SOOCHOW JOURNAL OF MATHEMATICS
Volume 24, No. 2, pp. 113-130, April 1998

MIDST-LANGUAGES

BY

H. J. SHYR AND S. S. YU

Abstract. A language L over a finite alphabet X is a midst-language if L is neither regular nor disjunctive. For the case $|X| = 1$, there are no midst-languages over X. For the case $|X| \geq 2$, every completely disjunctive language contains no midst-languages. On the other hand some well-known languages such as the Dyck language and the so called atom Fibonacci languages are midst-languages. In this paper we study some algebraic properties of midst-languages. In particular we characterize the languages of the form $L = \bigcup_{f\in\Lambda} f^+$ to be midst-languages, where Λ is a regular component. We derive that for every infinite regular language $\Lambda \subseteq Q$, the language $L = \bigcup_{f\in\Lambda} f^+$ is not regular. An example of two disjoint midst-languages of which the union is disjunctive is given in this paper. The catenation, complement and union of languages concerning midst-languages are investigated. We show that some particular right ideals of X^* are able to be expressed as a disjoint union of infinitely many midst-languages.

1. Introduction

It is known that there are disjunctive languages such that every infinite subset of them is disjunctive. These are the so called completely disjunctive languages ([3]). But in [3], it is also shown that every infinite regular language over a finite alphabet consisting of at least two letters contains a language which is neither regular nor disjunctive. The well known Dyck language is also a language which is neither regular nor disjunctive. In this paper, we are going to investigate the languages which are neither regular nor disjunctive. This is another approach to investigate the properties of regular languages.

Let X^* be the free monoid generated by a finite alphabet X. Every element of X^* is a *word* and every subset of X^* is a *language*. Let $X^+ = X^*\backslash\{1\}$, where

Received August 27, 1997.
AMS Subject Classification. 68R15.
This paper was supported by the National Science Council R.O.C. under Grant NSC 87-2115-M-005-003.

1 is the empty word. For a given language $L \subseteq X^*$, the *principal congruence* P_L determined by L is defined as follows:

$$u \equiv v(P_L) \iff (xuy \in L \iff xvy \in L \ \forall x, y \in X^*).$$

It is well known that the language L is accepted by an automation if and only if L has finite P_L congruence classes, that is, P_L has finite index. A language which is accepted by an automation is called a *regular language* ([8]). We call a language L *disjunctive* if P_L is the equality. A language L is *dense* if for every word $u \in X^*$, there exist two words $x, y \in X^*$ such that $X^*uX^* \cap L \neq \emptyset$. A language L being dense is equivalent to that L contains a disjunctive subset (see [9]).

We call a language $L \subset X^*$ a *midst-language* if L is neither regular nor disjunctive. In [2], Fan and Shyr defined two so called atom Fibonacci languages over $X = \{a, b\}$ as follows:

$$F^1_{a,b} = \{a, b, ab, bab, abbab, \cdots\}; \ F^0_{a,b} = \{a, b, ba, bab, babba, \cdots\}.$$

These two languages have the following properties: (1) both $F^1_{a,b}$ and $F^0_{a,b}$ are regular free in the sense that the language contains no infinite regular subsets, and (2) no words in $F^1_{a,b}$ or $F^0_{a,b}$ contain a subword of the form u^4 for any $u \in X^+$. From (1), it is clear that both atom fibonacci languages are not regular. Again from (2) we see that both languages are not dense and hence not disjunctive. Therefore both of the atom Fibonacci languages are midst-languages. It has been shown that for the one letter alphabet case, there is no midst-language ([10]). Completely disjunctive languages (see [3]) are a type of the languages which contain no midst-sublanguages.

For the case of the alphabet X consisting of more than one letter, the family of all languages over X is then divided into three subfamilies. They are the family of all regular languages $\mathcal{R}$, which included the family of all finite languages; the family of all disjunctive languages $\mathcal{D}$ and the family of all midst-languages $\mathcal{M}$. That is, $2^{X^*} = \mathcal{R} \cup \mathcal{D} \cup \mathcal{M}$ and $\mathcal{R}$, $\mathcal{D}$, $\mathcal{M}$ are disjoint. The family of all regular languages is the most important family of languages in the formal language theory. The properties of regular languages is well developed. For the disjunctive languages, many algebraic properties have been carried out for the last two decades. The purpose of this paper is a study of the algebraic properties of midst-languages.

We call a word $f \in X^+$ a *primitive word* if f is not a power of any other word in X^+, i.e., $f = g^n$, $g \in X^+$, implies $n = 1$ and $f = g$ (see [6]). The set of all primitive words over X will be denoted by Q.

In the rest of this paper, we let the alphabet X concerned consist of at least two letters if it is not mentioned specially. This paper is organized in this way: In Section 2, we characterize the midst-languages L of the form $L = \bigcup_{f \in \Lambda} f^+$, $\Lambda = uv^+w$, $u, v, w \in X^+$. Section 3 concerns the properties of dense regular languages. We conjecture that there exists no dense regular subset in Q. Section 4 is devoted to the study of some algebraic properties of midst-languages. An example that the union of two disjoint midst-languages is disjunctive is given in Section 4. Furthermore, in Section 5, the decomposition of languages into unions of midst-languages is investigated.

Items not defined here or in the subsequence sections can be found in books [8] and [9], which we use as standard references. The following lemmas concerning the basic properties of primitive words will be needed in the sequel.

Lemma 1.1.([6]) *Let $f, g \in Q$ with $f \neq g$. Then $f^m g^n \in Q$ for all $m, n \geq 2$.*

Lemma 1.2.([9]) *Let $uv = f^i$, $u, v \in X^+, f \in Q, i \geq 1$. Then $vu = g^i$ for some $g \in Q$.*

We remark here that in Lemma 1.2, g is a *conjugate* of f, that is, $g = f_2 f_1$ if $f = f_1 f_2$ for some $f_1, f_2 \in X^*$.

Lemma 1.3.([6]) *Let $u, v \in X^+$. If u and v have powers u^m and v^n with a common initial segment of length* $\lg(u) + \lg(v)$*, then u and v are powers of a common word.*

Lemma 1.4.([3]) *If $u \in X^+$, then there is a unique primitive word f and a unique positive integer $k \geq 1$ such that $u = f^k$.*

The primitive word f in the above lemma will be call the *primitive root* of the word u and is denoted by $\lambda(u)$. For a language $L \subseteq X^+$, by $\lambda(L)$ we mean the language $\lambda(L) = \{\lambda(u) \mid u \in L\}$.

