BULLETIN *of the*
MALAYSIAN MATHEMATICAL
SCIENCES SOCIETY
http://math.usm.my/bulletin

Bull. Malays. Math. Sci. Soc. (2) **37**(3) (2014), 727–746

On Some Decompositions of r-Disjunctive Languages

[1]DI ZHANG, [2]YUQI GUO AND [3]K. P. SHUM
[1,2]School of Mathematics and Statistics, Lanzhou University, Lanzhou, 730000, China
[3]Institute of Mathematics, Yunnan University, Kunming, 650091, China
[1]zhdi88@gmail.com, [2]yqguo259@swu.edu.cn, [3]kpshum@ynu.edu.cn

Abstract. Some kinds of decompositions of r-disjunctive languages on an arbitrary alphabet will be investigated. We will show that an f-disjunctive (t-disjunctive) language can be divided into two parts and either one part of them is an f-disjunctive (t-disjunctive) language or both parts are r-disjunctive but not f-disjunctive (t-disjunctive) languages. Finally, a relevant result of H. J. Shyr and S. S. Yu concerning the disjunctive languages will be improved.

2010 Mathematics Subject Classification: 20M35

Keywords and phrases: Disjunctive languages, f-disjunctive languages, t-disjunctive languages, r-disjunctive languages, r-regular languages.

1. Introduction and preliminaries

Let X be a nonempty finite set called alphabet in which the elements are called letters. Let X^* be the free monoid generated by an alphabet X. Then, the elements and subsets of X^* are called the *words* and *languages* over X. The identity of the free monoid X^* is called an empty word and is denoted by 1. Let $X^+ = X^* \setminus \{1\}$ be the free semigroup generated by X. The length of a word w over X is the number of letters occurring in w and is denoted by $lg(w)$. We denote the cardinality of a language L over X by $|L|$. For any two languages A, B over X, the *concatenation* AB of A and B is the language $\{xy | x \in A, y \in B\}$ over X. For a given language L over X, the relation P_L on X^* defined by

$$x \equiv y(P_L) \ \Leftrightarrow \ ''(\forall u, v \in X^*) \ uxv \in L \Leftrightarrow uyv \in L''$$

is a congruence on free monoid X^* and is known as the *principal congruence* determined by L. The quotient monoid X^*/P_L is called the *syntactic monoid* of L and is denoted by $Syn(L)$. For any word u over X, we often use $[u]_L$ to denote the P_L-class of X^* containing u. As usual, the set of all positive (nonnegative) integers is denoted by $\mathbb{N}(\mathbb{N}^0)$.

We call a language L over X *disjunctive* [9] if P_L is the equality relation on X^*. Let $\mathscr{D}$ be the class of all disjunctive languages over X. A language L over X is called *regular* [5, 10] if the index of P_L (i.e., the number of P_L-classes of X^*) is finite. Let $\mathscr{R}$ be the class of all regular languages over X. Then we call a language L over X a *midst-language* [12] if L is

Communicated by Lee See Keong.
Received: July 28, 2013; *Revised:* March 3, 2014.

neither regular nor disjunctive. Let $\mathscr{M}$ be the class of all midst-languages over X. Then, we have the following proposition.

Proposition 1.1. [11] *Let X be an alphabet with $|X| = 1$. Then $\mathscr{M} = \emptyset$, that is, over X, a language is disjunctive if and only if it is not regular.*

But when $|X| \geq 2$, the case is completely different from $|X| = 1$ [6]. In this case, $\mathscr{M} \neq \emptyset$, that is, $\{\mathscr{R} \cup \mathscr{D}\} \subsetneq 2^{X^*}$.

We call a language L over X *dense* if $X^*wX^* \bigcap L \neq \emptyset$ for any $w \in X^*$; otherwise, the language L is said to be *thin*. According to Reis and Shyr [10], a language L is dense if and only if L contains a disjunctive language. Denote the class of all dense languages over X by $\mathscr{D}^d$.

The generalized disjunctive languages have been considered by a number of authors in the literature, such as, Guo, Reis and Thierrin [1] in 1988 called a language L over X *relatively f-disjunctive* (*relatively disjunctive*), that is, *rf-disjunctive* for short (*r-disjunctive* for short), if there exists a dense language D over X such that for all $u \in X^*$, $|[u]_L \cap D| < \infty$ ($|[u]_L \cap D| \leq 1$). It has been shown in [1] that L is rf-disjunctive if and only if L is r-disjunctive, if and only if either X^* has no dense P_L-classes or has infinitely many dense P_L-classes. Let $\mathscr{D}_r$ be the class of all r-disjunctive languages over X. Then, the concept of *relatively regular* language was first introduced by Liu, Shum and Guo in 2008 (see [6]). They called a language L over X relatively regular, that is, *r-regular* for short, if $Syn(L)$ has a finite ideal. Let $\mathscr{R}_r$ be the class of all r-regular languages over X.

This paper is based on the following background.

(I) Obviously, when $|X| = 1$, $\mathscr{D}_r = \mathscr{D}$, $\mathscr{R}_r = \mathscr{R}$. In [6], the authors proved the following fact which forms a generalization of Proposition 1.1 to any alphabet X from $|X| = 1$.

Proposition 1.2. [6] *Let X be an alphabet. Then a language over X is r-disjunctive if and only if it is not r-regular.*

Leading up to [1], [6], some special cases of r-disjunctive languages have been defined. In particular, a language L over an alphabet X was first called by Guo, Shyr and Thierrin [2] *f-disjunctive* if each P_L-class of X^* is finite, and later, Mu [8] called a language L over X *t-disjunctive* if each P_L-class of X^* is thin. Denote the class of all f-disjunctive (t-disjunctive) languages over X by $\mathscr{D}_f(\mathscr{D}_t)$.

The following proposition is useful in this paper.

Proposition 1.3. [1,7]

(1) *If $|X| = 1$, then*

$$\mathscr{D} = \mathscr{D}_f = \mathscr{D}_t = \mathscr{D}_r.$$

(2) *If $|X| \geq 2$, then*

$$\mathscr{D} \subsetneq \mathscr{D}_f \subsetneq \mathscr{D}_t \subsetneq \mathscr{D}_r \subsetneq \mathscr{D}^d.$$

Some more characterizations of r-disjunctive languages can be found in [1,2,6,8,9,11].

(II) The following result is a known result on some decompositions of disjunctive languages.

Proposition 1.4. [10] *Let L be a disjunctive language over X, $L = L_1 \dot\cup L_2$ (i.e., $L = L_1 \cup L_2$ and $L_1 \cap L_2 = \emptyset$). Then, the following statements hold.*

(1) $\{L_1, L_2\} \cap \mathscr{D} \neq \emptyset$, *or*
(2) $\{L_1, L_2\} \subseteq \mathscr{D}^d \setminus \mathscr{D}$.

In the case (2) of Proposition 1.4, each one of L_1 and L_2 is not r-regular language, for if not, then another one of them must be disjunctive by [6]. This fact clearly contradicts to $L_1, L_2 \notin \mathscr{D}$. Hence, Proposition 1.4 can be modified to the following form.

Proposition 1.4′. *Let L be a disjunctive language over X, $L = L_1 \dot{\cup} L_2$. Then, the following statements hold.*

(1) $\{L_1, L_2\} \cap \mathscr{D} \neq \emptyset$, *or*
(2)′ $\{L_1, L_2\} \subseteq \mathscr{D}_r \setminus \mathscr{D}$.

(III) Proposition 1.4′ actually says that the disjoint union decompositions of disjunctive languages have two cases. Any disjunctive language has the decomposition of the case (1) in Proposition 1.4. As to the decomposition in the case (2) of Proposition 1.4, not every disjunctive language has this decomposition, for instance, *discrete disjunctive language*, that is, the disjunctive language L with $|L \cap X^n| \leq 1$ for any $n \in \mathbb{N}$ [10]. But Shyr and Yu have shown in [12] that there exists such a disjunctive language L over X with $|X| \geq 3$ so that $L = L_1 \dot{\cup} L_2$, $\{L_1, L_2\} \subseteq \mathscr{D}^d \setminus \mathscr{D}$, to be more precise, we have

$$\{L_1, L_2\} \subseteq \mathscr{D}_f \setminus \mathscr{D}. \tag{1.1}$$

In this paper, we will further discuss in Section 4 about the existence case mentioned by Shyr and Yu in [12]; we will also discuss the decompositions of languages in $\mathscr{D}_f, \mathscr{D}_t$ and $\mathscr{D}_r$ like Proposition 1.4′ in Section 2; and in Section 3, we will show that the languages in $\mathscr{D}_f, \mathscr{D}_t$ have similar decompositions just as the case (2)′ in Proposition 1.4′ with (1.1).

Making contact with the above background of this paper, we have started to apply some results of this paper to our following work to describe the disjunctive degree in some sense of languages, this shows one spot of the potential value of this paper.

For terminologies and notations not mentioned in this paper, the reader is referred to [4, 5, 10].

In the remaining part of the paper, we always assume that $|X| \geq 2$.

2. Some decompositions of r-disjunctive languages(I)

In the following theorem, we consider the decompositions of languages in $\mathscr{D}_f, \mathscr{D}_t$ and $\mathscr{D}_r$ which are similar to Proposition 1.4′.

Theorem 2.1. *Let $L \in \mathscr{D}_f$ $(\mathscr{D}_t,\ \mathscr{D}_r)$, $L = L_1 \dot{\cup} L_2$. Then the following statements hold:*

(1) $\{L_1, L_2\} \cap \mathscr{D}_f$ $(\mathscr{D}_t,\ \mathscr{D}_r)$ $\neq \emptyset$, *or*
(2) $\{L_1, L_2\} \subseteq \mathscr{D}_r \setminus \mathscr{D}_f$ $(\mathscr{D}_r \setminus \mathscr{D}_t,\ \mathscr{D}_r \setminus \mathscr{D}_r)$.

Remark 2.1. If we divide a language $L \in \mathscr{D}_r$ into L_1 and L_2, then L_1, L_2 must satisfy Theorem 2.1(1). Otherwise, both of L_1, L_2 are not in $\mathscr{D}_r$, by Proposition 1.2, they are in $\mathscr{R}_r$. Hence, L is in $\mathscr{R}_r$ because by [6], $\mathscr{R}_r$ is closed under the operation of union, this is clearly a contradiction. Here, we write $\{L_1, L_2\} \subseteq \mathscr{D}_r \setminus \mathscr{D}_r$ to seek a unity of expression with the $\mathscr{D}_f$ and $\mathscr{D}_t$ languages.

In proving the above theorem, we need the following proposition.

Proposition 2.1.
(1) *If* $L \in \mathscr{D}_t$ $(\mathscr{D}_f, \mathscr{D}_r)$, $R \subseteq L$ *and* $R \in \mathscr{R}_r$, *then* $L \setminus R \in \mathscr{D}_t$ $(\mathscr{D}_f, \mathscr{D}_r)$.
(2) *If* $L \in \mathscr{D}_t$ $(\mathscr{D}_f, \mathscr{D}_r)$, $R \in \mathscr{R}_r$ *and* $L \cap R = \emptyset$, *then* $L \cup R \in \mathscr{D}_t$ $(\mathscr{D}_f, \mathscr{D}_r)$.

Proof. By [6], Proposition 2.1 holds for $L \in \mathscr{D}_f$ and $L \in \mathscr{D}_r$. Here we just discuss about $L \in \mathscr{D}_t$.

