Acta Informatica (2010) 47:347–357
DOI 10.1007/s00236-010-0122-7

ORIGINAL ARTICLE

A note on pure codes

C. C. Huang

Received: 27 July 2009 / Accepted: 27 July 2010 / Published online: 14 August 2010

Abstract This study extends the understanding of two-element pure codes. Some characteristics of different length two-element pure codes are studied. It is shown that a language is a pure code which contains two distinct primitive words u and v with different lengths if and only if the regular expression u^+v^+ of the two distinct words u and v is primitive.

1 Introduction

Property-preserving iterated homomorphisms can be applied to generate words or languages. One can refer to [6] for definitions and notions of property-preserving iterated homomorphisms which are also related to OL schemes. Shyr and Thierrin have proposed some properties of homomorphisms which preserve primitive words in [6]. They argue that if an injective homomorphism $h : X^* \to X^*$ is such that $h(X)$ is a pure code, then h preserves primitive words. Using the definition of pure codes, to check whether a given language is a pure code is not easy. This motivates the investigation to discover a simple method for checking whether a given language is a pure code or not.

The notion of pure languages is introduced in [5]. In [3], Fan and Huang investigate some characteristics of pure codes. A pure code consisting of two distinct primitive words u and v with the same length can imply that uv is a primitive word. If there are two primitive words u and v with the same length such that uv is a primitive word, then the language $\{u, v\}$ is a pure code. This gives rise to a simple procedure to check whether a same length two-element language $\{u, v\}$ is a pure code or not. For example, since $ab, ba \in Q$, we have $abba \in Q$, which implies the language $\{ab, ba\}$ is a pure code. In this paper, we extend the research to find a general method for checking whether a given two-element language $\{u, v\}$ is a

C. C. Huang (✉)
School of Applied Information Sciences, Chung Shan Medical University, 402 Taichung, Taiwan
e-mail: cchuang@csmu.edu.tw

C. C. Huang
Information Technology Office, Chung Shan Medical University Hospital, 402 Taichung, Taiwan

pure code. The focus is on studying the characteristics between regular expression and pure codes with two different length elements.

This paper is organized into four sections. The first section is an overview of the paper. In the second section, some well-known definitions and properties applied in this paper are examined. In the third section, a two-element pure code $\{u, v\}$ consisting two distinct primitive words u and v with different lengths derives that regular expression u^+v^+ is primitive. That is, all elements of u^+v^+ are primitive words. Furthermore, for primitive words u and v with different lengths, the primitivity of a regular expression u^+v^+, that is, $u^+v^+ \subseteq Q$, derives that the language $\{u, v\}$ is a pure code. This provides a brief procedure to check whether a different length two-element language $\{u, v\}$ is a pure code. Because the process of derivative proof is fairly involved, the detailed proof is postponed until the final section.

2 Definitions and preliminaries

Let X be a finite alphabet and X^* be the free monoid generated by X. Any element of X^* is called a *word*. The length of a word w is the number of letters occurring in w and denoted by $\lg(w)$. Any subset of X^* is called a *language*. Let $X^+ = X^* \setminus \{\lambda\}$ where λ is the empty word. If u is a word such that $u = xwy$ where $w \in X^+, x, y \in X^*$, then the word w will be called a *subword* of u. A subword w of u is a *proper subword* of u if $u = xwy$ such that $xy \in X^+$. For $u \in X^+$, $E(u)$ and $\bar{E}(u)$ are denoted as the set of all subwords of u and the set of all proper subwords of u respectively. A word $w \in X^+$ is said to be *primitive* if $w = f^n$ with $f \in X^+$ always implies $n = 1$. Let Q denote the set of all primitive words and $Q^{(i)} = \{f^i | f \in Q\}$ for every $i \geq 2$. For a word $w \in X^+$, there exists a unique primitive word f and a unique integer $i \geq 1$ such that $w = f^i$. Let $f = \sqrt{w}$ and call f the *root* of w. For two words $u, v \in X^+$, it is denoted by $v \leq_{\mathrm{p}} u$ ($v <_{\mathrm{p}} u$) if v is the prefix (proper prefix) of u and denoted by $v \leq_{\mathrm{s}} u$ ($v <_{\mathrm{s}} u$) if v is the suffix (proper suffix) of u. A language $L \subseteq X^+$ is a *code* if $x_1x_2\cdots x_n = y_1y_2\cdots y_m$, $x_i, y_j \in L$ implies that $m = n$ and $x_i = y_i, i = 1, 2, \ldots, n$. A language $L \subseteq X^+$ is called *pure* if for any $x \in L^*$, $\sqrt{x} \in L^*$. The property of codes and the characteristic of pure languages are combined to derive the following definition.

Definition 2.1 *A language L is a pure code if it is a code such that for any $x \in L^*$, $\sqrt{x} \in L^*$.*

Next we list some results used in the paper.

Lemma 2.1 ([4]) *Let $u, v \in Q$ with $u \neq v$. Then $u^m v^n \in Q$ for all $m \geq 2, n \geq 2$.*

Lemma 2.2 ([4]) *If $uv = vu$, $u, v \in X^+$, then u, v are powers of a common word.*

Lemma 2.3 ([1]) *If $uv = vz$ where $u, v, z \in X^*$ and $u \neq \lambda$, then $u = (pq)^i$, $v = (pq)^j p$, and $z = (qp)^i$ for some $p, q \in X^*$, $i \geq 1$, $j \geq 0$, and $pq, qp \in Q$.*

Lemma 2.4 ([8]) *If $uq^m = g^k$ for some $m, k \geq 1$, $u \in X^+$, and $g \in Q$ with $u \notin q^+$, then $q \neq g$ and $\lg(g) > \lg(q^{m-1})$.*

Lemma 2.5 ([8]) *Let $p \neq q \in Q$. Then $|p^+q^+ \cap \cup_{i\geq 2} Q^{(i)}| \leq 1$.*

Lemma 2.6 ([8]) *Let $u \in X^+$ with $u \notin q^+$ and $\lg(u) \leq \lg(q)$. If $uq^m = g^k$ for some $m, k \geq 2$ and $g \in Q$, then $k = 2$, $m = 2$, $u \in Q$ and $u = yxxy$, $q = x(yx)^{j+1}$ for some $x \neq y \in X^+$, $j \geq 1$. Moreover, x and y are not powers of a common word.* □

Lemma 2.7 ([8]) *If $pq^m = g^k$ for some $m, k \geq 2$ and $g \in Q$, then one of the following two statements holds:*

(A) $p = (xq^m)^{k-1}x$ *for some* $x \in X^+$*;*
(B) $p = (yx(x(yx)^{j+1})^{m-1})^{k-2}yx(x(yx)^{j+1})^{m-2}xy$ *and* $q = x(yx)^{j+1}$ *for some* $x \neq y \in X^+$, $j \geq 0$.