Lemma 1.5.([9]) *Let* $u, v \in X^+$*,* u^m *and* v^n *have a common subword of length* $\lg(u) + \lg(v)$*. Then* $\lambda(u)$ *is a conjugate of* $\lambda(v)$*.*

Lemma 1.6.([6]) *Let* $u, v \in X^+$*. If* $uv = vu$*, then* u *and* v *are powers of a common primitive word, that is,* $\lambda(u) = \lambda(v)$*.*

2. Characterizations of Some Midst-Languages

A language L is called *global* (*co-global*) if $\lambda(L) = Q$ ($\lambda(L) = Q \backslash F$ where F is a finite set). It has been shown that for a coglobal language L, there are infinitely many primitive words f such that $f^+ \subset L$ ([1]). First let us see the following:

Proposition 2.1.([11]) *Let* $L \subseteq X^+$*. Then* L *is dense if and only if* $\lambda(L)$ *is dense.*

It is known that if a language L is disjunctive, then $\bar{L} = X^* \setminus L$ is disjunctive too. Since every disjunctive language is dense, by Proposition 2.1, $L = \bigcup_{f \in \Lambda} f^+$ being disjunctive implies that both Λ and $Q \setminus \Lambda$ are dense. It is interesting to investigate the properties of the language $L = \bigcup_{f \in \Lambda} f^+$ in terms of Λ, where Λ is a regular component, that is, Λ is of the form uv^+w for some $v \in X^+$, $u, w \in X^*$. In particular we like to know the case of $L = \bigcup_{f \in \Lambda} f^+$ being a midst-language.

Proposition 2.2. *Let* $\Lambda = uv^+w$*, where* $u, v, w \in X^+$*. Then the language* $L = \bigcup_{f \in \Lambda} f^+$ *is not dense.*

Proof. We show L not dense by showing that $\lambda(L) = \lambda(\Lambda) = \lambda(uv^+w)$ is not dense. Let $m = \lg(uvw)$ and let $q = a^{3m}b^{3m}$ for $a, b \in X$, $a \neq b$. Suppose there exist $x, y \in X^*$ such that $xqy = xa^{3m}b^{3m}y = uv^iw$ for some $i \geq 1$. As $\lg(xa^m) > \lg(u)$ and $\lg(b^my) > \lg(w)$, there exist $x', y' \in X^+$ such that $x'a^{2m}b^{2m}y' = v^j$ for some $1 \leq j \leq i$. Since $2m > \lg(v^2)$, v is a factor of a^{2m} and also a factor of b^{2m}. Thus $v = a^k = b^k$, where $k = \lg(v)$. This implies that $a = b$, a contradiction! Then clearly $X^*qX^* \cap uv^+w = \emptyset$. Thus uv^+w is not dense. By Proposition 2.1, $\lambda(uv^+w)$ is not dense. As $\lambda(L) = \lambda(uv^+w)$, again by Proposition 2.1, L is not dense.

Proposition 2.3. *Let $\Lambda = uv^+w$, where $u, v, w \in X^+$ and let $L = \bigcup_{f\in\Lambda} f^+$. Then the following statements hold true:*

(1) *If $v = g^i$ and $wu = g^{ri}$ for some $g \in Q$, $i \geq 1$, $r \geq 1$, then L is regular.*

(2) *If $v = g^i$ and $wu = g^{ri+s}$ for some $g \in Q$, $i > 1$, $r \geq 1$ and $1 \leq s < i$, then L is not regular.*

(3) *If $v = g^{ri}$ and $wu = g^i$ for some $g \in Q$, $i \geq 1$ and $r \geq 2$, then L is not regular.*

(4) *If $v = g^{ri+s}$ and $wu = g^i$, $g \in Q$, $i \geq 2$, $r \geq 1$, $1 \leq s < i$, then L is not regular.*

(5) *If $\lambda(v) \neq \lambda(wu)$, then L is not regular.*

Proof. Let $\Lambda' = v^+wu$ and let $L' = \bigcup_{f'\in\Lambda'}(f')^+$. In statements (1)–(4), as $\lambda(v) = \lambda(wu) = g$ for some $g \in Q$, for each $f' \in \Lambda'$, $f' = g^t$ for some $t \geq 1$. It is true that L' is regular if and only if L is regular. Thus in the following proof of statements (1)–(4), we only show whether L' is regular or not.

(1) Suppose $v = g^i$ and $wu = g^{ri}$ for some $g \in Q$ and $i, r \geq 1$. One must have that $L' = \bigcup_{k\geq 1}(g^{(k\cdot i)+ri})^+ = \bigcup_{k\geq 1}(g^{(k+r)i})^+ = \bigcup_{k\geq r+1}(g^{k\cdot i})^+ = \bigcup_{m\geq 1}(\bigcup_{k\geq r+1} g^{m\cdot k\cdot i}) = \bigcup_{k\geq r+1} g^{k\cdot i} = \bigcup_{k\geq 1} g^{(k+r)i} = (g^i)^+g^{ri}$. Therefore, L' is a regular language.

(2) Let $v = g^i$ and $wu = g^{ri+s}$ for some $g \in Q$, $i > 1$, $r \geq 1$ and $1 \leq s < i$. Then $L' = \bigcup_{k\geq 1}(g^{ki+ri+s})^+$. Let $d = \gcd(i, s)$. Then $i = di'$ and $s = ds'$ for some $i', s' \geq 1$ such that i' and s' are relatively prime. Thus $L' = \bigcup_{k\geq 1}((g^{ki'+ri'+s'})^d)^+ = \bigcup_{k\geq 1}((g^{ki'+ri'+s'})^+)^d$. As $i > s$, $i' > s' \geq 1$. Suppose there exist $k, m, n \geq 1$ such that $(i')^{n+1} + s'(i')^n = m(ki' + ri' + s')$. It is true that i' and $ki'+ri'+s'$ are relatively prime and $(i')^n$ is a factor of $(i')^{n+1}+s'(i')^n$. One must have that $m = (i')^n m'$ for some $m' \geq 1$. This implies that $i' + s' = m'(ki' + ri' + s') \geq 2i' + s'$ which is impossible. Thus $d((i')^{n+1} + s'(i')^n) \neq dm(ki' + ri' + s')$ for any $k, m, n \geq 1$. That is, $g^{i(i')^n} g^{s(i')^n} = g^{d(i'+s')(i')^n} \notin L'$ for every $n \geq 1$. Let $t \geq 1$ be such that $(i')^t > r$ and let $m \geq n + t$. Then $(i')^{m-n} \geq (i')^t > r$ and $i(i')^m + s(i')^n = (i')^n(i(i^{m-n} - r) + ri + s)$. That is, $g^{i(i')^m} g^{s(i')^n} = g^{i(i')^m+s(i')^n} \in L'$. This implies that if we consider $m = n + jt$ for any $j \geq 1$, then $g^{i(i')^n} \not\equiv g^{i(i')^{n+jt}} (P_{L'})$. In particular, if we take for some $j \geq 1$ and $n = 1,\ 1+t,\ 1+2t,\ 1+3t,\ \cdots$, then all the following are in different $P_{L'}$