(1) Suppose that $L \in \mathscr{D}_t$. Then for any dense language $\{x_1, x_2, \ldots, x_m, \ldots\}$, there exist x_i, x_j, $i \neq j$ such that

$$x_i \not\equiv x_j (P_L).$$

Since $R \in \mathscr{R}_r$, by Lemma 3.3 of [6], there exist a $w \in X^*$, and an $n \in \mathbb{N}$ such that $(wX^*w)^{(n)}$ is contained in a P_R-class, and whence

$$(wx_1w)^n \equiv (wx_2w)^n \equiv \cdots \equiv (wx_mw)^n \equiv \cdots (P_R).$$

Notice that $\{(wx_iw)^n | i = 1, 2, \ldots\}$ is dense and $(wx_iw)^n \neq (wx_jw)^n$, since $x_i \neq x_j$ when $i \neq j$. Then by the definition of t-disjunctive languages, there exist $(wx_iw)^n$ and $(wx_jw)^n$, $i \neq j$ such that

$$(wx_iw)^n \not\equiv (wx_jw)^n (P_L).$$

That is, for some $u, v \in X^*$, we have

$$u(wx_iw)^n v \in L,\ u(wx_jw)^n v \notin L,$$

or vice versa. We now suppose that the former case hold, so $u(wx_jw)^n v \notin R$. Then this result leads to $u(wx_iw)^n v \notin R$ since $(wx_iw)^n \equiv (wx_jw)^n (P_R)$. Thus $u(wx_iw)^n v \in L \setminus R$, but $u(wx_jw)^n v \notin L \setminus R$. This shows that

$$(wx_iw)^n \not\equiv (wx_jw)^n (P_{L\setminus R}),$$

and hence

$$x_i \not\equiv x_j (P_{L\setminus R}).$$

Thus, $L \setminus R \in \mathscr{D}_t$.

(2) This part follows directly from (1) because for any language L over X, $P_L = P_{\overline{L}}$, where $\overline{L}$ is the complement of L. ∎

Corollary 2.1. *If* $L \in \mathscr{D}_r \setminus \mathscr{D}_t$ *and* $R \subseteq L$ *and* $R \in \mathscr{R}_r$, *then* $L \setminus R \in \mathscr{D}_r \setminus \mathscr{D}_t$.

We now return to prove Theorem 2.1.

Proof of Theorem 2.1. We only prove that $L \in \mathscr{D}_f(\mathscr{D}_t)$. Assume that neither L_1 nor L_2 is f-disjunctive(t-disjunctive) language and suppose that L_2 is not an r-disjunctive language. Then L_2 is an r-regular language. Now, we see that L_1 is f-disjunctive(t-disjunctive)language by Proposition 2.1. This result contradicts to our assumption. Hence, L_2 is an r-disjunctive language. Similarly, L_1 is also an r-disjunctive language. ∎

Any f-disjunctive (t-disjunctive, r-disjunctive) language has the decomposition of the case (1) in Theorem 2.1 by Proposition 2.1. For the languages in $\mathscr{D}_f$ and $\mathscr{D}_t$, we naturally ask the question: Does the decomposition of the languages that satisfying Theorem 2.1(2) exist? More precisely, we ask whether the decomposition which is similar to (1.1) exists or not? We will give an affirmative answer to the above questions in section 3. In order to simplify our description, we let $\mathscr{D}_1 = \mathscr{D}, \mathscr{D}_2 = \mathscr{D}_f \setminus \mathscr{D}, \mathscr{D}_3 = \mathscr{D}_t \setminus \mathscr{D}_f, \mathscr{D}_4 = \mathscr{D}_r \setminus \mathscr{D}_t$. Hence, $\mathscr{D}_r$ is a disjoint union of $\mathscr{D}_1, \mathscr{D}_2, \mathscr{D}_3$ and $\mathscr{D}_4$.

3. Some decompositions of r-disjunctive languages(II)—Decomposition from $\mathscr{D}_i$ to $\mathscr{D}_{i+1}$, $i = 2,3$

First of all, we give the following preparations.

Let L be a nonempty language over an alphabet X. We call L contained in X^+ a *code* if for any $x_i, y_j \in L$, $i = 1,2,\dots,m$, $j = 1,2,\dots,n$, $x_1x_2\cdots x_m = y_1y_2\cdots y_n$ implies that $m = n$ and $x_i = y_i$, $i = 1,2,\dots,n$. We call L a *prefix language* (*suffix language*) if for any $x \in L, xy \notin L$ for all $y \in X^+$($yx \notin L$, for all $y \in X^+$). It is immediate to see that each prefix(suffix) language contained in X^+ is a code. Hence, we also call a prefix (suffix) language contained in X^+ a *prefix (suffix) code*. Obviously, a singleton-set of X^+ is a prefix code and is a suffix code as well.

For prefix codes and suffix codes, we have the following lemma.

Lemma 3.1. [2] *Let L be a language over X and P(S) a prefix(suffix) code over X. Then for any $u, v \in X^*$, $u \not\equiv v(P_L)$ implies $u \not\equiv v(P_{PL}(P_{LS}))$, i. e. $P_{PL}(P_{LS}) \subseteq P_L$.*

Corollary 3.1. *If L is a t-disjunctive language over X and P(S) is a prefix(suffix) code over X, then $PL(LS)$ is t-disjunctive.*

Similar consequence for disjunctive languages and f-disjunctive languages can be found in [10] and [2] respectively.

Proposition 3.1. *Let L be a language over X. If L is not f-disjunctive, then for any finite language F of X^*, the language FL and LF are not f-disjunctive.*

Proof. We just consider the language FL, the conclusion for LF can be dually obtained. Suppose that $F \neq \emptyset$ and $F \neq \{1\}$ (the conclusion is trivial when $F = \emptyset$ and $F = \{1\}$).

Let $m = max\{lg(x)|x \in F\} \in \mathbb{N}$. Suppose that L is not an f-disjunctive language. Then there exist an infinite language $\{x_1, x_2, \dots, x_n, \dots\}$ such that

$$x_1 \equiv x_2 \equiv \cdots \equiv x_n \equiv \cdots (P_L).$$

Since P_L is a congruence,

$$w^m x_1 \equiv w^m x_2 \equiv \cdots \equiv w^m x_n \equiv \cdots (P_L)$$

for any $w \in X^+$. We now show that

$$w^m x_1 \equiv w^m x_2 \equiv \cdots \equiv w^m x_n \equiv \cdots (P_{FL}).$$

Suppose that there exist $w^m x_i, w^m x_j \in \{w^m x_1, w^m x_2, \dots, w^m x_n, \dots\}$ such that $w^m x_i \not\equiv w^m x_j (P_{FL})$. Then, there exist $u, v \in X^*$ such that $uw^m x_i v \in FL$ and $uw^m x_j v \notin FL$ or vice verse. Without loss of generality, we may let $uw^m x_i v \in FL$ and $uw^m x_j v \notin FL$.

Consider the following two cases:

(1) $u = u_1u_2$, for some $u_1 \in X^+$, $u_2 \in X^*$ such that $u_1 \in F$, $u_2 w^m x_i v \in L$. Clearly, $u_2 w^m x_j v \notin L$. Hence, $x_i \not\equiv x_j(P_L)$, which is a contradiction.

(2) $uw^{k_1}w_1 \in F$, $w_2 w^{k_2} x_i v \in L$, where $w_1 \in X^+$, $w_2 \in X^*$, $w = w_1w_2$, $k_1, k_2 \in \mathbb{N}^0$ and $k_1 + k_2 + 1 = m$. Again $w_2 w^{k_2} x_j v \notin L$ and $x_i \not\equiv x_j(P_L)$, and so a contradiction.

This shows that the conclusion

$$w^m x_1 \equiv w^m x_2 \equiv \cdots \equiv w^m x_n \equiv \cdots (P_{FL})$$

holds and hence FL is not an f-disjunctive language. ∎

The following proposition is a similar proposition for t-disjunctive languages.

Proposition 3.2. *Let L be a language over X. If L is not a t-disjunctive language, then for any finite language F of X^*, the language FL and LF are not t-disjunctive.*

Proof. The proof of this proposition is similar to the proof of Proposition 3.1. ∎

Similar consequence of non-disjunctive languages can be found in [12].

In the following lemma, we will show that there are languages in $\mathscr{D}_2$ which are unions of two disjoint languages in $\mathscr{D}_3$, see the following Proposition 3.3 and Proposition 3.4.

We now use w_x to denote the number of letters x occurring in the word w over X. Then, we establish the following lemma.

Lemma 3.2. *Let $X = \{x_1, x_2, \dots, x_r\}$, $r \geq 2$,*

$$L_{x_i} = \{w \in X^* \mid w_{x_1}, w_{x_2}, \dots, w_{x_{i-1}}, w_{x_{i+1}}, \dots, w_{x_r} \in I\},$$

where $I = \{2^0, 2^1, 2^2, \dots, 2^n, \dots\}$. Then $L_{x_i} \in \mathscr{D}_3$, $i = 1, 2, \dots, r$.

Proof. For any $i \in \{1, 2, \dots, r\}$, we assert that the $P_{L_{x_i}}$-classes are

$$C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r} = \{w \in X^* \mid w_{x_1} = j_1, w_{x_2} = j_2, \dots, w_{x_{i-1}} = j_{i-1}, \\ w_{x_{i+1}} = j_{i+1}, \dots, w_{x_r} = j_r\},$$

$j_1, \dots, j_{i-1}, j_{i+1}, \dots, j_r = 0, 1, 2, \dots$. In fact, suppose $u, v \in X^*$, and

$$u_{x_s} = p,\ v_{x_s} = q,$$

for some letter x_s with $x_s \in X \setminus \{x_i\}$, where $p, q \in \mathbb{N}^0$, $p \neq q$. Since $p \neq q$, we may let $p - q = k$, $k \in \mathbb{N}$. Then for a sufficient large m, $m \in \mathbb{N}$, we can find $x_s^n \in X^*$, $n \in \mathbb{N}$ such that

$$(x_s^n u)_{x_s} = n + p = 2^m,\ \ 2^m - 2^{m-1} > k.$$

Hence, we have $(x_s^n v)_{x_s} = n + q = n + p - k = 2^m - k$, and $2^{m-1} < 2^m - k < 2^m$.

