On Exchange-Robust and Subst-Robust Primitive Partial Words

Ananda Chandra Nayak[1], Amit K. Srivastava[2] and Kalpesh Kapoor[1]

[1] Department of Mathematics
Indian Institute of Technology Guwahati, Guwahati, India
[2] Department of Computer Science & Engg.
Indian Institute of Technology Guwahati, Guwahati, India
{n.ananda,amit.srivastava,kalpesh}@iitg.ernet.in

Abstract. A partial word is a word that may have some unknown places known as “holes” and can be replaced by the symbols from the underlying alphabet. A partial word u is said to be primitive if there does not exist a word v such that u is contained in a nontrivial integer power of v. We study the preservation of primitivity in partial words by the effect of some point mutation operations. In this paper, we investigate the effect of exchanging two adjacent symbols and of substituting a symbol by another symbol from the alphabet. We characterize the classes of primitive partial words with one hole which are not exchange robust and not substitute robust. We prove that the language of exchange robust primitive partial words with one hole is not right 1-dense and also prove that the language of primitive partial words with one hole which is substitute robust is closed under conjugacy relation. We show that the language of non-exchange-robust primitive partial words is not context-free over a binary alphabet.

Keywords: Combinatorics on words, primitive word, partial word, exchange-robust partial word, subst-robust partial word

1 Introduction

Let V be a finite alphabet. A word is a sequence of symbols from the alphabet V. A partial word is a word that may have some unknown positions known as “holes” or “do not know” symbols and are represented by $\Diamond$. The holes in the partial words are place holder for the usual symbols in the alphabet. The study of partial words is motivated by its application in molecular biology [1]. For example, the alignment of two DNA sequences to recover as much information as possible can be seen as the construction of two compatible partial words. The combinatorial

V. Biló, A. Caruso (Eds.): ICTCS 2016, Proceedings of the 17th Italian Conference on Theoretical Computer Science, 73100 Lecce, Italy, September 7–9 2016, pp. 190–202 published in CEUR Workshop Proceedins Vol-1720 at http://ceur-ws.org/Vol-1720

properties of words play a vital role in the areas including formal languages [18], coding theory [2], string searching algorithms [6], and computational biology [10].

In study of an algebraic structure, it is interest to investigate the operators that preserves the algebraic structure. For example, homomorphisms are well-known structure-preserving transformations in algebra and word combinatorics. A word is said to be primitive if it cannot be represented as a power of a shorter word. The algorithmic, algebraic and applied combinatorial properties of primitive words have been extensively studied; see for example [11,12]. In [17,13], Păun et al. have studied the conditions to preserve primitivity under morphisms. Dassow et al. [7] investigated some operations where they proved that ww' is a primitive word where w is a given word and w' is a modified mirror image of w. This has been extended by Blanchet-Sadri et al. [5] for partial words. In [16], Păun et al. introduced the notion of robustness of primitive words as point mutation operations such as inserting a symbol, deletion of a symbol, substituting a symbol by another symbol in the primitive word which preserves the primitivity. In this paper, we discuss the robustness of primitive partial words with respect to substitution and exchange operations. These classes of primitive partial words are known as subst-robust and exchange-robust primitive partial words.

The rest of the paper is organized as follows. In Section 2 we discuss the basic concepts and related results that are required in later sections. We characterize the exchange-robust primitive partial words and identify some important properties in Section 3. We prove that the language of non-exchange-robust primitive partial words is not context-free over a given alphabet in Section 4. In Section 5, we characterize the primitive partial words which remain primitive on substituting a symbol by another symbol and identify some important results. Section 6 presents about the conclusion and future work.

2 Prerequisites

Let V be a finite and nontrivial alphabet. A sequence of symbols from V is called a word or string. A total word $w = a_0a_1 \dots a_{n-1}$ of length n is a total function $w : \{0,\ 1, \dots, n-1\} \mapsto V$ where $a_i \in V$ for $i = 0,\ 1, \dots, n-1$. The length of a word w is denoted by $|w|$ and defined as the number of symbols appearing in w. The empty word, λ, does not contain any symbol and $|\lambda| = 0$. The set of all strings over the alphabet V is denoted as V^*. The notation $\alpha(w)$ is used for the set of distinct symbols appearing in w and $|\alpha(w)|$ is the number of distinct symbols in w. A positive integer p is said to be a period of w if $a_i = a_{i+p}$ for $0 \le i \le n-p-1$. Let $w = xyz$ be a word. Then y is called a factor of w and x and z are called prefix and suffix of w, respectively. The reverse of a word w is written as $rev(w)$ and defined as $rev(w) = a_{n-1} \cdots a_1a_0$ when $w = a_0a_1 \cdots a_{n-1}$. A word w is primitive if there does not exist a word u such that $w = u^n$ with $n \ge 2$. The language of primitive and nonprimitive words are denoted as Q and Z, respectively [11]. For a language L, we define $length(L) = \{|w| : w \in L\}$.