Lemma 2.8 ([2]) *Let $x, y \in X^+$. Then $xy \neq yx$ if and only if $x(yx)^m \in Q$ for all $m \geq 2$.*

Lemma 2.9 *If $x_1x_2 = x_2x_3 = x_3x_4$, where $x_1, x_2, x_3, x_4 \in X^+$, then x_1, x_2, x_3 and x_4 are powers of a common word.*

Proof Since $x_1x_2 = x_2x_3 = x_3x_4$, this yields that $\lg(x_3) = \lg(x_1)$ and $\lg(x_4) = \lg(x_2)$; hence $x_3 = x_1, x_4 = x_2$. This implies that $x_1x_2 = x_2x_1$ and $x_2x_3 = x_3x_2$. By Lemma 2.2, x_1, x_2, x_3, x_4 are powers of a common word. □

Lemma 2.10 ([6]) *Let uv be a primitive word over X, where $u \neq \lambda$, and $v \neq \lambda$. Then $\{u, v\}$ is a code.*

Lemma 2.11 ([10]) *Let $x, y \in X^+$. If $xy <_p y^i$ for some $i \geq 2$, then $\sqrt{x} = \sqrt{y}$.*

Lemma 2.12 *Let $u \in Q$. If $u \leq_p (pq)^i$ and $u \leq_s (qp)^i$ with $pq \in Q$ for some $p, q \in X^+, i \geq 1$, then there exist $x, y \in X^+$ with $i_1, i_2 \geq 1,\ j_1 \geq 0$ and $xy,\ yx \in Q$ such that one of the following statements holds:*

(I) $u = (xy)^{j_1}x$ *and* $p = (xy)^{i_1+j_1}x$*;*
(II) $u = (xy)^{i_1+j_1}x$ *and* $pq = (xy)^{i_1+i_2+j_1}x$*;*
(III) $p = u^k$ *for some* $k \geq 2$.
(IV) $u = (pq)^k p$ *for some* $k \geq 0$.

Proof Let $u \in Q$. Let $u \leq_p (pq)^i$ and $u \leq_s (qp)^i$ for some $p, q \in X^+, i \geq 1$. Consider the following cases:

(1) $\lg(u) = \lg((pq)^k)$ for some $1 \leq k \leq i$. Then $u = (pq)^k = (qp)^k$. Since $u \in Q$, this implies that $k = 1$, that is, $pq = qp$. By Lemma 2.2, p, q are powers of a common word. Thus $pq \notin Q$, a contradiction.

(2) $\lg((pq)^k) < \lg(u) < \lg((pq)^k p)$ for some $0 \leq k < i$. Then there exist $p_1, p_2, p_3, p_4 \in X^+$ such that $u = (pq)^k p_1 = p_4(qp)^k$ and $p = p_1p_2 = p_3p_4$. Let $k \geq 1$. We have $p_3p_4q = p_4qp_3$. By Lemma 2.2, p_3, p_4q are powers of a common word. Thus $p_3p_4q = pq \notin Q$, a contradiction. Hence $k = 0$, that is, $u = p_1 = p_4$. This implies that $p = up_2 = p_3u$. By Lemma 2.3, we have $p_3 = (xy)^{i_1}, u = (xy)^{j_1}x$, and $p_2 = (yx)^{i_1}$ for some $x, y \in X^*, i_1 \geq 1, j_1 \geq 0$ and $xy, yx \in Q$. Let $j_1 = 0$. We consider the following subcases: $x = \lambda$, $y = \lambda$ or $x, y \in X^+$. If $x = \lambda$, then $u = \lambda$, a contradiction. If $y = \lambda$, then $u = x$ and $p = x^{i_1+1} = u^{i_1+1}$. Statement *(III)* holds. If $x, y \in X^+$, then $u = x$ and $p = (xy)^{i_1}x$. Statement *(I)* holds. Let $j_1 \geq 1$. Consider the following subcases: $x = \lambda$, $y = \lambda$, or $x, y \in X^+$. If $x = \lambda$, then $u = y^{j_1}$ and $p = y^{i_1+j_1}$. Hence Statement *(III)* holds. If $y = \lambda$, then $u = x^{j_1+1} \notin Q$, a contradiction. If $x, y \in X^+$, then $u = (xy)^{j_1}x$ and $p = (xy)^{i_1+j_1}x$. Statement *(I)* holds.

(3) $\lg(u) = \lg((pq)^k p)$ for some $0 \leq k < i$. Then $u = (pq)^k p$ and hence Statement *(IV)* holds.

(4) $\lg(u) < \lg((pq)^k pq)$ for some $0 \le k < i$. Then there exist $q_1, q_2, q_3, q_4 \in X^+$ such that $u = (pq)^k pq_1 = q_4 p(qp)^k$ and $q = q_1 q_2 = q_3 q_4$. Let $k \ge 1$. Then $pq_3 q_4 = q_4 pq_3$. By Lemma 2.2, pq_3, q_4 are powers of a common word. Thus $pq_3 q_4 = pq \notin Q$, a contradiction. Hence $k = 0$, that is, $u = pq_1 = q_4 p$. By Lemma 2.3, $q_4 = (xy)^{i_1}$, $p = (xy)^{j_1} x$, and $q_1 = (yx)^{i_1}$ for some $x, y \in X^*$, $i_1 \ge 1$, $j_1 \ge 0$ and $xy, yx \in Q$. Thus $u = (xy)^{i_1 + j_1} x$. This in conjunction with $pq = uq_2 = q_3 u$ and Lemma 2.3 yields that $q_2 = (yx)^{i_2}$ and $q_3 = (xy)^{i_2}$ for some $i_2 \ge 1$. Thus $pq = (xy)^{i_1 + i_2 + j_1} x$, and hence $u = (xy)^{i_1 + j_1} x$ for some $x, y \in X^*$. Next, we show that $x, y \in X^+$. If $x = \lambda$ or $y = \lambda$, then it is clear that $pq = (xy)^{i_1 + i_2 + j_1} \notin Q$. Both cases contradict $pq \in Q$, and hence Statement *(II)* holds. □

3 The properties of two-element pure codes

Recall that a language L is a pure code if it is a code with the property of pure languages. In the following, we give the characterization for $\{u, v\} \subseteq X^n$, which are pure codes. Some known results are needed.

Lemma 3.1 ([3]) *Let $L = \{u, v\} \subset Q$. If L is a pure code, then $uv \in Q$.*

Lemma 3.2 ([3]) *Let $u \ne v \in Q$ with $\lg(u) = \lg(v)$. Then $\{u, v\}$ is a pure code if and only if $uv \in Q$.*

In the following proposition, we examine the characteristics of the two-element pure code $L = \{u, v\}$. The elements of the regular expression $u^+ v^+$, formed by the two primitive words u and v, are not always primitive words. For example, given $u = (xy)^2 x$ and $v = y$ for some $x, y \in X^+$ with $\sqrt{x} \ne \sqrt{y}$, the word $uv = (xy)^3 \notin Q$. As a language containing two primitive words is a pure code, one can obtain that the regular expression $u^+ v^+$ formed by the different length primitive words u and v is primitive, as is done in the following proposition.

Proposition 3.1 *Let $u, v \in Q$ with $\lg(u) \ne \lg(v)$. If $L = \{u, v\}$ is a pure code, then $u^+ v^+ \subset Q$.*

Proof Let $L = \{u, v\}$ be a pure code for some $u, v \in Q$ with $\lg(u) > \lg(v)$. Without loss of generality, we assume that $\lg(u) > \lg(v)$. We claim that $u^+ v^+ \subset Q$. By Lemma 3.1, $uv \in Q$, and by Lemma 2.1, $u^m v^n \in Q, m, n \ge 2$. Furthermore, by Lemma 2.5, $|u^+ v^+ \setminus Q| \le 1$. Thus there are only two cases that can be considered either $uv^m = f^n$ or $u^m v = f^n$ for some $f \in Q$ where $m, n \ge 2$. Note that L is a pure code, which implies that $f \in L^* = \{u, v\}^*$. For $uv^m = f^n$, by Lemma 2.7, one of the following conditions holds:

(1) $u = (xv^m)^{n-1} x$ for some $x \in X^+$. Then $uv^m = (xv^m)^n = f^n$, and hence that $f = xv^m$, and $u = f^{n-1} x$. It follows that $\lg(v) < \lg(f) < \lg(u)$. This implies $f \notin L^*$, a contradiction.