Consider

$$z = x_{t_1}^{n_1} x_{t_2}^{n_2} \cdots x_{t_{r-2}}^{n_{r-2}} x_s^n u$$

and

$$z' = x_{t_1}^{n_1} x_{t_2}^{n_2} \cdots x_{t_{r-2}}^{n_{r-2}} x_s^n v,$$

where $t_1, t_2, \dots, t_{r-2}$ is an arrangement of $\{1, 2, \dots, r\} \setminus \{s, i\}$, and $n_1, n_2, \dots, n_{r-2} \in \mathbb{N}^0$. Choose $n_1, n_2, \dots, n_{r-2}$ such that $z_{x_{t_1}}, \dots, z_{x_{t_{r-2}}} \in I$, and by the above discussion, we have $z_{x_s} = 2^m$ and $z_{x_s} \in I$, but $z'_{x_s} = 2^m - k$, $z'_{x_s} \notin I$. By the construction of L_{x_i}, we have $z \in L_{x_i}$ and $z' \notin L_{x_i}$, and hence we conclude that $u \not\equiv v(P_{L_{x_i}})$. This result shows that each $P_{L_{x_i}}$-class is contained in some $C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r}$, $j_1, \dots, j_{i-1}, j_{i+1}, \dots, j_r \in \mathbb{N}^0$. On the other hand, for any $w, w' \in C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r}$, $C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r} \in \{C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r} \mid j_1, \dots, j_{i-1}, j_{i+1}, \dots, j_r \in \mathbb{N}^0\}$, if $u, v \in X^*$ and $uwv \in L_{x_i}$, then by the construction of L_{x_i}, we have $(uwv)_{x_s} = u_{x_s} + w_{x_s} + v_{x_s} \in I$, $s = 1, 2, \dots, i-1, i+1, \dots, r$. Since $w_{x_s} = w'_{x_s}$ for $s = 1, 2, \dots, i-1, i+1, \dots, r$, we have $(uw'v)_{x_s} = u_{x_s} + w'_{x_s} + v_{x_s} = u_{x_s} + w_{x_s} + v_{x_s} \in I$, $s = 1, 2, \dots, i-1, i+1, \dots, r$. This implies that $uw'v \in L_{x_i}$. Dually, for all $u, v \in X^*$, we can deduce that $uwv \in L_{x_i}$ from the fact $uw'v \in L_{x_i}$. Hence, $w \equiv w'(P_{L_{x_i}})$. Thus the assertion holds.

For each $C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r} \in \{C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r} \mid j_1, \dots, j_{i-1}, j_{i+1}, \dots, j_r \in \mathbb{N}^0\}$, we observe that for every word w in $C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r}$, the w_{x_i} is not restricted, and so w_{x_i} can be any number in $\mathbb{N}^0$. Since $\mathbb{N}^0$ is infinite, $C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r}$ is infinite. Moreover, by the definition of $C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r}$, for all $w \in C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r}$, $w_{x_1} = j_1$, where j_1 is a given number in $\mathbb{N}^0$, we have for any word $u \in X^*$ with u_{x_1} greater than j_1, u is not a subword of any word in $C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r}$. This shows that $C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r}$ is not dense. Thus $C_{j_1 j_2 \cdots j_{i-1} j_{i+1} \cdots j_r}$ is an infinite but not dense language. This shows that $L_{x_i} \in \mathscr{D}_3$. ∎

Consider the following languages.

$L_{x_r} = \{w \in X^* |\ w_{x_1}, w_{x_2}, \ldots,\ w_{x_{r-1}} \in I\}$, where $I = \{2^0, 2^1, 2^2, \ldots, 2^n, \ldots\}$,

$L_{x_1} = \{w \in X^* |\ w_{x_2}, \ldots,\ w_{x_{r-1}},\ w_{x_r} \in I\}$, where $I = \{2^0, 2^1, 2^2, \ldots, 2^n, \ldots\}$.

By Lemma 3.2, $L_{x_r}, L_{x_1} \in \mathscr{D}_3$. Let

$$L_1 = L_{x_r}\{x_r\},\ \ L_2 = L_{x_1}\{x_1\}.$$

Since $\{x_r\},\ \{x_1\}$ are suffix codes, by Corollary 3.1, $L_1, L_2 \in \mathscr{D}_t$ and $L_1, L_2 \notin \mathscr{D}_f$ by Proposition 3.1. Thus, we arrive at the following proposition.

Proposition 3.3. *$L_1, L_2 \in \mathscr{D}_3$ and $L_1 \cap L_2 = \emptyset$.*

We also have the following proposition.

Proposition 3.4. *Let $L = L_1 \cup L_2$. Then $L \in \mathscr{D}_2$.*

Proof. To proceed with the proof, we assert that if $u \equiv v(P_L)$ for $u, v \in X^*$, $u \neq v$, then $u_{x_i} = v_{x_i}$ for all $x_i \in X$, $i = 1, 2, \ldots, r$. In fact, if $u, v \in X^*$, $u_{x_s} \neq v_{x_s}$, $x_s \in X \setminus \{x_r\}$, then by the proof of Lemma 3.2, there exist $w_1, w_2 \in X^*$ such that $w_1 u w_2 \in L_{x_r}, w_1 v w_2 \notin L_{x_r}$, and so $w_1 u w_2 x_r \in L_1, w_1 v w_2 x_r \notin L_1$, also by the construction of L_2, $w_1 v w_2 x_r \notin L_2$, we have $u \not\equiv v(P_L)$. If $u, v \in X^*$, $u_{x_r} \neq v_{x_r}$, then by the proof of Lemma 3.2 again, there exist $w_1, w_2 \in X^*$ such that $w_1 u w_2 \in L_{x_1}, w_1 v w_2 \notin L_{x_1}$, and so we have $w_1 u w_2 x_1 \in L_2, w_1 v w_2 x_1 \notin L_2$. Now, by the construction of L_1, $w_1 v w_2 x_1 \notin L_1$, we see that $u \not\equiv v(P_L)$. This shows that each P_L-class is contained in some $C_{j_1 j_2 \cdots j_r}$ with $j_1, j_2, \ldots, j_r \in \mathbb{N}^0$, where $C_{j_1 j_2 \cdots j_r} = \{w \in X^* |\ w_{x_1} = j_1, \ldots, w_{x_i} = j_i, \ldots, w_{x_r} = j_r\}$.

Moreover, for $C_{j_1 j_2 \cdots j_r} \in \{C_{j_1 j_2 \cdots j_r} | j_1, \ldots, j_r \ \in \mathbb{N}^0\}$ with $|C_{j_1 j_2 \cdots j_r}| = 1$, by the above discussions, we have $C_{j_1 j_2 \cdots j_r}$ is a P_L-class. For $C_{j_1 j_2 \cdots j_r} \in \{C_{j_1 j_2 \cdots j_r} | j_1, \ldots, j_r \ \in \mathbb{N}^0\}$ with $|C_{j_1 j_2 \cdots j_r}| \geq 2$. Suppose that $u, v \in C_{j_1 j_2 \cdots j_r}$, u ends at letter x_1, and v does not end at letter x_1. Then we have the following two cases:

(1) v ends at letter x_r,

(2) v does not end at letter x_r.

For case (1), in view of the proof in Lemma 3.2, we are able to find some word $w_1 \in X^*$ such that

$$(w_1 u)_{x_i} \in I,\ i = \ 2,\ \ldots,\ r,\ \ (w_1 u)_{x_1} \notin I.$$

This shows that $w_1 u \in L_{x_1}$, by the above assumption and the construction of L_2, $w_1 u \in L_2$. On the other hand, by $u, v \in C_{j_1 j_2 \cdots j_r}$, we have

$$(w_1 v)_{x_i} \in I,\ i = \ 2,\ \ldots,\ r,\ \ (w_1 v)_{x_1} \notin I.$$

This shows that $w_1 v \in L_{x_1}$ and $w_1 v \notin L_{x_r}$. Since v does not end at x_1, we have $w_1 v \notin L_2$, and by the construction of L_1, $w_1 v \notin L_1$. Hence, we have $u \not\equiv v(P_L)$.

For case (2), by using similar arguments as those in case (1), we can obtain $u \not\equiv v(P_L)$. Similarly, if $u, v \in C_{j_1 j_2 \cdots j_r}$, u ends at x_r, and v does not end at x_r, then we also have $u \not\equiv v(P_L)$. Thus, for $C_{j_1 j_2 \cdots j_r} \in \{C_{j_1 j_2 \cdots j_r} | j_1, \ldots, j_r \ \in \mathbb{N}^0\}$ with $|C_{j_1 j_2 \cdots j_r}| \geq 2$, $C_{j_1 j_2 \cdots j_r}$ can be divided into three parts:

$$C^I_{j_1 j_2 \cdots j_r} = \{w \in X^* |\ w_{x_1} = j_1, \ldots, w_{x_i} = j_i, \ldots, w_{x_r} = j_r, \\ \textit{and } w \textit{ ends at } x_i, i = 2, \ldots, r-1\},$$

$$C^{II}_{j_1 j_2 \cdots j_r} = \{w \in X^* | \ w_{x_1} = j_1, \ldots, w_{x_i} = j_i, \ldots, w_{x_r} = j_r, \\ \textit{and } w \textit{ ends at } x_1\},$$

$$C^{III}_{j_1 j_2 \cdots j_r} = \{w \in X^* | \ w_{x_1} = j_1, \ldots, w_{x_i} = j_i, \ldots, w_{x_r} = j_r, \\ \textit{and } w \textit{ ends at } x_r\},$$

and if P_L-class is contained in $C_{j_1 j_2 \cdots j_r}$ with $|C_{j_1 j_2 \cdots j_r}| \geq 2$, $j_1, j_2, \ldots, j_r \in \mathbb{N}^0$, then it must be in some $C^J_{j_1 j_2 \cdots j_r}$, $J \in \{I, II, III\}$. On the other hand, if u, v are any two words in $C^I_{j_1 j_2 \cdots j_r}$, $|C_{j_1 j_2 \cdots j_r}| \geq 2$, $j_1, \ldots, j_r \in \mathbb{N}^0$, that is, $u, v \in C_{j_1 j_2 \cdots j_r}$ and both of u and v end neither at x_1 nor at x_r, then for any $w_1, w_2 \in X^*$, $(w_1 u w_2)_{x_i} = (w_1 v w_2)_{x_i}$, $i = 1, 2, \ldots, r$. So $w_1 u w_2 \in L_{x_1}$ if and only if $w_1 v w_2 \in L_{x_1}$, and $w_1 u w_2 \in L_{x_r}$ if and only if $w_1 v w_2 \in L_{x_r}$. If $w_2 = 1$, then by previous assumption, both of $w_1 u w_2$ and $w_1 v w_2$ belong neither to L_1, nor to L_2. If $w_2 \neq 1$, then $w_1 u w_2 \in L_1$ if and only if $w_1 v w_2 \in L_1$, and $w_1 u w_2 \in L_2$ if and only if $w_1 v w_2 \in L_2$. This implies that $w_1 u w_2 \in L$ if and only if $w_1 v w_2 \in L$. Both of the two cases show that $u \equiv v(P_L)$. So $C^I_{j_1 j_2 \cdots j_r}$, $|C_{j_1 j_2 \cdots j_r}| \geq 2$, $j_1, \ldots, j_r \in \mathbb{N}^0$, is a P_L-class. Similarly, we can show that $C^{II}_{j_1 j_2 \cdots j_r}$ and $C^{III}_{j_1 j_2 \cdots j_r}$ are P_L-classes, $|C_{j_1 j_2 \cdots j_r}| \geq 2$, $j_1, \ldots, j_r \in \mathbb{N}^0$.