A partial word u of length n over an alphabet V can be defined by a partial function $u : \ \{0, \dots, n-1\} \mapsto V$ [4]. A partial word u may contain some '*do not know*' symbols known as holes along with the symbols from the underlying alphabet. A "hole" or "do not know" symbol is represented by $\Diamond$. For $0 \leq i < n$, if $u(i)$ is defined then we say that $i \in D(u)$ (the domain of u), otherwise $i \in H(u)$ (the set of holes) [3]. A total word is a partial word with empty set of holes. For example, $u = \Diamond bac\Diamond a$ is a partial word of length 6 and $D(u) = \{1, 2, 3, 5\}$ and $H(u) = \{0, 4\}$. We use $V_\Diamond$ to denote enlarged alphabet $V \cup \{\Diamond\}$. If u is a partial word of length n over V, then the companion of u is the total function $u_\Diamond : \{0, \ldots, n-1\} \rightarrow V \cup \{\Diamond\}$.

Let u and v be two partial words of equal length. Then u is said to be *contained* in v, if all elements in $D(u)$ are also in the set $D(v)$ and $u(i) = v(i)$ for all $i \in D(u)$ and denoted by $u \sqsubset v$. A partial word u is said to be *compatible* to a partial word v if there exists a partial word w such that $u \sqsubset w$ and $v \sqsubset w$ and is denoted by $u \uparrow v$. A partial word u is said to be primitive if there does not exist a word v such that $u \sqsubset v^n$ with $n \geq 2$ [4]. For example, $w = abb\Diamond$ is a primitive partial word and $u = ab\Diamond b$ is a nonprimitive partial word over $V = \{a,\ b\}$. The languages of primitive partial words and nonprimitive partial words are denoted as Q_p and Z_p, respectively. In particular, Q_p^i denotes the language of primitive partial words with at most i holes. A local period of a partial word u is a positive integer p such that $u(i) = u(i+p)$ whenever $i,\ i+p \in D(u)$ [4].

Let $w = uv$ be a nonempty partial word. Then, the partial words u and v are said to be prefix and suffix of w, respectively. A partial word y is said to be a factor of a word w if w can be written as xyz, where $x,\ z \in V_\Diamond^*$ and $y \in V_\Diamond^+$. The set V_i^* contains all partial words with exactly i-holes. The partial word y is said to be proper factor if $x \neq \lambda$ or $z \neq \lambda$. A prefix (suffix) of length k of a partial word w is denoted as $\text{pref}(w, k)$ ($\text{suff}(w, k)$), respectively, where $k \in \{0, 1, \ldots, |w|\}$ and $\text{pref}(w, 0) = \text{suff}(w, 0) = \lambda$.

We now recall some results from the literature that will be useful later in this paper.

Theorem 1 (Fine and Wilf's Theorem [9]). *Let u and v be nonempty words over V. Suppose u^h and v^k, for some h and k, have a common prefix (or suffix) of length $|u| + |v| - gcd(|u|, |v|)$. Then there exists $z \in V^*$ of length $gcd(|u|, |v|)$ such that $u, v \in z^*$.*

Berstel and Boasson revisited Fine and Wilf's theorem for partial words with one hole.

Theorem 2 (Berstel and Boasson [1]). *Let w be a partial word with one hole which is locally p-periodic and locally q-periodic. If $|w| \geq p + q$ then w is $gcd(p, q)$-periodic.*

Definition 3 (Reflective Language [19]). *A language L is called reflective if $xy \in L$ implies $yx \in L$ for all $x,\ y \in V^*$.*

Lemma 4. *[3] Let u and v be partial words. If there exists a primitive word x such that $uv \sqsubset x^n$ for some positive integer n, then there exists a primitive word y such that $vu \sqsubset y^n$. Moreover, if uv is primitive then vu is primitive.*

Corollary 5. *The languages Q_p and Z_p are reflective.*

Proposition 6. *[3] Let w be a partial word with one hole such that $|\alpha(w)| \geq 2$. If a is any letter, then either w or wa is primitive.*

The following result shows that we can obtain at least one primitive word by substituting a symbol by another symbol.

Proposition 7. *[16] Let $|V| \geq 3$. For any word $x \in V^*$ and for each decomposition $x = x_1 a x_2,\ x_1, x_2 \in V^*,\ a \in V$, there is $b \in V$, $b \neq a$ such that $x_1 b x_2$ is primitive.*

But the result is not true in case of partial words. For example, consider $w = x\Diamond xa$. Substituting a by any letter from V will generate nonprimitive partial words.

Proposition 8. *[3] Let u be a partial word with one hole which is not of the form $x\Diamond x$ where $x \in V^+$. Then for $a \in V$, at most one of the partial words ua is not primitive.*

3 Exchange-Robust Primitive Partial Words with One Hole

In this section, we consider a new formal language known as exchange-robust primitive partial words with one hole which remains primitive when any two consecutive symbols in a partial word are exchanged. In particular, we consider $a \neq \Diamond$ for $a \in V$ because exchanging a and $\Diamond$ will generate different partial words. For example, consider $w = aba\Diamond aa \in Q_p$, but exchanging position 3 and 4 we have $w' = abaa\Diamond a \notin Q_p$.

