language $X \subseteq \Sigma^+$ is an *infix code* (resp. *outfix code*) if no word $x \in X$ has a proper infix (a proper outfix) in X.

A language $X \subseteq \Sigma^+$ is *in-catenatable* (resp. *out-catenatable*) if a catenation of two words in X has a proper infix (proper outfix) in X which is neither a proper prefix nor a proper suffix. Formally, X is in-catenatable if there exist $u_1, u_2, u_3, u_4 \in \Sigma^+ - X$ such that u_1u_2, u_3u_4 and u_2u_3 is in X, X is out-catenatable if there exist $u_1, u_2, u_3, u_4 \in \Sigma^+ - X$ such that u_1u_2, u_3u_4 and u_1u_4 is in X with $u_1u_2 \neq u_3u_4$. A language $X \subseteq \Sigma^+$ is a *strongly infix code* (resp. *strongly outfix code*) if X is an infix code (outfix code) and is not in-catenatable (out-catenatable).

Let M be a monoid and let N be a submonoid of M. Then N is *right unitary(resp. left unitary)* in M if for all $u, v \in M$, $u \in N$ and $uv \in N$ ($vu \in N$) together imply $v \in N$. The submonoid N is *biunitary* if it is both left and right unitary. The submonoid N is *double unitary* in M if for all $u, x, y \in M$, $u \in N$ and $xuy \in N$ together imply x and $y \in N$. The submonoid N is *mid-unitary* in M if for all $u, x, y \in M$, $xy \in N$ and $xuy \in N$ together imply $u \in N$.

Proposition 1 ([1]) *Let $X \subseteq \Sigma^+$ be a code. A language X is a prefix code (resp., suffix code, bifix code) iff X^* is right unitary (left unitary, biunitary).*

Proposition 2 ([5]) *Let $X \subseteq \Sigma^+$ be a code. A language X is a strongly infix code iff X^* is double unitary.*

Let $Z \subseteq \Sigma^*$ and $Y \subseteq \Delta^*$ be two codes with $\Delta = alph(Y)$. Then the codes Y and Z are called *composable* (through β) if there is a bijection β from Δ onto Z. Then β defines a morphism $\Delta^* \to \Sigma^*$ which is injective since Z is a code. The set $X = \beta(Y) \subseteq Z^* \subseteq \Sigma^*$ is obtained by *composition* of Y and Z. We denote it by $X = Y \circ_\beta Z$ or $X = Y \circ Z$ when the context permits it.

For a language L over Σ and u in Σ^*, let

$$L..u = \{(x, y) | x, y \in \Sigma^* \text{ and } xuy \in L\}.$$

The *syntactic congruence* P_L is defined by

$$u \equiv v(P_L) \text{ iff } L..u = L..v.$$

The *syntactic monoid* $Syn(L)$ of L is the quotient monoid Σ^*/P_L.

3 Closure properties under composition

In this section we consider the closure properties of some classes of codes under composition.

Proposition 3 ([1, 5]) *The class of prefix (suffix, and bifix) codes is closed under composition.*

Theorem 4 ([5]) *The class of strongly infix codes is closed under composition.*

The class of outfix codes is not closed under composition, as that of infix codes is not closed [5]. For outfix codes $Z = \{aba, abb, aaa\}$ and $Y = \{c, de\}$, β is defined by $\beta(c) = aba$, $\beta(d) = abb$, and $\beta(e) = aaa$. Then $X = Y \circ_\beta Z = \{aba, abbaaa\}$ is not an outfix code.

Next we consider composition of the class of strongly outfix codes.