In view of the above facts, we deduce that for $C_{j_1 j_2 \cdots j_r} \in \{C_{j_1 j_2 \cdots j_r} | j_1, \ldots, j_r \in \mathbb{N}^0\}$ with $|C_{j_1 j_2 \cdots j_r}| \geq 2$, the P_L-classes are

$$C^{I}_{j_1 j_2 \cdots j_r} = \{w \in X^* | \ w_{x_1} = j_1, \ldots, w_{x_i} = j_i, \ldots, w_{x_r} = j_r, \\ \textit{and } w \textit{ ends at } x_i, i = 2, \ldots, r-1\},$$

$$C^{II}_{j_1 j_2 \cdots j_r} = \{w \in X^* | \ w_{x_1} = j_1, \ldots, w_{x_i} = j_i, \ldots, w_{x_r} = j_r, \\ \textit{and } w \textit{ ends at } x_1\},$$

$$C^{III}_{j_1 j_2 \cdots j_r} = \{w \in X^* | \ w_{x_1} = j_1, \ldots, w_{x_i} = j_i, \ldots, w_{x_r} = j_r, \\ \textit{and } w \textit{ ends at } x_r\}.$$

Clearly,

$$|C^{I}_{j_1 j_2 \cdots j_r}| + |C^{II}_{j_1 j_2 \cdots j_r}| + |C^{III}_{j_1 j_2 \cdots j_r}| = |C_{j_1 j_2 \cdots j_r}|$$

and

$$|C_{j_1 j_2 \cdots j_r}| \leq r^{j_1 + j_2 + \cdots + j_r}.$$

This shows that $L \in \mathscr{D}_2$. ∎

We will construct a language of the form $L = L_1 \cup L_2$, where L is in $\mathscr{D}_3$ and L_1, L_2 are two disjoint languages in $\mathscr{D}_4$. To this aim, we need some preparations. Recall that a nonempty language L over X is an *infix language* if for all $x, y, u \in X^*$, $u \in L$ and $xuy \in L$ together imply $x = y = 1$. Clearly, each infix language contained in X^+ is a code, we usually call this code an *infix code*.

For any $x \in X^+$ with $|X| \geq 2$, we let

$$P_{nt}(x) = \{w \in X^+ \mid x = wu \textit{ for some } u \in X^+\},$$

$$S_{nt}(x) = \{w \in X^+ \mid x = uw \textit{ for some } u \in X^+\},$$

and

$$I(x) = \{w \in X^* \mid x = uwv \textit{ for some } u, v \in X^*\}.$$

Remark 3.1. We denote the set of all non-trivial prefixes(suffixes) of word x by $P_{nt}(x)$ ($S_{nt}(x)$) and the set of all infixes of word x by $I(x)$.

Definition 3.1. [7,13] *Let $L \subseteq X^+$, $L \neq \emptyset$. Then, we call L a* solid code *if L is an infix code and $P_{nt}(u) \cap S_{nt}(v) = \emptyset$ for every $u, v \in L$.*

Corollary 3.2. [13] *Any nonempty subset of a solid code is also a solid code.*

Definition 3.2. [13] *Let $L \subseteq X^+$, $L \neq \emptyset$ and $w \in X^*$. Then, we call the factorization*

$$w = x_1y_1x_2y_2\cdots x_ny_nx_{n+1} \tag{3.1}$$

an L-representation *of w if $y_i \in L$, $I(x_j) \cap L = \emptyset$, for $i = 1,2,\ldots,n$, $j = 1,2,\ldots,n+1$, $n \in \mathbb{N}^0$. And we call the case $n = 0$ in* (3.1)*, that is, $w = x_1$, the* trivial L-representation *of w. Obviously, w has the trivial L-representation if and only if $I(w) \cap L = \emptyset$, and at this time, w has only the trivial L-representation.*

We call n in (3.1) *an L*-length *of w, denote the set of L-lengths of w by $L_L(w)$ and call $(x_1,x_2,\ldots,x_n,x_{n+1})$ in* (3.1) *an L*-coefficient *of w, and the set of L-coefficients of w is denoted by $C_L(w)$.*

Proposition 3.5. [12,13] *Let L be a nonempty language in X^+. Then the following statements hold.*

(1) *Any word w over X has L-representation.*
(2) *Any word w over X has unique L-representation if and only if L is a solid code.*

Obviously, if L is a solid code, then for any $w \in X^*$, the L-length of w and L-coefficient of w are unique, at this time, we denote them by $l_L(w)$ and $c_L(w)$ respectively.

In the following, we let $|X| \geq 2$, and $\{a,b\} \subseteq X$. Consider $\{ba\}$, by the definition of solid code, $\{ba\}$ is a solid code. Now, by Proposition 3.5, for any word $w \in X^*$, w has unique $\{ba\}$-representation

$$w = x_1y_1\cdots x_ny_nx_{n+1},$$

where $y_i = ba, i = 1,2,\ldots,n$, and $ba \notin I(x_j), j = 1,2,\ldots,n+1$. Let $\overline{w} = x_1x_2\cdots x_nx_{n+1}$, where $(x_1,x_2,\ldots,x_{n+1})$ is the $\{ba\}$-coefficient of w. Notice that ba may be in $I(\overline{w})$, for example, if $w = bbaa$, then $\overline{w} = ba$. We use $\overline{w}_a$ and $\overline{w}_b$ to denote the numbers of the letter a and b occurring in word $\overline{w}$ respectively. Clearly, for any word $w \in X^*$, $\overline{w}$ is unique, and hence $\overline{w}_a$ and $\overline{w}_b$ are unique.

Proposition 3.6. *Let $L_1 = \{w \in X^* \mid \overline{w}_a = \overline{w}_b\}$. Then $L_1 \in \mathscr{D}_4$.*

Proof. We assert that the P_{L_1}-classes are

$$C_i = \{w \in X^* | \overline{w}_a = \overline{w}_b + i\},\ \ i = 0,\ \pm 1,\ \pm 2,\ldots.$$

In fact, if u, v are any two words over X with $\overline{u}_a - \overline{u}_b = i$, $\overline{v}_a - \overline{v}_b = j$, $i \neq j$, without loss of generality, we may let $i \geq 0$, then, obviously, for $x = 1, y = b^i$, we have

$$xuy = ub^i,\ xvy = vb^i,$$

by the definition of $\overline{xuy}, \overline{xvy}$, we have

$$\overline{xuy} = \overline{u}b^i,\ \overline{xvy} = \overline{v}b^i.$$

And so

$$\overline{xuy}_a = \overline{u}_a,\ \overline{xuy}_b = \overline{u}_b + i, \text{ and } \overline{xuy}_a = \overline{xuy}_b \text{ by assumption},$$

$$\overline{xvy}_a = \overline{v}_a,\ \overline{xvy}_b = \overline{v}_b + i,\ \textit{and}\ \overline{xvy}_a \neq \overline{xvy}_b \text{ by assumption.}$$

By the construction of L_1, we have $xuy \in L_1, xvy \notin L_1$. Hence, $u \not\equiv v(P_{L_1})$.

On the other hand, for every two words $u, v \in C_i$, $i = 0, \pm 1, \pm 2, \ldots$, by the definition of C_i, we have

$$\overline{u}_a - \overline{u}_b = \overline{v}_a - \overline{v}_b = i.$$

Consider wu and wv, $w \in X^*$. Then we divide our discussion into the following cases:

(1) Both of u, v begin with letter a. In this case, let $u = au', v = av'$, where $u', v' \in X^*$. If $w = w'b, w' \in X^*$, then $\overline{wu} = \overline{w'bau'}$ and $\overline{wv} = \overline{w'bav'}$, hence

$$\overline{wu}_a = \overline{w'}_a + \overline{u'}_a = \overline{w'}_a + \overline{u}_a - 1,$$
$$\overline{wu}_b = \overline{w'}_b + \overline{u'}_b = \overline{w'}_b + \overline{u}_b,$$

and

$$\overline{wv}_a = \overline{w'}_a + \overline{v'}_a = \overline{w'}_a + \overline{v}_a - 1,$$
$$\overline{wv}_b = \overline{w'}_b + \overline{v'}_b = \overline{w'}_b + \overline{v}_b,$$

so

$$\overline{wu}_a - \overline{wu}_b = \overline{w'}_a - \overline{w'}_b + i - 1,$$
$$\overline{wv}_a - \overline{wv}_b = \overline{w'}_a - \overline{w'}_b + i - 1.$$

If $w = w'a, w' \in X^*$, and $\overline{w}_a = \overline{w'}_a + 1$, then $\overline{wu} = \overline{w'aau'}$ and $\overline{wv} = \overline{w'aav'}$, hence

$$\overline{wu}_a = \overline{w'}_a + \overline{u}_a + 1,$$
$$\overline{wu}_b = \overline{w'}_b + \overline{u}_b,$$

and

$$\overline{wv}_a = \overline{w'}_a + \overline{v}_a + 1,$$
$$\overline{wv}_b = \overline{w'}_b + \overline{v}_b,$$

so

$$\overline{wu}_a - \overline{wu}_b = \overline{w'}_a - \overline{w'}_b + i + 1,$$
$$\overline{wv}_a - \overline{wv}_b = \overline{w'}_a - \overline{w'}_b + i + 1.$$

If $w = w'x$, where $w' \in X^*$ and $x \in (X \cup \{ba\}) \setminus \{a, b\}$, then $\overline{wu} = \overline{w'xau'}$ and $\overline{wv} = \overline{w'xav'}$, hence

$$\overline{wu}_a = \overline{w'}_a + \overline{u}_a,$$
$$\overline{wu}_b = \overline{w'}_b + \overline{u}_b,$$

and

$$\overline{wv}_a = \overline{w'}_a + \overline{v}_a,$$
$$\overline{wv}_b = \overline{w'}_b + \overline{v}_b,$$

so

$$\overline{wu}_a - \overline{wu}_b = \overline{w'}_a - \overline{w'}_b + i,$$
$$\overline{wv}_a - \overline{wv}_b = \overline{w'}_a - \overline{w'}_b + i.$$

Therefore, in every case, we have $\overline{wu}_a - \overline{wu}_b = \overline{wv}_a - \overline{wv}_b$, and hence $\overline{wu}_a = \overline{wu}_b$ if and only if $\overline{wv}_a = \overline{wv}_b$, for any $w \in X^*$.

(2) u begins with letter a and v does not begin with letter a or vice verse. Without loss of generality, suppose that $u = au'$, where $u' \in X^*$. In this case, if $w = w'b, w' \in X^*$, then for $wu = w'bau'$, we have $\overline{wu}_a - \overline{wu}_b = \overline{w'}_a - \overline{w'}_b + i - 1$; for $wv = w'bv$, we have $\overline{wv}_a - \overline{wv}_b =$

$\overline{w'}_a - \overline{w'}_b + i - 1$. If $w = w'a, w' \in X^*$ and $\overline{w}_a = \overline{w'}_a + 1$, then for $wu = w'aau'$, we have $\overline{wu}_a - \overline{wu}_b = \overline{w'}_a - \overline{w'}_b + i + 1$; for $wv = w'av$, we have $\overline{wv}_a - \overline{wv}_b = \overline{w'}_a - \overline{w'}_b + i + 1$. If $w = w'x$, where $w' \in X^*$ and $x \in (X \cup \{ba\}) \setminus \{a, b\}$, then for $wu = w'xau'$, we have $\overline{wu}_a - \overline{wu}_b = \overline{w'}_a - \overline{w'}_b + i$; for $wv = w'xv$, we have $\overline{wv}_a - \overline{wv}_b = \overline{w'}_a - \overline{w'}_b + i$. Therefore, in every case, we also have $\overline{wu}_a - \overline{wu}_b = \overline{wv}_a - \overline{wv}_b$, and hence $\overline{wu}_a = \overline{wu}_b$ if and only if $\overline{wv}_a = \overline{wv}_b$, for any $w \in X^*$.

(3) u, v do not begin with letter a. Similar to case (1), we obtain that $\overline{wu}_a - \overline{wu}_b = \overline{wv}_a - \overline{wv}_b$, and hence $\overline{wu}_a = \overline{wu}_b$ if and only if $\overline{wv}_a = \overline{wv}_b$, for any $w \in X^*$.