Definition 9 (Exchange-Robust Partial Words). *A primitive partial word $w = a_0 a_1 \cdots a_{i+1} a_{i+2} \cdots a_{n-1}$ of length n with one hole is said to be exchange-robust if and only if*

$$\text{pref}(w, i)\ .\ a_{i+1} a_i\ .\ \text{suff}(w, n - i - 2)$$

is a primitive partial word for all $i \in \{0,\ 1, \ldots n - 2\}$.

Remark : If a symbol a and a hole $\Diamond$ are adjacent, we exchange a and $\Diamond$.

We denote Q_p^{1X} as the set of all primitive partial words with one hole which are exchange-robust over the alphabet V. Clearly, the set of all exchange-robust primitive partial words with one hole is a subset Q_p. There are infinitely many primitive partial words with one hole which are exchange-robust. For example, $a^n \Diamond b^n a,\ n \geq 2$ is exchange-robust.

Our next result concerns the exchange of two different symbols at consecutive places in a nonprimitive total word. We prove that the new word we obtained by exchanging any two distinct consecutive symbols at any position in a nonprimitive word results in a primitive word.

Lemma 10. *Let w be a total word with $|\alpha(w)| \geq 2$. If $w = x_1abx_2 \in Z$ with $a \neq b$ then $x_1bax_2 \in Q$.*

Proof. We prove it by contradiction. Let w be a nonprimitive word. Then there exists a unique primitive word u such that $w = u^m,\ m \geq 2$. We can express $w = u^{m_1}u_1abu_2u^{m_2}$ where $u_1abu_2 = u$ and $m_1,\ m_2 \geq 0,\ m_1 + m_2 \geq 1$. Assume for contradiction that $w' = u^{m_1}u_1bau_2u^{m_2} \notin Q$. As we know the languages Q and Z are reflective, then it is enough to consider the word $abu_2u^{m_2}u^{m_1}u_1$. Suppose $abu_2u^{m_2}u^{m_1}u_1 = v^m$ and $bau_2u^{m_2}u^{m_1}u_1 = y^n,\ m,\ n \geq 2$ and $y \in Q$. Let p be the common suffix of v^m and y^n. The words v^m and y^n have common suffix of length $m|v| - 2$ and $n|y| - 2$, respectively. We have $|p| = m|v| - 2 = n|y| - 2$. It is not possible to have $m = n = 2$ which is not feasible.

So at least one of m and n is strictly greater than 2. Without loss of generality, let us assume that $m \geq 3$ and $n \geq 2$. Now,

$$\begin{aligned} &2|p| = m|v| + n|y| - 4 \\ &\Rightarrow |p| = \tfrac{m}{2}|v| + \tfrac{n}{2}|y| - 2 \\ &\Rightarrow |p| \geq |y| + |v| + \tfrac{1}{2}|v| - 2 \ (\because m \geq 3 \textit{ and } n \geq 2) \end{aligned}$$

Since $|v| \geq 2$, we obtain that $|p| \geq |y| + |v| - 1$. Hence by Theorem 1, v and y are powers of the same primitive word which is a contradiction. Thus $bau_2u^{m_2}u^{m_1}u_1 \in Q$ which implies that $w' = u^{m_1}u_1bau_2u^{m_2} \in Q$. □

The above result does not hold for partial words. Consider the partial word $w = a\Diamond baab \in Z_p$. if we exchange b and a, we have $w' = a\Diamond abab \notin Q_p$.

Next we study the primitive partial words with one hole in which exchange of two distinct consecutive symbols results in a nonprimitive partial word. We call this set of partial words as non-exchange-robust primitive partial words with one hole. We denote the set of non-exchange-robust primitive partial words with one hole over an alphabet V as $Q_p^{1\overline{X}}$. It is easy to see that $Q_p^{1\overline{X}} \cup Q_p^{1X} = Q_p^1$.

Definition 11 (Non-exchange-robust Primitive Partial Words). *A primitive partial word with one hole is said to be non-exchange-robust if and only if exchange of two distinct consecutive symbols results in a nonprimitive partial word.*

We give the structural characterization of non-exchange-robust primitive partial words with one hole.

Theorem 12. *A primitive partial word w with one hole is non-exchange-robust if and only if w is contained in some word of the form $u^{k_1}u_1abu_2u^{k_2},\ a,\ b \in V,\ a \neq b$ where $u_1xyu_2 \sqsubset u_1abu_2$ for $x,\ y \in V_\Diamond$ such that $u_1yxu_2 \sqsubset u_1bau_2 = u^m$ with $m \geq 2$.*

Proof. We prove the necessary and sufficient conditions as follows:
($\Rightarrow$) Let w be a primitive partial word with one hole. Suppose $w = v_1xyv_2 \sqsubset u^{k_1}u_1abu_2u^{k_2}$ where $a \neq b$ such that $v_1 \sqsubset u^{k_1}u_1,\ v_2 \sqsubset u_2u^{k_2},\ xy \sqsubset ab$. If we exchange x and y, we get $w' = v_1yxv_2 \sqsubset u^{k_1}u_1bau_2$ such that $u_1bau_2 = u^m$ for

$m \geq 2$. Hence $w' \sqsubset u^k,\ k \geq 2$ where $k_1 + m + k_2 = k$ and thus w is not an exchange-robust primitive partial word.

