<table>
<thead>
<tr>
<th>Title</th>
<th>Syntactic Congruences of some Codes (Algebraic Systems, Formal Languages and Computations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Moriya, Tetsuo</td>
</tr>
<tr>
<td>Citation</td>
<td>数理解析研究所講究録 (2000), 1166: 191-199</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2000-08</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/2433/64340">http://hdl.handle.net/2433/64340</a></td>
</tr>
<tr>
<td>Type</td>
<td>Departmental Bulletin Paper</td>
</tr>
<tr>
<td>Textversion</td>
<td>publisher</td>
</tr>
</tbody>
</table>

Kyoto University
Syntactic Congruences of some Codes

Tetsuo MORIYA

Department of Electrical Engineering, Faculty of Engineering
Kokushikan University
Setagaya 4-28-1, Setagaya-ku, Tokyo 154-8515, Japan
Electric Mail: moriya@kokushikan.ac.jp

Abstract

We consider syntactic congruences of some codes. As a main result, for an infix code $L$, it is proved that the following (i) and (ii) are equivalent and that (iii) implies (i), where $P_L$ is the syntactic congruence of $L$.

(i) $L$ is a $P_{L^2}$-class.

(ii) $L^m$ is a $P_{L^k}$-class, for two integers $m$ and $k$ with $1 \leq m \leq k$.

(iii)$L^*$ is a $P_{L^1}$-class.

Next we show that every (i), (ii) and (iii) holds for a strongly infix code $L$. Moreover we consider properties of syntactic congruences of a residue $W(L)$ for a strongly outfix code $L$.

Keywords: prefix code, suffix code, infix code, 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. A lot of classes of codes have been defined and studied ([1], [2]). Among those codes, prefix codes, suffix code, bifix codes, infix codes and outfix codes have many remarkable algebraic properties ([2], [3], [4]). Recently a strongly infix code and a strongly outfix code were defined and the closure property under composition operation for these code was proved ([5][6]).

In this paper we study syntactic congruences of some codes, especially, (strongly) infix codes and (strongly) outfix codes. Several properties of the syntactic congruence $P_L$ of $L$, for $L$ infix or outfix, have been presented in [2] and [3] and moreover some interesting characterizations have been presented on the syntactic monoid and the syntactic congruence $P_L$ of $L$ for an infix code $L([7])$. We mainly deal with the syntactic congruence $P_{L^n}$ of $L^n$, $n > 1$, and $P_{L^*}$ of $L^*$ in this paper below.

In section 2 some basic definitions and results are presented.

In section 3, first we prove that the following (i) and (ii) are equivalent for an infix code $L$, and that (iii) implies (i), where $P_L$ is the syntactic congruence of $L$.

(i) $L$ is a $P_{L^2}$-class.
(ii) $L^m$ is a $P_{L^k}$-class, for two integers $m$ and $k$ with $1 \leq m \leq k$.
(iii) $L^*$ is a $P_{L^n}$-class.

Next we show that every (i), (ii) and (iii) holds for a strongly infix code $L$, and moreover we show that $L^*$ is contained in a $P_{W(L)}$-class, where $W(L)$ is a residue of $L$. Last we consider a relation between $P_{L^n}$-class and $W(L)$ for a strongly outfix code $L$.

2 Preliminaries

Let $\Sigma$ be an alphabet. $\Sigma^*$ denotes the free monoid generated by $\Sigma$, that is, the set of all finite words over $\Sigma$, including the empty word 1, and $\Sigma^+ = \Sigma^* - 1$. For $w$ in $\Sigma^*$, $|w|$ 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 $xy \neq w$. The set of all left factors (resp. right factors) of a word $x$ is denoted by $\text{Pref}(x)(\text{Suf}(x))$.

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

A language $L \subseteq \Sigma^+$ is *in-catenatable* (resp. *out-catenatable*) if a catenation of two words in $L$ has a proper infix (proper outfix) in $L$ which is neither a proper prefix nor a proper suffix. Formally, $L$ 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 $L$, and $L$ 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 $L$ with $u_1u_2 \neq u_3u_4$. A language $L \subseteq \Sigma^+$ is a *strongly infix code* (resp. *strongly outfix code*) if $L$ is an infix code (outfix code) and is not in-catenatable (out-catenatable). A strongly infix (resp. outfix) code may be abbreviated to an *s-infix* (s-outfix) code.

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 $L \subseteq \Sigma^+$ be a code. A language $L$ is a prefix code (resp., suffix code, bifix code, s-infix code) iff $L^*$ is right unitary (left unitary, biunitary, double unitary).

