行政院國家科學委員會專題研究計畫 成果報告

自動稠密性及反自動稠密性言語的研究
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 98-2115-M-040-001-
執 行 期 間 ：98 年 08 月 01 日至 99 年 07 月 31 日
執 行 單 位 ：中山醫學大學應用資訊科學學系（所）

計畫主持人：黃政治

計畫參與人員：碩士級-專任助理人員：葉信賢

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 99 年 10 月 11 日

Acta Informatica (2010) 47:209–219
DOI 10.1007/s00236-010-0116-5

ORIGINAL ARTICLE

A note on autodense related languages

Chen-Ming Fan · C. C. Huang · H. J. Shyr · Kuo-Hsiang Chen

Received: 26 February 2009 / Accepted: 23 March 2010 / Published online: 9 May 2010

Abstract In this paper, some algebraic properties of autodense languages and pure autodense languages are studied. We also investigate the algebraic properties concerning anti-autodense languages. The family of anti-autodense languages contains infix codes, comma-free codes, and some subfamilies of new codes which are anti-autodense prefix codes, anti-autodense suffix codes and anti-autodense codes. The relationships among these subfamilies of new codes are investigated. The characterization of $L^n, n \geq 2$, which are anti-autodense is studied.

1 Introduction

Both regular languages and disjunctive languages are especially important applications in the field of formal languages. Recall that every regular language is accepted by a finite automaton [10]. It is the union of the equivalence classes of a congruence relation of finite index. Moreover, from the definition of disjunctive language, every disjunctive language

This work was supported by the National Science Council R.O.C. under Grant NSC 98-2115-M-040-001.

C.-M. Fan
Department of Information Management,
National Chin-Yi University of Technology, Taichung 411, Taiwan

C. C. Huang
School of Applied Information Sciences, Chung Shan Medical University, Taichung 402, Taiwan

C. C. Huang (✉)
Information Technology Office, Chung Shan Medical University Hospital, Taichung 402, Taiwan
e-mail: cchuang@csmu.edu.tw

H. J. Shyr
Department of Applied Mathematic, National Chung-Hsing University, Taichung 402, Taiwan

K.-H. Chen
Liberal Arts Center, Da-Yeh University, Chang-Hwa 515, Taiwan

has infinitely many congruence classes. This yields that no disjunctive language is regular. For the definition and properties of disjunctive languages, one is referred to [1,5,11,13] or [20]. To simplify the judgment of disjunctivity of a language immediately, Ries and Shyr [11] indicate that every disjunctive language is dense. The characteristic of dense provides the method to check whether a language is disjunctive. We firstly check whether a language is dense because checking whether it is dense or not is much easier than checking whether it is disjunctive. This study motivates the investigation of denseness property of a language. There are some researches related to dense languages. The definition of dense is given in [7]. In [18] the authors consider the subset of X^* named dense domain. One has the interesting result that is a language is a dense domain if and only if it is dense. Some characterizations of dense languages have been studied in [15]. Recently, the investigation concerning the classifications of dense languages has been studied in [8]. More properties of dense languages can be found in [14] or [16]. Furthermore, we study algebraic properties of autodense languages and anti-autodense languages in this paper.

This paper is organized into several sections. The first section introduces the overview of this paper. In the second section, we display some well-known definitions and properties applied in this paper. Moreover, the definitions of autodense languages and anti-autodense languages are given. The relationships of the families of languages concerning autodense property are presented in the section. In the third section, we study some algebraic properties concerning autodense languages and pure autodense languages. There is a subset of an autodense language that is not autodense. The union of autodense languages is autodense and the finite union of pure autodense languages is pure autodense. It can be shown that the family of all autodense languages is not closed under catenation. Moreover, we provide a method to construct a union of infinitely many pure autodense languages which is dense. In the meanwhile, a procedure is provided to construct a discrete autodense language contained in a dense language. In the final section, some algebraic properties of anti-autodense languages are studied. Union and catenation of two anti-autodense languages may not be an anti-autodense language. However, the families of all anti-autodense prefix codes and all anti-autodense suffix codes are closed under catenation. Moreover, we also investigate the characterization of L^n, $n \geq 2$ which are anti-autodense languages.

2 Definitions and preliminaries

In this paper the alphabet X containing more than one letter is assumed. Let X^* be the free monoid generated by X. Every element of X^* is a *word* and every subset of X^* is a *language*. Let 1 denote the empty word, and $X^+ = X^* \backslash \{1\}$. A language $L \subseteq X^*$ is *dense* if for any $w \in X^*$, there exist $x, y \in X^*$ such that $xwy \in L$. That is, for every $w \in X^*$, $X^*wX^* \cap L \neq \emptyset$. A set $S \subseteq X^*$ is called *dense domain*, if for any $D \subseteq X^*$, the property "$X^*uX^* \cap D \neq \emptyset$, for all $u \in S$" implies that D is dense. A *primitive word* is a word which is not a power of any other word. Let Q be the set of all primitive words over X. Every word $u \in X^+$ can be expressed as a power of a primitive word in a unique way, that is, for any $u \in X^+$, $u = f^n$ for a unique $f \in Q$ and $n \geq 1$. In this manner, f is the primitive root of u and denoted by $f = \sqrt{u}$. For a language L, let $\sqrt{L} = \{\sqrt{u} \mid u \in L\}$. A language L is a global (coglobal) language if $\sqrt{L} = Q(\sqrt{L} = Q\backslash F$, where F is a finite language). A *forest language* [2] is a language L such that $L = \bigcup_{f \in \Lambda} f^+$ for some $\Lambda \subseteq Q$.