From the above three cases, we deduce that if $u, v \in C_i$, $i = 0, \pm 1, \pm 2, \ldots$, then for any $w \in X^*$,

$$\overline{wu}_a - \overline{wu}_b = \overline{wv}_a - \overline{wv}_b, \tag{3.2}$$

and

$$\overline{wu}_a = \overline{wu}_b \Leftrightarrow \overline{wv}_a = \overline{wv}_b. \tag{3.3}$$

Similarly, if $u, v \in C_i$, $i = 0, \pm 1, \pm 2, \ldots$, then for any $w' \in X^*$,

$$\overline{uw'}_a - \overline{uw'}_b = \overline{vw'}_a - \overline{vw'}_b, \tag{3.4}$$

and

$$\overline{uw'}_a = \overline{uw'}_b \Leftrightarrow \overline{vw'}_a = \overline{vw'}_b. \tag{3.5}$$

Then, for every $u, v \in C_i$, $i = 0, \pm 1, \pm 2, \ldots$, for any $w, w' \in X^*$, consider wuw' and wvw'. Since $u, v \in C_i$, by (3.2) and by the definition of C_i, we see immediately that both wu and wv are in some $C_{i'}$, $i' \in \{0, \pm 1, \pm 2, \ldots\}$. Hence, by (3.5), we have $\overline{wuw'}_a = \overline{wuw'}_b$ if and only if $\overline{wvw'}_a = \overline{wvw'}_b$, that is, $wuw' \in L_1$ if and only if $wvw' \in L_1$. This result implies that $u \equiv v(P_{L_1})$. In view of the above facts, we have P_{L_1}-classes are

$$C_i = \{w \in X^* | \overline{w}_a = \overline{w}_b + i\}, \ \ i = 0, \ \pm 1, \ \pm 2, \ldots.$$

It can be easily verified that each C_i is dense, $i = 0, \ \pm 1, \ \pm 2, \ldots$. Hence $L_1 \in \mathscr{D}_4$. ∎

Proposition 3.7. *Let* $L_2 = \{w \in X^* \mid \overline{w}_a = 2\overline{w}_b\}$. *Then* $L_2 \in \mathscr{D}_4$.

Proof. We first show that $\overline{u}_a - 2\overline{u}_b \neq \overline{v}_a - 2\overline{v}_b$ implies $u \not\equiv v(P_{L_2})$, for any two words $u, v \in X^*$. In fact, if $\overline{u}_a - 2\overline{u}_b = i$, $\overline{v}_a - 2\overline{v}_b = j$, $i \neq j$, then without loss of generality, we may let $i \geq 0$. Then we consider $a^s ub^t$ and $a^s vb^t$, where $s, t \in \mathbb{N}^0$, by the definition of $\overline{a^s ub^t}$, we have

$$\overline{a^s ub^t}_a = s + \overline{u}_a, \ \overline{a^s ub^t}_b = t + \overline{u}_b.$$

Choose s, t such that $s + i = 2t$, then by assumption, we have

$$s + \overline{u}_a = s + 2\overline{u}_b + i = 2(t + \overline{u}_b),$$

so $\overline{a^s ub^t}_a = 2\overline{a^s ub^t}_b$, by the construction of L_2, $a^s ub^t \in L_2$, while

$$\overline{a^s vb^t}_a = s + \overline{v}_a = s + 2\overline{v}_b + j,$$

$$2\overline{a^s vb^t}_b = 2t + 2\overline{v}_b,$$

clearly $\overline{a^s vb^t}_a \neq 2\overline{a^s vb^t}_b$ since $s + j \neq 2t$. This shows that $a^s vb^t \notin L_2$. Hence, $u \not\equiv v(P_{L_2})$. So let

$$C_j = \{w \in X^* \mid \overline{w}_a = 2\overline{w}_b + j\}, \ \ j = 0, \ \pm 1, \ \pm 2, \ldots,$$

we have $u \equiv v(P_{L_2})$ implies $u, v \in C_j$ for some $j \in \{0, \ \pm 1, \ \pm 2, \ldots\}$.

Next we will show that the P_{L_2}-classes are not analogous to the P_{L_1}-classes. In fact, if $u,v \in C_j$ for some $j \in \{0, \pm 1, \pm 2, \ldots\}$, u begins with letter a and v does not begin with a or vice verse, then without loss of generality, we may let $u = au'$, $u' \in X^*$. Consider bu and bv, since $\overline{bu}_a = \overline{u'}_a = \overline{u}_a - 1$, $\overline{bu}_b = \overline{u'}_b = \overline{u}_b$, $\overline{bv}_a = \overline{v}_a$ and $\overline{bv}_b = \overline{v}_b + 1$, we have $\overline{bu}_a - 2\overline{bu}_b \neq \overline{bv}_a - 2\overline{bv}_b$. By discussion in the previous paragraph, $bu \not\equiv bv(P_{L_2})$, so $u \not\equiv v(P_{L_2})$. Moreover, if $u,v \in C_j$ for some $j \in \{0, \pm 1, \pm 2, \ldots\}$, both of u and v begin with letter a and u ends at letter b, v does not end at letter b or vice verse, then without loss of generality, we may let $u = u'b$, $u' \in X^*$. Consider ua and va, then we have $\overline{ua}_a = \overline{u}_a$, $\overline{ua}_b = \overline{u}_b - 1$, $\overline{va}_a = \overline{v}_a + 1$ and $\overline{va}_b = \overline{v}_b$, so $\overline{ua}_a - 2\overline{ua}_b \neq \overline{va}_a - 2\overline{va}_b$, by the discussion in the previous paragraph again, we have $ua \not\equiv va(P_{L_2})$, and so $u \not\equiv v(P_{L_2})$. Similarly, if $u,v \in C_j$ for some $j \in \{0, \pm 1, \pm 2, \ldots\}$, u and v do not begin with letter a and u ends at the letter b, v does not end at the letter b or vice verse, then we also have $u \not\equiv v(P_{L_2})$. Hence, each $C_j, j = 0, \pm 1, \pm 2, \ldots$, can be divided into four parts, say

$$C_j^I = \{w \in X^* \mid \overline{w}_a = 2\overline{w}_b + j,\ w \textit{ begins with letter } a \textit{ and ends at letter } b\},$$

$$C_j^{II} = \{w \in X^* \mid \overline{w}_a = 2\overline{w}_b + j,\ w \textit{ begins with letter } a \textit{ and does not end at letter } b\},$$

$$C_j^{III} = \{w \in X^* \mid \overline{w}_a = 2\overline{w}_b + j,\ w \textit{ does not begin with letter } a \textit{ and ends at letter } b\},$$

$$C_j^{IV} = \{w \in X^* \mid \overline{w}_a = 2\overline{w}_b + j,\ w \textit{ neither begins with letter } a \textit{ nor ends at letter } b\},$$

and we have $u \equiv v(P_{L_2})$ implies $u,v \in C_j^J$ for some $j \in \{0, \pm 1, \pm 2, \ldots\}$, $J \in \{I, II, III, IV\}$. It is routine to check that the converse implication holds for each j and J. Hence, P_{L_2}-classes are

$$C_j^I = \{w \in X^* \mid \overline{w}_a = 2\overline{w}_b + j,\ w \textit{ begins with letter } a \textit{ and ends at letter } b\},$$

$$C_j^{II} = \{w \in X^* \mid \overline{w}_a = 2\overline{w}_b + j,\ w \textit{ begins with letter } a \textit{ and does not end at letter } b\},$$

$$C_j^{III} = \{w \in X^* \mid \overline{w}_a = 2\overline{w}_b + j,\ w \textit{ does not begin with letter } a \textit{ and ends at letter } b\},$$

$$C_j^{IV} = \{w \in X^* \mid \overline{w}_a = 2\overline{w}_b + j,\ w \textit{ neither begins with letter } a \textit{ nor ends at letter } b\},$$

where $j = 0, \pm 1, \pm 2, \ldots$. It is easy to check that each P_{L_2}-class is dense. Hence, $L_2 \in \mathscr{D}_4$. ∎

We now construct a language in $\mathscr{D}_3$ which is a disjoint union of the above two languages L_1 and L_2 in $\mathscr{D}_4$.

Consider $L_1 \cap L_2$. If $w \in L_1 \cap L_2$, then by the constructions of L_1 and L_2, we have that $\overline{w}_a = \overline{w}_b$, $\overline{w}_a = 2\overline{w}_b$, so $\overline{w}_a = \overline{w}_b = 0$. This result implies that $w \in ((X \cup \{ba\}) \setminus \{a,b\})^*$, where $(X \cup \{ba\}) \setminus \{a,b\}$ is a finite language over X since X is finite, and

$$((X \cup \{ba\}) \setminus \{a,b\})^* = \{1\} \cup ((X \cup \{ba\}) \setminus \{a,b\}) \cup ((X \cup \{ba\}) \setminus \{a,b\})^2 \cup \cdots .$$

Let $C = ((X \cup \{ba\}) \setminus \{a,b\})^*$. By the above discussion, we have $L_1 \cap L_2 \subseteq C$, and $L_1 \cap L_2 \supseteq C$ is obviously. Hence, we have $L_1 \cap L_2 = C$. By the definition of *rational language* [4], C is a rational language over X. Then by Theorem 4.2.9 of [4], C is regular. Hence $L_1 \setminus C \in \mathscr{D}_4$ by Corollary 2.1. Let $L_1' = L_1 \setminus C$. Then, it is clear that $L_1' \cap L_2 = \emptyset$. We have

$$L_1 \cup L_2 = L_1' \,\dot{\cup}\, L_2 = \{w \in X^* \mid \overline{w}_a = \overline{w}_b\} \cup \{w \in X^* \mid \overline{w}_a = 2\overline{w}_b\}.$$

We have the following proposition.

Proposition 3.8. *$L = L_1' \cup L_2$ is in $\mathscr{D}_3$.*

Proof. We first assert that if $u,v \in X^*$ and $u \equiv v(P_L)$, then $u,v \in C_{ij}$ for some $i,j \in \mathbb{N}^0$, where

$$C_{ij} = \{w \in X^* \mid \overline{w}_a = i, \overline{w}_b = j\},\ i,\ j \in \mathbb{N}^0.$$

In fact, for any two words u,v over X, suppose that u,v are not in the same C_{ij}, that is, $\overline{u}_a \neq \overline{v}_a$ or $\overline{u}_b \neq \overline{v}_b$, we only discuss the case $\overline{u}_a \neq \overline{v}_a$, the case for $\overline{u}_b \neq \overline{v}_b$ can be similarly obtained. Now we divide our discussion into the following two cases:

(1) $\overline{u}_a - \overline{u}_b = \overline{v}_a - \overline{v}_b$,
(2) $\overline{u}_a - \overline{u}_b \neq \overline{v}_a - \overline{v}_b$.