**Proposition 2** [6] Let $L \subseteq \Sigma^+$ be a code. If a language $L$ is a strongly outfix code, then $L^*$ is mid-unitary.

**Proposition 3** Let $L \subseteq \Sigma^+$ be a code. If $L^*$ is mid-unitary, then $L$ is an outfix code.
\textbf{Proof.} Suppose that $L$ would not be outfix with $L^*$ mid-unitary. There exist $x,y \in \Sigma^*$ and $u \in \Sigma^+$ such that both $xuy$ and $xy$ are in $L$. Since $L^*$ is mid-unitary, we have that $u \in L^*$, and thus $u \in L^+$. It is easily obtained that both $uyx$ and $yux$ are in $L^*$, since both $xuy$ and $xuyxuy$ are in $L^*$. Thus $uyxu$ has two factorization. This contradicts the fact that $L$ is a code. \hfill \Box

For a language $L$ over $\Sigma$ and $u$ in $\Sigma^*$, let

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

The \textit{syntactic congruence} $P_L$ is defined by

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

The \textit{syntactic monoid} $\text{Syn}(L)$ of $L$ is the quotient monoid $\Sigma^*/P_L$. For any language $L \subseteq \Sigma^*$, let $W(L)$ denote the \textit{residue} of $L$, that is,

$$W(L) = \{u \in \Sigma^* | L..u = \phi\}.$$  

\section{3 Syntactic congruences of some codes}

In this section we consider properties of syntactic congruences of some codes.

Before discussing, we give some basic results.

**Proposition 4** [3] \textit{Every infix code $L$ is a $P_L$-class.}

**Proposition 5** [3] \textit{Let $L$ be an outfix code. Then every $P_L$-class different from $W(L)$ is an outfix code.}

**Lemma 6** \textit{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 $u \equiv v(P_L)$. \hfill \Box
Lemma 7 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 $x, y \in L^\star$, then $x \in L^i$ and $y \in L^j$ for integers $i, j \geq 0$ such that $i + j = k - m$.

Proof. Let $u = u_1 \ldots u_m; u_1, \ldots, u_m \in L$, $xuy = v_1 \ldots v_k; v_1, \ldots, v_k \in L$, $x = a_1 \ldots a_i; a_1, \ldots, a_i \in L$, and $y = b_1 \ldots b_j; b_1, \ldots, b_j \in L$. Since $L$ is a code, $a_1 = v_1, \ldots, a_i = v_i; u_1 = v_{i+1}, \ldots, u_m = v_{i+m-1}; b_1 = v_{i+m}, \ldots, b_j = v_{i+m+j}$. It is obvious that $i + m + j = k$. Thus the result holds.

Lemma 8 For a languages $L$ and $K$, if $P_L \subseteq P_K$ and $K$ is contained in a $P_L$-class, then $K$ is equal to a $P_L$-class.

Proof. It is obvious from the fact that $L$ is a union of $P_L$-classes.

Now we consider properties of a syntactic congruence $P_{L^n}$ of $L^n$ and a syntactic congruence $P_{L^*}$ of $L^*$ for an infix code $L$ and a positive integer $n$. The first result holds for a prefix code or a suffix code.

Proposition 9 Let $L$ be a prefix code or a suffix code. For an integer $n \geq 2$, $P_{L^n} \subseteq P_{L^{n-1}}$.

Proof. Let $L$ be a prefix code. Suppose that $u \equiv v(P_{L^n})$ and $xuy \in L^{n-1}$. Taking an arbitrary word $w \in L$, we have that $wxuy \in L^n$. It follows that $wxuy \in L^n$, by $u \equiv v(P_{L^n})$. Hence $xuy$ is in $L^*$ since $L^*$ is right unitary. By Lemma 7, $xuy$ is in $L^{n-1}$. Similarly we have that $xuy \in L^{n-1}$ implies $xuy \in L^{n-1}$. Thus $u \equiv v(P_{L^{n-1}})$. In the case of a suffix code, we can similarly prove the result.

Proposition 10 Let $L$ be an infix code. Then the following conditions are equivalent:

(i) $L$ is a $P_{L^2}$-class.

(ii) $L^m$ is a $P_{L^k}$-class, for two integers $m$ and $k$ with $1 \leq m \leq k$.