equivalent classes:

$$g^{ii'},\ g^{i(i')^{1+t}},\ g^{i(i')^{1+2t}},\ g^{i(i')^{1+3t}},\ g^{i(i')^{1+4t}},\ldots.$$

Hence there are infinitely many $P_{L'}$ classes and L' is not regular.

(3) Let $v = g^{ri}$; $wu = g^i$ for some $g \in Q$, $i \geq 1$ and $r \geq 2$. Then $L' = \bigcup_{k\geq 1}(g^{i(kr+1)})^+$. Suppose there exist $k, m, n \geq 1$ such that $ir^n + ir^n = mi(kr+1)$. Then $2r^n = m(kr+1)$. Since r and $kr+1$ are relatively prime, $m = r^n m'$ for some $m' \geq 1$. This implies that $2 = m'(kr+1) \geq 3$ which is impossible. Thus $ir^n + ir^n \neq mi(kr+1)$ for any $k, m, n \geq 1$. Therefore, $g^{ir^n} g^{ir^n} \notin L'$ for any $n \geq 1$. For every $j \geq 1$, $ir^{n+j} + ir^n = r^n i(r^j + 1)$. That is, $g^{ir^{n+j}} g^{ir^n} \in L'$. This implies that $g^{ir^n} \not\equiv g^{ir^t}(P_{L'})$ for every $t > n$. Clearly, all the following are in different P'_L classes:

$$g^{ir},\ g^{ir^2},\ g^{ir^3},\ \ldots.$$

Hence there are infinitely many $P_{L'}$ classes and L' is not regular.

(4) Let $v = g^{ri+s}$; $wu = g^i$ for some $g \in Q$, $i \geq 2$, $r \geq 1$, $1 \leq s < i$. Then $L' = \bigcup_{k\geq 1}(g^{k(ri+s)+i})^+$. Let $d = \gcd(i, s)$. Then $i = i'd$ and $s = s'd$ for some $i', s' \geq 1$ such that i' and s' are relatively prime. Thus $L' = \bigcup_{k\geq 1}((g^{k(ri'+s')+i'})^+)^d$. As $i > s$, $i' > s' \geq 1$. For every $n, j \geq 1$, $(ri' + s')^{n+j} + (ri' + s')^n i' = (ri' + s')^n((ri' + s')^j + i')$. Thus $g^{d((ri'+s')^{n+j}} g^{di'(ri'+s')^n} = g^{(ri'+s')^n((ri'+s')^{j-1}(ri+s)+i)} \in L'$. Suppose that there exist $m, k \geq 1$ such that $(ri' + s')^n + (ri' + s')^n i' = m(k(ri' + s') + i')$. Since $ri' + s'$ and $k(ri' + s') + i'$ are relatively prime and $(ri' + s')^n$ is a factor of $(ri' + s')^n(1 + i')$, $m = (ri' + s')^n m'$ for some $m' \geq 1$. One must have that $1 + i' = m'(r(i' + s') + i') \geq s' + 2i'$ which is impossible. Thus $(ri' + s')^n + (ri' + s')^n i' \neq m(k(ri' + s') + i')$ for every $k, m \geq 1$. That is, $d(ri' + s')^n + di'(ri' + s')^n \neq dm(k(ri' + s') + i')$. It follows that $g^{d(ri'+s')^n} g^{di'(ri'+s')^n} \notin L'$ for any $n \geq 1$. This implies that $g^{d(ri'+s')^n} \not\equiv g^{d(ri'+s')^{n+j}}(P_{L'})$ for every $n, j \geq 1$. Clearly, all the following are in different P'_L classes:

$$g^{ri+s},\ g^{(ri+s)(ri'+s')},\ g^{(ri+s)(ri'+s')^2},\ \ldots.$$

Hence there are infinitely many $P_{L'}$ classes and L' is not regular.

(5) Consider $i < j$. Clearly $(uv^iw)^n \in L$ for any $n \geq 1$. Then $uv^jw(uv^iw)^{n-1} \notin L$ if $n-1 > \lg(uv^jw) \geq 2$. It follows that $uv^iw \not\equiv uv^jw(P_L)$, if $i \neq j$. Thus, there are infinitely many different P_L classes. This implies that L is not regular.

Propositions 2.2 and 2.3 yield the following characterization of the language $L = \bigcup_{f\in\Lambda} f^+$ been midst-language, where Λ is a regular component.

Theorem 2.4. *Let $\Lambda = uv^+w$, where $u, v, w \in X^+$ and let $L = \bigcup_{f\in\Lambda} f^+$. Then the following statements are equivalent:*
(1) *L is a midst-language.*
(2) *Either $\lambda(v) \neq \lambda(wu)$ or $\lambda(v) = \lambda(wu)$ and $\lg(wu) \neq r\lg(v)$ for any $r \geq 1$ holds true.*
(3) *L is not regular.*

Example. Let $X = \{a, b\}$, $\Lambda = ab^+a$ and let $L = \bigcup_{f\in\Lambda} f^+$. It is easy to see that $ab^i \not\equiv ab^j(P_L)$, if $i \neq j$. Thus there are infinitely many P_L classes and, hence, L is not regular. As $L \cap X^*a^2X^* = \emptyset$, L is not dense. Hence, L is not disjunctive. One must have that L is a midst-language. This fact can easily be seen from (2) of Theorem 2.4; in fact $\lambda(v) = \lambda(b)$ and $\lambda(wu) = \lambda(a^2) = a$ and $a \neq b$.

