Robustness of Primitive and L-Primitive Words

Thesis submitted in partial fulfilment of the requirements
for the award of the degree of

Doctor of Philosophy

in

Computer Science and Engineering

by

Amit Kumar Srivastava

Under the supervision of

Dr. Benny George Kenkireth
Dr. Kalpesh Kapoor

Department of Computer Science and Engineering

Indian Institute of Technology Guwahati

Guwahati - 781039 Assam India

SEPTEMBER, 2017

Dedicated to

My beloved Family

For their love, care and support

Acknowledgements

While a completed dissertation bears the single name of the student, the process that leads to its completion is always accomplished in combination with the dedicated work of other people. I wish to acknowledge my appreciation to certain people.

First and foremost I want to thank my advisor Dr. Kalpesh Kapoor. He has taught me, both consciously and unconsciously, how to think and move forward in theoretical computer science. I appreciate all his contributions of time and ideas to make my Ph.D. work good. The joy and enthusiasm he has for his research was contagious and motivational for me, even during tough times in the Ph.D. pursuit.

I would like to thank my administrative advisor Dr. Benny George K. for his important initial time to discuss basic ideas to start and improve my work. I would also like to thank him for his support and encouragement. I would like to thank to Professor G. Sajith, Dr. Deepanjan Kesh and Dr. K.V. Krishna, for their support and encouragement.

I wish to express my unqualified thanks to my wife, Priyanka Amit Srivastava. I could never have accomplished this dissertation without her love, support, and understanding. I also wish to say thanks to my son, Shivam and I am sorry that I could not give him time when he needed most because of my research work and to complete of my dissertation. I am extremely grateful to my father, Shri Brindaban Bihari Lal, and my mother Smt. Kusum Srivastava, who raised me and taught me to study hard and to give priority in my life to the quest for knowledge.

Last but certainly not least, thanks to people and the government of my country, Bharat, and faculties of CSE department, IITG. I would like to express my full appreciation to the Technical Officers of the Department of Computer Science and Engineering and the medical staff of IIT Guwahati hospital to support me every time I needed.

September 2017 Amit Kumar Srivastava

Declaration

I certify that

- The work contained in this thesis is original and has been done by myself and under the general supervision of my supervisors.
- The work reported herein has not been submitted to any other Institute for any degree or diploma.
- Whenever I have used materials (concepts, ideas, text, expressions, data, graphs, diagrams, theoretical analysis, results, etc.) from other sources, I have given due credit by citing them in the text of the thesis and giving their details in the references. Elaborate sentences used verbatim from published work have been clearly identified and quoted.
- I also affirm that no part of this thesis can be considered plagiarism to the best of my knowledge and understanding and take complete responsibility if any complaint arises.
- I am fully aware that my thesis supervisors are not in a position to check for any possible instance of plagiarism within this submitted work.

September 2017 Amit Kumar Srivastava

Department of Computer Science and Engineering
Indian Institute of Technology Guwahati
Guwahati - 781039 Assam India

Dr. Benny George Kenkireth
Assistant Professor
Email : ben@iitg.ernet.in

Dr. Kalpesh Kapoor
Professor
Email : kalpesh@iitg.ernet.in

Certificate

This is to certify that this thesis entitled "**Robustness of Primitive and L-Primitive Words**" submitted by **Amit Kumar Srivastava**, in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy, to the Indian Institute of Technology Guwahati, Assam, India, is a record of the bonafide research work carried out by him under our guidance and supervision at the Department of Computer Science and Engineering, Indian Institute of Technology Guwahati, Assam, India. To the best of my knowledge, no part of the work reported in this thesis has been presented for the award of any degree at any other institution.

Date: September 2017
Place: Guwahati

Dr. Kalpesh Kapoor

Dr. Benny George Kenkireth

Abstract

Word combinatorics is a field which aims to study words and formal languages over some alphabet containing symbols, to understand their properties, with respect to operations such as concatenation, insertion, deletion and exchange of symbols. One of the most important study in the field of Word combinatorics is of primitive words, their properties and robustness. A word is said to be primitive if this can not be written as proper power of a smaller word. We investigate the effect on primitive words of point mutations (inserting or deleting symbols, substituting a symbol for another one), of morphisms, and of the operation of taking prefixes.