Proof. (i) $\implies$ (ii) : Suppose that $L$ is a $P_{L^2}$-class. First we prove that $L$ is a $P_{L^k}$-class for every $k \geq 2$. Let $u$ and $v$ be in $L$ and $xuy \in L^k$ for $x, y \in \Sigma^\star$. If one of
the two words \(x\) and \(y\) is in \(L^*\), then the other is also in \(L^*\), since \(L\) is an infix code. Then \(xvy\) is in \(L^k\) by Lemma 7. So assume that neither \(x\) nor \(y\) is in \(L^*\). Since \(L\) is infix, the word \(u\) has no proper factor in \(L\). Then there exist \(u_1, u_2, z, w \in \Sigma^+\) such that \(wu_1u_2z \in L, u = u_1u_2, w \in Suf(x), z \in Pre(y)\). We have that \(wuz\) is in \(L^2\), so \(xvy\) is in \(L^k\) since \(L\) is a \(P_{L^2}\)-class. Similarly we have that \(xvy \in L^k\) implies \(xuvy \in L^k\). Hence \(L\) is contained in a \(P_{L^k}\)-class for \(k \geq 2\). Since \(P_{L^k} \subseteq P_L, L\) is a \(P_{L^k}\)-class by Lemma 8.

Next suppose that \(u, v \in L^m\) and \(xvy \in L^k\) with \(m \leq k\) for \(x, y \in \Sigma^*\). Let \(u = u_1 \ldots u_m\) for \(u_1, ..., u_m \in L\) and \(v = v_1 \ldots v_m\) for \(v_1, ..., v_m \in L\). Since \(L\) is a \(P_{L^k}\)-class, \(xv_1u_2 \ldots u_my\) is in \(L^k\) for \(v_1 \in L\). Furthermore, for \(v_2 \in L, xv_1v_2u_3 \ldots u_my \in L^k\). Continueing this process, we can prove that for \(v \in L^m, xvy \in L^k\). Similarly as above, we have that \(L^m\) is contained in a \(P_{L^k}\)-class. By Lemma 8, \(L^m\) is a \(P_{L^k}\)-class since \(P_{L^k} \subseteq P_{L^m}\).

\((ii) \implies (i) :\) trivial.

**Proposition 11** For an infix code \(L\), if \(L^*\) is a \(P_{L^*}\)-class, then \(L\) is a \(P_{L^2}\)-class.

**Proof.** Let \(u, v \in L,\) and \(xvy \in L^2\). There exist \(u_1\) and \(u_2 \in \Sigma^+\) such that \(u_1u_2 = u, xu_1u_2y \in L\). By the hypothesis, we have that \(xvy \in L^*\). Suppose that \(xvy \in L^k\) for \(k > 2\). Let \(xvy = w_1 \ldots w_k\) for \(w_1, ..., w_k \in L\). Since \(L\) is infix, we have that \(|x| < |w_1| < |xv|\) and \(|y| < |w_k| < |vy|\). Hence \(w_2 \ldots w_{k-1}\) is a proper factor of \(v\). This is a contradiction. Thus \(xvy \in L^2\). By symmetry, we have that \(xvy \in L^2\) implies \(xuvy \in L^2,\) and thus \(L\) is contained in a \(P_{L^2}\)-class. By Lemma 8 and the fact that \(P_{L^2} \subseteq P_L,\) the result holds.

Unfortunately, the converse of Proposition 11 does not holds. For an alphabet \(\Sigma = \{a_1^{(1)}, a_1^{(2)}, a_2, b_1, b_2, c_1^{(1)}, c_1^{(2)}, c_2, d_1, d_2\}\), consider the infix code \(L = xx_1\Sigma \cup x\Sigma y_1 \cup x\{x_1, u, v_1\} \cup x_2x_2\Sigma y \cup x_2\Sigma y_1y \cup x_2\{x_1, u, v_1\}y \cup \Sigma yy \cup \{uv, vy\},\) where \(x_1 = a_1^{(1)}, a_1^{(2)}, x_2 = a_2, u = b_1b_2, v_1 = c_1^{(1)}, c_1^{(2)}, y_1 = c_2, y = d_1d_2, x = x_1x_2\). It can be easily checked that \(L\) an infix code, and \(L\) is a \(P_{L^2}\)-class. Although both \(uvuv\) and \(xuvy\) are in \(L^2, xuvuvy\) is not in \(L^2\) since \(vu\) is not in \(L\). Alternatively, \(xuvy\) and \(xuvuvy\) are not in the same class of \(P_{L^*}\).

Next we consider \(P_{L^n}, n \geq 1,\) and \(P_{L^*}\) for s-infix code \(L\).
Proposition 12 For every s-infix code \( L \), \( L \) is a \( P_{L^2} \)-class.