For case (1), we consider $a^s(ba)u(ba)b^t$ and $a^s(ba)v(ba)b^t$, $s,t \in \mathbb{N}^0$. Choose s,t such that $s,t > 0$ and $s + \overline{u}_a = 2(t + \overline{u}_b)$, then by the construction of L_2, we have $a^s(ba)u(ba)b^t \in L_2$ and hence $a^s(ba)u(ba)b^t \notin L_1'$. Since $s,t > 0$, $a^s(ba)u(ba)b^t \notin C$. We have $a^s(ba)u(ba)b^t \notin L_1$. In view of the proof of Proposition 3.6, we have $u \equiv v(P_{L_1})$ because of the fact $\overline{u}_a - \overline{u}_b = \overline{v}_a - \overline{v}_b$, thus $a^s(ba)v(ba)b^t \notin L_1$. And we have

$$\begin{aligned}
\overline{a^s(ba)v(ba)b^t}_a - 2\overline{a^s(ba)v(ba)b^t}_b &= s + \overline{v}_a - 2\,(t + \overline{v}_b) \\
&= \overline{v}_a - 2\overline{v}_b - (\overline{u}_a - 2\overline{u}_b) \\
&= \overline{v}_a - \overline{v}_b - (\overline{u}_a - \overline{u}_b) + \overline{u}_b - \overline{v}_b \\
&= \overline{u}_b - \overline{v}_b \\
&= \overline{u}_a - \overline{v}_a \\
&\neq 0.
\end{aligned}$$

So $a^s(ba)v(ba)b^t \notin L_2$. Hence $a^s(ba)v(ba)b^t \notin L$. This shows $u \not\equiv v(P_L)$.

For case (2), if $\overline{u}_a - 2\overline{u}_b = \overline{v}_a - 2\overline{v}_b$, then similar to the discussion in the case (1), we choose s,t such that $s,t > 0$ and $s + \overline{u}_a = t + \overline{u}_b$. Then by the construction of L_1', $a^s(ba)u(ba)b^t \in L_1'$ and hence we deduce that $a^s(ba)u(ba)b^t \notin L_2$. This implies that $s + \overline{u}_a - 2t - 2\overline{u}_b \neq 0$. This shows that $\overline{a^s(ba)v(ba)b^t}_a - 2\overline{a^s(ba)v(ba)b^t}_b = s + \overline{v}_a - 2t - 2\overline{v}_b = s - 2t + \overline{u}_a - 2\overline{u}_b \neq 0$, $a^s(ba)v(ba)b^t \notin L_2$. Also $\overline{a^s(ba)v(ba)b^t}_a - \overline{a^s(ba)v(ba)b^t}_b = s + \overline{v}_a -$

$t-\overline{v}_b=\overline{v}_a-\overline{v}_b-(\overline{u}_a-\overline{u}_b)\neq 0$, so $a^s(ba)v(ba)b^t\notin L_1$. Hence, $a^s(ba)v(ba)b^t\notin L$. We have $u\not\equiv v(P_L)$.

If $\overline{u}_a-2\overline{u}_b\neq\overline{v}_a-2\overline{v}_b$, then choose s,t such that $s,t>0$ and $s+\overline{u}_a=t+\overline{u}_b$. Then by the discussion in the previous paragraph, we conclude that $a^s(ba)u(ba)b^t\in L_1'$ and $a^s(ba)v(ba)b^t\notin L_1$. Consider

$$\overline{a^s(ba)v(ba)b^t}_a-\ 2\overline{a^s(ba)v(ba)b^t}_b=\ s+\overline{v}_a-\ 2t-\ 2\overline{v}_b,$$

if $s+\overline{v}_a-\ 2t-\ 2\overline{v}_b\neq 0$, then $a^s(ba)v(ba)b^t\notin L_2$. So we have $a^s(ba)v(ba)b^t\notin L$. This shows that $u\not\equiv v(P_L)$. If $s+\overline{v}_a-\ 2t-\ 2\overline{v}_b=0$, then, we choose another $t'\in\mathbb{N}^0$ such that $s+\overline{v}_a=t'+\overline{v}_b$. Then $t'=2t+\overline{v}_b$, and $t'>0$ by $t>0$, and so $a^s(ba)v(ba)b^{t'}\in L_1'$. On the other hand, we have $\overline{a^s(ba)u(ba)b^{t'}}_a-\ \overline{a^s(ba)u(ba)b^{t'}}_b=s+\overline{u}_a-t'-\overline{u}_b$. Since $s+\overline{u}_a=t+\overline{u}_b$, we have $s+\overline{u}_a-t'-\overline{u}_b=t-t'$, so $s+\overline{u}_a-t'-\overline{u}_b\neq 0$ by $t'=2t+\overline{v}_b$ and $t>0$. Hence, $a^s(ba)u(ba)b^{t'}\notin L_1$. And $\overline{a^s(ba)u(ba)b^{t'}}_a-\ 2\overline{a^s(ba)u(ba)b^{t'}}_b=s+\overline{u}_a-2t'-2\overline{u}_b=-3t-2\overline{v}_b-\overline{u}_b\neq 0$. Clearly, we see that $a^s(ba)u(ba)b^{t'}\notin L_2$. Hence, $a^s(ba)u(ba)b^{t'}\notin L$. We also have $u\not\equiv v(P_L)$. Thus, we have shown that our assertion holds.

By the above assertion, we see immediately that each P_L-class is contained in some C_{ij}, $i,j\in\mathbb{N}^0$. Since C_{ij} is thin (if otherwise, we let $w=(ba)a^{i+1}b^{j+1}(ba)$. Then for any $u,v\in X^*$, we have $\overline{uwv}_a\geq i+1,\overline{uwv}_b\geq j+1$. This shows that $uwv\notin C_{ij}$, a contradiction) for $i,j=0,1,2,\cdots$, we easily see that each P_L-class is thin.

Next we continue to show that there are infinite P_L-classes. Consider

$$C_{00}=\ \{w\in X^*|\overline{w}_a=0,\overline{w}_b=0\}.$$

Clearly, $1\in C_{00}$. For any word $w\in C_{00}\setminus\{1\}$, we shall show that $1\not\equiv w(P_L)$. Take $x=b^2,y=a^3$. Then

$$x1y=\ b^2 1a^3=\ b^2a^3,\ xwy=\ b^2wa^3,$$

by the definition of $\overline{x1y}$, we have $\overline{x1y}=ba^2$, hence $\overline{x1y}_a=2$, $\overline{x1y}_b=1$ and $x1y\in L_2$. While $\overline{xwy}=b^2\overline{w}a^3$ because $w\neq 1$ and w neither begins with a nor ends at b. This leads to $\overline{xwy}_a=3$, $\overline{xwy}_b=2$. Hence, we have $xwy\notin L_1$, $xwy\notin L_2$, and so $xwy\notin L$. Hence, we have $1\not\equiv w(P_L)$. On the other hand, for any two words $u,v\in C_{00}\setminus\{1\}$, by the definition of C_{00}, we have $\overline{u}_a=\overline{u}_b=0,\overline{v}_a=\overline{v}_b=0$, and so for any $x,y\in X^*$,

$$\overline{xuy}_a=\ \overline{x}_a+\ \overline{y}_a,\ \overline{xuy}_b=\ \overline{x}_b+\ \overline{y}_b,$$

$$\overline{xvy}_a=\ \overline{x}_a+\ \overline{y}_a,\ \overline{xvy}_b=\ \overline{x}_b+\ \overline{y}_b.$$

This shows that

$$xuy\ \in L_1'\ \textit{if and only if}\ xvy\ \in L_1',$$

and

$$xuy\ \in L_2\ \textit{if and only if}\ xvy\ \in L_2.$$

Thus, we have

$$xuy\ \in L\ \textit{if and only if}\ xvy\ \in L.$$

Hence, we have proved that $u\equiv v(P_L)$.

Now, C_{00} can be divided into two parts, namely, $C_{00}\setminus\{1\}$ and $\{1\}$. In view of the above facts, we see that $C_{00}\setminus\{1\}$ and $\{1\}$ are both P_L-classes. Notice that $(ba)^+\subseteq C_{00}\setminus\{1\}$, where $(ba)^+=(ba)^*\setminus\{1\}=\{ba,(ba)^2,(ba)^3,\ldots\}$, so $C_{00}\setminus\{1\}$ is infinite. Therefore, $L\in\mathscr{D}_3$. ∎

4. The decomposition from $\mathscr{D}_1$ to $\mathscr{D}_2$

In [12], Shyr and Yu have shown the existence of a disjunctive language which can be partitioned into two parts such that both of them are midst-languages. The disjunctive language can be constructed on X with $|X| \geq 3$. In this section, we will improve their construction on X with $|X| \geq 2$ and simplify the construction of the disjunctive language as well. We will show that the disjunctive language is not only a disjoint union of midst-languages but is also a disjoint union of languages in $\mathscr{D}_2$.

In our study, the free monoid X^* sometimes needs to be equipped with a total order. In this paper, we only adopt the *standard total order* $\leq$ which is defined on X^* as follows [10]: For any $u,v \in X^*$, if $lg(u) < lg(v)$, then $u < v$; if $lg(u) = lg(v)$, then $\leq$ is the lexicographical order on X^n for all $n \geq 1$. For a word $x \in X^*$, we write $\sharp x = m$ if x stands at the mth position in this order.

Recall that $C_L(w)$ is the set of L-coefficients of w, for any word w over X and any nonempty language L in X^+.

We begin with the following definition.

Definition 4.1. *Let $L \subseteq X^+$, $L \neq \emptyset$. Then, we define a binary relation σ_L on X^* as follows:*

$$(w_1,\ w_2) \in \sigma_L \ \Leftrightarrow\ C_L(w_1) \cap C_L(w_2) \neq \emptyset.$$

If $(w_1,w_2) \in \sigma_L$, then w_1 and w_2 are said to be L-related .

Corollary 4.1. *σ_L is left compatible and right compatible with the operation on free monoid X^*.*

Proof. We only consider the left compatibility. For the case of right compatibility, it can be proved analogously. Suppose that $w_1 \sigma_L w_2$ holds. Then, by the definition of σ_L, there exist $(x_1,x_2,\ldots,x_{n+1})$ such that w_1 has an L-representation

$$w_1 = x_1y_1x_2y_2\cdots x_ny_nx_{n+1},$$

and w_2 has L-representation

$$w_2 = x_1y_1^{'}x_2y_2^{'}\cdots x_ny_n^{'}x_{n+1}.$$

Now, for any $x \in X^*$, we have

$$xw_1 = xx_1y_1x_2y_2\cdots x_ny_nx_{n+1},$$
$$xw_2 = xx_1y_1^{'}x_2y_2^{'}\cdots x_ny_n^{'}x_{n+1}.$$

If $I(xx_1) \cap L = \emptyset$, then, we let $xx_1 = x_1^{'}$. Now, we see that

$$xw_1 = x_1^{'}y_1x_2y_2\cdots x_ny_nx_{n+1},$$
$$xw_2 = x_1^{'}y_1^{'}x_2y_2^{'}\cdots x_ny_n^{'}x_{n+1}$$

are the L-representations of xw_1 and xw_2 respectively, so $xw_1\ \sigma_L\ xw_2$. If $I(xx_1) \cap L \neq \emptyset$, then xx_1 has an L-representation

$$xx_1 = u_1v_1\cdots u_mv_mu_{m+1},$$

where $v_i \in L$, $I(u_j) \cap L = \emptyset$, $i = 1,2,\ldots,m$, $j = 1,2,\ldots,m+1$. Clearly,

$$xw_1 = u_1v_1\cdots u_mv_mu_{m+1}y_1x_2y_2\cdots x_ny_nx_{n+1},$$
$$xw_2 = u_1v_1\cdots u_mv_mu_{m+1}y_1^{'}x_2y_2^{'}\cdots x_ny_n^{'}x_{n+1}$$

are L-representations of xw_1 and xw_2 respectively. Hence, we have proved that $xw_1 \ \sigma_L \ xw_2$. ∎

Lemma 4.1. *If L is a solid code, then σ_L is a congruence on X^*.*

Proof. By the definition of σ_L, σ_L is clearly reflexive and symmetric. If L is a solid code, then σ_L is transitive by Proposition 3.5 since, at this point, $(w_1,w_2) \in \sigma_L$ if and only if $c_L(w_1) = c_L(w_2)$. Hence, σ_L is an equivalence relation on X^*. By Corollary 4.1, we have proved that σ_L is a congruence on X^*. ∎

For the solid codes, we have the following lemma.