3. Dense Regular Languages

The purpose of this section is to investigate some properties of dense regular languages. It is clear that if a dense language is expressed as a finite union of subsets, then at least one of the subsets is dense. For every $L \subseteq X^*$ and $u \in X^*$, the *left quotient* $u^{-1}L$ and the *right quotient* Lu^{-1} of L by u are defined as

$$u^{-1}L = \{x \in X^* \mid ux \in L\};$$

$$Lu^{-1} = \{x \in X^* \mid xu \in L\}.$$

It is known that the family of all regular languages is closed under left quotient, right quotient and catenation (see [8]). Hence for any regular language L, both $u(u^{-1}L)$ and $(Lu^{-1})u$ are regular subsets of L.

If a deterministic finite automaton (in short, dfa) M has the minimal number of states of the dfas' which accept the same regular language, then the dfa M is called a *minimal dfa* (in short, an *mdfa*). It is known that a language is regular iff it is accepted by an mdfa. Consider a dfa $M = (S, X, \delta, q_0, F)$, and a word $u = a_1 a_2 \cdots a_m \in X^*$ where $a_1, a_2, \ldots, a_m \in X$. Then a sequence of states $q_{i_0}, q_{i_1}, \ldots, q_{i_m} \in S$ is a *path* of u on M iff $\delta(q_{i_{j-1}}, a_j) = q_{i_j}$, $1 \le j \le m$. A path of u on M is *simple* iff for $0 \le r < s \le m$, $q_{i_r} = q_{i_s}$ implies $r = 0$ and $s = m$. That is, there do not exist two states in a simple path being the same state except the case that the first state may equal to the last state. If $q_{i_0}, q_{i_1}, \ldots, q_{i_m}$ is a path of u with $q_{i_0} = q_0$ and $q_{i_m} \in F$, then the path of u on M is called a *successful path* with label of u or a *successful path* of u. Since M is deterministic, the successful path of a word is unique and only words in the language accepted by M have successful paths.

Proposition 3.1. *Let $L \subseteq X^+$ be a dense regular language. Then for every $p \in Q$, there exist $x, y \in X^*$, $n, k \ge 1$ such that $\lg(x) < n$, $\lg(y) < n$ and $x(p^k)^* y \subseteq L$.*

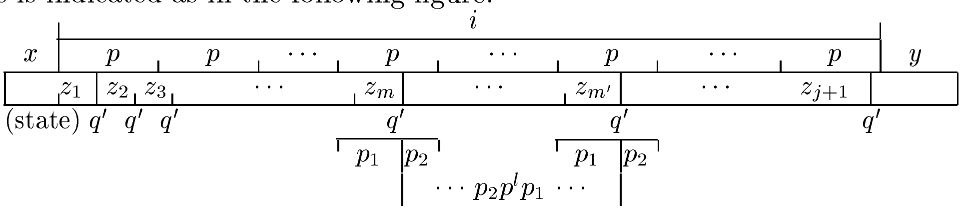
Proof. Let L be a dense regular language. Let M be a dfa with n states which accepts the language L. As L is dense, for every $p \in Q$ and for any integer $i \ge 1$, there are words $x, y \in X^*$ such that $xp^i y \in L$. Take i very big, say $i \ge 3(lg(p) + 1)n$ and $xp^i y \in L$. Let $q_0, q_1, q_2, \cdots, q_r$ be the successful path of $xp^i y$. Since i is sufficnetly big, there exist a state q' and words $z_1, z_2, \cdots, z_j, z_{j+1}$ with $lg(z_k) \le n$ for $k = 1, 2, \cdots, j+1$ and $j \ge 3lg(p) + 1$, having the transitions

$$\delta(q_0, xz_1) = \delta(q_0, xz_1 z_2) = \ldots = \delta(q_0, xz_1 z_2 \cdots z_{j+1}) = q'.$$

There is no difficult to realize that there at least two q' will locate at the same position of the word p. Let $p_1 p_2 = p$ be such that

$$\delta(q_0, xz_1 z_2 \cdots z_m) = \delta(q_0, xz_1 z_2 \cdots z_m z_{m+1} \cdots z_{m'}) = q',$$

where $z_1 z_2 \cdots z_m = p^r p_1$, $z_{m+1} \cdots z_{m'} = p_2 p^l p_1$, $p_1, p_2 \in X^*$, for some $r, l \ge 0$. This is indicated as in the following figure.



It follows that $x(p^r p_1)(p_2 p^l p_1)^k (p_2 p^{i-r-l-2})y = xp^r (p^{l+1})^k p^{i-r-l-1} y \in L$ for every $k \geq 0$. Now if $lg(xp^r) \geq n$, or $lg(p^{i-r-l-1}y) \geq n$, we can always choose $x', y' \in X^*$ with $lg(x') < n,\ lg(y') < n$ such that $x'(p^{l+1})^* y' \subseteq L$.

Proposition 3.2. *Let $L \subseteq X^+$ be a dense regular language and let M be a dfa which accepts L. Then there exist an $n \geq 1$ and $x, y \in X^*$ with $0 \leq \lg(xy) \leq n$ such that $xy \in L$, the path of xy on M is a simple successful path and $x((x^{-1}L)y^{-1})y$ is a dense regular subset of L.*

Proof. Let $L \subseteq X^+$ be a dense regular language and let M be an mdfa which accepts L with n states for some $n \geq 1$. Then for every final state q_f, there exists a word $u \in L$ such that the path of u starting from the initial state q_0 and ending in the state q_f is a simple successful path. We call such a path of u, P_f path. As there are finitely many states and $|X| < \infty$, there are only finitely many simple successful paths, say,

$$P_{f_1}, P_{f_2}, \ldots, P_{f_m},\ m \geq 1,$$

on an mdfa. For $1 \leq i \leq m$, let L_{f_i} be the set of all words in L which contains a sequence of subwords such that the related path with label of this sequence of subwords is the simple successful path P_{f_i}. Then L_{f_i} is accepted by the dfa M' which is the same as M except that M' has only one final state q_{f_i} that is the last state of P_{f_i}. Thus $L = \bigcup_{i \geq 1}^{m} L_{f_i}$ and each L_{f_i} is regular. This implies that at least one of L_{f_i} is a dense regular language. For this P_{f_i}, there are finitely many $u \in L$ such that the successful path with label of u is P_{f_i}. Let $u_1, u_2, \ldots, u_r$ be the words such that the successful paths of them are P_{f_i} and for each u_j, $1 \leq j \leq r$, let $L_{u_j} = \bigcup_{x,y \in X^*, xy = u_j} x((x^{-1}L_{f_i})y^{-1})y$. As L_{f_i} is regular, L_{u_j} is regular for each $1 \leq j \leq r$. From Proposition 3.1, one must have that the set $L' = \bigcup_{j \geq 1}^{r} L_{u_j}$ is a dense subset of L_{f_i}. Again, at least one of L_{u_j} is dense. Clearly, L_{u_j} is a dense regular subset of L. From the definition of L_{f_i}, it follows that $L_{u_j} = \bigcup_{x,y \in X^*, xy = u_j} x((x^{-1}L_{f_i})y^{-1})y = \bigcup_{x,y \in X^*, xy = u_j} x((x^{-1}L)y^{-1})y$. Again, there exist $x, y \in X^*$, $xy = u_j$, such that $x((x^{-1}L)y^{-1})y$ is a dense regular subset of L_{u_j} and, hence, a dense regular subset of L.