Proof. Let \( u, v \in L \). Suppose that \( xuy \in L^2 \). 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 \( i + j = 1 \) by Lemma 7. That is, either \( x = 1 \) and \( y \in L \), or \( y = 1 \) and \( x \in L \). Thus \( xuy \in L^2 \). Similarly, it is easily obtained that \( xvy \in L^2 \) implies \( xuy \in L^2 \). Thus \( u \equiv v(P_{L^2}) \). Hence \( L \) is contained in a \( P_{L^2} \)-class. By Lemma 8 and Proposition 9, \( L \) is a \( P_{L^2} \)-class. \( \square \)

Corollary 13 For every s-infix code \( L \), and two integers \( m \) and \( k \) with \( 1 \leq m \leq k \), \( L^m \) is a \( P_{L^k} \)-class.

Proof. It is obvious by Propositions 12 and 14. \( \square \)

Proposition 14 Let \( L \) be a s-infix code over \( \Sigma \). Then \( L^* \) is a \( P_{L^*} \)-class.

Proof. Let \( u, v \in L^* \). Suppose that \( xuy \) is in \( L^* \) for \( x, y \in \Sigma^* \). Since \( L^* \) is double-unitary, both \( x \) and \( y \) are in \( L^* \). Hence \( xvy \) is in \( L^* \). Similarly we have that \( xvy \in L^* \) implies \( xuy \in L^* \). Thus \( u \equiv v(P_{L^*}) \), and so \( L^* \) is contained in a \( P_{L^*} \)-class. Since \( L^* \) is a union of \( P_{L^*} \)-classes, the result holds. \( \square \)

Proposition 15 Let \( L \) be a s-infix code over \( \Sigma \). Then \( L^* \) is contained in a \( P_{W(L^*)} \)-class.

Proof. Let \( u, v \in L^* \). Suppose that \( xuy \notin W(L^*) \), that is, \( L^* \cdot xuy \neq \phi \). Then immediately we have that \( \Sigma^* x \cap L^* \neq \phi \) and \( y \Sigma^* L^* \neq \phi \) since \( L^* \) is double unitary. Hence \( xvy \notin W(L^*) \). Similarly we can obtained that \( xvy \notin W(L^*) \) implies \( xuy \notin W(L^*) \). Thus the result holds. \( \square \)

Remark 1 The result such as Proposition 12 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 \( P_{L^m} \subseteq P_{L^{m-1}} \). However \( L \) is not a \( P_{L^2} \)-class since the two words \( aba \) and \( bab \) are not in the same class of \( P_{L^2} \).
Last we consider the syntactic congruence $P_{L^n}$ of $L^n$ for a strongly outfix code $L$.

**Proposition 16** Let $L$ be a s-outfix code over $\Sigma$. Then every $P_{L^n}$-class $(1 \leq n)$ not contained in $W(L)$ is a s-outfix code.

**Proof.** Since the class of outfix codes is closed under concatenation [2], we have that $P_{L^n}$-class different from $W(L^n)$ is an outfix code by Proposition 5. Moreover it follows that $P_{L^n}$-class not contained in $W(L)$ is an outfix code by that $W(L^n) \subseteq W(L)$.

Suppose that such a $P_{L^n}$-class is not s-outfix, that is, there exist $x_1, x_2, z_1, z_2 \in \Sigma^+$ such that $x_1z_1 \equiv x_2z_2 \equiv x_1z_2 (P_L)$ and $x_1z_1 \neq x_2z_2$. Since $P_{L^n} \subseteq P_L$, these three words are in the same $P_L$-class different from $W(L)$. So there exist $w_1, w_2 \in \Sigma^*$ such that $w_1x_1z_1w_2 \in L, w_1x_2z_2w_2 \in L$ and $w_1x_1z_2w_2 \in L$. Then we have that $w_1x_1z_1w_2w_1x_2z_2w_2 \in L^2, w_1x_1z_1w_2 \in \Sigma^+$ and $w_1x_1z_1w_2 \neq w_1x_2z_2w_2$. This contradicts the fact that $L$ is s-outfix. Thus the result holds. \hfill \Box

**Remark 2** In Proposition 16, a similar result as Proposition 5 for an s-outfix code $L$ does not hold. That is, $P_{L^n}$-class different from $W(L^n)$, but contained in $W(L)$, is not necessarily s-outfix. For an s-outfix code $L = \{abbb, baaab, caaac\}$, let $w_1 = abbbabaa, w_2 = caaacaab, and w_3 = baabaa$. Then $w_1 \equiv w_2 \equiv w_3 (P_L)$, but $w_1w_2$ has a proper outfix $abbbabaa = w_1$ in $L$. Thus the class which contains $w_1, w_2$ and $w_3$ is not s-outfix.

**References**