($\Leftarrow$) Let $w \in Q_p^1$ which is not an exchange-robust partial word. Then there exists at least one consecutive positions where exchanging them makes the partial word nonprimitive. The partial word w can be written as either v_1abv_2 where $v_1,\ v_2 \in V_\Diamond^*$ or $v_1a\Diamond v_2$ or $v_1\Diamond av_2$ where $v_1,\ v_2 \in V^*$. In first case, as we have exactly one hole, it is exactly in one among v_1 or v_2. Let $w' = v_1bav_2 \in Z_p$ that is $w' = v_1bav_2 \sqsubset u^m$ for $m \geq 2$. Now $v_1 \sqsubset u^iu_1$ and $v_2 \sqsubset u_2u^j$ for $i,\ j \geq 0$. Combining both we have $v_1bav_2 \sqsubset u^iu_1bau_2u^j$ where $u_1bau_2 = u^k$ for $k \geq 2$.

The other two cases can be handled similarly. $\square$

Let us define $Q_p^{1\overline{X}} = Q_p^1 \setminus Q_p^{1X}$ where '$\setminus$' is the set minus operator. There are primitive partial words of arbitrary length which are non-exchange-robust; for example, $(ab)^n\Diamond a(ab)^n$ for $n \geq 1$. We denote the set of exchange-robust (non-exchange-robust, respectively) primitive partial words with arbitrary number of holes by Q_p^X ($Q_p^{\overline{X}}$, respectively). The set of $Q_p^{1\overline{X}}$ is not closed under the cyclic permutation unlike the language of del-robust primitive partial words with one hole [14]. For example, consider the partial word $abbabb\Diamond ab \in Q_p^{1\overline{X}}$. One of the cyclic permutation of the partial word is $ababbabb\Diamond$, which is exchange-robust.

Definition 13 ([16]). *A language L is said to be dense if for any word y there exist $x,\ z \in V^*$ such that $xyz \in L$. Let k be a positive integer. If for every $u \in V^*$ there exists a word $x \in V^*,\ |x| \leq k$ such that $ux \in L$ then L is said to be right k-dense.*

Next we prove that the language of primitive partial words with at most one hole is dense over the alphabet $V_\Diamond$. We show that the language of primitive partial words with one hole which are exchange-robust is not dense.

Lemma 14. *The language Q_p^1 is dense over alphabet $V_\Diamond$ in V_1^*.*

Proof. Consider a partial word w with at most one hole . Let $|w| = n,\ n \geq 1$. There are two different cases depending upon whether w is a primitive partial word or a nonprimitive partial word.

Case A. Let w is a primitive partial word. By choosing $x = \lambda$ and $y = \lambda$ according to definition we have $xwy \in Q_p^1$.

Case B. Let w be a nonprimitive partial word. Here we consider two subcases depending on whether w is contained in power of a symbol from the alphabet or power of a word having at least two different letters.

Case B.1 Let $w \sqsubset a^n,\ n \geq 2$ for some symbol $a \in V$. It can be easily seen that $wb^n \in Q_p^1$ where a and b are two distinct letters. Here $x = \lambda$ and $y = b^n$ for some $b \neq a$ such that $xwy \in Q_p^1$.

Case B.2 Let $w \sqsubset u^k$, where $|\alpha(u)| \geq 2$ and $k|n$. By choosing $x = \lambda$ and $y = b^n$, we have $xwy \in Q_p^1$.

Hence, for every $w \in V_1^*$, there exist $x,\ y \in V_1^*$ such that $xwy \in Q_p^1$. So Q_p^1 is dense over the alphabet $V_\Diamond$ in V_1^*. □

We prove the following proposition.

Proposition 15. *The language Q_p^{1X} is not right 1-dense.*

Proof. It is sufficient to find one partial word for which we cannot find any word which satisfies the condition. Let $w = x\Diamond x$ be a primitive partial word with one hole where $x \in V^*$. Let us assume that w is not an exchange-robust primitive partial word. Here both wa and wb are not primitive. Hence we cannot find a word z with $|z| \leq 1$ for w such that $wz \in Q_p^{1X}$. Thus Q_p^{1X} is not right 1-dense. □

For the above proposition, such partial word exists. For example, take $w = aaba\Diamond aaba$. We have $w \notin Q_p^{1X}$ and if we concatenate a or b at the right end of w then we obtain a nonprimitive partial word.

4 $Q_p^{\overline{X}}$ is not context-free

In this section we prove that the language of non-exchange-robust primitive partial words is not a context-free language over a given alphabet. In our proof, we use the fact that intersection of a CFL and a regular language is also context-free. We also use the result that the family of context-free languages are closed under generalized sequential machine(gsm) mapping, and for details see [8].