Moreover, some classes of codes in this paper are defined as follows. A language $L \subseteq X^+$ is called a *code* if $x_1x_2 \ldots x_m = y_1y_2 \ldots y_n$ and $x_i, y_j \in X$ for all $1 \leq i \leq m, 1 \leq j \leq n$

imply that $m = n$ and $x_i = y_i$ for all $1 \leq i \leq n$. For any two words $u, v \in X^*$, $u \leq_p v (u \leq_s v)$ if $v = ux (v = xu)$ for some $x \in X^*$. Meanwhile, $u <_p v (u <_s v)$ denotes that $u \leq_p v (u \leq_s v)$ and $u \neq v$ for $u, v \in X^*$. A language $L \subseteq X^+$ is an *infix code* if for all $x, y, u \in X^*$, $u, xuy \in L$ together imply $x = y = 1$. A language $L \subseteq X^+$ is a *prefix code* (*suffix code*) if $L \cap LX^+ = \emptyset (L \cap X^+L = \emptyset)$. A language $L \subseteq X^+$ is a *bifix code* if it is both a prefix code and a suffix code. It follows immediately that an infix code is a bifix code. A language $L \subseteq X^+$ is an *intercode of index m* if $L^{m+1} \cap X^+L^mX^+ = \emptyset$, for $m \geq 1$. The family of intercodes is an important subfamily of bifix codes [19]. A *comma-free code* is an intercode of index 1. Some algebraic properties of intercodes and comma-free codes can be found in [4] or [19].

Definition 2.1 Let $L \subseteq X^+$. A language L is autodense if for any $w \in L$, there exist $x, y \in X^+$ such that $xwy \in L$.

No infix code is an autodense language. Since a comma-free code is an infix code [4], a comma-free code can never be an autodense language. Remark that every dense language is autodense. Let $\mathcal{L}_d$ be the family of all dense languages and let $\mathcal{L}_{au}$ be the family of all autodense languages over X. Therefore $\mathcal{L}_d \subset \mathcal{L}_{au}$.

Definition 2.2 A language is *pure autodense* if it is autodense but not dense.

An intercode of index 2 can be a pure autodense language. The well known context-free language $C = \{a^nb^n | n \geq 1\}$ is an example of pure autodense language.

If the set $L \subseteq X^+$ is not an autodense language, then there exists $w \in L$ such that $xwy \notin L$ for all $x, y \in X^+$. Such a language is *non-autodense*. Moreover, we consider a stronger version of this language which will cover all the infix codes and hence cover all of comma-free codes. We have the definition of *anti-autodense language* as follow.

Definition 2.3 A language L is anti-autodense if $L \cap X^+LX^+ = \emptyset$. Moreover, if a language L satisfies $L \cap X^+LX^+ = \emptyset$, then we say that L satisfies *anti-autodense condition.*

Every infix code has such property. Since a comma-free code is an infix code, every comma-free code is an anti-autodense language.

In the following, we will investigate some characterization between anti-autodense language and another codes. Firstly, we study some examples as follow.

In Guo et al. [3], it was shown that a maximal prefix code L is a right semaphore code if and only if $L \cap X^+LX^+ = \emptyset$. However, an anti-autodense language may not be a code. For example, the language $L = \{ab, aba, bab\}$ over $X = \{a, b\}$ satisfies the anti-autodense condition $L \cap X^+LX^+ = \emptyset$. But L is not a code because $(ab)(ab)(ab) = (aba)(bab)$.

Recall that an infix code is a prefix code and also a suffix code. Every infix code satisfies the anti-autodense condition. But a language satisfying the anti-autodense condition may not be an infix code. For example, the language $L = \left\{ab, ba^2, ba^2b\right\}$ satisfies the condition $L \cap X^+LX^+ = \emptyset$, but it is neither a prefix code nor a suffix code. By this example, we point out that there is a code L satisfying the condition $L \cap X^+LX^+ = \emptyset$ but L is not a bifix code.

Let $X = \{a, b\}$ and $L_1 = a^+b$, $L_2 = ab^+$, $L_3 = \left(ba^+b \backslash \left\{ba^3b\right\}\right) \cup \left\{ab, ba^2\right\}$, $L_4 = ab^+a \cup \{a\}$. All L_1, L_2, L_3 and L_4 satisfy the condition $L \cap X^+LX^+ = \emptyset$. Here, L_1 is a prefix code but not a suffix code. L_2 is a suffix code but not a prefix code. L_3 and L_4 are anti-autodense codes but neither prefix codes nor suffix codes. It is easy to construct an anti-autodense language which is not a code. The language $\left\{a, a^2\right\}$ is one. One of the same kind of infinite language is $\left\{a, a^2\right\} \cup ab^+$.

Beside the definition of an anti-autodense language, we give the definitions of codes related to anti-autodense. Let $L \subseteq X^+$.

Definition 2.4

(i) L is an *anti-autodense prefix code* if L is a prefix code, satisfying the condition $L \cap X^+LX^+ = \emptyset$.
(ii) L is an *anti-autodense suffix code* if L is a suffix code, satisfying the condition $L \cap X^+LX^+ = \emptyset$.
(iii) L is an *anti-autodense bifix code* if L is an anti-autodense prefix code and also an anti-autodense suffix code.
(iv) L is an *anti-autodense code* if L is a code, satisfying the condition $L \cap X^+LX^+ = \emptyset$.

From the above definitions, these families of languages exist and all are different families of languages. Let $\mathcal{L}_{aa}$, $\mathcal{L}_{aab}$, $\mathcal{L}_{aap}$, $\mathcal{L}_{aas}$ and $\mathcal{L}_{aac}$ represent them, that is,

- $\mathcal{L}_{aab}$: the family of all anti-autodense bifix codes over X.
- $\mathcal{L}_{aap}$: the family of all anti-autodense prefix codes over X.
- $\mathcal{L}_{aas}$: the family of all anti-autodense suffix codes over X.
- $\mathcal{L}_{aac}$: the family of all anti-autodense codes over X.

For convenience, the following notations are used.

- $\mathcal{L}_{aa}$: the family of all anti-autodense languages over X.
- $\mathcal{L}_p$: the family of all prefix codes over X.
- $\mathcal{L}_s$: the family of all suffix codes over X.
- $\mathcal{L}_b$: the family of all bifix codes over X.
- $\mathcal{L}_i$: the family of all infix codes over X.