We characterise the subset of

- Primitive words that remains primitive on the operations, viz. substitution of any arbitrary symbol from the primitive words, deletion or insertion of a symbol in the primitive words or exchange of consecutive symbols. The properties of the languages of such primitive words are also discussed.

2 We find a property of language L such that the set QL, language of L-primitive words over an alphabet is reflective. We also find the smallest language L such that $QL = Q$. We examine the robustness of the language of L-primitive words.

3 We next examine the robustness of the language of pseudo-primitive words with a morphic involution. It is proved that a language of ins-robust pseudo-primitive words is not regular for an involution morphism.

Contents

List of Symbols

Symbols	Description
λ	Empty word
V	A finite Alphabet containing at least two symbols
$\mathbb{N}$	Set of natural numbers
$pref(w)$	The set of all prefixes of word w
$pref_k(w)$	The prefix of length k of word w
$suf(w)$	The set of all suffixes of word w
w^R	Reverse of word w
$per(w)$	Period of word w
$alph(w)$	Alphabet of word w, that is, set of the symbols present in word w
$\|w\|_a$	number of the symbol a in the word w
$Fact(w)$	Collection of factors of word w.
Q	Language of primitive words over an alphabet V

Z	Language of non-primitive words over an alphabet V
$Q(n)$	Language of primitive words of length n over an alphabet V
$Z(n)$	Language of non-primitive words of length n over an alphabet V
Q_S	Language of subst-robust primitive words over an alphabet V
$Q_{\overline{S}}$	Language of non-subst-robust primitive words over an alphabet V
Q_D	Language of del-robust primitive words over an alphabet V
$Q_{\overline{D}}$	Language of non-del-robust primitive words over an alphabet V
Q_I	Language of ins-robust primitive words over an alphabet V
$Q_{\overline{I}}$	Language of non-ins-robust primitive words over an alphabet V
Q_X	Language of exchange-robust primitive words over an alphabet V
$Q_{\overline{X}}$	Language of non-exchange-robust primitive words over an alphabet V
QL	Set of primitive words with respect to a language $L \subseteq V^+$ over an alphabet V
ZL	Set of non-primitive words with respect to a language $L \subseteq V^+$ over an alphabet V
QL_I	Ins-robust L-primitive words
$QL_{\overline{I}}$	Non-ins-robust L-primitive words
QL_D	Del-robust L-primitive words
QL_X	Exchange-robust L-primitive words

Q_θ	Language of θ-primitive words
$Q_{\theta I}$	Language of ins-robust θ-primitive words
$Q_{\theta \bar{I}}$	Language of non-ins-robust θ-primitive words
$T_\theta(x)$	The θ-border of word x
$t_{\theta x}$	θ-quasiperiod of word x.

"You have to grow from the inside out. None can teach you, none can make you spiritual. There is no other teacher but your own soul."

Swami Vivekananda (1863 - 1902) World spiritual leader

Chapter 1

Introduction

Extensive research has been done over the past three decades on ***Combinatorics on Words***. Despite the fact that there has been important contributions on words starting from the last century, they were usually needed as tools in computer science and mathematics. The main objects of automata theory are words, and in fact in any standard model of computing, words are main entity. Even when computing on numbers computers operate on words, i.e., representations of numbers as words. Consequently, on one hand, it is natural to study algorithmic properties of words. The collective works by several people under the pseudonym of Lothaire that has been documented in the form of series of books; namely algorithmic [1], algebraic [2], and applied combinatorics on words [3] gives an account of it. Several classical notions on properties of words have been explored. The wide range of applications of combinatorial properties of words in the subject of formal language theory [4], coding theory [5], computational biology [6], DNA computing [7], string matching [8, 9] etc. have drawn a lot of attention. One of the interesting problem which is still unsolved is whether the language of primitive words is context-free [10].

Primitive words play an important role in formal language theory [4], coding theory [5], combinatorics on words [1]. The theory of primitive words has been extensively studied and many combinatorial properties have been unveiled; see for example [11, 12, 13, 10, 14, 15].