Lemma 4.2. *Let L be a solid code. Then the following statements are equivalent:*

(1) *L is finite.*
(2) *every σ_L-class is finite.*

Proof. $(1) \Rightarrow (2)$. Suppose that $|L| = m$, $m \in \mathbb{N}$. For any $w \in X^*$, if $I(w) \cap L \neq \emptyset$ and w has the unique L-representation $w = x_1y_1x_2y_2\cdots x_ny_nx_{n+1}$, then $|w\sigma_L| = m^n$. Moreover, if $I(w) \cap L = \emptyset$ and w has the trivial L-representation $w = x_1$, then $|w\sigma_L| = 1$.

$(2) \Rightarrow (1)$. Observe that L is a σ_L-class. Hence, every σ_L-class is finite implies L is finite. ∎

Now, let $|X| \geq 2$ and $X = \{a,b,\ldots\}$. Consider $u_1 = a^3b^3, u_2 = a^2bab^2, v = a^2b^2ab$. We have

$$P_{nt}(u_1) = \{a,a^2,a^3,a^3b,a^3b^2\},\ S_{nt}(u_1) = \{b,b^2,b^3,ab^3,a^2b^3\},$$
$$P_{nt}(u_2) = \{a,a^2,a^2b,a^2ba,a^2bab\},\ S_{nt}(u_2) = \{b,b^2,ab^2,bab^2,abab^2\},$$
$$P_{nt}(v) = \{a,a^2,a^2b,a^2b^2,a^2b^2a\},\ S_{nt}(v) = \{b,ab,bab,b^2ab,ab^2ab\}.$$

Clearly $\{u_1,u_2,v\}$ is a solid code. By Corollary 3.2, we obtain that $\{u_i,v\}$ is also a solid code, for $i = 1,2$. Thus, the $\{u_i,v\}$-representation of every $w \in X^+$ is unique, for $i = 1,2$.

We define the following sets, where $s \in \mathbb{N}$:

$A_{u_1,v}(s) = \{w \in X^+ \mid l_{\{u_1,v\}}(w) = s\}$.

$\overline{A}_{u_1,v}(s) = \{ab^6wa^{6+q_1}ba^{6+q_2}\cdots ba^{6+q_{s+1}} \mid w \in A_{u_1,v}(s)$ and $\sharp x_i = q_i, i = 1,2,\ldots,s+1$, when the $\{u_1,v\}-$representation of w is $w = x_1y_1x_2y_2\cdots x_sy_sx_{s+1}\}$.

$A_{u_1,v} = \bigcup_{s\geq 1} \overline{A}_{u_1,v}(2^s)$.

The following lemma is a crucial lemma. The proof can be found in [13]. However, we notice that the proof given in [13] has a possible gap (that is, let $\{u\}$ be a solid code with $u \in \{aX^+b\}$ and $u \notin \{a^+b,ab^+\}$, $lg(u) = n, n \geq 3$. If $u \in I(a^kb^nw)$ for any $k \in \mathbb{N}^0$, and any $w \in X^*$, then $u \notin I(a^kb^n)$. But we notice that there exist such solid codes $\{u\}$ such that $u \in I(a^kb^n)$, for example, let $n = 6, k \geq 3$. Then by the above discussion, $\{u = a^3b^3\}$ is a solid code, and clearly $u \in I(a^kb^6)$). Therefore, in the following revised proof of this lemma, it contains the consideration for $u \in I(a^kb^n)$.

Lemma 4.3. [13, Lemma 3.11] *Let $\{u,v\}$ be a solid code, $\{u,v\} \subseteq (aX^+b \cap X^n)$ for some $n \geq 3$, and $\{u,v\} \cap \{a^+b,ab^+\} = \emptyset$, $w_1,w_2 \in X^*$. If $(w_1,w_2) \notin \sigma_{\{u,v\}}$, then for any $i,j,k \in \mathbb{N}^0$,*

$$(u^iv^ja^kb^nw_1, u^iv^ja^kb^nw_2) \notin \sigma_{\{u,v\}}.$$

Proof. (revised) Let $z = u^i v^j a^k b^n w_1$ and $z' = u^i v^j a^k b^n w_2$. We first suppose that $(z, z') \in \sigma_{\{u,v\}}$. Then, the $\{u,v\}$-representations of z and z' are

$$z = x_1 y_1 x_2 y_2 \cdots x_m y_m x_{m+1},$$

and

$$z' = x_1 y_1' x_2 y_2' \cdots x_m y_m' x_{m+1}.$$

Thus, $x_1 = x_2 = \cdots = x_{i+j} = 1$, and the $\{u,v\}$-representations of $a^k b^n w_1$ and $a^k b^n w_2$ are

$$a^k b^n w_1 = x_{i+j+1} y_{i+j+1} \cdots x_m y_m x_{m+1},$$

and

$$a^k b^n w_2 = x_{i+j+1} y_{i+j+1}' \cdots x_m y_m' x_{m+1}.$$

On the one hand, if $I(a^k b^n) \cap \{u,v\} \neq \emptyset$, then by our hypothesis and the $\{u,v\}$-representations of $a^k b^n w_1$ and $a^k b^n w_2$, we have $a^k b^n = x_{i+j+1} y_{i+j+1} b^p$ with $1 < p < n-1$, and b^p is a prefix of x_{i+j+2} (that is, $x_{i+j+2} = b^p x, x \in X^*$). Hence, the $\{u,v\}$-representations of w_1 and w_2 have the forms

$$w_1 = x_{i+j+2}' y_{i+j+2} \cdots x_m y_m x_{m+1}$$

and

$$w_2 = x_{i+j+2}' y_{i+j+2}' \cdots x_m y_m' x_{m+1},$$

with $b^p x_{i+j+2}' = x_{i+j+2}, x_{i+j+2}' \in X^*$. On the other hand, if $I(a^k b^n) \cap \{u,v\} = \emptyset$, then from the $\{u,v\}$-representations of $a^k b^n w_1$ and $a^k b^n w_2$ again, $a^k b^n$ is a prefix of x_{i+j+1} or x_{i+j+1} is a proper prefix of $a^k b^n$ (that is, $x_{i+j+1} x = a^k b^n, x \in X^+$). We now claim that x_{i+j+1} is not the proper prefix of $a^k b^n$, for otherwise, if $x_{i+j+1} a^{k'} b^n = a^k b^n$, $k' \geq 1$, then by the hypothesis $lg(u) = lg(v) = n$, we have $y_{i+j+1} x = a^{k'} b^n, x \in X^+$. This result contradicts to $I(a^k b^n) \cap \{u,v\} = \emptyset$; if $x_{i+j+1} b^{n'} = a^k b^n$, $1 \leq n' \leq n$, then y_{i+j+1} begins with letter b, this contradicts to $\{u,v\} \subseteq aX^+b$. This result hence shows that $a^k b^n$ is the prefix of x_{i+j+1}. Hence, the $\{u,v\}$-representations of w_1 and w_2 are

$$w_1 = x_{i+j+1}' y_{i+j+1} \cdots x_m y_m x_{m+1} \textit{ and } w_2 = x_{i+j+1}' y_{i+j+1}' \cdots x_m y_m' x_{m+1},$$

where $a^k b^n x_{i+j+1}' = x_{i+j+1}, x_{i+j+1}' \in X^*$.

Both of the above two cases imply that $(w_1, w_2) \in \sigma_{\{u,v\}}$. Thus, we arrive at a contradiction and our proof is completed. ∎

We state the following proposition.

Proposition 4.1. $P_{A_{u_1,v}} = \sigma_{\{u_1,v\}}$.

Proof. Let w_1 and w_2 be two words over X, $(w_1, w_2) \notin \sigma_{\{u_1,v\}}$. Then, we will show that $w_1 \not\equiv w_2 (P_{A_{u_1,v}})$. Consider $z_1 = u_1^i v^j ab^6 w_1$ and $z_2 = u_1^i v^j ab^6 w_2$. By Lemma 4.3, we have $(z_1, z_2) \notin \sigma_{\{u_1,v\}}$. Now, we choose i and j such that $l_{\{u_1,v\}}(z_1) = 2^t$ for some $t \in \mathbb{N}$. Let the $\{u_1,v\}$-representation of z_1 be

$$z_1 = x_1 y_1 x_2 y_2 \cdots x_n y_n x_{n+1},$$

$n = 2^t$ and let the $\{u_1,v\}$-representation of z_2 be

$$z_2 = x_1' y_1' x_2' y_2' \cdots x_m' y_m' x_{m+1}',$$

for some $m \in \mathbb{N}^0$. Since $(z_1, z_2) \notin \sigma_{\{u_1,v\}}$, we have $n \neq m$, or $n = m$ but $x_h \neq x_h'$ for some h, $1 \leq h \leq n+1$. Let $\sharp x_p = q_p$, $p = 1, 2, \cdots, n+1$. Then by the construction of $A_{u_1,v}$, we have

$$\overline{z_1} = ab^6 z_1 a^{6+q_1} b a^{6+q_2} b \cdots a^{6+q_{n+1}} \in \overline{A}_{u_1,v}(2^t),$$

and

$$\overline{z_2} = ab^6 z_2 a^{6+q_1} b a^{6+q_2} b \cdots a^{6+q_{n+1}} \notin \overline{A}_{u_1,v}(2^s), \text{for each } s \in \mathbb{N}.$$

Hence $\overline{z_1} \not\equiv \overline{z_2}(P_{A_{u_1,v}})$. It follows that $z_1 \not\equiv z_2(P_{A_{u_1,v}})$, and hence $w_1 \not\equiv w_2(P_{A_{u_1,v}})$. So

$$P_{A_{u_1,v}} \subseteq \sigma_{\{u_1,v\}}.$$

We now proceed to prove the converse statement of the above proposition. We first suppose that w_1 and w_2 are two different words over X, $(w_1, w_2) \in \sigma_{\{u_1,v\}}$. Since $\sigma_{\{u_1,v\}}$ is a congruence, $(xw_1y, xw_2y) \in \sigma_{\{u_1,v\}}$ for any $x, y \in X^*$. From the definition of $\sigma_{\{u_1,v\}}$, we have

$$xw_1y = x_1y_1x_2y_2 \cdots x_ny_nx_{n+1},$$
$$xw_2y = x_1y_1'x_2y_2' \cdots x_ny_n'x_{n+1}.$$