Proposition 5 *Let $X \subseteq \Sigma^+$ be a code. If a language X is a strongly outfix code, then X^* is mid-unitary.*

Proof. Suppose that X is strongly outfix. Let $x, u, y \in \Sigma^*$ be such that xy, $xuy \in X^*$. Let $xy = u_1...u_n$; $xuy = v_1...v_m$ for $u_1, ..., u_n; v_1, ..., v_m \in X$. Suppose that u is not in X^*.
(Case 1) $|v_1| < |xu|$
(1.1) $|xu| < |v_1v_2|$. There exists an integer $k > 0$ such that $|v_1...v_k| < |xu| < |v_1...v_{k+1}|$. Moreover there exist $v_1^{(1)}, v_1^{(2)}$, and $v_{k+1}^{(1)}, v_{k+1}^{(2)}$ in Σ^+ such that $v_1^{(1)}v_1^{(2)} = v_1$, $v_{k+1}^{(1)}v_{k+1}^{(2)} = v_{k+1}$ and $v_1^{(2)}v_2...v_kv_{k+1}^{(1)} = u$. We have that $v_1^{(1)}v_{k+1}^{(2)}v_{k+2}...v_m = xy$. Since X is suffix, and X^* is left unitary, $v_1^{(1)}v_{k+1}^{(2)}$ is in X^*. If $v_1^{(1)}v_{k+1}^{(2)} \in X^n$ for $n > 1$, either $v_1^{(1)}$ has a proper left factor in X or $v_2^{(k+1)}$ has a proper right factor in X. Moreover $v_1^{(1)}v_{k+1}^{(2)}$ is not in X since X is not out-catenatable. Hence $v_1^{(1)}v_{k+1}^{(2)} = 1$. This is a contradiction.
(1.2) $|xu| < |v_1|$. There exists v_1' in Σ^+ such that $xuv_1' = v_1$. We have that $xv_1'v_2...v_m = xy$. Since X is suffix, and X^* is left unitary, xv_1' is in X^*. We get that $xv_1' = 1$ in the same way as in (1.1). Thus $u = v_1 \in X$.
(Case 2) $|v_1| < |x|$
There exists an integer $k > 0$ such that $|v_1...v_k| < |x| < |v_1...v_{k+1}|$.
(2.1) $|xu| < |v_1...v_{k+1}|$. There exists $v_{k+1}^{(1)}, v_{k+1}^{(2)} \in \Sigma^+$ such that $v_{k+1}^{(1)}uv_{k+1}^{(2)} = v_{k+1}$. We have that $x = v_1...v_kv_{k+1}^{(1)}$, $v_{k+1}^{(2)}v_{k+2}...v_m = y$. Since xy is in X^*, and X is prefix and suffix, $v_{k+1}^{(1)}v_{k+1}^{(2)}v_{k+1}...v_m$ is in X^*. Moreover, since $v_{k+1}...v_m$

is in X^*, $v_k^{(1)}v_k^{(2)}$ is in X^*. If $v_k^{(1)}v_k^{(2)}$ is in X^n for $n>1$, either $v_{k+1}^{(1)}$ has a proper left factor in X or $v_{k+1}^{(2)}$ has a proper right factor in X. Since X is outfix, $v_{k+1}^{(1)}v_{k+1}^{(2)}$ is not in X. Hence $v_{k+1}^{(1)}v_{k+1}^{(2)}=1$. Thus $u=v_2\in X$. This is a contradiction.
(2.2) $|v_1...v_{k+1}|<|xu|$. There exists an integer $j>k$ such that $|v_1...v_j|<|xu|<|v_1...v_{j+1}|$. Then there exist $v_{k+1}^{(1)},v_{k+1}^{(2)}$, and $v_{j+1}^{(1)},v_{j+1}^{(2)}$ such that $v_{k+1}^{(1)}v_{k+1}^{(2)}=v_{k+1}$, $v_{j+1}^{(1)}v_{j+1}^{(2)}=v_{j+1}$. We have that $v_{k+1}^{(1)}v_{j+1}^{(2)}v_{j+2}...v_m=xy$. We get that $v_{k+1}^{(1)}v_{j+1}^{(2)}=1$ in the same way as in (1.1). ❏

The converse of the previous proposition does not hold. Let a language X be $\Sigma^3=\{aaa,aab,aba,abb,baa,bab,bba,bbb\}$, for $\Sigma=\{a,b\}$. Then X^* is mid unitary since for every word $w\in\Sigma^*$, $|w|=3n$ for an integer n iff $w\in X^*$. It is obvious that X is not a strongly outfix code.