(2) $u = (yx(x(yx)^{j+1})^{m-1})^{n-2} yx(x(yx)^{j+1})^{m-2} xy$ and $v = x(yx)^{j+1}$ for some $x \ne y \in X^+, j \ge 0$. Since $u \ne v \in Q$, it implies that $\sqrt{x} \ne \sqrt{y}$. From $uv^m = (yx(x(yx)^{j+1})^{m-1})^n = f^n$, it follows that $u = f^{n-2} yxv^{m-2} xy$ and $f = yxv^{m-1}$. If $n \ge 3$, then $\lg(v) < \lg(f) < \lg(u)$. This implies that $f \notin \{u, v\}^*$, a contradiction. If $n = 2$, then $u = yxv^{m-2} xy$, $v = x(yx)^{j+1}$, and $f = ux(yx)^j = yxv^{m-1}$. This yields that $\lg(u) < \lg(f) < \lg(uv)$, which implies that $f \notin u^* \cup v^*$ and $uv, vu \notin E(f)$. Hence $f \notin \{u, v\}^*$, a contradiction.

Therefore $u^m v = f^n$. Since $\lg(u) > \lg(v)$ and $u \ne v \in Q$, by Lemma 2.6, it follows that $m = n = 2, u = x(yx)^{j+1}$, and $v = yxxy$, where $j \ge 1$ and $x, y \in X^+$

with $\sqrt{x} \neq \sqrt{y}$; hence $u^2v = (x(yx)^{j+1}xy)^2$ and $f = uxy = x(yx)^j v$. This yields that $\lg(u) < \lg(f) < \lg(uv)$, which implies that $f \notin u^* \cup v^*$ and $uv, vu \notin E(f)$. Hence $f \notin \{u, v\}^*$, a contradiction. This completes the proof. □

Next, we find a procedure to check whether a different length two-element language is a pure code. The primitivity of the regular expression u^+v^+ determines that the two-element language $L = \{u, v\}$ is a pure code. For instance, since $u = aba$ and $v = ba \in Q$, it follows that $u^+v^+ \subset Q$ and hence $\{aba, ba\}$ is a pure code. Furthermore, $u = (ab)^2a$, $v = baab$. It follows that $u^2v = \big((ab)^2aab\big)^2 \notin Q$. By the following proposition, one can obtain that $\{(ab)^2a, baab\}$ is not a pure code.

Proposition 3.2 *Let $u, v \in Q$ with $\lg(u) \neq \lg(v)$. Then $u^+v^+ \subset Q$ implies that $\{u, v\}$ is a pure code.*

Proof The complete proof is found in the Section 4. □

From Lemma 3.1 and Propositions 3.1 and 3.2, the final result is as follows.

Theorem 3.1 *Let $u \neq v \in Q$. Then $L = \{u, v\}$ is a pure code if and only if $u^+v^+ \subset Q$.*

4 Proof of the main result

The process of detailed proof concerning Proposition 3.2 is studied in this section. Since the proof is involved, some lemmata are considered first. The conclusion is presented in Propositions 4.2 and 4.3.

Lemma 4.1 ([7]) *Let $uv = f^i$, $u, v \in X^+$, $f \in Q$, $i \geq 1$. Then $vu = g^i$ for some $g \in Q$.*

Lemma 4.2 *Let $u, v \in Q$ and $u^+v^+ \subset Q$. Then $u^+v^+u^+ \subset Q$.*

Proof Let $u, v \in Q$ and $u^+v^+ \subset Q$. Since $u^+u^+v^+ \subset u^+v^+$, we have $u^+u^+v^+ \subset Q$. By Lemma 4.1, it implies that $u^+v^+u^+ \subset Q$. □

Lemma 4.3 *Let $u, v \in Q$ and $u^+v^+ \subset Q$. Let $u^{i_1}v^{j_1} \cdots u^{i_r}v^{j_r} = f^n$, where $f \in Q$, $r, n \geq 2$ and $i_l, j_l \geq 1$ for all $l = 1, 2, \ldots, r$ and $u^{i_1}v^{j_1} \cdots u^{i_k}v^{j_k} \notin f^+$ for all $k < r$. If $\lg(f^m) > \lg(u^{i_1}v^{j_1} \cdots u^{i_{k-1}}v^{j_{k-1}})$ for some $1 \leq m < n$, $1 \leq k < r$, then one of the following statements is true*:

(1) $\lg(f^m) < \lg(u^{i_1}v^{j_1} \cdots v^{j_{k-1}}u^{i_k})$ *imply that* $f^m = u^{i_1}v^{j_1} \cdots u^{i_k - 1}u_1$, *where* $u_1 \in X^+$ *with* $u_1 <_p u$;

(2) $\lg(f^m) < \lg(u^{i_1}v^{j_1} \cdots v^{j_{k-1}}u^{i_k}v^{j_k})$ *imply that* $f^m = u^{i_1}v^{j_1} \cdots u^{i_k}v_1$, *where* $v_1 \in X^+$ *with* $v_1 <_p v$.

Proof (1) If $i_k = 1$, then the result is clear. Let $i_k \geq 2$. There exist $u_1, u_2 \in X^+$ with $u = u_1u_2$ and $i \geq 0$ such that $f^m = u^{i_1}v^{j_1} \cdots u^{i_{k-1}}v^{j_{k-1}}u^i u_1$ and $u_2u^{i_k - i - 1}v^{j_k} \cdots u^{i_r}v^{j_r} = f^{n-m}$. If $i_k - i - 1 \geq 1$, then it follows that $u_1u_2 <_p f^m$ and $u_2u_1 <_p f^{n-m}$. For $u^{i_1}v^{j_1} \cdots u^{i_r}v^{j_r} = f^n$, by Lemma 4.1, there exists $g \in Q$ with $\lg(f) = \lg(g)$ such that $v^{j_k}u^{i_k+1}v^{j_k+1} \cdots u^{i_r}v^{j_r}u^{i_1}v^{j_1} \cdots v^{j_{k-1}}u^{i_k} = g^n$. Since $i_k \geq 2$, by Lemma 2.4, $\lg(g) > \lg(u)$; hence $\lg(f) > \lg(u)$. This implies that u_1u_2, u_2u_1 are prefixes of f. Thus $u_1u_2 = u_2u_1$. By Lemma 2.2, u_1, u_2 are powers of a common word; hence $u = u_1u_2 \notin Q$, a contradiction. Therefore $i_k - i - 1 = 0$. This in conjunction with $f^m = u^{i_1}v^{j_1} \cdots u^{i_{k-1}}v^{j_{k-1}}u^i u_1$ and $u_2u^{i_k - i - 1}v^{j_k} \cdots u^{i_r}v^{j_r} = f^{n-m}$ yields that $f^m = u^{i_1}v^{j_1} \cdots u^{i_k - 1}u_1$.

(2) The proof is similar to (1). □

Proposition 4.1 *Let $u, v \in Q$ and $u^+v^+ \subset Q$. Let $u^{i_1}v^{j_1}\cdots u^{i_r}v^{j_r} = f^n$, where $f \in Q$, $r, n \geq 2$ and $i_k, j_k \geq 1$ for all $k = 1, 2, \ldots, r$. Then $\lg(f) > \lg(u^{i_{max}-1})$ and $\lg(f) > \lg(v^{j_{max}-1})$, where $i_{max} = \max\{i_1, i_2, \ldots, i_r\}$ and $j_{max} = \max\{j_1, j_2, \ldots, j_r\}$.*

Proof By Lemma 2.1 and Lemma 2.4, the result is clear. □

Lemma 4.4 *Let $u, v \in Q$ with $\lg(u) > \lg(v)$ and $u^+v^+ \subset Q$. Assume that $u^{i_1}v^{j_1}\cdots u^{i_r}v^{j_r} = f^n$, where $f \in Q$, $r, n \geq 2$ and $i_k, j_k \geq 1$ for all $k = 1, 2, \ldots, r$ and $u^{i_1}v^{j_1}\cdots u^{i_k}v^{j_k} \notin f^+$ for all $k < r$. If $\lg(f) \leq \lg(u^{i_{min}}v)$, where $i_{min} = \min\{i_1, i_2, \ldots, i_r\}$, then it implies that $i_k \leq i_{min} + 1$ and the following two statements are true*:

(1) *there exists $t \in X^+$ with $t <_p v$ such that $f = u^{i_{min}}t$;*

(2) *for some $k = 2, 3, \ldots, r$, $u^{i_k}v^{j_k}\cdots u^{i_r}v^{j_r}u^{i_1}v^{j_1}\cdots u^{i_{k-1}}v^{j_{k-1}} = g_k^n$, $g_k \in Q$, there exists $z \in X^+$ with $z <_p u$ such that $g_k = u^{i_{min}}z$.*

Proof Let $u, v \in Q$ with $\lg(u) > \lg(v)$ and $u^+v^+ \subset Q$. Let

$$u^{i_1}v^{j_1}\cdots u^{i_r}v^{j_r} = f^n, \tag{4–1}$$

where $f \in Q$, $r, n \geq 2$ and $i_k, j_k \geq 1$ for all $k = 1, 2, \ldots, r$. Without loss of generality, by Lemma 4.1, let $i_1 = i_{min}$, that is, $i_1 \leq i_k$ for all $k = 1, 2, \ldots, r$. Furthermore, for some $k = 2, 3, \ldots, r$, by Lemma 4.1 again, there exists $g_k \in Q$ such that

$$u^{i_k}v^{j_k}\cdots u^{i_r}v^{j_r}u^{i_1}v^{j_1}\cdots u^{i_{k-1}}v^{j_{k-1}} = g_k^n. \tag{4–2}$$

Then there exist $f_1, f_2 \in X^+$ with $f = f_1f_2$ such that $g_k = f_2f_1$. It is clear that $\lg(g_k) = \lg(f_2f_1) = \lg(f_1f_2) = \lg(f)$. Since $\lg(f) \leq \lg(u^{i_1}v)$, we have $\lg(g_k) \leq \lg(u^{i_1}v)$. This in conjunction with $i_1 \leq i_k$ yields that $\lg(g_k) \leq \lg(u^{i_k}v)$. From Eq. (4–1), by Lemma 4.3, either $f = u^{i_1-1}u_1$ for some $u_1 <_p u$ or $f = u^{i_1}v_1$ for some $v_1 <_p v$. In the meanwhile, from Eq. (4–2), by Lemma 4.3, either $g_k = u^{i_k-1}u_3$ for some $u_3 <_p u$ or $g_k = u^{i_k}v_3$ for some $v_3 <_p v$. Then there are the following four cases:

(1) $f = u^{i_1-1}u_1$ and $g_k = u^{i_k-1}u_3$. Since $\lg(f) = \lg(g_k)$ and u_1, u_3 are prefixes of u, we have $i_k = i_1$ and $u_1 = u_3$. This in conjunction with $f = f_1f_2$ and $g_k = f_2f_1$ yields that $f_1f_2 = f_2f_1$. By Lemma 2.2, f_1, f_2 are powers of a common word. This implies that $f = f_1f_2 \notin Q$, a contradiction.

(2) $f = u^{i_1-1}u_1$ and $g_k = u^{i_k}v_3$. Since $i_k \geq i_1$, we have $(i_1 - 1)\lg(u) + \lg(u_1) < i_1 \lg(u) < i_k \lg(u) + \lg(v_3)$; hence $\lg(f) < \lg(g_k)$. This contradicts that $\lg(f) = \lg(g_k)$.

(3) $f = u^{i_1}v_1$ and $g_k = u^{i_k-1}u_3$. If $i_k = i_1$, then $\lg(f) = i_1 \lg(u) + \lg(v_1) = i_k \lg(u) + \lg(v_1) > (i_k - 1)\lg(u) + \lg(u_3) = \lg(g_k)$. This contradicts that $\lg(f) = \lg(g_k)$. Moreover, if $i_k \geq i_1 + 2$, then $\lg(f) = i_1 \lg(u) + \lg(v_1) \leq (i_k - 2)\lg(u) + \lg(v_1) < (i_k - 1)\lg(u) < (i_k - 1)\lg(u) + \lg(u_3) = \lg(g_k)$. This also contradicts that $\lg(f) = \lg(g_k)$. Therefore $i_k = i_1 + 1$, that is, $i_k = i_{min} + 1$. Let $t = v_1$, $z = u_3$. Then we have $f = u^{i_{min}}t$, $g_k = u^{i_{min}}z$.

(4) $f = u^{i_1}v_1$ and $g_k = u^{i_k}v_3$. This proof is similar to case (1). □

To examine Proposition 3.2, we will prove that the language $\{u, v\}$ is a pure code by the mathematical induction on r in the word $u^{i_1}v^{j_1}\cdots u^{i_r}v^{j_r}$ for all $i_k, j_k \geq 1$ where $k = 1, 2, \ldots, r$. The following lemma is the case when $r = 2$. The complete proof of Proposition 3.2 is illustrated in Propositions 4.2 and 4.3.

Lemma 4.5 *Let* $u, v \in Q$ *with* $\lg(u) \neq \lg(v)$ *and* $u^+v^+ \subset Q$. *Then* $(u^+v^+)^2 \backslash (u^+v^+)^{(2)} \subset Q$.

Proof Let $u, v \in Q$ with $\lg(u) \neq \lg(v)$ and $u^+v^+ \subset Q$. Without loss of generality, we let $\lg(u) > \lg(v)$. To show $(u^+v^+)^2 \setminus (u^+v^+)^{(2)} \subset Q$. The case $u^{i_1}v^{j_1}u^{i_1}v^{j_2} \in Q$ for all $i_1, j_1, j_2 \geq 1$ and $j_1 \neq j_2$, is considered firstly. This implies that $u^{i_1}v^+u^{i_1}v^+ \subset Q \cup (u^+v^+)^{(2)}$. Next, the case $u^{i_1}v^{j_1}u^{i_2}v^{j_2} \in Q$ for all $i_1, i_2, j_1, j_2 \geq 1$ and $i_1 \neq i_2$, is considered as well. The detail proof is studied as follow.
(1) $u^{i_1}v^{j_1}u^{i_1}v^{j_2} \in Q$ for all $i_1, j_1, j_2 \geq 1$ and $j_1 \neq j_2$. Suppose that

$$u^{i_1}v^{j_1}u^{i_1}v^{j_2} = f^m \tag{4–3}$$

where $f \in Q$ and $m \geq 2$. Without loss of generality, by Lemma 4.1, let $j_1 > j_2$. Since $uv \in Q$, by Lemma 2.10, $\{u, v\}$ is a code. It implies that $f \notin \{u, v\}^+$. This in conjunction with $u^+v^+ \subset Q$ yields that

$$u^{i_1}v^{j_1} = f^i f_1 \quad \text{and} \quad u^{i_1}v^{j_2} = f_2 f^{m-i-1} \tag{4–4}$$

where $f_1, f_2 \in X^+$ with $f = f_1 f_2$ and $i \geq 0$. If $i = 0$, then $u^{i_1}v^{j_1} = f_1$, $u^{i_1}v^{j_2} = f_2 f^{m-1}$. This implies that $\lg(f_1) = \lg(u^{i_1}v^{j_1}) > \lg(u^{i_1}v^{j_2}) = \lg(f_2 f^{m-1})$, a contradiction. Hence $i \geq 1$. Consider Eq. (4–3), by Proposition 4.1, $\lg(f) > \lg(u^{i_1-1})$. If $\lg(u^{i_1-1}) < \lg(f) < \lg(u^{i_1})$, then there exists $u_1 \in X^+$ such that $f = u^{i_1-1}u_1$. Since $u^{i_1}v^{j_2} = f_2 f^{m-i-1}$, we have $f_2 f_1 <_p u^{i_1-1}u_1$. Thus $f_1 f_2 = f_2 f_1$. By Lemma 2.2, f_1, f_2 are powers of a common word. This implies that $f \notin Q$, a contradiction. If $\lg(u^{i_1} < \lg(f) \leq \lg(u^{i_1}v)$, then, from Eq. (4–4), $f_1 f_2 <_p u^{i_1}v$ and $f_2 f_1 <_p u^{i_1}v$; hence $f_1 f_2 = f_2 f_1$, a contradiction. Hence $\lg(f) > \lg(u^{i_1}v)$. Moreover, by Lemma 4.3 and Eq. (4–3), we have the following two subcases:

(1-1) $\lg(f) < \lg(u^{i_1}v^{j_1}u^{i_1})$. Then $f = u^{i_1}v^{j_1}u^{i_1-1}u_1$ for some $u_1 \in X^+$ with $u_1 <_p u$ and $u_2 v^{j_2} = f^{m-1}$, where $u_2 \in X^+$ with $u = u_1 u_2$. By Proposition 4.1, we have $\lg(f) > \lg(v^{j_1-1}) \geq \lg(v^{j_2})$. This in conjunction with $u_2 v^{j_2} = f^{m-1}$ and $\lg(f) > \lg(u^{i_1-1})$ yields that $m = 2$, i.e., $f = u_2 v^{j_2}$; hence $u_2 v^{j_2} = u^{i_1}v^{j_1}u^{i_1-1}u_1$. If $u^{i_1} = u_2 v^k$ for some $1 \leq k < j_2$, then $v^{j_2-k} = v^{j_1}u^{i_2-1}u_1$. Since $j_1 > j_2$, we have $\lg(v^{j_2-k}) < \lg(v^{j_1}u^{i_2-1}u_1)$, a contradiction. If $u^{i_1} = u_2 v^k v_3$ for some $0 \leq k < j_2$ and $v_3 <_p v$, then $v^{j_1}u^{i_1-1}u_1 = v_4 v^{j_2-k-1}$, where $v_4 \in X^+$ with $v = v_3 v_4$. Thus $v_3 v_4 = v_4 v_3$. By Lemma 2.2, v_3, v_4 are powers of a common word. That is, $v \notin Q$, a contradiction.

(1-2) $\lg(f) < \lg(u^{i_1}v^{j_1}u^{i_1}v^{j_2})$. Then $f = u^{i_1}v^{j_1}u^{i_1}v_1$ for some $v_1 \in X^+$ with $v_1 <_p v$ and $v_2 v^{j_2-1} = f^{m-1}$, where $v_2 \in X^+$ with $v = v_1 v_2$. Recall that $j_1 > j_2$. $\lg(f) = \lg(u^{i_1}v^{j_1}u^{i_1}v_1) > \lg(v^{j_2}) > \lg(v_2 v^{j_2-1}) = \lg(f^{m-1})$. This contradicts to $m \geq 2$.
(2) $u^{i_1}v^{j_1}u^{i_2}v^{j_2} \in Q$ for all $i_1, i_2, j_1, j_2 \geq 1$ and $i_1 \neq i_2$. The proof of this case is similar to case (1).

By above discussion, we have $(u^+v^+)^2 \setminus (u^+v^+)^{(2)} \subset Q$. □

Proposition 4.2 *Let* $u, v \in Q$ *with* $\lg(u) > \lg(v)$. *Then* $u^+v^+ \subset Q$ *implies that* $\{u, v\}$ *is a pure code.*

Proof Let $u, v \in Q$ with $\lg(u) > \lg(v)$. Since $uv \in Q$, by Lemma 2.10, $\{u, v\}$ is a code. For convenience, let $L = \{u, v\}$. Now, we prove that L is a pure language. That is, $\sqrt{w} \in L^*$ for every word $w \in L^*$. If $w = \lambda \in L^*$, then $\sqrt{w} = \lambda \in L^*$. Hence we consider $w \in L^+$. Let $w = u^k \in L^+$ for some $k \geq 1$. If $w \in Q$, then it implies that $w = u$ because $u \in Q$; hence $\sqrt{w} = w = u \in L^+$. If $w \notin Q$, then it is clear that $u = \sqrt{w} \in L^+$. Note that for $w = v^k \in L^+$ for some $k \geq 1$, it is similar to $w = u^k$. Furthermore, let

$w = u^{i_0}v^{j_0}u^{i_1}v^{j_1}\cdots u^{i_r}v^{j_r}$ for $i_0, j_0 \geq 0$ and $i_k, j_k \geq 1$ where $k = 1, 2, \ldots, r$. We consider the following subcases: Case(I) $i_0 = j_0 = 0$. It follows that $w = u^{i_1}v^{j_1}\cdots u^{i_r}v^{j_r}$. Case($II$)$i_0 = 0$ and $j_0 \neq 0$. Then $w = v^{j_0}u^{i_1}v^{j_1}\cdots u^{i_r}v^{j_r}$. By Lemma 4.1, we have $w_1 = u^{i_1}v^{j_1}\cdots u^{i_r}v^{j_r+j_0}$. Case($III$)$i_0 \neq 0$ and $j_0 = 0$. Then $w = u^{i_0+i_1}v^{j_1}\cdots u^{i_r}v^{j_r}$. From the above three subcases, we prove that L is a pure language by the mathematical induction on r in the word $w = u^{i_1}v^{j_1}\cdots u^{i_r}v^{j_r}$ for all $i_k, j_k \geq 1$, where $k = 1, 2, \ldots, r$. If $w = u^{i_1}v^{j_1} \in Q$, then $\sqrt{w} \in L^+$. By Lemma 4.5, the result is true when $r = 2$. Suppose that the result is true for $2 \leq r \leq n$. We want to show that if $w = u^{i_1}v^{j_1}\cdots u^{i_{n+1}}v^{j_{n+1}}$, then $\sqrt{w} \in L^+$. If $w = u^{i_1}v^{j_1}\cdots u^{i_{n+1}}v^{j_{n+1}} \in Q$, then it is clear that $\sqrt{w} \in L^+$. Thus let $w = u^{i_1}v^{j_1}\cdots u^{i_{n+1}}v^{j_{n+1}} \notin Q$. By Lemma 4.1, without loss of generality, assume that $i_{n+1} \leq i_k$ for all $k = 1, 2, \ldots, n$ and

$$u^{i_{n+1}}v^{j_{n+1}}u^{i_1}v^{j_1}\cdots u^{i_n}v^{j_n} = f^m \tag{4–5}$$

for some $f \in Q$ and $m \geq 2$. If $f \in \{u, v\}^+$, then it is clear that $\{u, v\}$ is a pure language. Hence $f \notin \{u, v\}^+$ is considered. Note that f satisfies Lemma 4.3. From Eq. (4–5) and by Lemma 2.1, since $u^{i_{n+1}}v^{j_{n+1}} \in Q$, two cases $u^{i_1}v^{j_1}\cdots u^{i_n}v^{j_n} \in Q$ or $u^{i_1}v^{j_1}\cdots u^{i_n}v^{j_n} \notin Q$ need to be considered.