Here $\mathcal{L}_{aab}$ is the family of all infix codes over X, that is, $\mathcal{L}_{aab} = \mathcal{L}_{aa} \cap \mathcal{L}_b = \mathcal{L}_i$. And $\mathcal{L}_{aap} = \mathcal{L}_{aa} \cap \mathcal{L}_p$; $\mathcal{L}_{aas} = \mathcal{L}_{aa} \cap \mathcal{L}_s$. The relationships of the families of languages are presented as follows. $\mathcal{L}_{aab} \subset \mathcal{L}_{aap} \subset \mathcal{L}_{aac} \subset \mathcal{L}_{aa}$; $\mathcal{L}_{aab} \subset \mathcal{L}_{aas} \subset \mathcal{L}_{aac} \subset \mathcal{L}_{aa}$.

Furthermore, there are some results used in the rest of this paper as follow.

Lemma 2.1 ([18]) *Q is a dense domain.*

Lemma 2.2 ([18]) *Let $L \subseteq X^+$. Then the following are equivalent*

(i) *L is dense.*
(ii) *$\sqrt{L}$ is dense.*
(iii) *L is a dense domain.*

From the previous lemma, we have that L is a dense language if and only if L is a dense domain. From now on, we will speak about dense languages instead of dense domains.

Lemma 2.3 *Let $A, B \subseteq X^*$. Then $A \cup B$ is dense if and only if either A or B is dense.*

Proof Let $A, B \subseteq X^*$. If A is dense, then clearly $A \cup B$ is also dense, for any B. Conversely, if neither A nor B is dense, then there exist $u, v \in X^+$ such that $X^*uX^* \cap A = \emptyset$ and $X^*vX^* \cap B = \emptyset$. We have $X^*uvX^* \cap A = \emptyset$ and $X^*uvX^* \cap B = \emptyset$. This implies that $X^*uvX^* \cap (A \cup B) = \emptyset$; hence $A \cup B$ is not dense. □

Lemma 2.4 ([18]) *Let L be a dense domain and let F be a finite subset of L. Then $L \backslash F$ is a dense domain.*

Lemma 2.5 *Let $L \subseteq X^*$. If L is dense, then L^+ is dense.*

Proof Let $L \subseteq X^*$ be a dense language. Note that if A is dense and $A \subseteq B$ for some $A, B \in X^*$, then B is dense. Since $L \subset L^+$, it is clear that L^+ is dense. □

Lemma 2.6 ([19]) *Let $L \subseteq X^*$. If L is an intercode, then $L \subseteq Q$.*

3 The autodense languages

In this section, we study some algebraic properties of autodense languages. Beside the language is autodense, we also study some languages which are pure autodense or which are not pure autodense.

Lemma 3.1 *Every autodense language is infinite.*

Proof Suppose that L is a finite autodense language. Let $u \in L$ be one of the words with maximal length in L. By the definition of the autodense language, there exist $x, y \in X^+$ such that $xuy \in L$. This contradicts that $u \in L$ is one of the words with maximal length in L. Thus an autodense language must be infinite. □

From the above lemma, we conclude that a subset of an autodense language may not be an autodense language. Since the intersection of two languages can be finite, an intersection of two autodense languages may not be autodense. Furthermore, the following example confirms us in our claim. For instance, an infinite subset of an autodense language may not be autodense. Let $L = a^*ba^2a^* \cup b^*ab^*$. Then L is an autodense language because L is the union of two autodense languages. The language $L' = L \setminus X^+ba^2X^+ = a^*ba^2 \cup ba^2a^* \cup b^*ab^*$ is an infinite subset of L. But L' is not autodense for $L' \cap X^+ba^2X^+ = \emptyset$. Beside the union of autodense language is autodense, we prove that $L^+ = L \cup L^2 \cup L^3 \cup \cdots$ is autodense in the following proposition.

Proposition 3.1 *For any nonempty language $L \subseteq X^+$, the language L^+ is autodense.*

Proof Let $z \in L^+ = L \cup L^2 \cup L^3 \cup \cdots$. Then $z \in L^r$ for some $r \geq 1$. This implies that $z^3 = zzz \in L^{3r} \subset L^+$. Let $x = y = z \in X^+$. It follows that $xzy = z^3 \in L^{3r}$. Therefore, L^+ is autodense. □

From Proposition 3.1, if $L = \{f\},\ f \in X^+$, then $L^+ = f^+$ is autodense.

Proposition 3.2 *The following are true*:

(i) $\mathcal{L}_{au} \cdot \mathcal{L}_d \subset \mathcal{L}_d$.
(ii) $\mathcal{L}_d \cdot \mathcal{L}_{au} \subset \mathcal{L}_d$.

Proof Both (i) and (ii) are immediate. □

From Proposition 3.2, $\mathcal{L}_d$ is an ideal of $\mathcal{L}_{au}$. But $\mathcal{L}_{au}$ is not closed under catenation. Indeed, let $L_1 = \{b^iab^i | i \geq 1\}$ and $L_2 = \{a^jba^j | j \geq 1\}$. Then L_1, L_2 are both autodense and $L_1L_2 = \{b^iab^ia^jba^j | i, j \geq 1\}$ is not autodense. It is obviously that $(bab)(aba) \in L_1L_2$, but $X^+(bab)(aba)X^+ \cap L_1L_2 = \emptyset$.

Lemma 3.2 ([17]) *Let $L \subseteq X^+$. If L contains a maximal code, then L^+ is dense.*

Corollary 3.1 *Let $L \subseteq X^+$. If L contains a maximal code, then L^+ is autodense.*

Proof Since a dense language is an autodense language, by Lemma 3.2, the corollary is clear. □

The converse of Corollary 3.1 is not true. For instance, let L_1 be any dense code. If L_1 is not maximal, then let $L = L_1$. If L_1 is maximal, then let $L = L_1 \setminus \{w\}$, where $w \in L_1$. This in conjunction with Lemma 2.4 yields that L is a dense language. Moreover by Lemma 2.5, we have that the language L^+ is dense. Therefore L^+ is an autodense language, but L does not contain a maximal code.