Proposition 6 *Let Y and Z be composable codes, and let $X=Y\circ Z$. If Y is an outfix code and Z is a strongly outfix code, then X is an outfix code.*

Proof. Assume that Y is an outfix code, and that Z is a strongly outfix code. Consider $xy,xuy\in X$ with $u\in\Sigma^*$. Since $X\subseteq Z^*$, we have that $xy,xuy\in Z^*$, and since Z^* is mid unitary, this implies $u\in Z^*$.

Let $w=\beta^{-1}(xy)$, $z=\beta^{-1}(xuy)$. Then, we have that $z=\beta^{-1}(xuy)=\beta^{-1}(x)\beta^{-1}(u)\beta^{-1}(y)\in Y$, and $w=\beta^{-1}(xy)=\beta^{-1}(x)\beta^{-1}(y)\in Y$. Since Y is outfix, $\beta^{-1}(u)=1$. Consequently $u=1$. This shows that X is outfix. ❏

Lemma 7 *Let X be a code and Y, Z be outfix codes with $X=Y\circ Z$. If X is out-catenatable, then either Y or Z is out-catenatable.*

Proof. Let $\Delta=alph(Y)$. Suppose that X is out-catenatable. There exist $u_1,u_4\in\Sigma^+-X,u_2,u_3\in\Sigma^+$ such that $u_1u_2\in X,u_3u_4\in X$, $u_1u_4\in X$. Let $u_1u_2=z_1^{(1)}...z_n^{(1)}$, $u_3u_4=z_1^{(2)}...z_m^{(2)}$, and $u_1u_4=z_1^{(3)}...z_l^{(3)}$ for $z_1^{(1)},...,z_n^{(1)}$; $z_1^{(2)},...,z_m^{(2)}$; $z_1^{(3)},...,z_l^{(3)}\in Z$.
(Case 1) $u_1\in Z^*$.
There exists an integer $i>0$ and $j>0$ such that $u_1=z_1^{(1)}...z_i^{(1)}=z_1^{(3)}...z_j^{(3)}$. Since X is a (prefix) code, we have that $i=j$, and $z_1^{(1)}=z_1^{(3)},...,z_i^{(1)}=z_i^{(3)}$. It follows that $u_2=z_{i+1}^{(1)}...z_n^{(1)}$, $u_3=z_1^{(2)}...z_{m-l+i}^{(2)}$, and $u_4=z_{m-l+i+1}^{(2)}...z_m^{(2)}=z_{i+1}^{(3)}...z_l^{(3)}$. Similarly we have that $z_{m-l+i+1}^{(2)}=z_{l-i}^{(3)}$, ... , $z_m^{(2)}=z_l^{(3)}$. Let $\beta(a_1^{(1)})=z_1^{(1)},...,\beta(a_n^{(1)})=z_n^{(1)},\beta(a_1^{(2)})=z_1^{(2)},...,\beta(a_m^{(2)})=z_m^{(2)},\beta(a_1^{(3)})=z_1^{(3)},.....,\beta(a_l^{(3)})=z_l^{(3)}$ for $a_1^{(1)},...,a_n^{(1)}$;