(1) $u^{i_1}v^{j_1}\cdots u^{i_n}v^{j_n} \in Q$. There exist $f_1, f_2 \in X^+$ with $f = f_1f_2$ such that $u^{i_{n+1}}v^{j_{n+1}} = f^if_1,\ u^{i_1}v^{j_1}\cdots u^{i_n}v^{j_n} = f_2f^{m-i-1}$, where $0 \leq i < m$. If $i \geq 1$, by Lemma 4.3, either $f^i = u^{i_{n+1}-1}u_1$ for some $u_1 <_p u$ or $f^i = u^{i_{n+1}}v_1$ for some $v_1 <_p v$. Both cases imply that $\lg(f^i) < \lg(u^{i_{n+1}}v)$, that is, $i\lg(f) < \lg(u^{i_{n+1}}v)$. By Lemma 4.4 and $i_{n+1} \leq i_k$ for all $k = 1, 2, \ldots n$, it follows that $\lg(f) > \lg(u^{i_{n+1}})$ and $\lg(f) > \lg(v^{j_{n+1}})$; hence $2\lg(f) > \lg(u^{i_{n+1}}) + \lg(v^{j_{n+1}})$. This in conjunction with $i\lg(f) < \lg(u^{i_{n+1}}v)$ yields that $i = 1$; that is, $\lg(f) \leq \lg(u^{i_{n+1}}v)$. From Eq. (4–5), by Lemma 4.1, there exist $g_1, g_2 \in Q$ with $\lg(g_1) = \lg(g_2) = \lg(f)$ such that

$$u^{i_1}v^{j_1}\cdots u^{i_{n+1}}v^{j_{n+1}} = g_1^m$$

and

$$u^{i_2}v^{j_2}\cdots u^{i_{n+1}}v^{j_{n+1}}u^{i_1}v^{j_1} = g_2^m.$$

This in conjunction with $\lg(f) \leq \lg(u^{i_{n+1}}v)$, $i_{n+1} \leq i_k$ for all $k = 1, 2, \ldots n$, and Lemma 4.4 yields that $i_1 = i_2 = i_{n+1} + 1$ and $g_1 = g_2 = u^{i_{n+1}}z$ for some $z <_p u$. Then $u^{i_1}v^{j_1}\cdots u^{i_{n+1}}v^{j_{n+1}} = g_1tv^{j_1}g_1tv^{j_2}u^{i_3}\cdots u^{i_{n+1}}v^{j_{n+1}} = g_1^m$, where $t \in X^+$ with $zt = u$. Since $g_1 \in Q$, we have $tv^{j_1} = g_1$. Thus $u^{i_{n+1}}ztv^{j_1} = u^{i_{n+1}+1}v^{j_1} = g_1^2$. This contradicts $u^+v^+ \subset Q$. Therefore, $i = 0$. Then $u^{i_{n+1}}v^{j_{n+1}} = f_1,\ u^{i_1}v^{j_1}\cdots u^{i_n}v^{j_n} = f_2f^{m-1}$. By Lemma 4.3, there are the following subcases:

(1-1) $f = u^{i_{n+1}}v^{j_{n+1}}u^{i_1}v^{j_1}\cdots u^{i_k-1}u_1$ with $u_1 <_p u$ for some $u_1 \in X^+$ and $k \geq 1$. Then $f_2 = u^{i_1}v^{j_1}\cdots u^{i_k-1}u_1,\ f^{m-1} = u_2v^{j_k}\cdots u^{i_n}v^{j_n}$ and $u = u_1u_2$ for some $u_2 \in X^+$. It follows that $u_2v^{j_k}\cdots u^{i_n}v^{j_n} = (u^{i_{n+1}}v^{j_{n+1}}u^{i_1}v^{j_1}\cdots u^{i_1-1}u_1)^{m-1}$. If $\lg(u_2) = \lg(u_1)$, then $u_2 = u_1$; hence $u = u_1u_2 \notin Q$, a contradiction. Hence $\lg(u_1) \neq \lg(u_2)$. We only consider the subcase $\lg(u_1) < \lg(u_2)$. The proof of the subcase $\lg(u_1) > \lg(u_2)$ is similarly. As $\lg(u_1) < \lg(u_2)$, there exist $u_{21}, u_{22} \in X^+$ such that $u_2 = u_{21}u_{22},\ u_{21} = u_1$ and

$$u_{22}v^{j_1}\cdots u^{i_n}v^{j_n} = u_2u^{i_{n+1}-1}v^{j_{n+1}}u^{i_1-1}u_1f^{m-2}. \tag{4–6}$$

This yields that $u_{22} <_p u_2$; hence $u_2 = u_{22}u_{23} = u_{21}u_{22}$ for some $u_{23} \in X^+$. By Lemma 2.3, we have $u_{21} = (pq)^i, u_{22} = (pq)^j p, u_{23} = (qp)^i$ for some $p, q \in X^*, i \geq 1, j \geq 0$ and $pq, qp \in Q$. Then $u = u_{21}u_{22}u_{23} = (pq)^{2i+j}p$. Note that $p, q \in X^+$ because $p = \lambda$ or $q = \lambda$ imply that $u \notin Q$, a contradiction. Furthermore, we consider Eq. (4–6) again. Since $m \geq 2$ and $f = u^{i_{n+1}}v^{j_{n+1}}u^{i_1}v^{j_1}\cdots u^{i_k-1}u_1$, there are the following two subcases:

(1-1-1) $v \leq_s u_1$. There exists $u_{11} \in X^*$ such that $u_1 = u_{11}v$. Since $u_1 = u_{21} = (pq)^i$, this implies that $v \leq_s (pq)^i$. In the Meanwhile, from Eq. (4–6), it follows that

$$v^{j_1}\cdots u^{i_n}v^{j_n} = u_{23}u^{i_{n+1}-1}v^{j_{n+1}}u^{i_1-1}u_1f^{m-2}. \tag{4–7}$$

Thus $v \leq_p (qp)^i$. Therefor, we have $v \leq_s (pq)^i$ and $v \leq_p (qp)^i$. By Lemma 2.12, there are the following four cases for some $x, y \in X^+$ with $i'_1, i'_2 \geq 1, j'_1 \geq 0$ and $xy, yx \in Q.(I)v = (xy)^{j'_1}x$ and $q = (xy)^{i'_1+j'_1}x$. This in conjunction with $v^{j_1}\cdots u^{i_n}v^{j_n} = (qp)^i u^{i_{n+1}-1}v^{j_{n+1}}u^{i_1}v^{j_1}\cdots u^{i_1-1}u_1f^{m-2}$ yields that $(xy)^{j'_1}x <_p (yx)^{i'_1+j'_1}xp(qp)^{i-1}$. If $j'_1 \geq 1$, then $xy = yx$. By Lemma 2.2, x, y are powers of a common word. Thus $xy \notin Q$, a contradiction. Hence, $j'_1 = 0$ and $x^{j_1}u^{i_2}\cdots u^{i_n}v^{j_n} = (yx)^{i'_1}x(qp)^{i-1}u^{i_{n+1}-1}v^{j_{n+1}}u^{i_1}v^{j_1}\cdots u^{i_k-1}u_1f^{m-2}$. It is true that the integer j_1 is enough lange such that $yx <_p x^{j_1}$. By Lemma 2.11, we have $\sqrt{x} = \sqrt{y}$. Thus $xy \notin Q$, a contradiction. $(II)v = (xy)^{i'_1+j'_1}x$ and $qp = (xy)^{i'_1+i'_2+j'_1}x$. This subcase is similar to $(I).(III)q = v^{i'_1+j'_1+1}$. This in conjunction with $u = (pq)^{2i+j}p$ yields that $uv^{i'_1+j'_1+1} = (pq)^{2i+j+1} \notin Q$, a contradiction. $(IV)v = (qp)^{i'_1+j'_1+1}q$. This also in conjunction with $u = (pq)^{2i+j}p$ yields that $uv = (pq)^{2i+j+i'_1+j'_1+2} \notin Q$, a contradiction.