Theorem 16. *The language of non-exchange robust partial words is not context-free over the alphabet $V = \{a,\ b\}$.*

Proof. Consider the regular language $R = ba^+ba^+ba^+ba^+$. Consider the language

$$L = \{ba^{n_1}ba^{n_2}ba^{n_3}ba^{n_4} \mid n_1,\ n_2,\ n_3,\ n_4 \geq 1,\ (|n_1 - n_3| \leq 1,\ |n_2 - n_4| \leq 1,\ |(n_1 + n_2) - (n_3 + n_4)| = 0 \text{ or } 2) \text{ and } (n_1 \neq n_3 \text{ or } n_2 \neq n_4)\} \quad (1)$$

We claim that $Q_p^{\overline{X}} \cap R = L$.
The inclusion $Q_p^{\overline{X}} \cap R \supseteq L$ is easy to observe. For the converse, let us take a word $w = ba^{n_1}ba^{n_2}ba^{n_3}ba^{n_4} \in Q_p^{\overline{X}} \cap R$. As $w \in Q_p^{\overline{X}}$, then w can be represented as $w = u_1abu_2$ such that $u_1bau_2 \in Z$. We have the following possibilities of exchanging.

Case 1. $aba^{n_1-1}ba^{n_2}ba^{n_3}ba^{n_4}$
Case 2. $ba^{n_1-1}ba^{n_2+1}ba^{n_3}ba^{n_4}$
Case 3. $ba^{n_1+1}ba^{n_2-1}ba^{n_3}ba^{n_4}$
Case 4. $ba^{n_1}ba^{n_2-1}ba^{n_3+1}ba^{n_4}$
Case 5. $ba^{n_1}ba^{n_2+1}ba^{n_3-1}ba^{n_4}$
Case 6. $ba^{n_1}ba^{n_2}ba^{n_3-1}ba^{n_4+1}$
Case 7. $ba^{n_1}ba^{n_2}ba^{n_3+1}ba^{n_4-1}$

It is easy to see that all the above cases are in the language $Q_p^{\overline{X}}$ only if we have

(1) $n_1 \neq n_3$ or $n_2 \neq n_4$ (otherwise $ba^{n_1}ba^{n_2}ba^{n_1}ba^{n_2} \notin Q$)
(2) $|n_1 - n_3| \leq 1,\ |n_2 - n_4| \leq 1,\ |(i + j) - (k + l)| = 0$ or 2 (otherwise the word $w' \in Q_p^{\overline{X}}$

Hence the inclusion $Q_p^{\overline{X}} \cap R \subseteq L$.

As we know that a CFL is closed under the gsm mapping then using a sequential transducer (a gsm), the language $Q_p^{\overline{X}} \cap R$ can be translated into a new language

$$L' = \{a^{n_1}b^{n_2}c^{n_3}d^{n_4} \mid n_1,\ n_2,\ n_3,\ n_4 \geq 1,\ |n_1 - n_3| \leq 1,\ |n_2 - n_4| \leq 1,$$
$$|(n_1 + n_2) - (n_3 + n_4)| = 0 \text{ or } 2 \text{ and } (n_1 \neq n_3 \text{ or } n_2 \neq n_4)\} \quad (2)$$

Now we prove that L' is not a context-free language. Assume for contradiction that L' is context-free. Suppose there exist a constant $N > 0$ which must exist by Ogden's lemma. As L' satisfies Ogden's lemma (see Appendix), then every $w \in L',\ |w| \geq N$ can be decomposed into $w = uvxyz$ such that the following conditions hold: (i) vxy contains at most N marked symbols, (ii) v and y have at least one marked symbol, (iii) and $uv^ixy^iz \in L'$ for all $i \geq 0$.

Consider a string $w = a^{n_1}b^{n_2}c^{n_3}d^{n_4}$ such that $n_1 = N,\ n_2 = N,\ n_3 = N + 1$ and $n_4 = N - 1$. As $|n_1 - n_3| \leq 1,\ |n_2 - n_4| \leq 1,\ |(n_1 + n_2) - (n_3 + n_4)| = 0$ and $n_1 \neq n_3,\ n_2 \neq n_4$ then $w \in L'$. Let us mark all the occurrences of b which are at least N of them. Now we can decompose $w = uvxyz$ in such a way that all the conditions of Ogden's lemma are satisfied.

Clearly, neither v nor y contain two different symbols. There are two cases depending on whether vy contains an occurrence of a or not.

(I) Suppose vy does not contain any occurrence of a. In this case, we have $u = a^Nb^{i_1},\ v = b^{m_1},\ x = b^{m_2},\ y = b^{m_3}$ such that $m_1+m_3 \geq 1, k_1 = m_1+m_2+m_3$ and $z = b^{N-(k_1+i_1)}c^{N+1}d^{N-1}$. For $i = 2$, $uvxyz = a^Nb^{N+(m_1+m_3)}c^{N+1}d^{N-1}$ $= a^{p_1}b^{p_2}c^{p_3}d^{p_4}$ which is a contradiction as $|p_2 - p_4| \geq 2$.

(II) Suppose vy contains occurrences of a. Let $v = a^j$ and $y = b^k$ for $j,\ k \geq 1$. If $j < k$, then for a large value of i, we can have $w' = uv^ixy^iz = a^{p_1}b^{p_2}c^{p_3}d^{p_4}$ such that $|p_1 - p_3| > 1$ which is a contradiction. Therefore we must have $j \geq k$. Consider the word uv^ixy^iz which becomes $a^{N-j+ji}b^{N-k+ki}c^{N+1}d^{N-1}$. For $i = 5$, we have $w'' = a^{N+4j}b^{N+4k}c^{N+1}d^{N-1}$ where $|(N+4j)-(N+1)| = 4j - 1 \geq 3,\ |(N + 4k) - (N - 1)| = 4k + 1 \geq 5$ and $|(N + 4j + N + 4k) - (N + 1 + N - 1)| = 4(j + k) \geq 8$ which is a contradiction.