The following proposition generalize Theorem 2.4.

Proposition 3.3. *Let $\Lambda \subseteq Q$ be an infinite regular language. Then $L = \bigcup_{f \in \Lambda} f^+$ is not regular.*

Proof. Let $\Lambda \subseteq Q$ be an infinite regular language. Suppose that $L = \bigcup_{f \in \Lambda} f^+$ is regular. Since Λ is infinite, there exist $x, y, z \in X^*$ such that $y \neq 1$ and $xy^+z \subseteq \Lambda$. As $\Lambda \subseteq Q$, $zx \neq 1$ and $\lambda(y) \neq \lambda(zx)$. By the condition that L is regular, there are infinitely many $xy^{i_1}z$, $xy^{i_2}z, \ldots$ in a same P_L class. Since $\lambda(y) \neq \lambda(zx)$, for every $j \neq k$, $\lambda(xy^{i_j}z) \neq \lambda(xy^{i_k}z)$. By Lemma 1.1, $(xy^{i_j}z)^2(xy^{i_k}z)^2 \in Q$. Since P_L is a congruence, $xy^{i_j}z \equiv xy^{i_k}z(P_L)$ implies that $(xy^{i_j}z)^4 \equiv (xy^{i_j}z)^2(xy^{i_k}z)^2(P_L)$. As $xy^{i_j}z \in \Lambda \subseteq L$ and $xy^{i_j}z \equiv xy^{i_k}z(P_L)$ for every $j, k \geq 1$, the fact that $(xy^{i_j}z)^4 \in L$ implies that $(xy^{i_j}z)^2(xy^{i_k}z)^2 \in L$. Thus $(xy^{i_j}z)^2(xy^{i_k}z)^2 \in \Lambda$ for every $j \neq k$. But for any $j \geq 1$, $(xy^{i_j}z)^4 \notin \Lambda$. Now if $xy^{i_j}z \equiv xy^{i_k}z(P_\Lambda)$ for $j \neq k$, then $(xy^{i_j}z)^4 \equiv (xy^{i_j}z)^2(xy^{i_k}z)^2(P_\Lambda)$. Since Λ is a union of P_Λ-classes, it is not true that $(xy^{i_j}z)^2(xy^{i_k}z)^2 \in \Lambda$ and $(xy^{i_j}z)^4 \notin \Lambda$. Thus for every $j \neq k$, $xy^{i_j}z \not\equiv xy^{i_k}z(P_\Lambda)$. This contradicts the fact that Λ is regular. Therefore, L can not be regular.

Conjecture 3.4. The set Q contains no dense regular subsets.

If Conjecture 3.4 holds, then together with Proposition 3.3, one has the following result:

Conjecture 3.5. Let $\Lambda \subseteq Q$ be a regular language. Then $L = \bigcup_{f \in \Lambda} f^+$ is a midst-language if and only if Λ is infinite.

4. Some Algebraic Properties of Midst-Languages

In this section, we consider the catenation, complement and union of languages related to midst-property. First, we realize that the catenation of two disjunctive languages can be a midst-language. In fact, Reis and Shyr constructed such two languages in [7]. The languages are $D_1 = (aQ^ca \cup X^*b^2)$ and $D_2 = (aQ^ca \cup b^2X^*)$, where $Q^c = X^* \setminus Q$ and $a \neq b \in X$.

A language L is said to be an *infix code* (*prefix code*) if for every $x, y, u \in X^*$, the conditions $u \in L$ and $xuy \in L$ ($uy \in L$) togather imply $xy = 1$ ($y = 1$) (see [9]).

Lemma 4.1. *Let* L_1 *be a language which is not disjunctive. Then for any infix code* L_2 *the language* L_2L_1 *is not disjunctive.*

Proof. Let L_2 be an infix code and suppose L_1 is not disjunctive with $u \equiv v(P_{L_1})$. We show that the language L_2L_1 is not disjunctive by showing that $wu \equiv wv(P_{L_2L_1})$ for every $w \in L_2$. Suppose for some $x, y \in X^*$ such that $xwuy \in L_2L_1$. Then $xwuy = w'r$ for some $w' \in L_2$ and $r \in L_1$.

(1) $w' = x_1, r = x_2wuy$ for some $x_1x_2 = x, x_1, x_2 \in X^*$. Since $u \equiv v(P_{L_1})$, we have $x_2wvy \in L_1$ and so $xwvy \in L_2L_1$.

(2) $w' = xw_1, r = w_2uy, w_1w_2 = w$, for some $w_1, w_2 \in X^*$. Again we have $w_2vy \in L_1$ and so $xwvy \in L_2L_1$.

(3) That $w' = xwz$ for some $z \in X^*$ implies that L_2 is not an infix code. Thus case (3) can not happen.

Similary, $xwvy \in L_2L_1$ for some $x, y \in X^*$ imply that $xwuy \in L_2L_1$. We then conclude that $wu \equiv wv(P_{L_2L_1})$ and the proof is completed.

Let $L \subseteq X^+$ be a given non-empty language. A word $w \in L$ is a *left singular word* if the set $\{w, x\}$ is a prefix code for any $x \in L$. We call a language L a *left singular language* if L cantains at least one left singular word (see [9]). If L is a prefix code, then every word in L is a left singular word. Thus the family of all left singular languages contains properly the family of all prefix codes (see [9]). Since every infix code is a prefix code and every prefix code is a left singular language, we have the following:

Corollary 4.2. *If* L_1 *is a midst-language and* L_2 *is an infix code, then the language* L_2L_1 *is a midst-language.*

For $a \neq b \in X$, let $L_1 = a^+\{b^m \mid m \geq 1, m \neq 2^k, k \geq 2\}$; $L_2 = \{b^m \mid m \geq 1, m \neq 2^k, k \geq 2\}a^+$. Then it is easy to see that L_1 and L_2 are midst-languages while $L_1L_2 = a^+bb^+a^+$ is regular. Thus the family of midst-languages is not closed under catenation.