From the definition of pure autodense language, the pure autodense language is autodense. This implies that the union of pure autodense languages is autodense. Furthermore, it can be derived that the finite union of pure autodense languages is pure autodense.

Proposition 3.3 *Let* $n \geq 1$ *and* $L = \bigcup_{i=1}^{n} A_i$, *where* A_i *is a pure autodense language. Then* L *is pure autodense.*

Proof Let $n \geq 1$ and $L = \bigcup_{i=1}^{n} A_i$, where A_i is a pure autodense language. Then L is autodense. Since a finite union of non-dense language is also not dense, this yields that L is not dense; hence L is pure autodense. □

In the following propositions, we study some languages which are not pure autodense. Recall that L is a global (coglobal) language if $\sqrt{L} = Q$ ($\sqrt{L} = Q \backslash F$, where F is a finite language.)

Proposition 3.4 *Every coglobal language is dense and hence is not pure autodense.*

Proof Let $L \subseteq X^+$ be a coglobal language. Then there exists a finite language $F \subset X^+$ such that $\sqrt{L} = Q \backslash F$. By Lemmata 2.1 and 2.2, Q is dense. These in conjunction with Lemma 2.4 yield that $\sqrt{L}$ is dense. By Lemma 2.2 again, L is dense and hence is not pure autodense. □

Proposition 3.5 *The complement of a pure autodense language in* X^* *is dense and hence is not pure autodense.*

Proof Let L be a pure autodense language and let $\bar{L} = X^+ \backslash L$. From the definition of autodense language, L is not dense. Because L is not dense, there is $w \in X^*$ such that $X^* w X^* \subseteq \bar{L}$. Then $\bar{L}$ is dense; hence $\bar{L}$ is not pure autodense. □

There are some examples of pure autodense languages in the following propositions. For any word $w = a_1 a_2 \ldots a_r$, where $a_i \in X$, $i = 1, 2, \ldots, r$, let the mirror image of w be $w^R = a_r a_{r-1} \ldots a_2 a_1$. For a language L, the mirror image of L is defined as $L^R = \{w^R \mid w \in L\}$. It is clear that $\left(L^R\right)^R = L$.

Proposition 3.6 *For any* $L \subseteq X^+$, L *is pure autodense if and only if its mirror image* L^R *is pure autodense.*

Proof Since $\left(L^R\right)^R = L$, we need only show the necessary condition. Assume that L is not dense. Let w be a word such that $xwy \notin L$ for every $x, y \in X^*$. Then $y^R w^R x^R \notin L^R$. Since $x, y \in X^*$ are arbitrary, hence L^R is not dense. Moreover, if L is autodense, then clearly L^R is autodense. Therefore for a pure autodense language L, the mirror image L^R is a pure autodense language. □

Let $w \in X^+$ and A_w be a pure autodense language containing the word w. Then $X^+ = \bigcup_{w \in X^+} A_w$ is dense. Thus, an infinite union of pure autodense languages may be dense. However, the pure autodense language $\bigcup_{i \in N} a^+ b^i a^+$ is not dense because $X^* bab^2 ab X^* \cap \bigcup_{i \in N} a^+ b^i a^+ = \emptyset$. From this observation, not all infinite unions of pure autodense languages are dense. We will provide a method to construct an infinite union of pure autodense languages which is dense in the following proposition.

Lemma 3.3 *Every word $u \in X^+$ is contained in a pure autodense language f^+ where $f \in Q$.*

Proof Let $u \in X^+$ be a given word. To find that there exists a pure autodense language L such that $u \in L$, we let $u = f^n$, where $f \in Q$ and $n \geq 1$. If $L = f^+$, then $u \in L$. By the definition of autodense language, clearly $L = f^+$ is an autodense language. Since $\sqrt{L} = \{f\}$ is not dense, by Lemma 2.2, L is not dense. Therefore, the language f^+ is a pure autodense language containing the word u. □

Proposition 3.7 *For a finite subset $\Lambda \subset Q$, the forest language $L = \bigcup_{f \in \Lambda} f^+$ is a pure autodense language.*

Proof Clearly, this proof follows from Proposition 3.3 and Lemma 3.3. □

From the above proposition, the language $\bigcup_{f \in Q} f^+ = X^+$ is dense where f^+ is pure autodense.

Corollary 3.2 *Let $L = \bigcup_{i \in I} A_i$, where I is an index set, finite or infinite, and each A_i be a pure autodense language. Then the language L is an autodense language.*

Proof This result is immediately. □

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

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

A language $L \subseteq X^*$ is discrete if for every two words x and y in L, $\lg(x) \neq \lg(y)$. It implies that there exists a unique word $u \in L \cap X^m$ for some m.

Proposition 3.8 *Let L be a discrete autodense language. Then for any $n \geq 1$, there exists an integer $m > n$ such that at least two words $u \neq v \in X^m$ with the property $u \not\equiv v\,(P_L)$.*

Proof Assume that L is an autodense language. By Lemma 3.1, L is infinite. Then for any $n \geq 1$, there exists an integer $m > n$ such that $|L \cap X^m| \neq \emptyset$. Since L is discrete, there are two words $u \neq v \in X^m$ such that $u \in L$ and $v \notin L$. By the assumption that L is an autodense language, there exist $x, y \in X^+$ such that $xuy \in L$. Since L is discrete, it implies that $xvy \notin L$. Therefore $u \not\equiv v\,(P_L)$. □

In the following proposition, we provide a procedure to construct a discrete autodense language contained in a dense language. For simplifying our procedure, we need the following lemma.