If xw_1y is in some $\overline{A}_{u_1,v}(2^t)$, $t \in \mathbb{N}$, then $xw_1y = x_1y_1x_2y_2 \cdots x_ny_nx_{n+1}$ has the form $ab^6w'a^{6+q_1} ba^{6+q_2} b \cdots a^{6+q_{2^t+1}}$, where $w' \in X^*$, $q_i \in \mathbb{N}$, $i = 1, 2, \ldots, 2^t+1$. Since $I(ab^6) \cap \{u_1, v\} = \emptyset$ by the definitions of u_1 and v, similar to the proof of Lemma 4.3, we have $ab^6x_1' = x_1$, $x_1' \in X^*$. Similarly, we have $x_{n+1}'a^{6+q_1} ba^{6+q_2} b \cdots a^{6+q_{2^t+1}} = x_{n+1}$, $x_{n+1}' \in X^*$. Thus, xw_1y and xw_2y can be written as

$$ab^6x_1'y_1x_2y_2 \cdots x_ny_nx_{n+1}'a^{6+q_1} ba^{6+q_2} b \cdots a^{6+q_{2^t+1}}, \tag{4.1}$$

$$ab^6x_1'y_1'x_2y_2' \cdots x_ny_n'x_{n+1}'a^{6+q_1} ba^{6+q_2} b \cdots a^{6+q_{2^t+1}}, \tag{4.2}$$

respectively. By our assumption that $xw_1y \in \overline{A}_{u_1,v}(2^t)$, we have $n = 2^t$ in (4.1), and $\sharp x_1' = q_1, \sharp x_i = q_i, i = 2, \ldots, 2^t, \sharp x_{2^t+1}' = q_{2^t+1}$. This means that $xw_2y \in \overline{A}_{u_1,v}(2^t)$. Dually, it can be proved that $xw_1y \in \overline{A}_{u_1,v}(2^t)$ from the fact that $xw_2y \in \overline{A}_{u_1,v}(2^t)$, $t \in \mathbb{N}$. This shows that $xw_1y \in \overline{A}_{u_1,v}(2^t)$ if and only if $xw_2y \in \overline{A}_{u_1,v}(2^t)$. Hence, $w_1 \equiv w_2(P_{A_{u_1,v}})$ and so

$$\sigma_{\{u_1,v\}} \subseteq P_{A_{u_1,v}}.$$ ∎

Proposition 4.2. *The following statements always hold.*

(1) $u_1^i v^j \equiv v^j u_1^i (P_{A_{u_1,v}})$, *for all* $i, j \in \mathbb{N}^0$.

(2) $A_{u_1,v}$ *is an f-disjunctive language.*

Proof. (1) Clearly, $(u_1^i v^j, v^j u_1^i) \in \sigma_{\{u_1,v\}}$, for all $i, j \in \mathbb{N}^0$, by Proposition 4.1, the result holds.

(2) By Proposition 4.1, $P_{A_{u_1,v}} = \sigma_{\{u_1,v\}}$ and by Lemma 4.2, every $\sigma_{\{u_1,v\}}$-class contains only finite elements. This shows that $A_{u_1,v}$ is f-disjunctive. ∎

By Proposition 4.2, we have $A_{u_1,v} \in \mathscr{D}_2$, it is clear that $A_{u_1,v}$ is a midst-language. Similarly, by replacing u_1 with u_2, the languages $A_{u_2,v}(s), \overline{A}_{u_2,v}(s), A_{u_2,v}$ are defined respectively. Clearly, Proposition 4.2 is also valid for $A_{u_2,v}$. Thus, $A_{u_2,v} \in \mathscr{D}_2$.

Let $B_1 = A_{u_1,v}\{b\}, B_2 = \{b\}A_{u_2,v}$. Then, by Lemma 3.1 and Lemma 4.3 of [12], B_1 and B_2 are both in $\mathscr{D}_2$. Clearly, B_1 and B_2 are disjoint. We now show that $B_1 \cup B_2$ is a disjunctive language.

Proposition 4.3. *The language* $B_1 \cup B_2$ *is disjunctive.*

Proof. Let w_1, w_2 be two words over X such that $(w_1, w_2) \notin \sigma_{\{u_1,v\}}$ and $(w_1, w_2) \notin \sigma_{\{u_2,v\}}$. Then by Proposition 4.1, we have $w_1 \not\equiv w_2(P_{A_{u_1,v}})$, $w_1 \not\equiv w_2(P_{A_{u_2,v}})$. This means that there exist $x, y \in X^*$ such that $xw_1y \in A_{u_1,v}$, $xw_2y \notin A_{u_1,v}$ or vice versa. Hence, by the construction of B_1, $xw_1yb \in B_1$, $xw_2yb \notin B_1$ or vice versa. Moreover, by the construction of B_2, $xw_1yb \notin B_2$ and $xw_2yb \notin B_2$. We have $xw_1yb \in B_1 \cup B_2$, $xw_2yb \notin B_1 \cup B_2$ or vice versa. This implies that $w_1 \not\equiv w_2(P_{B_1 \cup B_2})$. Similarly, when $w_1 \not\equiv w_2(P_{A_{u_2,v}})$, we also have $w_1 \not\equiv w_2(P_{B_1 \cup B_2})$.

Let w_1, w_2 be two words over X such that $(w_1, w_2) \in \sigma_{\{u_1,v\}}$ or $(w_1, w_2) \in \sigma_{\{u_2,v\}}$. Then by the definition of u_1, u_2, v, there exist no $w_1, w_2 \in X^*$, $w_1 \neq w_2$ such that $(w_1, w_2) \in \sigma_{\{u_1,v\}}$ and $(w_1, w_2) \in \sigma_{\{u_2,v\}}$. Assume that $(w_1, w_2) \in \sigma_{\{u_1,v\}}$ but $(w_1, w_2) \notin \sigma_{\{u_2,v\}}$. Now, by applying Proposition 4.1, we see that $w_1 \not\equiv w_2(P_{A_{u_2,v}})$. Hence, there exist $x, y \in X^*$ such that $xw_1y \in A_{u_2,v}$, $xw_2y \notin A_{u_2,v}$ or vice versa. Thus, we have $bxw_1y \in B_2$, $bxw_2y \notin B_2$ or vice versa. By the construction of B_1 again, we also have $bxw_1y \notin B_1$ and $bxw_2y \notin B_1$. Hence, we deduce that $bxw_1y \in B_1 \cup B_2$, $bxw_2y \notin B_1 \cup B_2$ or vice versa. This shows that $w_1 \not\equiv w_2(P_{B_1 \cup B_2})$. Similarly, if $(w_1, w_2) \in \sigma_{\{u_2,v\}}$ and $(w_1, w_2) \notin \sigma_{\{u_1,v\}}$, one also has $w_1 \not\equiv w_2(P_{B_1 \cup B_2})$.

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

5. Some questions related to the decompositions of r-disjunctive languages

We first observe that Proposition 1.4′ can be modified into the following form.

Proposition 1.4″. *Let L be a disjunctive language over X, $L = L_1 \dot{\cup} L_2$. Then the following statements hold.*

(1) $|\{L_1, L_2\} \cap \mathscr{D}| = 1$, *or*
(2) $|\{L_1, L_2\} \cap \mathscr{D}| = 2$, *or*
(3) $\{L_1, L_2\} \subseteq \mathscr{D}_r \setminus \mathscr{D}$.

By the main result in [6], we see immediately that any disjunctive language L has a decomposition of $L = L_1 \dot{\cup} L_2$ such that $L_1 \in \mathscr{R}_r, L_2 \in \mathscr{D}$. At this time, we have $|\{L_1, L_2\} \cap \mathscr{D}| = 1$; and any disjunctive language has the decomposition of case (2), in fact, the disjunctive language L is dense and so by [3] (a dense language can be divided into two disjoint disjunctive languages), L has the decomposition $L = L_1 \dot{\cup} L_2$ such that $|\{L_1, L_2\} \cap \mathscr{D}| = 2$; for the case (3), not every disjunctive language has this decomposition, but it has been shown that there exist such languages, see Section 4. Meanwhile, for those r-disjunctive languages, we see that any f-disjunctive (t-disjunctive, r-disjunctive) language L has the decomposition of $L = L_1 \dot{\cup} L_2$ and $|\{L_1, L_2\} \cap \mathscr{D}_f| = 1$ ($|\{L_1, L_2\} \cap \mathscr{D}_t| = 1$, $|\{L_1, L_2\} \cap \mathscr{D}_r| = 1$) by Proposition 2.1. Now, we can also see that there exist f-disjunctive(t-disjunctive) languages L such that $L = L_1 \dot{\cup} L_2$ and $\{L_1, L_2\} \subseteq \mathscr{D}_r \setminus \mathscr{D}_f (\mathscr{D}_r \setminus \mathscr{D}_t)$, see Section 3. But we still do not know whether every f-disjunctive(t-disjunctive, r-disjunctive) language L has the decomposition $L = L_1 \dot{\cup} L_2$ with $|\{L_1, L_2\} \cap \mathscr{D}_f| = 2$ ($|\{L_1, L_2\} \cap \mathscr{D}_t| = 2$, $|\{L_1, L_2\} \cap \mathscr{D}_r| = 2$)?

In closing this paper, we point out that the above question also leads to a more special question.

Does every f-disjunctive (t-disjunctive, r-disjunctive) language L can be decomposed into $L = L_1 \dot{\cup} L_2$ such that $P_L = P_{L_1} = P_{L_2}$?

We remark that for disjunctive languages, the above two questions are actually the same question, but for f-disjunctive (t-disjunctive, r-disjunctive) languages, they are different questions.

Acknowledgement. Yuqi Guo is the corresponding author of the article. The research was supported by the National Natural Science Foundation of China (No. 11371177). The authors would like to thank the referees for their revision suggestions.

References

[1] Y. Q. Guo, C. M. Reis and G. Thierrin, Relatively f-disjunctive languages, *Semigroup Forum* **37** (1988), no. 3, 289–299.
[2] Y. Q. Guo, H. J. Shyr and G. Thierrin, F-disjunctive language, *Internat. J. Comput. Math.* **18** (1986), 219–237.
[3] Y. Q. Guo, G. W. Xu and G. Thierrin, Disjunctive decomposition of languages, *Theoret. Comput. Sci.* **46** (1986), no. 1, 47–51.
[4] J. M. Howie, *Automata and Languages*, Oxford Science Publications, Oxford Univ. Press, New York, 1991.
[5] G. Lallement, *Semigroups and Combinatorial Applications*, John Wiley & Sons, New York, 1979.
[6] Y. Liu, K. P. Shum and Y. Q. Guo, Relatively regular languages and thin codes, *European J. Combin.* **29** (2008), no. 1, 261–267.
[7] Y. Liu, Studies on some languages and their syntactic semigroups, Doctoral Dissertation of Zhongshan University, 2006.
[8] L. W. Mu, Some kinds of generalized disjunctive languages, Dissertation of Harbin Institute of Technology, 1988.
[9] C. M. Reis and H. J. Shyr, Some properties of disjunctive languages on a free monoid, *Inform. and Control* **37** (1978), no. 3, 334–344.
[10] H. J. Shyr, *Free Monoids and Languages*, third edition, Hon Min Book Co., Taichung, 2001.
[11] H. J. Shyr, Disjunctive languages on a free monoid, *Information and Control* **34** (1977), no. 2, 123–129.
[12] H. J. Shyr and S. S. Yu, Midst-languages, *Soochow J. Math.* **24** (1998), no. 2, 113–130.
[13] H. J. Shyr and S. S. Yu, Solid codes and disjunctive domains, *Semigroup Forum* **41** (1990), no. 1, 23–37.
[14] H. J. Shyr, A characterization of dense languages, *Semigroup Forum* **30** (1984), no. 2, 237–240.