Lemma 4.3. *Let* L *be an infinite language. If* L *is not disjunctive, then for any finite language* F*, the language* FL *is not disjunctive.*

Proof. Let F be a finite language such that the maximal length of words in F is m. Suppose L is not disjunctive and $u, v \in X^+$, $u \neq v$, such that $u \equiv v(P_L)$.

Then since P_L is a congruence, for any $w \in X^+$, $w^m u \equiv w^m v(P_L)$. We now show that $w^m u \equiv w^m v(P_{FL})$ and FL is not disjunctive. Suppose on the contrary that $w^m u \not\equiv w^m v(P_{FL})$. Then there exist $x, y \in X^*$ such that, (say) $xw^m uy \in FL$ and $xw^m vy \notin FL$. Consider the following two cases:

(1) $x = x_1 x_2$, for some $x_1 \in X^+, x_2 \in X^*$ such that $x_1 \in F$ and $x_2 w^m uy \in L$. It is clear that $x_2 w^m vy \notin L$ and $u \not\equiv v(P_L)$, a contradiction.

(2) $xw^{k_1} w_1 \in F$, $w_2 w^{k_2} uy \in L$, where $w_1, w_2 \in X^*$, $w = w_1 w_2$ and $k_1 + k_2 - 1 = m$. Again, clearly $w_2 w^{k_2} vy \notin L$. It follow that $u \not\equiv v(P_L)$, a contradiction.

This shows that the condition $w^m u \equiv w^m v(P_{FL})$ is true and hence FL is not disjunctive.

Lemma 4.4. *Let L_1 be an infinite language. If L_1 is not regular, then for any left-singular language L_2, the language $L_2 L_1$ is not regular.*

Proof. Suppose L_1 is not a regular language. Then there exist infinitely many words $u_1, u_2, u_3, \ldots$ which are in different P_{L_1} classes. Let L_2 be a left-singular language and $w \in L_2$ be a left-singular word in L_2. As $u_i \not\equiv u_j(P_{L_1})$ for $i \neq j$, then there exist $x, y \in X^*$ such that $xu_i y \in L_1$ and $xu_j y \notin L_1$ or vice versa. Since w is a left-singular word in L_2, $wxu_i y \in L_2 L_1$ and $wxu_j y \notin L_2 L_1$ or vice versa. Thus $u_i \not\equiv u_j(P_{L_2 L_1})$ for $i \neq j$. This implies that $L_2 L_1$ is not regular.

Proposition 4.5.([15]) *For a language $L \subseteq X^+$ and an integer $i \geq 1$, if $L \subseteq X^i(X^* \backslash \{a^m \mid a \in X,\ m \geq 2\})$ and $X^i \cup X^{i+1} \subseteq L$, then $LQ = X^i(X^+ \backslash \{a^m \mid a \in X,\ m \geq 3\})$.*

From Proposition 4.5, one must have that $(X \cup X^2)Q$ is a regular language. From this fact we see that the condition 'left-singular' on L_2 in Lemma 4.4 can not be replaced by the condition 'finite'. However, from Lemmas 4.1 and 4.4, one has the following:

Corollary 4.6. *For a midst-language L and an infix code I, we have that IL is a midst-language.*

From Lemmas 4.3 and 4.4, one has the following result immediately.

Corollary 4.7. *Let L be a midst-language and let F be a finite left-singular language. Then FL is a midst-language.*

Since the complement of a regular language is regular and the complement of a disjunctive language is disjunctive, we have the following:

Proposition 4.8. *The complement* L^c *of a midst-language* $L \subseteq X^*$ *is also a midst-language, where* $L^c = X^* \setminus L$.

It has been shown that there are disjunctive languages D such that LD is disjunctive for any non-empty language L. Such a language is called a *disjunctive core*. Examples for disjunctive cores are $\bar{Q} = X^+ \backslash Q$ and $D(i)$ for $i \geq 4$ (see [15]). We then have the following:

Proposition 4.9. *In the semigroup* 2^{X^*} *the subset* $\mathcal{M}$ *of all midst-languages contains no ideal of* 2^{X^*}.

It is known that for a disjunctive language L and a regular language R, $L \setminus R$ is disjunctive. Thus if a disjunctive language $L = L_1 \cup L_2$, where L_1 and L_2 are two disjoint languages and both of them are not disjunctive, then L_1 and L_2 are not regular. This observation yields:

Corollary 4.10. *Let* L *be a disjunctive language and* $L = L_1 \cup L_2$ *with* $L_1 \cap L_2 = \emptyset$. *Then both of* L_1 *and* L_2 *are not disjunctive if and only if they are midst-languages.*

Furthermore, concerning the decomposition of a disjunctive language into a union of two disjoint subsets, we have the following result:

Proposition 4.11. *Let* L *be a disjunctive language and* $L = L_1 \cup L_2$ *with* $L_1 \cap L_2 = \emptyset$. *Then* L_1 *being not dense implies that* L_2 *is disjunctive.*

Proof. Let L_1 be not dense. Then there exists a word $w \in X^*$ such that $L_1 \cap X^*wX^* = \emptyset$. For any two distinct words $u, v \in X^*$, consider the words uw and vw. As L is a disjunctive language, there exist $x, y \in X^*$ such that $xuwy \in L$ and $xvwy \notin L$ or vice versa. This implies that $xuwy \in L_2$ and $xvwy \notin L_2$ or vice versa. That is, L_2 is a disjunctive language.

In order to construct languages of the form $L = L_1 \cup L_2$, where L is disjunctive and L_1 and L_2 are two disjoint non-disjunctive languages, we consider the following definitions.

For $u \in X^+$, we define the following two sets:

$$\mathrm{P}(u) = \{v \in X^+ \mid u = vw \text{ for some } w \in X^+\};$$

$$\mathrm{S}(u) = \{v \in X^+ \mid u = wv \text{ for some } w \in X^+\}.$$

Let X be an ordered alphabet, that is, $X = \{a_1, a_2, \ldots, a_n\}$ with $a_i < a_j$ if $i < j$. Then, on X^*, the *standard total order* $\leq$ is defined as follows ([9]): For any $u, v \in X^*$, $u < v$ if $\lg(u) < \lg(v)$ and $<$ is the lexicographic order on X^n for all $n \geq 1$. For a word $u \in X^*$, let $\#(u) = m$ if u stands at the m-th position in the ordered set $X^* = \{w_1 < w_2 < w_3 < \ldots\}$. We note that $w_1 = 1$ and $\#(w_1) = 1$.