Lemma 3.4 *Let $L \subseteq X^*$. The following are equivalent:*

(i) *For every $w \in X^*$, $X^*wX^* \cap L \neq \emptyset$.*
(ii) *For every $w \in X^+$, $X^*wX^* \cap L \neq \emptyset$.*
(iii) *For every $w \in X^+$, $X^+wX^+ \cap L \neq \emptyset$.*

Proof (i) $\Rightarrow$ (ii) Immediate. (ii) $\Rightarrow$ (i) Only to show that $X^*(1)X^* \cap L \neq \emptyset$ holds true, where 1 is the empty word. But this is always true, for $X^*(1)X^* \cap L = X^*X^* \cap L = X^* \cap L = L$, which is certainly not empty. (ii) $\Rightarrow$ (iii) Suppose that (ii) holds. Let $w \in X^+$ and consider the word z_1wz_2, where $z_1, z_2 \in X^+$. By (ii) $X^*(z_1wz_2)X^* \cap L \neq \emptyset$. It is clear that $X^+wX^+ \cap L \neq \emptyset$ and (iii) holds. (iii) $\Rightarrow$ (ii) Immediate. □

Proposition 3.9 *Every dense language contains a discrete autodense language.*

Proof Let L be a dense language. We construct a discrete autodense language A contained in L inductively in the following way. Start with a word, say, $u_1 \in L$ and put u_1 into A. Since L is dense, by Lemma 3.4, there exist $v_1, v_1' \in X^+$ such that $u_2 = v_1 u_1 v_1' \in L$. Note that $\lg(u_1) < \lg(u_2)$. Put the word u_2 into A; hence u_2 is the second word in A. Now inductively, suppose that the word $u_n \in A$. There exist $v_n, v_n' \in X^+$ such that $u_{n+1} = v_n u_n v_n' \in L$. Then naturally, put $u_{n+1} = v_n u_n v_n'$ into A. Note that $\lg(u_1) < \lg(u_2) < \cdots < \lg(u_{n+1})$. Since L is a dense language, this procedure can be infinitely continued without any problem. The final set A is clearly a discrete autodense language contained in L. □

4 The anti-autodense languages

We study some algebraic properties of the families of languages related to anti-autodense in this section. Firstly, we deal with the closure property of the union and catenation of two languages on the families of languages $\mathcal{L}_{aa}$, $\mathcal{L}_{aab}$, $\mathcal{L}_{aap}$, $\mathcal{L}_{aas}$ and $\mathcal{L}_{aac}$.

(I) Union of two languages.

Union of two anti-autodense languages may not be an anti-autodense language. For example, let $a \in X$. The languages $L_1 = \{a, a^2\}$ and $L_2 = \{a^2, a^3\}$ are both in $\mathcal{L}_{aa}$ while the union $L_1 \cup L_2 = \{a, a^2, a^3\}$ is not in $\mathcal{L}_{aa}$. Since $\mathcal{L}_p$ and $\mathcal{L}_s$ are not closed under union, both $\mathcal{L}_{aap}$ and $\mathcal{L}_{aas}$ are not closed under union of two languages. Similarly, $\mathcal{L}_{aab}$ and $\mathcal{L}_{aac}$ are also not closed under union of two languages.

(II) Catenation of two languages.

It is known that $\mathcal{L}_p$, $\mathcal{L}_s$, and $\mathcal{L}_b$ are closed under catenation. Moreover, $<\mathcal{L}_p, \cdot>$, $<\mathcal{L}_s, \cdot>$ and $<\mathcal{L}_b, \cdot>$ are all free semigroups [9,20]. It is also known that $\mathcal{L}_i$, the family of all infix codes, is closed under catenation. That is, $\mathcal{L}_i$ is a semigroup. But $<\mathcal{L}_i, \cdot>$ is not free [6]. In general, catenation of two anti-autodense languages is not an anti-autodense language. For example, let $a \neq b \in X$. The languages $L_1 = \{a, a^2\}$ and $L_2 = \{b, b^2\}$ are both in $\mathcal{L}_{aa}$ while $L_1 L_2 = \{ab, ab^2, a^2 b, a^2 b^2\} \notin \mathcal{L}_{aa}$. Therefore $\mathcal{L}_{aa}$ is not closed under catenation. In the following proposition, we will investigate the characterization concerning the catenation of two anti-autodense languages.

Proposition 4.1 *Let L_1, L_2 be two anti-autodense languages. Then $L_1 L_2$ is an anti-autodense language if and only if one of the following conditions holds:*

(i) *L_1 is a suffix code.*
(ii) *L_2 is a prefix code.*

Proof The necessary condition is clear. Now we prove the sufficient condition. Since both L_1 and L_2 are anti-autodense languages, $L_1 \cap X^+ L_1 X^+ = \emptyset$ and $L_2 \cap X^+ L_2 X^+ = \emptyset$. Suppose that $(L_1 L_2) \cap X^+ (L_1 L_2) X^+ \neq \emptyset$. Then there exist $u_1, u_2 \in L_1$, $v_1, v_2 \in L_2$ such that $u_1 v_1 = x u_2 v_2 y$ for some $x, y \in X^+$. Since L_1 and L_2 are anti-autodense languages, this yields that $u_1 = x u_2$ and $v_1 = v_2 y$. From $u_1, u_2 \in L_1$ and $u_1 = x u_2, x \in X^+$, L_1 is not a suffix code, a contradiction. Similarly, from $v_1, v_2 \in L_2$ and $v_1 = v_2 y$, $y \in X^+$, L_2 is not a prefix code, a contradiction. □

Corollary 4.1 *The families of languages $\mathcal{L}_{aap}$ and $\mathcal{L}_{aas}$ are both closed under catenation.*

Corollary 4.2 *$<\mathcal{L}_{aap}, \cdot>$ is a subsemigroup of the free semigroup $<\mathcal{L}_p, \cdot>$ and $<\mathcal{L}_{aas}, \cdot>$ is a subsemigroup of the free semigroup $<\mathcal{L}_s, \cdot>$.*