(1-1-2) $u_1 <_s v$. For Eq. (4–7), there exist $v_1, v_2, v_3, v_4 \in X^+$ such that $v = v_1v_2 = v_3v_4$, $v_1 = u_{23}$ and $v_4 = u_1$, that is, $v = (qp)^i v_2 = v_3(pq)^i$. We consider the following four subcases: $(I)\ \lg(v_3) < \lg(q)$. There exist $q_1, q_2 \in X^+$ such that $v_3 = q_1$ and $q = q_1q_2$. This implies that $q_2p(qp)^{i-1}v_2 = pq_1q_2(pq)^{i-1}$. If $i > 1$, then we have $q_2pq_1 = pq_1q_2$. By Lemma 2.2, pq_1, q_2 are powers of a common word. Thus $pq \notin Q$, a contradiction. Therefore, $i = 1$; hence $v = qpv_2 = q_1pq$. By Lemma 2.3, we have $q_1p = (xy)^{i'_1}, q = (xy)^{j'_1}x, pv_2 = (yx)^{i'_1}$ for some $x, y \in X^*, i'_1 \geq 1, j'_1 \geq 0$ and $xy, yx \in Q$. For $pv_2 = (yx)^{i'_1}$, this in conjunction with $q_2pv_2 = pq = pq_1q_2$ and Lemma 2.3 yields that $pq_1 = (xy)^{i'_1}, q_2 = (xy)^{j'_2}x$ for $j'_2 \geq 0$. Then $q_1p = (xy)^{i'_1} = pq_1$. By Lemma 2.2, p, q_1 are powers of a common word. Since $xy \in Q$, we have $p = (xy)^{k_1}$ and $q_1 = (xy)^{k_2}$ with $k_1 + k_2 = i'$. From $pv_2 = (yx)^{i'_1}$, we have $p <_p (yx)^{i'_1}$. This implies that $xy = yx$. By Lemma 2.2, x, y are powers of a common word. Thus $xy \notin Q$, a contradiction. $(II)\ \lg(v_3) = \lg(q)$. Then $v = q(pq)^i$. This in conjunction with $u = (pq)^{2i+j}p$ yields that $uv = (pq)^{3i+j+1} \notin Q$, a contradiction. $(III)\ \lg(v_3) < \lg(qp)$. There exist $p_1, p_2 \in X^+$ such that $v_3 = qp_1$ and $p = p_1p_2$. This implies that $p_2(qp_1p_2)^{i-1}v_2 = (p_1p_2q)^i$. If $i > 1$, then we have $p_2qp_1 = p_1p_2q$. By Lemma 2.2, p_1, p_2q are powers of a common word. Thus $pq \notin Q$, a contradiction. Therefore, $i = 1$; hence $v = qpv_2 = qp_1pq$. Note that $\lg(v_2) = \lg(qp_1)$. There exist $p_3, p_4 \in X^+$ such that $v_2 = p_4q$ and $p = p_3p_4$. Then $v = qp_1p_2p_4q = qp_1p_3p_4q$; hence $p_2 = p_3$. Since $p = p_1p_2 = p_3p_4$, by Lemma 2.3, we have $p_1 = (xy)^{i'_1}, p_2 = (xy)^{j'_1}x, p_4 = (yx)^{i'_1}$ for some $x, y \in X^*, i'_1 \geq 1, j'_1 \geq 0$ and $xy, yx \in Q$. Then $p = (xy)^{i'_1+j'_1}x$. Since $v^{j_1}u^{i_2}\cdots u^{i_n}v^{j_n} = pqu^{i_{n+1}-1}v^{j_{n+1}}u^{i_1-1}u_1f^{m-2}$, it implies that $p_4q <_p u = (pq)^{2i+j}p$; hence $yx = xy$. By Lemma 2.2, x, y are powers of a common word. Thus $xy \notin Q$, a contradiction. $(IV)\ \lg(v_3) = \lg(qp)$. Then $v_3 = qp$. Since $v = (qp)^i v_2 = qp(pq)^i$ and $\lg(v_2) = \lg(v_3)$, it implies that $pq = qp$. By Lemma 2.2, p, q are powers of a common word. Thus $pq \notin Q$, a contradiction. (1-2) $f = u^{i_{n+1}}v^{j_{n+1}}u^{i_1}v^{j_1}\cdots u^{i_k}v_1$ for some $v_1 <_p v$ and $k \geq 1$. The proof of this subcase is similar to (1-1).

(2) $u^{i_1}v^{j_1}\cdots u^{i_n}v^{j_n} \notin Q$. There exists a primitive word $q \in (u^+v^+)^+$ and $k \geq 2$ such that $q^k = u^{i_1}v^{j_1}\cdots u^{i_n}v^{j_n}$; hence $w = u^{i_{n+1}}v^{j_{n+1}}q^k = f^m$. By Lemma 2.7, there are the following two cases:

(2-1) $u^{i_{n+1}}v^{j_{n+1}} = (xq^k)^{m-1}x$ for some $x \in X^+$. Since $q \in (u^+v^+)^+$ and $\{u, v\}$ is a code, this implies that $u^{i_1}v^{j_1} \leq_p q$. If $q = u^{i_1}v^{j_1}$, then

$$u^{i_{n+1}}v^{j_{n+1}} = xu^{i_1}v^{j_1}u^{i_1}v^{j_1}(u^{i_1}v^{j_1})^{k-2}(xq^k)^{m-2}x. \tag{4–8}$$

Since $m, k \geq 2$ and $i_{n+1} \leq i_1$, we consider the following two subcases:

(2-1-1)$xu^{i_1} = u^{i_{n+1}}v^{k_0}$ for some $k_0 \geq 0$. Then from Eq. (4–8), we have $v^{j_{n+1}-k_0-j_1} = u^{i_1}v^{j_1}(u^{i_1}v^{j_1})^{k-2}(xq^k)^{m-2}x$. If $u^{i_1} \in v^+$, then since $u, v \in Q$, it implies that $u = v$, a contradiction. Thus $u^{i_1} = v^{k_1}v_1$ and $v_2v^{j_{n+1}-k_0-j_1-k_1-1} = v^{j_1}(u^{i_1}v^{j_1})^{k-2}(xq^k)^{m-2}x$, where $v_1, v_2 \in X^+$ with $v = v_1v_2$ and $k_1 \geq 1$. This implies that $v_1v_2 = v_2v_1$. By Lemma 2.2, v_1, v_2 are powers of a common word. Thus $v = v_1v_2 \notin Q$, a contradiction.

(2-1-2)$xu^{i_1} = u^{i_{n+1}}v^{k_0}v_1$ for some $k_0 \geq 0$ and $v_1 <_p v$. We have $v_2v^{j_{n+1}-k_0-1} = v^{j_1}u^{i_1}v^{j_1}(u^{i_1}v^{j_1})^{k-2}(xq^k)^{m-2}x$, where $v_2 \in X^+$ with $v = v_1v_2$. This implies that $v_1v_2 = v_2v_1$. By Lemma 2.2, v_1, v_2 are powers of a common word. Thus $v = v_1v_2 \notin Q$, a contradiction.
By above discussion, $u^{i_1}v^{j_1} = q$ is impossible. Hence $u^{i_1}v^{j_1} <_p q$. Again, since $q \in (u^+v^+)^+$ and $\{u, v\}$ is a code, there exists $1 < l < n$ such that $q = u^{i_1}v^{j_1}\cdots u^{i_l}v^{j_l}$. Then

$$u^{i_{n+1}}v^{j_{n+1}} = xu^{i_1}v^{j_1}u^{i_2}v^{j_2}\cdots u^{i_l}v^{j_l}q^{k-1}(xq^k)^{m-2}x.$$

Since $m, k \geq 2$ and $i_{n+1} \leq i_1$, there are the following two subcases: $xu^{i_1} = u^{i_{n+1}}v^{k_0}$ for some $k_0 \geq 0$, and $xu^{i_1} = u^{i_{n+1}}v^{k_0}v_1$ for some $k_0 \geq 0$, $v_1 <_p v$. The proofs of these subcases are similar to the situation when $q = u^{i_1}v^{j_1}$.