Hence L' is not context-free. As we know that the family of context-free languages is closed under sequential transducers and intersection with regular languages, we conclude that $Q_p^{\overline{X}}$ is also not context-free. □

5 Subst-Robust Primitive Partial Words

In this section we study the set of primitive partial words that remains primitive on substitution of a symbol by another symbol. We refer to the definition of substitute robust total words [16] and define symbol substitution in partial words as follows. Consider a partial word $x \in V_1^+$. We define $one(x) = \{x_1bx_2 \mid x = x_1ax_2,\ x_1,\ x_2 \in V_1^*,\ a, b\ \in V,\ a \neq b\}$. Let $L \subseteq V_\Diamond^*$ and $x \in L$. Then x is called subst-robust (w.r.to L) if $one(x) \subseteq L$.

Definition 17 (Subst-Robust Primitive Partial Words). *A primitive partial word w with one hole is said to be subst-robust if and only if $one(w) \subseteq Q_p^1$.*

Remark: Since $\Diamond$ is considered as a place holder for any of the symbol from the given alphabet, only a symbol $a \in V$ can be substituted by another symbol $b \in V$ such that $a \neq b$.

Proposition 18 ([16]). *If L consists of only subst-robust words, then $L = \{w \in V^* \mid |w| \in length(L)\}$.*

The above proposition is not true in case of partial words with one hole. For example, let $L' = \{a\Diamond b, b\Diamond b, b\Diamond a, a\Diamond a\}$. Though L' is subst-robust, it does not contain all the partial words with one hole of length 3. Next we extend the result of Păun et al. [16] in case of partial words.

Lemma 19. *Let $x = x_1\alpha\beta x_2$ be a partial word with one hole where $\alpha, \beta \in V$ and $|x| \geq 4$. Then at least one of the partial words $x_1\alpha'\beta x_2$, $x_1\alpha\beta' x_2$ is primitive where $\alpha \neq \alpha'$ and $\beta \neq \beta'$.*

Proof. We prove it by contradiction. Consider a partial word x with one hole and $|x| \geq 4$. As $|x| \geq 4$, then x can be written as $x = x_1\alpha\beta x_2$ where $\alpha,\ \beta \in V$, $|x_1x_2| \geq 2$ and either x_1 or x_2 contains a hole. As Q_p^1 is reflective, then to prove the lemma it is sufficient to prove that at least one of the partial word $x_2x_1\alpha'\beta$ or $x_2x_1\alpha\beta'$ is primitive.

Assume the contrary. Let $x_2x_1\alpha'\beta \sqsubset u^m$ and $x_2x_1\alpha\beta' \sqsubset v^n$ for $m,\ n \geq 2$ and $u,\ v \in Q$. It is not possible to have $m = n = 2$ otherwise $u = v$ which is a contradiction. So at least one of $m,\ n$ is greater than 2. without loss of generality, let us assume that $m \geq 3,\ n \geq 2$. Similarly, we cannot have $|u| = 1$; otherwise $x_2x_1\alpha'\beta \sqsubset u^m$ implies that $u \in \{a,\ b\}$. Since $\alpha \neq \alpha',\ \beta \neq \beta'$ then $x_2x_1\alpha\beta'$ is primitive which is a contradiction to the assumption. Hence we have $|u| \geq 2$.

Now, we have $|x_2x_1| = m|u| - 2$ and $|x_2x_1| = n|v| - 2$ which implies that

$$\begin{aligned} 2|x_2x_1| &= m|u| + n|v| - 4 \\ \Rightarrow |x_2x_1| &= \tfrac{m}{2}|u| + \tfrac{n}{2}|v| - 2 \end{aligned}$$

As $m \geq 3$ and $n \geq 2$, we can write that $|x_2x_1| \geq |u| + |v| + \frac{1}{2}|u| - 2$. Also $|u| \geq 2$ implies that $|x_2x_1| \geq |u| + |v| - 1$. We consider the following cases.
(a) If $|x_2x_1| = |u| + |v| - 1$ then $m = n = 2$ which leads to a contradiction.
(b) If $|x_2x_1| > |u|+|v|-1$ then by Theorem 2, there exist a word y such that $u = y^k$ and $v = y^l$ for some integers k and l. Hence $x_2x_1\alpha'\beta \sqsubset y^{km}$ and $x_2x_1\alpha\beta' \sqsubset y^{ln}$ which is a contradiction. Thus at least one of the $x_2x_1\alpha'\beta,\ x_2x_1\alpha\beta'$ is a primitive partial word. $\square$

In the above lemma, $|x| \geq 4$ is necessary. For example, let $x = \Diamond ab$ and substituting a by b or b by a will generate nonprimitive partial words. Lemma 19 does not hold for partial words with at least two holes. For example, $w = \Diamond aa\Diamond$ over $V = \{a,\ b\}$. Substituting first occurrence of a by b or last occurrence of a by b will generate nonprimitive partial words.