However, neither $<\mathcal{L}_{aap}, \cdot>$ nor $<\mathcal{L}_{aas}, \cdot>$ are free semigroups. These results will be derived in Proposition 4.2. In order to prove Proposition 4.2, the following notations and Lemma 4.1, the well-known result given by Schützenberger [12], are needed. Let M be a semigroup. For any two subsets A and B of M, we define

$$A^{-1}B = \{x \in M \mid Ax \cap B \neq \emptyset\};$$
$$BA^{-1} = \{x \in M \mid xA \cap B \neq \emptyset\}.$$

Lemma 4.1 ([12]) *Let S be a subsemigroup of a free semigroup M. Then S is free if and only if $S^{-1}S \cap SS^{-1} \subseteq S$.*

Proposition 4.2 *Neither the subsemigroup $<\mathcal{L}_{aap}, \cdot>$ of the free semigroup $<\mathcal{L}_p, \cdot>$ nor the subsemigroup $<\mathcal{L}_{aas}, \cdot>$ of the free semigroup $<\mathcal{L}_s, \cdot>$ are free.*

Proof Since the family of all prefix codes $\mathcal{L}_p$ and the family of all suffix codes $\mathcal{L}_s$ are free semigroups, Lemma 4.1 is applicable. Firstly, we show that $<\mathcal{L}_{aap}, \cdot>$ is not free. Let $L_1 = \{ab, b\}$, $L_2 = \{b, aba\}$. Then $L_1L_2 = \{ab^2, ababa, b^2, baba\}$ and $L_2L_1 = \{bab, aba^2b, b^2, abab\}$. Both L_1L_2 and L_2L_1 are in $\mathcal{L}_{aap}$. These in conjunction with $L_1 \in \mathcal{L}_{aap}$ yield that $L_2 \in \mathcal{L}_{aap}^{-1}\mathcal{L}_{aap} \cap \mathcal{L}_{aap}\mathcal{L}_{aap}^{-1}$. But $L_2 = \{b, aba\}$ implies that $L_2 \cap X^+L_2X^+ \neq \emptyset$. It follows that $L_2 \notin \mathcal{L}_{aap}$. Therefore, by Lemma 4.1, $\mathcal{L}_{aap}$ is not free. Next, $<\mathcal{L}_{aas}, \cdot>$ is not free either, by symmetry. □

By Corollary 4.1, the families of languages $\mathcal{L}_{aap}$ and $\mathcal{L}_{aas}$ are closed under catenation. Moreover, the family of infix codes $\mathcal{L}_i$ is a subfamily of $\mathcal{L}_{aap}$ and $\mathcal{L}_{aas}$, that is, $\mathcal{L}_i \subset \mathcal{L}_{aap} \cap \mathcal{L}_{aas}$. It is known that an infix code can never be dense. We would like to generalize this result in the following proposition.

Proposition 4.3 *No anti-autodense language is dense.*

Proof By Lemma 3.4, we have that a language $L \subseteq X^*$ is dense if for any $w \in X^+$, $L \cap X^+wX^+ \neq \emptyset$. Thus if there exists a word $w \in X^+$ such that $L \cap X^+wX^+ = \emptyset$, then L is not dense. Hence a language with the condition $L \cap X^+LX^+ = \emptyset$ certainly ensures that L is not dense. Recall that a language with the condition $L \cap X^+LX^+ = \emptyset$ is an anti-autodense language. Thus the result is immediate. □

Corollary 4.3 *Infix codes, anti-autodense prefix codes, anti-autodense suffix codes, and anti-autodense codes are non-dense languages.*

Proof The results follow directly from Proposition 4.3 and the facts, $\mathcal{L}_i \subset \mathcal{L}_{aap} \subset \mathcal{L}_{aac} \subset \mathcal{L}_{aa}$, $\mathcal{L}_i \subset \mathcal{L}_{aas} \subset \mathcal{L}_{aac} \subset \mathcal{L}_{aa}$. □

Corollary 4.4 *The complement of an anti-autodense language is autodense.*

Proof Let $L \subseteq X^+$ be an anti-autodense language and let $\bar{L} = X^*\backslash L$ be the complement of L. By Proposition 4.3, L is not dense. Moreover, since X^* is dense, by Lemmata 2.2 and 2.3, $\bar{L} = X^*\backslash L$ is dense. Hence, the complement of an anti-autodense language is autodense. □

Remark Let $X = \{a, b\}$ and let $L \subseteq X^+$ be an anti-autodense language. If there exists a word $z \in \bar{L} = X^+ \backslash L$ such that $xzy \in L$ for some $x, y \in X^+$, then $|L \cap X^+ z X^+|$ may be infinite. For example, let $L = ab^+$. It is clear that $L \cap X^+ L X^+ = \emptyset$. Then $L = ab^+$ is an anti-autodense language. It is also clear that $z = b \in X^+ \backslash L$, $azb = abb \in L$, and $L \cap X^+ b X^+ = L$ is infinite.

In the following propositions, we will study some anti-autodense properties of languages $L^n, n \geq 1$. Let $L \subseteq X^+$ be an anti-autodense language. Then it can be derived that $X^+ L^{n-1} X^+ \subset X^+ L X^+$ for every $n \geq 3$. This result can be found from the following example. Let $L = \{a, ab\}$. Then $L^{n-1} = \left\{a^{n-1}, a^{n-2}ab, \ldots, (ab)^{n-1}\right\}$. Hence the number of a in every word in $X^+ L^{n-1} X^+$ is greater than $n-1$. Thus $b^n ab^n \in X^+ L X^+ \backslash X^+ L^{n-1} X^+$ for $n \geq 3$. This implies that $X^+ L^{n-1} X^+ \subset X^+ L X^+$.