$a_1^{(2)}, ..., a_m^{(2)}; a_1^{(3)}, ..., a_l^{(3)} \in \Delta$. We have that $a_1^{(1)}...a_n^{(1)} \in Y$, $a_1^{(2)}...a_m^{(2)} \in Y$, and $a_1^{(3)}...a_l^{(3)} = a_1^{(1)}...a_i^{(1)}a_{l-i}^{(3)}...a_l^{(3)} \in Y$. Thus Y is out-catenatable.
(Case 2) $u_1 \notin Z^*$.
There exist an integer $i > 0$ and $u, v \in \Sigma^+ - Z$ such that $u_1 = z_1^{(1)}...z_i^{(1)}u$ and $uv = z_{i+1}^{(1)}$. Since $u_1, u_4 \in X \subseteq Z^*$, it is obvious that u_4 is not in Z^*. Then there exist $u', v' \in \Sigma^+ - Z$ and $j > 0$ such that $u_3 = z_1^{(2)}...z_j^{(2)}u'$, $u_4 = v'z_{j+2}^{(2)}...z_l^{(2)}$, $u'v' = z_{j+1}^{(2)}$. Since $u_1u_4 = z_1^{(3)}...z_l^{(3)} = z_1^{(1)}...z_i^{(1)}uv'z_{j+1}^{(2)}...z_l^{(2)}$, and $uv' = z_i^{(3)}$, we have that the catenation $z_{i+1}^{(1)}z_{j+1}^{(2)} = uvu'v'$ of two words $z_{i+1}^{(1)} = uv$ and $z_{j+1}^{(2)} = u'v'$ has a proper outfix $uv' = z_i^{(3)} \in Z$ which is neither a proper prefix nor a proper suffix. Thus Z is out-catenatable. ❑

Theorem 8 *The class of strongly outfix codes is closed under composition.*

Proof. The result is immediate by Proposition 6 and Lemma 7. ❑

Next we consider whether properties of codes are inherited by a component of a decomposition or not.

Proposition 9 ([1]) *Let X, Y and Z be codes with $X = Y \circ Z$. If X is prefix (suffix, bifix), then also is Y .*

Proposition 10 ([5]) *Let X, Y, and Z be codes with $X = Y \circ Z$. If X is infix (strongly infix), then also is Y .*

Proposition 11 *Let X, Y, and Z be codes with $X = Y \circ Z$. If X is an outfix, then also is Y .*

Proof. Assume that X and $Z \subseteq \Sigma^*$, $Y \subseteq \Delta^*$, and β: $\Delta^* \to \Sigma^*$ be an injective morphism with $\beta(\Delta) = Z$ and $\beta(Y) = X$. Let $xuy, xy \in Y$. Then $\beta(xuy) = \beta(x)\beta(u)\beta(y)$ and $\beta(xy) = \beta(x)\beta(y)$ are in X. Since X is outfix, $\beta(u) = 1$. Thus $u = 1$. ❑

Lemma 12 *Let X be an outfix code, and Y, Z be codes with $X = Y \circ Z$. If Y is out-catenatable, then also is X.*

Proof. Suppose that Y is out-catenatable. There exist $u_1, u_2, u_3, u_4 \in \Delta^+ - Y$ such that u_1u_2, u_3u_4, and $u_1u_4 \in Y$. Let $\beta(u_i) = v_i$ for $i = 1, ..., 4$. Hence $v_1v_2 = \beta(u_1u_2)$, $v_2v_3 = \beta(u_2u_3)$, and $v_1v_4 = \beta(u_1u_4) \in X$. It is obvious that $v_i = \beta(u_i) \in \Sigma^+ - X$ for $i = 1, ..., 4$. Thus X is out-catenatable. ❑

Proposition 13 *Let X, Y, and Z be codes with $X = Y \circ Z$. If X is a strongly outfix code, then also is Y .*

Proof. It is obvious by Proposition 11 and Lemma 12. ❑

4 Syntactic congruences of some codes

In this section we consider properties of syntactic congruences of infix codes and strongly infix codes. More precisely we study syntactic congruences of L^n, where L is a (strongly) infix code L and n is a positive integer.

Before discussing, we give some basic results.

Proposition 14 ([3]) *Every infix code L is a P_L-class.*

Lemma 15 *For languages L, $K \subseteq \Sigma^*$, if L is a P_K-class, then $P_K \subseteq P_L$.*

Proof. Suppose that L is a P_K-class, and that $u \equiv v(P_K)$. Then one has that $xuy \equiv xvy(P_K)$ for every x, y. If xuy is in L, then it is in a class of P_K. Thus xvy is in the same class of P_K, that is, in L. Similarly we can easily obtained that $xvy \in L$ implies $xuy \in L$. Hence $u \equiv v(P_L)$. ❑

Lemma 16 *Let L be a code, and let m and k be integers with $1 \leq m \leq k$. If $u \in L^m$, $xuy \in L^k$ and $x, y \in L^*$, then $x \in L^i$ and $y \in L^j$ for integers $i, j \geq 0$ such that $i + j = k - m$.*

Proof. Let $u = u_1...u_m; u_1, ..., u_m \in L$, $xuy = v_1...v_k; v_1, ..., v_k \in L$, $x = a_1...a_i; a_1, ..., a_i \in L$, and $y = b_1...b_j; b_1, ..., b_j \in L$.