Outline of Thesis

The thesis comprises six chapters. The chapter wise organization of the thesis is given below:

Chapter 1: This chapter discusses the motivation behind the research, followed by a survey of the state-of-art. It also briefly describes the contributions made in this thesis.

Chapter 2: In this chapter, we review the basic concepts on words. We mention some important results which are required to analyse the work in this thesis.

Chapter 3: This chapter presents the contributions made in some special type of primitive words which remain primitive after substitution of a symbol. This type of primitive words are called substitute robust primitive words. We discuss the characteristics and properties of these words. We also discuss the relation of the language of substitute robust primitive words, language of non-substitute robust primitive words with some formal languages. We discuss about del-robust primitive words, the primitive words which remain primitive after deletion of any one symbol. Next we study about the characteristics, properties and algorithmic method to identify such words. We show that the language of non-del-robust primitive words is not context free. We also discuss relation of the language of del-robust primitive words with other formal languages. The similar results are discussed for the ins-robust and exchange-robust primitive words.

Chapter 4: This chapter presents the contributions made in L-Primitive Words and discuss on various point mutations on these words. We discuss the word primitivity over some language L called L-primitive word in this chapter and characterise them. We discuss various robustness of L-primitive words. The relation of language of L-primitive words with the language of primitive words is discussed with the various languages which is subset of V^*. We also discuss some robustness of L-primitive words.

Chapter 5: This chapter presents the contributions made in Pseudo Quasiperiodic Words. In this chapter, we study some robustness of θ-primitive words that remains θ-primitive on insertion of any arbitrary symbol from the alphabet. Recall that θ is a morphic involution on V^*. We also discuss the θ-superprimitive words and pseudo L-primitive words and their characterization.

Chapter 6: This chapter presents the future works.

"You cannot believe in God until you believe in yourself."

Swami Vivekananda (1863 - 1902): A chief disciple of the 19th-century Indian mystic Ramakrishna.

Chapter 2

Background and Literature Survey

We introduce words and give some basic results on words, morphism and primitive words and formal languages which will be used in later chapters.

2.1 Words

An alphabet is a finite non-empty set V. The elements of V are called symbols or letters of V. A finite word over an alphabet V is a finite sequence of letters drawn from V. We assume that an alphabet V contains at least two elements. The empty word is denoted by λ. Concatenation or product of words is defined as $(a_1 \ldots a_n).(b_1 \ldots b_m) = a_1 \ldots a_n.b_1 \ldots b_m$. Clearly, this operation is associative and the empty word is the unit element with respect to this operation. The set of all words of length n over V is denoted by V^n. We define $V^* = \bigcup_{n \in \mathbb{N}} V^n$, where $V^0 = \{\lambda\}$ and, $V^+ = V^* \setminus \{\lambda\}$. A language L over V is a subset of V^*. Consequently, $V^* = (V^*, .)$ and $V^+ = (V^+, .)$ are a monoid and a semigroup respectively. Recall that a monoid S is called *free* if it has a subset B such that each element of S can be uniquely expressed as the finite sequences of zero or more elements of B. Such a B is referred to as a free generating set of S, or a base of S.

We denote length of word w as $|w|$ which is the total number of letters in w. The notation $|w|_a$ denotes the number of letter a in w. The letters that appear in w is $Alph(w) = \{a \mid |w|_a \geq 1\}$. A *power* of a word u is a word of the form u^k for some $k \in \mathbb{N}$. It is convenient to set $u^0 = \lambda$, for each word u. When $k \in \mathbb{N} \setminus \{0, 1\}$, we say that u^k is a proper power of u. A word u is said to be a *prefix* (resp. suffix, resp. factor) of a word v if there exists a word t (resp. t, resp. t and s) such that $ut = v$ (resp. $tu = v$, resp. $tus = v$). All these are said to be *proper* if they are different from v. The set of all prefixes of w is denoted by $pref(w)$, while $pref_k(w)$ means the prefix of length k of w (or w if $|w| < k$). Similarly, by $suf(w)$,for instance, we mean the set of suffixes of w and $Fact(w)$ is set of factors of w. The *reverse* of a word $w = a_1 \dots a_n$ with $a_i \in V$ is $w^R = a_n \dots a_1$. A language L is called *reflective* if $uv \in L$ implies $vu \in L$, for all $u, v \in V^*$.