Proposition 4.4 *Let $L \subseteq X^+$ be an anti-autodense language. Then $L^n \not\subset L$ for any $n \geq 2$.*

Proof Let $L \subseteq X^+$ be an anti-autodense language. Then $L \cap X^+ L X^+ = \emptyset$. Suppose that $L^n \subset L$ for any $n \geq 2$. This in conjunction with $L \cap X^+ L X^+ = \emptyset$ yields that $L^n \cap X^+ L X^+ = \emptyset$. Since $X^+ L^{n-1} X^+ \subset X^+ L X^+$ is true, this implies that $L^n \cap X^+ L^{n-1} X^+ = \emptyset$. Then L is an intercode of index $n-1$ with $n \geq 2$. But, by Lemma 2.6, any intercode contains only primitive words, the condition $L^n \subset L$ is impossible. Therefore $L^n \not\subset L$ for any $n \geq 2$. □

The converse of Proposition 4.4 is not true. For example, the language $L = a^+ \cup b^+$ over $X = \{a, b\}$ has the property $L^2 \not\subset L$ while L is an autodense language. An autodense language may not have the above property. For example, a^+ is an autodense language over $X = \{a, b\}$. Here $\left(a^+\right)^2 \subset a^+$.

For an anti-autodense language L, by Proposition 4.1, it is clear that L^2 is an anti-autodense language if and only if L is a prefix code or a suffix code. Now we extend this result to L^n for $n \geq 3$.

Proposition 4.5 *Let L be an anti-autodense language. Then for $n \geq 2$, L^n is an anti-autodense language if and only if one of the following conditions hold: L is a prefix code or L is a suffix code.*

Proof Let L be an anti-autodense language. Then $L \cap X^+ L X^+ = \emptyset$. Assume that L is a prefix code or L is a suffix code. We want to show that $L^n \cap X^+ L^n X^+ = \emptyset$. We prove the result by induction on $n \geq 2$. It is clear that L^n is an anti-autodense language by Proposition 4.1. Recall that $<\mathcal{L}_p, \cdot>$, $<\mathcal{L}_s, \cdot>$ are all free semigroups. By assumption, $L^n, n \geq 2$ is a prefix code or a suffix code. Thus both L, L^n are prefix codes or suffix codes. By Proposition 4.1 again, $L^{n+1} = L\left(L^n\right)$ is an anti-autodense language. Hence by induction, L^n is an anti-autodense language. Now we assume that L^n is an anti-autodense language. If L is neither a prefix code nor a suffix code, then there exist $u, v, w, z \in L$ such that $w = xu$, $z = vy$ for some $x, y \in X^+$. For $n \geq 2$, we consider $wv^{n-2}z \in L^n$. Then $wv^{n-2}z = xuv^{n-2}vy = xuv^{n-1}y$. This implies that $wv^{n-2}z \in X^+ L^n X^+$. That is, $L^n \cap X^+ L^n X^+ \neq \emptyset$. This contradicts that L^n is an anti-autodense language. Hence L is a prefix code or a suffix code. □

It is easy to split X^+ into a disjoint union of two autodense languages. For example, let $X = \{a, b\}$.

(i) $X^+ = a^+ \cup \left(X^+ \backslash a^+\right)$.
(ii) $X^+ = aX^* \cup bX^*$.
(iii) $X^+ = \left(a^+ \cup b^+\right) \cup \left(X^+ \backslash \left(a^+ \cup b^+\right)\right)$.

However, the same situation is not true for the case of anti-autodense languages.

Proposition 4.6 *Let $X^+ = L_1 \cup L_2 \cup \cdots \cup L_r$, where $r \geq 2$ and the union is a disjoint union. It is impossible that all the L_i are anti-autodense languages.*

Proof This proof follows by induction from Lemmata 2.2 and 2.3, and Proposition 4.3. □

Remark For any language $L \subseteq X^+$, it is impossible that both L and $X^+ \backslash L$ are anti-autodense languages.

Corollary 4.5 *Let $X^+ = L_1 \cup L_2 \cup \cdots \cup L_r$, where $r \geq 2$ and the union is a disjoint union. Then at least one of the component L_i in the decomposition is not an infix code.*

Proof Since every infix code is an anti-autodense language, the corollary follows directly from the above proposition. □

Corollary 4.5 tells us that X^+ is not a finite disjoint union of infix codes. Moreover, there are some other languages with this property. In fact, if L is not a finite disjoint union of infix code, then for any infix code $L_1 \subset L$, the language $L \backslash L_1$ is also not a finite disjoint union of infix codes.

Corollary 4.6 *Let A be a finite disjoint union of infix codes and let $L = X^+ \backslash A$. Then L is not a finite disjoint union of infix codes.*

Acknowledgments The authors would like to thank the referees for their careful reading of the manuscript and useful suggestions.