A language $L \subseteq X^+$ is called a *solid code* ([14]) if L is an infix code and $\mathrm{P}(u) \cap \mathrm{S}(v) = \emptyset$ for every $u, v \in L$. Here we also give the following definition:

For $u \in X^+$, let $\mathrm{E}(u) = \{v \in X^+ \mid u = xvy \text{ for some } x, y \in X^*\}$.

For a language $L \subseteq X^+$ and for any $u \in X^+$, any factorization representation

$$u = x_1 y_1 x_2 y_2 \cdots x_n y_n x_{n+1},$$

such that $y_j \in L$ and $\mathrm{E}(x_i) \cap L = \emptyset$, where $j = 1, 2, \ldots, n$, $i = 1, 2, \ldots, n+1$, is called an *L-representation* of u. By [4], if L is a solid code, then the L-representation of every word $u \in X^+$ is unique.

Let $u, v \in X^+$. For $w_1, w_2 \in X^+$, (w_1, w_2) is said to be a *(u, v)-related pair* if there exist some $\{u, v\}$-representations of w_1, w_2 such that

$$w_1 = x_1 y_1 x_2 y_2 \cdots x_n y_n x_{n+1},$$

$$w_2 = x_1 y_1' x_2 y_2' \cdots x_n y_n' x_{n+1},$$

where $y_i, y_i' \in \{u, v\}$ and $u, v \notin \mathrm{E}(x_j)$, for $1 \leq i \leq n$, $1 \leq j \leq n+1$.

In the following, let $\{a, b, c\} \subseteq X$ and fix u_1, u_2, v as $u_1 = a^3ba^2b^3$, $u_2 = a^3c^3b^3$, $v = a^3b^2ab^3$. Then it is not difficult to show that $\{u_1, u_2, v\}$ is a solid code. Thus the $\{u_i, v\}$-representation of every $w \in X^+$ is unique, $i = 1, 2$. For $w \in X^+$, let $\mathrm{N}_{u_i}(w)$ and $\mathrm{N}_v(w)$ denote the number occurrences of u_i and v in the $\{u_i, v\}$-representation of w, respectively, $i = 1, 2$. That is, if $\mathrm{N}_{u_i}(w) + \mathrm{N}_v(w) = s$, then $w = x_1 y_1 x_2 y_2 \cdots x_s y_s x_{s+1}$ is the $\{u_i, v\}$-representation of w, $i = 1, 2$.

For any $w \in X^+$ with $u_1, v \notin \mathrm{E}(w)$, let $\bar{w} = b^m ab^9 wa^9 ba^m$ if $\#w = m$, and define the set H to be

$$H_1 = \{\bar{w} \mid w \in X^+,\ u_1, v \notin \mathrm{E}(w)\}.$$

For $0 \le r \le s$, we also define the following sets:

$$A_1(u_1, r, s) = \{w \mid w \in X^+,\ \mathrm{N}_{u_1}(w) = r,\ \mathrm{N}_{u_1}(w) + \mathrm{N}_v(w) = s\}.$$
$$\overline{A}_1(u_1, r, s) = \{ab^9wa^{9+q_1}ba^{9+q_2}\cdots ba^{9+q_{s+1}} \mid w \in A_1(u_1, r, s)$$

and $\#x_i = q_i$, $1 \le i \le s+1$, when the $\{u_1, v\}$-representation of w is $w = x_1y_1x_2y_2\cdots x_sy_sx_{s+1}\}$.

$A_1 = \bigcup_{r,s\ge 1} \overline{A}_1(u_1, 2^r, 2^s)$. Let $B_1 = (H_1 \cup A_1)u_1$.

Then from [14], we have the following result: (We make a minor change of B_1. The original definition of B_1 is $B_1 = H_1 \cup A_1$. However, it can be shown that this does not affect the result.)

Lemma 4.12.([14]) *The following are true*:

(1) $u_1v \equiv vu_1(P_{B_1})$.

(2) *If* $w_1, w_2 \in X^*$, $w_1 \ne w_2$, (w_1, w_2) *is not a* (u_1, v)*-related pair, then* $w_1 \not\equiv w_2(P_{B_1})$.

Thus B_1 is a midst-language over X. Similarly, let $u_1 \to u_2$. The languages H_2, $A_2(u_2, r, s)$, $\overline{A}_2(u_2, r, s)$, A_2, C_2 and B_2 are defined, respectively. Clearly, B_1 and B_2 are disjoint. Lemma 4.12 also holds true for the replacements of $u_1 \to u_2$ and $B_1 \to B_2$. Again, B_2 is a midst-language. We now show that $B_1 \cup B_2$ is a disjunctive language.

Proposition 4.13. *The language* $B_1 \cup B_2$ *is disjunctive.*

Proof. Let $w_1, w_2 \in X^*$, $w_1 \ne w_2$, be such that (w_1, w_2) is neither a (u_1, v)-related pair nor a (u_2, v)-related pair. Let $w_3 = u_1w_1u_1$ and let $w_4 = u_1w_2u_1$. Then (w_3, w_4) is neither a (u_1, v)-related pair nor a (u_2, v)-related pair. By Lemma 4.12, $w_3 \not\equiv w_4(P_{B_1})$ and $w_3 \not\equiv w_4(P_{B_2})$. Moreover, from the definitions of B_1 and B_2, for $x, y \in X^*$, $xw_3y \in B_1$ implies $xw_4y \notin B_2$ and vice versa. Similarly, $xw_3y \in B_2$ implies $xw_4y \notin B_1$ and vice versa. This implies that $w_3 \not\equiv w_4(P_{B_1\cup B_2})$. Hence $w_1 \not\equiv w_2(P_{B_1\cup B_2})$.