We denote Q_p^{1S} as the set of primitive partial words with one hole which remains primitive on substitution of a symbol by another symbol from the given alphabet. There are infinitely many primitive partial words with one hole which are subst-robust. For example, $w = (ab)^n\Diamond$ for $n \geq 2$ is subst-robust. It is worth mentioning here that there are primitive partial words with one hole which are at the same time exchange-robust and subst-robust. An example of such partial word is

$$w_m = \Diamond aba^2b^2 \ldots a^m b^m$$

for $m \geq 2$ over $V = \{a,\ b\}$.

Let $w = ab\Diamond a$ be a primitive partial word over $V = \{a,\ b\}$. Substituting last occurence of a by b will generate a nonprimitive partial word. We call the set of primitive partial words as non-subst-robust primitive partial words with one hole which on substitution of a symbol by another symbol results in a nonprimitive partial word and denote by $Q_p^{1\overline{S}}$.

$$Q_p^{1\overline{S}} = \{w \mid w = u_1au_2 \in Q_p^1 \textit{ and } w' = u_1bu_2 \notin Q_p^1\}$$

Observe that $Q_p^{1S} = Q_p^1 \setminus Q_p^{1\overline{S}}$. Next, we characterize the set of primitive partial words with one hole which are not subst-robust.

Theorem 20. *A primitive partial word with one hole $w = xay$ where $x,\ y \in V_\Diamond^*,\ a \in V$ is not subst-robust if and only if it is contained in a word of the form $u^{k_1}u_1au_2u^{k_2}$ where $x \sqsubset u^{k_1}u_1,\ y \sqsubset u_2u^{k_2}$ with $k_1 + k_2 \geq 1$ such that $u_1bu_2 = u$ where $a \neq b$.*

Proof. ($\Rightarrow$): Let us assume that $w = xay$ is a non-subst-robust primitive partial word with one hole. Then there exist a position in w in which a symbol can be substituted by another symbol and makes it nonprimitive. Let $w' = xby$ be the nonprimitive partial word and hence $w' = xby \sqsubset u^m,\ m \geq 2$ for some $u \in Q$. Therefore, we have $x \sqsubset u^{k_1}u_1,\ y \sqsubset u_2u^{k_2}$ such that $u_1bu_2 = u$ for $b \neq a$. Hence $w \sqsubset u^{k_1}u_1au_2u^{k_2}$.

($\Leftarrow$): Let w be a primitive partial word with one hole and $w = xay \sqsubset u^{k_1}u_1au_2u^{k_2}$, $k_1 + k_2 \geq 1$ where $x,\ y \in V_\Diamond^*$, $a \in V$ with $x \sqsubset u^{k_1}u_1$ and $y \sqsubset u_2u^{k_2}$. Also it is given that substituting a symbol $b \neq a$, we have $u_1bu_2 = u$. If we substitute a symbol $b \neq a$ in $w = xay$, we get $w' = xby$ such that $w' = xby \sqsubset u^{k_1}uu^{k_2} = u^{k_1+k_2+1}$ and $k_1 + k_2 + 1 \geq 2$. Hence $w = xay$ is not subst-robust.

Next we prove that the language of subst-robust primitive words with one hole is closed under cyclic permutation. We know that two partial words x and y are conjugate if there exist partial words u and v such that $x \sqsubset uv$ and $y \sqsubset vu$. A language L is closed under conjugacy relation if the cyclic permutations of all the words are in L.

Lemma 21. *The language Q_p^{1S} is closed under conjugacy relation.*

Proof. We prove it by contradiction. Let $w = v_1v_2$ be a primitive partial word with one hole such that $w \in Q_p^{1S}$. Suppose $w' = v_2v_1 \notin Q_p^{1S}$. $w = v_1v_2 \in Q_p^{1S}$ implies $w = v_1v_2 \in Q_p$. Since $w' = v_2v_1 \notin Q_p^{1S}$ then we can write $w' = v_2v_1 \sqsubset u^{k_1}u_1au_2u^{k_2}$ such that $u_1bu_2 = u$. We consider two cases depending on whether a is in v_1 or in v_2.

Case A. If the symbol a in contained in v_2 then we consider the following possibilities.

Case A.1 If entire u_1au_2 is from v_2 then $v_2 \sqsubset u^{k_1}u_1au_2u^ru'_1$ and $v_1 \sqsubset u'_2u^s$ where $u = u'_1u'_2$ and $r + s + 1 = k_2$. Now $v_1v_2 \sqsubset u'_2u^su^{k_1}u_1au_2u^ru'_1$. Substituting a by b we obtain a nonprimitive partial word which is a contradiction that $v_1v_2 \in Q_p^{1S}$.

Case A.2 If a portion of u_2 is from v_2 then $v_2 \sqsubset u^{k_1}u_1au'_2$ and $v_1 \sqsubset u''_2u^{k_2}$ where $u_2 = u'_2u''_2$. Now $v_1v_2 \sqsubset u''_2u^{k_2}u^{k_1}u_1au'_2$ which will result a nonprimitive partial word after a is substituted by the letter b. Moreover $v_1v_2 \sqsubset (u''_2u_1bu'_2)^{k_1+k_2+1}$ and v_1v_2 is not subst-robust primitive partial word.