References

1. Calude, C., Yu, S.: Language-theoretic complexity of disjunctive sequences. Discrete Appl. Math. **80**, 203–209 (1997)
2. Fan, C.-M., Shyr, H.J.: Catenation Closed Pairs and Forest Languages. Words, Semigroups, & Transductions, World Scientific, pp. 115–127 (2000)
3. Guo, Y.Q., Thierrin, G., Zhang, S.H.: Semaphore codes and ideals. J. Inform. Optim. Sci. **9**(1), 73–83 (1988)
4. Hsieh, C.Y., Hsu, S.C., Shyr, H.J.: Some algebraic properties of comma-free codes. RIMS Kenkyuroku, (Japan) **697**, 57–65 (1989)
5. Hsu, S.C., Ito, M., Shyr, H.J.: Some properties of overlapping order and related languages. Soochow J. Math. **15**, 29–45 (1989)
6. Ito, M., Jürgensen, H., Shyr, H.J., Thierrin, G.: Outfix and infix codes and related classes of languages. J. Comput. Syst. Sci. **43**(3), 484–508 (1991)
7. Lallement, G.: Semigroups and Combinatorial Applications. Wiley, New York (1979)
8. Li, Z.-Z., Shyr, H.J., Tsai, Y.S.: Classifications of dense languages. Acta Inform. **43**, 173–194 (2006)
9. Perrin, D.: Codes conjugués. Inform. Control **20**, 221–231 (1972)
10. Rabin, M.O., Scott, D.: Finite automata and their decision problem. IBM J. Res. Dev. **3**, 114–125 (1959)
11. Reis, C.M., Shyr, H.J.: Some properties of disjunctive languages on a free monoid. Inform. Control **37**(3), 334–344 (1978)
12. Schützenberger, M.P.: On an application of semigroup methods to some problems in coding. IRE Trans. Inform. Theor. **IT-2**, 47–60 (1956)
13. Shyr, H.J.: Disjunctive languages on a free monoid. Inform. Control **34**(2), 123–129 (1977)
14. Shyr, H.J.: A characterization of dense languages. Semigroup Forum **30**, 237–240 (1984)
15. Shyr, H.J.: Characterization of right dense languages. Semigroup Forum **33**, 23–30 (1986)
16. Shyr, H.J., Tsai, Y.S.: Note on languages which are dense subsemigroups. Soochow J. Math. **11**, 117–122 (1988)
17. Shyr, H.J., Tsai, Y.S.: Disjunctive context-free languages. In: Proceedings of the International Symposium on the Semigroup Theory, pp. 213–223. Kyoto, Japan (1990)
18. Shyr, H.J., Tseng, D.C.: Some properties of dense languages. Soochow J. Math. **10**, 127–131 (1984)
19. Shyr, H.J., Yu, S.S.: Intercodes and some related properties. Soochow J. Math. **16**(1), 95–107 (1990)
20. Yu, S.S.: Languages and Codes. Tsang Hai Book Publishing Co., Taichung (2005)

無研發成果推廣資料

98年度專題研究計畫研究成果彙整表

<table>
<tr><td colspan="3">計畫主持人：黃政治</td><td colspan="5">計畫編號：98-2115-M-040-001-</td></tr>
<tr><td colspan="8">計畫名稱：自動稠密性及反自動稠密性言語的研究</td></tr>
<tr><td colspan="3" rowspan="2">成果項目</td><td colspan="3">量化</td><td rowspan="2">單位</td><td rowspan="2">備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）</td></tr>
<tr><td>實際已達成數（被接受或已發表）</td><td>預期總達成數(含實際已達成數)</td><td>本計畫實際貢獻百分比</td></tr>
<tr><td rowspan="14">國內</td><td rowspan="4">論文著作</td><td>期刊論文</td><td>0</td><td>0</td><td>0%</td><td rowspan="3">篇</td><td></td></tr>
<tr><td>研究報告/技術報告</td><td>0</td><td>0</td><td>0%</td><td></td></tr>
<tr><td>研討會論文</td><td>0</td><td>0</td><td>0%</td><td></td></tr>
<tr><td>專書</td><td>0</td><td>0</td><td>0%</td><td></td><td></td></tr>
<tr><td rowspan="2">專利</td><td>申請中件數</td><td>0</td><td>0</td><td>0%</td><td rowspan="2">件</td><td></td></tr>
<tr><td>已獲得件數</td><td>0</td><td>0</td><td>0%</td><td></td></tr>
<tr><td rowspan="2">技術移轉</td><td>件數</td><td>0</td><td>0</td><td>0%</td><td>件</td><td></td></tr>
<tr><td>權利金</td><td>0</td><td>0</td><td>0%</td><td>千元</td><td></td></tr>
<tr><td rowspan="4">參與計畫人力（本國籍）</td><td>碩士生</td><td>0</td><td>0</td><td>0%</td><td rowspan="4">人次</td><td></td></tr>
<tr><td>博士生</td><td>0</td><td>0</td><td>0%</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td>0</td><td>0%</td><td></td></tr>
<tr><td>專任助理</td><td>1</td><td>1</td><td>30%</td><td></td></tr>
<tr><td rowspan="14">國外</td><td rowspan="4">論文著作</td><td>期刊論文</td><td>1</td><td>1</td><td>100%</td><td rowspan="3">篇</td><td>論文成果發表於 Acta Informatica Vol.47, No.4, 209--219(2010).</td></tr>
<tr><td>研究報告/技術報告</td><td>0</td><td>0</td><td>0%</td><td></td></tr>
<tr><td>研討會論文</td><td>0</td><td>0</td><td>0%</td><td></td></tr>
<tr><td>專書</td><td>0</td><td>0</td><td>0%</td><td>章/本</td><td></td></tr>
<tr><td rowspan="2">專利</td><td>申請中件數</td><td>0</td><td>0</td><td>0%</td><td rowspan="2">件</td><td></td></tr>
<tr><td>已獲得件數</td><td>0</td><td>0</td><td>0%</td><td></td></tr>
<tr><td rowspan="2">技術移轉</td><td>件數</td><td>0</td><td>0</td><td>0%</td><td>件</td><td></td></tr>
<tr><td>權利金</td><td>0</td><td>0</td><td>0%</td><td>千元</td><td></td></tr>
<tr><td rowspan="4">參與計畫人力（外國籍）</td><td>碩士生</td><td>0</td><td>0</td><td>0%</td><td rowspan="4">人次</td><td></td></tr>
<tr><td>博士生</td><td>0</td><td>0</td><td>0%</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td>0</td><td>0%</td><td></td></tr>
<tr><td>專任助理</td><td>0</td><td>0</td><td>0%</td><td></td></tr>
</table>

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	無

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

   ■達成目標

   □未達成目標（請說明，以 100 字為限）

   □實驗失敗

   □因故實驗中斷

   □其他原因

   說明：

2. 研究成果在學術期刊發表或申請專利等情形：

   論文：■已發表 □未發表之文稿 □撰寫中 □無

   專利：□已獲得 □申請中 ■無

   技轉：□已技轉 □洽談中 ■無

   其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

   本研究主要是探究自動稠密言語及純自動稠密言語之代數特性，除此，反自動稠密言語的相關代數性質也一併加以討論。其中於反自動稠密言語家族中發現包含有中置數碼、自由逗點數碼等特殊數碼言語，對於密碼科學有學術理論研究上之貢獻。