(2-2) $u^{i_{n+1}}v^{j_{n+1}} = (yx(x(yx)^{j+1})^{k-1})^{m-2}yx(x(yx)^{j+1})^{k-2}xy$ and $q = x(yx)^{j+1}$ for some $x \neq y \in X^+$, $j \geq 0$. Then $w = (yx(x(yx)^{j+1})^{k-1})^m = f^m$; hence $f = yx(x(yx)^{j+1})^{k-1}$. Let $m \geq 3$. We have $u^{i_{n+1}}v^{j_{n+1}} = f^{m-2}yx(x(yx)^{j+1})^{k-2}xy$. This implies that $\lg(f) < \lg(u^{i_{n+1}}v^{j_{n+1}})$. By Lemma 4.3, $f = u^{i_{n+1}-1}u_1$ with $u_1 <_p u$ for some $u_1 \in X^+$ or $f = u^{i_{n+1}}v_1$ with $v_1 <_p v$ for some $v_1 \in X^+$. If $f = u^{i_{n+1}-1}u_1$, then it implies that $\lg(yx) < \lg(u^{i_{n+1}})$. This in conjunction with $i_{n+1} \leq i_1$ and $u^{i_1}v^{j_1}\cdots u^{i_n}v^{j_n} = q^k = (x(yx)^{j+1})^k$ yields that $\lg(xy) < \lg(u^{i_1})$. This implies that xy, yx are prefixes of $u^{i_{n+1}}$; hence that $xy = yx$. By Lemma 2.2, x, y are powers of a common word. Thus $\sqrt{x} = \sqrt{y}$; hence $q \notin Q$, a contradiction. If $f = u^{i_{n+1}}v_1$, then we have $u^{i_{n+1}}v_1 <_p u^{i_{n+1}}v^{j_{n+1}} = (yx(x(yx)^{j+1})^{k-1})^{m-2}yx(x(yx)^{j+1})^{k-2}xy$. Since $k \geq 2$, $m \geq 3$, $\lg(yx) < \lg(u^{i_{n+1}}v_1)$. This in conjunction with $i_{n+1} \leq i_1$ and $u^{i_1}v^{j_1}\cdots u^{i_n}v^{j_n} = q^k = (x(yx)^{j+1})^k$ yields that $\lg(xy) < \lg(u^{i_1}v_1)$. This implies that xy, yx are prefixes of $u^{i_{n+1}}v_1$; hence that $xy = yx$. By Lemma 2.2, x, y are powers of a common word. Thus $\sqrt{x} = \sqrt{y}$; hence $q \notin Q$, a contradiction. Let $m = 2$. We have $u^{i_{n+1}}v^{j_{n+1}} = yx(x(x(yx)^{j+1})^{k-2}xy$. If $\lg(yx) \leq \lg(u^{i_{n+1}})$, then $yx \leq_p u^{i_{n+1}}$. Now consider $u^{i_1}v^{j_1}\cdots u^{i_n}v^{j_n} = (x(yx)^{j+1})^k$. Since $x(yx)^{j+1} = q \in (u^+v^+)^+$, there exists $1 \leq l < n$ such that $x(yx)^{j+1} = u^{i_1}v^{j_1}\cdots u^{i_l}v^{j_l}$. This in conjunction with $\lg(xy) = \lg(yx)$ and $i_{n+1} \leq i_1$, we get $xy \leq u^{i_{n+1}}$. Thus $xy = yx$. By lemma 2.2, x, y are powers of a common word. This implies that $\sqrt{x} = \sqrt{y}$, a contradiction. Hence $\lg(yx) > \lg(u^{i_{n+1}})$. From $u^{i_{n+1}}v^{j_{n+1}} = yx(x(x(yx)^{j+1})^{k-2}xy$, we consider the following two cases:

(2-2-1) $yx = u^{i_{n+1}}v^{k_0}$ for some $k_0 \geq 1$. If $k = 2$, then $u^{i_{n+1}}v^{j_{n+1}} = yxxy$. Since $yx = u^{i_{n+1}}v^{k_0}$, we have $xy = v^{j_{n+1}-k_0}$. This in conjunction with $\lg(xy) = \lg(yx)$

and $\lg(u) > \lg(v)$ yields that $j_{n+1} - k_0 \geq 2$. That is, $xy \notin Q$. By Lemma 4.1, we have $yx \notin Q$. This implies that $u^{i_{n+1}}v^{k_0} \notin Q$. This contradicts to $u^+v^+ \subseteq Q$. If $k \geq 3$, then $(x(yx)^{j+1})^{k-2}xy = v^{j_{n+1}-k_0}$. We consider the following two subcases: $(I)(x(yx)^{j+1})^{k-2} = v^{k_1}$, $xy = v^{j_{n+1}-k_0-k_1}$, where $1 \leq k_1 < j_{n+1} - k_0$. Since $\lg(xy) = \lg(yx)$ and $\lg(u) > \lg(v)$, we get $j_{n+1} - k_0 - k_1 \geq 2$. That is, $xy \notin Q$. By Lemma 4.1, we have $yx \notin Q$. This implies that $u^{i_{n+1}}v^{k_0} \notin Q$. It contradicts to $u^+v^+ \subseteq Q$. $(II)(x(yx)^{j+1})^{k-2} = v^{k_1}v_1$, $xy = v_2v^{j_{n+1}-k_0-k_1-1}$, where $1 \leq k_1 < j_{n+1} - k_0$ and $v_1, v_2 \in X^+$ with $v = v_1v_2$. Since $v_2v_1 <_p xy <_p (x(yx)^{j+1})^{k-2} = v^{k_1}v_1$, we have $v_2v_1 = v_1v_2$. By Lemma 2.2, v_1, v_2 are powers of a common word. Then $v = v_1v_2 \notin Q$, a contradiction.

(2-2-2) $yx = u^{i_{n+1}}v^{k_0}v_1$ for some $k_0 \geq 0$ and $v_1 <_p v$. The proof of this subcase is similar to (2-2-1). □

Proposition 4.3 *Let $u, v \in Q$ with $\lg(u) < \lg(v)$. Then $u^+v^+ \subset Q$ implies that $\{u, v\}$ is a pure code.*

Proof By Lemma 4.1, The proof is similar to Proposition 4.2. □

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

References

1. De Luca, A.: On the combinatorial problems in free monoids. Discrete Math. **38**, 207–225 (1982)
2. Fan, C.-M., Shyr, H.J.: δ-codes and δ-languages. J. Discrete Math. Cryptogr. **8**(3), 381–394 (2005)
3. Fan, C.-M., Huang, C.C.: Properties of pure codes. Int. J. Comput. Math. **85**(1), 9–17 (2008)
4. Lyndon, R.C., Schützenberger, M.P.: On the equation $a^M = b^Nc^P$ in a free group. Michigan Math. J. **9**, 289–298 (1962)
5. Restivo, A.: On a question of McNaughton and Papert. Inform. Control **25**, 93–101 (1974)
6. Shyr, H.J., Thierrin, G.: Codes, languages and MOL schemes. RAIRO Inform. Théor./Theoret. Comput. Sci. **1**(4), 293–301 (1997)
7. Shyr, H.J., Thierrin, G.: Disjunctive languages and codes, fundamentals of computation theory. In: Proceedings of the 1977 International FCT-Conference. Lecture Notes in Computer Science, Springer, Poznan Poland, vol. 5, pp. 171–176 (1997)
8. Shyr, H.J., Yu, S.S.: Non-primitive words in the language p^+q^+. Soochow J. Math. **20**(4), 535–546 (1994)
9. Yu, S.S.: Palindrome words and reverse closed languages. Int. J. Comput. Math. **75**, 389–402 (2000)
10. Yu, S.S.: Factorization in the monoid of languages. Soochow J. Math. **30**(1), 103–125 (2004)