Case B. If the symbol a is not contained in v_2 then we can handle two different cases as the previous one.

□

The following corollary is a consequence of Theorem 20.

Corollary 22. *A primitive partial word with one hole $w \in Q_p^{1\overline{S}}$ if and only if it is either contained in $u^nu'a$ or it's cyclic permutation for some $u \in Q_p$ and $u'b = u$ for $b \neq a$.*

Proof. The proof of necessary and sufficient conditions are as follow:
(Necessary Part:) Let $w = xay$ be a primitive partial word with one hole which is not subst-robust.Then $w = xay \sqsubset u^{k_1}u_1au_2u^{k_2}$ for some $u \in Q$ such that $u_1bu_2 = u$ and $k_1 + k_2 + 1 \geq 2$. As the language $Q_p^{1\overline{S}}$ is reflective, then $yxa \sqsubset u_2u^{k_2}u^{k_1}u_1a = (u_2u_1b)^{k_1+k_2}u_2u_1a = x^{k_1+k_2}x'a$ where $x = u_2u_2$ and $x' = u_2u_1$.

(Sufficient Part:) Let w be a partial word. If w is contained in $u^nu'a$ where $u'b = u$ or it's cyclic permutation then by substituting a by b where $b \neq a$, we obtain $w' \in u^{n+1}$ which is a nonprimitive partial word. Hence w is not a subst-robust primitive partial word.

6 Conclusion

We investigated exchange-robust and substitute robust primitive partial words with one hole. The structural characterization of each of the class of primitive partial words with one hole have been discussed and also some important combinatorial properties related to each of the class have been identified. We have shown that the language of non-exchange-robust primitive partial words is not a context-free language. We mention some of the interesting questions that are still unanswered. (1) The notion of robustness can be studied further for partial words with at least two holes. (2) Is the language of primitive partial words Q_p^X context-free? (3) One can consider to exchange two symbols at any positions and preserve primitivity.

References

1. Berstel, J., Boasson, L.: Partial words and a theorem of Fine and Wilf. Theoretical Computer Science 218(1), 135–141 (1999)
2. Berstel, J., Perrin, D., et al.: Theory of codes, vol. 22. Citeseer (1985)
3. Blanchet-Sadri, F.: Primitive partial words. Discrete Applied Mathematics 148(3), 195–213 (2005)
4. Blanchet-Sadri, F.: Algorithmic combinatorics on partial words. CRC Press (2007)
5. Blanchet-Sadri, F., Nelson, S., Tebbe, A.: On operations preserving primitivity of partial words with one hole. In: AFL. pp. 93–107 (2011)
6. Crochemore, M., Rytter, W.: Jewels of stringology: text algorithms. World Scientific (2002)
7. Dassow, J., Martin, G.M., Vico, F.J.: Some operations preserving primitivity of words. Theoretical Computer Science 410(30), 2910–2919 (2009)
8. Dömösi, P., Ito, M.: Context-free languages and primitive words. World Scientific (2014)
9. Fine, N.J., Wilf, H.S.: Uniqueness theorems for periodic functions. Proceedings of the American Mathematical Society 16(1), 109–114 (1965)
10. Gusfield, D.: Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology. Cambridge University Press (1997)
11. Lothaire, M.: Combinatorics on Words. Cambridge University Press (1997)
12. Lothaire, M.: Applied Combinatorics on Words. Cambridge University Press (2005)
13. Mitrana, V.: Primitive morphisms. Information processing letters 64(6), 277–281 (1997)
14. Nayak, A.C., Srivastava, A.K.: On del-robust primitive partial words with one hole. In: Language and Automata Theory and Applications, pp. 233–244. Springer (2016)
15. Ogden, W.: A helpful result for proving inherent ambiguity. Theory of Computing Systems 2(3), 191–194 (1968)
16. Păun, G., Santean, N., Thierrin, G., Yu, S.: On the robustness of primitive words. Discrete applied mathematics 117(1), 239–252 (2002)
17. Paun, G., Thierrin, G.: Morphisms and primitivity. Bulletin-European Association For Theoretical Computer Science 61, 85–88 (1997)
18. Rozenberg, G., Salomaa, A.: Handbook of Formal Languages: Volume 1. Word, Language, Grammar, vol. 1. Springer (1997)
19. Shallit, J.: A Second Course in Formal Languages and Automata Theory. Cambridge University Press (2008)

A Ogden's Lemma for CFL

For completeness we recall the Ogden's lemma to prove that the language of non-exchange-robust partial words is not context-free.

Lemma 23 (Ogden's Lemma [15]). *Let L be a context-free language. The there exists a constant $N > 0$ such that every string $w \in L$ that has at least N marked symbols can be decomposed in the form $w = uvxyz$ such that the following conditions hold:*

(i) *together v and x have at least one marked symbol.*
(ii) *vxy contains at most N marked symbols.*
(iii) *$uv^ixy^iz \in L$ for all $i \geq 0$.*