A word w is said to be a *conjugate* of a word x if w is a cyclic shift of x, that is, if $w = uv$ and $x = vu$ for some u, $v \in V^*$ [16]. A language $L \subseteq V^*$ is called a k-*dense* language if for every word $w \in V^*$, there exist words x, $y \in V^*$ where $|xy| \leq k$ such that $xwy \in L$. L is said to be *dense* if it is k-dense for all $k \geq 1$. A language $L \subseteq V^*$ is called *right* k-*dense* if for every $u \in V^*$ there exists a word $x \in V^*$ where $|x| \leq k$ such that $ux \in L$. The language is said to be *right dense* if it is right k-dense for every $k \geq 1$.

Let $w = a_1 a_2 \dots a_n$ with $a_i \in V$. A number p is *period* of w if $a_i = a_{i+p}$ for $i = 1, \dots, n-p$. A word can have several periods. For example words $abcabcabca$ and $aabaabbaabaa$ have periods 3, 6, 9 and 7, 10, 11 respectively. We define the minimum of all the periods of a word w as $per(w)$. Moreover, any number greater than $|w|$ is always a period of w.

The rational $|w|/per(w)$ is called the exponent of w. If the exponent is an integer number $k > 1$, w can be simply written as u^k and is called an integer power (or k-power). A repetition is a word of exponent 2 or more, that is, a word with the period of at most half the word length. A maximal repetition at a position i in a word is a factor $w(i, j)$ which is a repetition such that its extension by one letter to the right or to the left yields a word with a larger period, that is,

- $per(w(i, j)) < per(w(i, j + 1))$
- $per(w(i, j)) < per(w(i - 1, j))$

where $per(w)$ is period of word w [17, 18]. For example, the factor $ababa$ in the word $w = abaababaabaab$, is a maximal repetition at fourth position with period 2, while the factor $abab$ is not a maximal repetition at this position. Section (4) [17] discuss the linear-time algorithm for finding all maximal repetitions in a word together with their periods.

Several facts about word combinatorics are known, we recall some of them.

Theorem 2.1 ([19])**.** *Let w and x be conjugates. Then w is a power if and only if x is a power. Furthermore, if $w = y^k$, $k \geq 2$, then $x = z^k$ where z is a conjugate of y.*

Lemma 2.1. *Let L be a reflective language. Then $\overline{L}$ is also reflective.*

Lemma 2.2. *Let L be a reflective language. Further, let $S \subseteq L$ is reflective. Then $L \setminus S$ is also reflective.*

Theorem 2.2 ([20])**.** *If words u, w, x and y over V satisfy $uw = xy$, then there exists a unique word t such that either*

(a) $u = xt$ *and* $y = tw$, *or*

(b) $x = ut$ *and* $w = ty$.

Theorem 2.3 ([21]). *Let* $u, v \in V^+$. *The following conditions are equivalent :*

(a) u *and* v *are conjugates,*

(b) *there exists a word* z *such that* $uz = zv$,

(c) *there exists words* z, p *and* q *such that* $u = pq$, $v = qp$ *and* $z \in p(qp)^*$.

Theorem 2.4 (Fine and Wilf, 1956, [22]). *Let* $u, v \in V^+$. *Then the words* u *and* v *are powers of a same word if and only if the words* u^r *and* v^r *have a common prefix of length* $|u| + |v| - gcd(|u|, |v|)$.

There exists an obvious reformulation of Theorem 2.4.

Corollary 2.1. *If a word has two periods* p *and* q, *and if it is of length at least* $p+q-gcd(p, q)$, *then it has also a period* $gcd(p, q)$.