Let $w_1, w_2 \in X^*$, $w_1 \ne w_2$, be such that (w_1, w_2) is a (u_1, v)-related pair or a (u_2, v)-related pair. However, from the definition of u_1, u_2, v, there exist no $w_1, w_2 \in X^*$, $w_1 \ne w_2$ such that (w_1, w_2) is a (u_1, v)-related pair and also a

(u_2, v)-related pair. Let (w_1, w_2) be a (u_1, v)-related pair but not a (u_2, v)-related pair. Let $w_3 = u_2 w_1 u_1$ and let $w_4 = u_2 w_2 u_1$. Then (w_3, w_4) is a (u_1, v)-related pair but not a (u_2, v)-related pair. By Lemma 4.12, $w_3 \not\equiv w_4(P_{B_2})$. Moreover, from the definitions of B_1 and B_2, for $x, y \in X^*$, $xw_3y \in B_2$ implies $xw_4y \notin B_1$ and vice versa. Thus $w_3 \not\equiv w_4(P_{B_1 \cup B_2})$. Hence $w_1 \not\equiv w_2(P_{B_1 \cup B_2})$. Similarly, if (w_1, w_2) be a (u_2, v)-related pair but not a (u_1, v)-related pair, one must also have $w_1 \not\equiv w_2(P_{B_1 \cup B_2})$.

Therefore, $B_1 \cup B_2$ is a disjunctive language.

5. Midst-Language-Decompositions of X^*

Languages that can be expressed as a disjoint union of infinitely many disjunctive languages are called *disjunctive splittable languages*. Decomposition of languages into a disjoint union of disjunctive languages has been studied by several peoples. See for example [5] and [12]. On the other hand the authors have studied the so called regular component splittable languages ([13]). We now at the stage to deal with the decomposition of some languages into infinitely many disjoint union of midst-languages.

Let D be the Dyck language over $X = \{a, b\}$. Clearly the complement $D^c = X^* \setminus D$ of D is also a midst-language. Since for any $n \geq 1$, $\{a^n b^n\}$ is an infix code, each language $a^n b^n D \subset D$ is a midst-language. It is also clear that for any $n, m \geq 1$, with $n \neq m$, $a^n b^n D \cap a^m b^m D = \emptyset$. Thus the following is a disjoint union:

$$D' = abD \cup a^2b^2D \cup a^3b^3D \cup \cdots \cup a^nb^nD \cup \cdots.$$

If we let $D_1 = D \backslash D'$, then $D = D' \cup D_1$ is a disjoint union and D_1 is a midst-language. Indeed, for $i \neq j$, $a^i bab^i \in D_1$, while $a^j bab^i \notin D_1$ and hence $a^i \not\equiv a^j(P_{D_1})$. Thus D_1 is not regular. It is easy to see that $a^2bab^2 \equiv a^3bab^3(P_{D_1})$, and so D_1 is not disjunctive.

Now we are able to decompose X^* into a disjoint union of infinitely many midst-languages for the case $X = \{a, b\}$ as follows:

Proposition 5.1. *For $X = \{a, b\}$, X^* is a disjoint union of infinitely many midst-languages.*

Proof. From the above discussion we see that the following decomposition is a disjoint union and each component is a midst-language.

$$X^* = D^c \cup (\bigcup_{n\geq 1} a^n b^n D) \cup D_1.$$

A language $I \subseteq X^*$ is called a *right ideal* if $IX^* \subseteq I$. It is well known that every right ideal I of X^* can be expressed as $I = PX^*$ where P is a prefix code. In fact $P = I \backslash IX^+$ (see [9]). We have the following:

Proposition 5.2. *Let $X = \{a, b\}$. If I is a right ideal of X^* such that $I = PX^*$ and P is an infix code, then I is a disjoint union of infinitely many midst-languages.*

Proof. From the proof of Proposition 5.1, one must have $X^* = D^c \cup (\bigcup_{n\geq 1} a^n b^n D) \cup D_1$. Since $I = PX^*$ and P is an infix code, by assumption, we have

$$\begin{aligned} I = PX^* &= P(D^c \cup (\bigcup_{n\geq 1} a^n b^n D) \cup D_1) \\ &= PD^c \cup (\bigcup_{n\geq 1} Pa^n b^n D) \cup PD_1. \end{aligned}$$

By Lemma 4.1, each component in the above is a midst-language and the proposition holds.

References

[1] R. K. Chang and H. J. Shyr, *Global and Coglobal Languages*, Algebra Colloq., 2:1(1995), 11-22.

[2] Chen-Ming Fan and H. J. Shyr, *Some properties of fibonacci languages*, Tamkang Journal of Mathematics, 27:2(1996), 165-182.

[3] S. W. Jiang, H. J. Shyr and S. S. Yu, *Completely disjunctive languages*, Periodica Polytechnica Transportation Engineering, (Proceeding of the second international mathematical mini conference, Budapest 1988, Part II), 19:1-2(1991), 101-110.

[4] H. Jürgensen and S. S. Yu, *Solid Codes*, J. Inf. Process. Cybern., EIK 26:10(1990), 563-574.

[5] Masashi Katsura and H. J. Shyr, Decomposition of languages into disjunctive outfix codes, Semigroups, Theory and Application, Proceedings, Oberwolfach, (1986), Edited by H. Jürgensen, G. Lallement, H.J. Weiert, Lecture Note in Mathematics 1320, Springer-Verlag, Berlin, 1988, 172-175.

[6] R. C. Lyndon and M. P. Schützenberger, *The equation $a^M = b^N c^P$ in a free group*, Michigan Math. J., 9(1962), 289-298.

[7] C. M. Reis and H. J. Shyr, *Some properties of disjunctive languages on a free monoid*, *Information and Control*, 37:3(1978), 334-344.
[8] A. Salomaa, Formal Languages, Academic Press, New York, 1973.
[9] H. J. Shyr, Free Monoids and Languages, Lecture Notes, Hon Min Book Company, Taichung, 1991.
[10] H. J. Shyr, *Disjunctive languages on a free monoid*, Information and Control, 34(1977), 123-129.
[11] H. J. Shyr and Din-Chang Tseng, *Some Properties of dense languages*, Soochow Journal of Mathematics, 10(1984), 127-131.
[12] H. J. Shyr and Y. S. Tsai, Disjunctive context-free languages, International Symposium on the Semigroup Theory and Its Related Fields, Proceedings, Kyoto, Japan, Edited by M. Yamada and H. Tominaga, 1990, 213-223.
[13] H. J. Shyr and S. S. Yu, *Regular component splittable languages*, Acta Math. Hungar. (To appear).
[14] Shyr, H. J. and S. S. Yu, *Solid codes and disjunctive domains*, Semigroup Forum, 41(1990), 23-37.
[15] S. S. Yu, *Properties of annihilators of languages*, Semigroup Forum, 56(1998), 49-69.

Department of Applied Mathematics, National Chung-Hsing University, Taichung, Taiwan.