Since L is a code, $a_1 = v_1, ..., a_i = v_i; u_1 = v_{i+1}, ..., u_m = v_{i+m-1}; b_1 = v_{i+m}, ..., b_j = v_{i+m+j}$. It is obvious that $i + m + j = k$. Thus the result holds. ❑

Now we present properties of syntactic congruences of L^n for a (strongly) infix code L.

Proposition 17 *For every infix code L and an integer $n \geq 2$, $P_{L^n} \subseteq P_{L^{n-1}}$.*

Proof. Suppose that $u \equiv v(P_{L^n})$ and $xuy \in L^{n-1}$. Taking an arbitrary word $w \in L$, we have that $xuyw \in L^n$. It follows that $xvyw \in L^n$, by $u \equiv v(P_{L^n})$. Hence xvy is in L^* since L^* is leftunitary . By Lemma 16, xvy is in L^{n-1} Similarly we have that $xvy \in L^{n-1}$ implies $xuy \in L^{n-1}$. Thus $u \equiv v(P_{L^{n-1}})$. ❑

Proposition 18 *For every strongly infix code L, and two integers m and k with $1 \leq m \leq k$, L^m is a P_{L^k}-class.*

Proof. Let $u, v \in L^m$. Suppose that $xuy \in L^k$. Since L^* is double unitary, one has that both x and y are in L^*. Then it follows that $x \in L^i$ and $y \in L^j$ with $i + j = k - m$ by Lemma 16. Thus $xvy \in L^i.\ L^m.L^j = L^k$. Similarly, it is easily obtained that $xvy \in L^k$ implies $xuy \in L^k$. Thus $u \equiv v(P_{L^k})$. Hence L^m is contained in a P_{L^k}-class.

Suppose that $u \equiv v(P_{L^k})$ with $u \in L^m$. Take a word $x \in L^{k-m}$. Then $xu \in L^{k-m}.L^m = L^k$. By $u \equiv v(P_{L^k})$, we have that $xv \in L^k$. Since L^* is right unitary, we have that v is in L^*. By Lemma 16, v is in L^m. Hence the result holds. ❑

Corollary 19 *For every strongly infix code L and integers m and k with $1 \leq m \leq k$, $P_{L^k} \subseteq P_{L^m}$.*

Proof. The result can be easily obtained by Lemma 15 and Proposition 18. ❑

The result such as Proposition 18 does not hold in general for an infix code: For an infix code $L = \{aba, bab\}$, which is not a strongly infix code, we have that $P_{L^n} \subseteq P_{L^{n-1}}$. However L is not a P_{L^2}-class since the two words aba and bab are not in the same class of P_{L^2}.

References

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

[2] M. Ito, H. Jürgensen, H. J. Shyr and G. Thierrin, Outfix and infix codes and related classes of languages, Journal of Computer and System Sciences 43 (1991), 484–508.

[3] H. Jürgensen and G. Thierrin, Infix Codes, in "Proceedings, 4th Hungarian Comp. Sci. Conf. Györ, 1985", 25–29.

[4] H. Jürgensen and G. Thierrin, Codes, and compatible partial orders on free monoids, in "Algebra and order, Proc.of the First International Symposium on Ordered Algebraic Structures, Luminy-Marseilles, 1984" (S.Wolfenstein, Ed.), pp. 323–333, Heldermann Verlag, Berlin, 1986.

[5] T. Moriya, Composition of Strongly Infix Codes, IEICE Trans. Inf. and Syst. E81-D No.6.(1998), 609–611.