Theorem 2.5 (Fine and Wilf, [1]). *Let* u *and* w *be words over an alphabet* V. *Suppose* u^h *and* w^k, *for some* h *and* k, *have a common prefix of length* $|u| + |w| - gcd(|u|, |w|)$. *Then there exists* $z \in V^*$ *of length* $gcd(|u|, |w|)$ *such that* $u, w \in z^*$. *The value* $|u| + |w| - gcd(|u|, |w|)$ *is also the smallest one that makes the theorem true.*

Lemma 2.3 (Lyndon-Schutzenberger [11]). *Let* $u, v \in V^*$ *with* $uv = vu$. *Then there exists a word* t *such that* $u, v \in t^* := \{t^n \mid n \in \mathbb{N}\}$.

Lemma 2.4 (Lyndon-Schutzenberger [11]). *Let* $u \in V^+$. *Then there exist a unique primitive word* z *and a unique integer* $k \geq 1$ *such that* $u = z^k$.

2.2 Finite Automata

An *automaton* over an alphabet V, is a composition $A = \langle S, E, I, T \rangle$, consisting of a finite set of states S, a finite set of edges or transitions $E \subset S \times V^* \times S$, set of an initial state $I \subset S$ and set of terminal states $T \subset S$. For an edge $e = (p, u, q)$, p is the origin state, u is the label and q is the end state. A path between two states is *successful* if it starts in an initial state and ends in a terminal state. The set recognized by the automaton is the set of labels of its successful paths. A set is *recognizable* or *regular* if it is the set of words recognized by some automaton (Kleene's Theorem). The symbol processing neural networks have a rich history [23, 24, 25]), and that networks of string-processing finite automata have appeared in many contexts ([26, 27, 28]).

Regular expressions over an alphabet V

(a) each symbol $a \in V$ is a regular expression.

(b) the empty string λ is a regular expression.

(c) the null set ϕ is a regular expression.

(d) if r and s are regular expressions, then so is $(r|s)$, where | represents union.

(e) if r and s are regular expressions, then so is rs.

(f) if r is a regular expression, then so is r^*.

Every regular expression is built up inductively, by finitely many applications of the above rules. A regular language is a formal language that can be expressed using a regular expression.

Lemma 2.5. *(Pumping lemma for regular languages [29]) For a regular language $L \subseteq V^*$, there exists an integer $p \geq 1$ such that for every word $w \in L$ with $|w| \geq p$, there is a factorization $w = xyz$ in V^* satisfying $y \neq \lambda$, $|xy| \leq p$ and $xy^n z \in L$ for all $n \in \mathbb{N}$.*

The integer p in the statement of the lemma is called the pumping length of L.

A *context-free grammar* $\mathcal{G} = \langle N, T, P \rangle$ consists of an alphabet N of variables, an alphabet T of terminal letters, which is disjoint from N , and a finite set $P \subseteq N \times (N \cup T)^*$ of productions. A language $L \subseteq T^*$ is a *context-free language* if there exists a context-free grammar $G = \langle N, T, P \rangle$ and a variable $v \in N$ such that $L = L(G, v) = \{w \in T^* \mid v \rightarrow^* w\}$.

Lemma 2.6. *(Pumping Lemma for Context-Free Languages [30, 31]) Let $L \subseteq V^*$ be a context-free language. There exists $p \in \mathbb{N}$ such that if $w \in L$ and $|w| \geq p$, then there exists a factorization $w = (u, v, x, y, z)$ satisfying $|v|, |y| > 0$, $|vxy| \leq p$, and $uv^i xy^i z \in L$ for each $i \geq 0$.*

Definition 2.1. *A pushdown automaton PDA is defined as $P = \langle S, I, T, \delta, s_0, s_t \rangle$, where*

(a) S is a finite set of states.

(b) I is the input alphabet

(c) T is the pushdown list alphabet

(d) δ is a mapping from $S \times (I \cup \{\lambda\}) \times T$ to $S \times T^*$. The value of $\delta(s, a, A)$ is, if defined, is of the form $(s^{'}, B)$ where $s^{'} \in S$, A on the top of the stack which is in T, is replaced by $B \in T^*$ and $a \in I \cup \{\lambda\}$.

(e) *$s_0 \in S$ is the initial state of the finite control.*

(f) *s_t is one of the designated final state.*

The language accepted by *PDAs* are exactly the context-free languages [32].

Definition 2.2. *A PDA, $P = \langle S, I, T, \delta, s_0, s_t \rangle$, is deterministic if*

(a) *for each $p \in S$, each $a \in I$, and each $A \in T$, δ does not contain both, an instruction $(p, \lambda, A)(q, B)$ and an instruction $(p, a, A)(q, B)$.*

(b) *for each $p \in S$, each $a \in I \cup \{\lambda\}$, and each $A \in T$, there is at most one instruction $(p, a, A)(q, B)$ in δ.*

2.3 Primitive Words

A word $w \in V^+$ is said to be primitive if w cannot be written as the integral power of a shorter word. Formally, w is primitive if $w = v^n$ implies $w = v$ and $n = 1$. The languages of primitive and non-primitive words are denoted by Q and Z, respectively [33]. We denote the set of primitive words of length n as $Q(n)$ and the set of non-primitive words of length n as $Z(n)$. Several facts are known about the languages Q and Z. We mention some of them below which will be used later.

Lemma 2.7. *A word w is primitive if and only if w is not an internal factor of its square ww, that is, $ww = xwy$ implies that either $x = \lambda$ or $y = \lambda$ [1, 40].*

Theorem 2.6 ([11]). *If $u \neq \lambda$, then there exist a unique primitive word p and a unique integer $k \geq 1$ such that $u = p^k$.*

Our next lemmas give an alternative condition for primitivity:

Lemma 2.8. *A nonempty word $w \in V^*$ is primitive if and only if it cannot be factored into two nonempty commuting words: $w \in Q \iff w \neq \lambda \wedge \forall u, v \in V^*, w = uv = vu \implies \lambda \in \{u, v\}$.*

Proof. If $w \in Q$ and $w = uv = vu$ for some non-empty words u and v then by lemma (2.3) there exists a word $t \in Q$ such that $u, v \in t^+$ which is a contradiction as $w = t^k$ for some $k \geq 2$. Therefore $\lambda \in \{u, v\}$.

Conversely, let $w \neq \lambda \wedge \forall u, v \in V^*, w = uv = vu \implies \lambda \in \{u, v\}$, and $w \notin Q$ then $w = t^k$ for some $t \in Q$ and $k \geq 2$. Suppose $u = t^r$ and $v = t^s$ such that $r, s \geq 1$ and $r + s = k$. In this case $w = uv = vu$ which is a contradiction. $\square$

Proposition 2.1 ([15]). *For every word $u \in V^+$ and every symbols $a, b \in V$, $a \neq b$, at least one of the words ua, ub is primitive.*

The above proposition says that the language of primitive words, Q is right 1-dense and therefore right k-dense for every k.

The next result has several rather interesting consequences, proving in some sense that "there are very many primitive words".

Corollary 2.2 ([34]). *Let V be an alphabet containing at least two symbols.*

(a) For every word $u \in V^*$, at most one of the words ua, with $a \in V$, is not primitive.

(b) For all words $u_1, u_2 \in V^*$, at most one of the words $u_1 a u_2$, with $a \in V$, is not primitive.

Lemma 2.9 ([15]). *The languages Q and Z are reflective.*

Theorem 2.7 ([34]). *Let $uv = f^i$, $u, v \in V^+$, $f \in Q$, $i \geq 1$. Then $vu = g^i$ for some $g \in Q$.*

Lemma 2.10 ([15]). *Let V be an alphabet containing at least two symbols.*

(a) If $w, wa \notin Q$ where $w \in V^+$ and $a \in V$, then $w \in a^+$.

(b) If $u_1, u_2 \in V^+$, $u_1 u_2 \neq a^n$, for any $a \in V$, $n \geq 1$ then at least one of the words among $u_1 u_2$, $u_1 a u_2$ is primitive.

Later we shall use several times the following two known results without explicitly mentioning them:

(a) If $f, g \in Q$, $f \neq g$, then for any $m, n \geq 2$, $f^m g^n \in Q$. [33]

(b) If $u, v \in V^+$, $uv \in Q$ and $n \geq 2$, then both $u(uv)^n$ and $v(uv)^n$ are in Q. [35]

For a morphism θ over an alphabet V, a word $w \in V^*$ is called pseudo-power of a word $t \in V^*$ if $w \in t\{t, \theta(t)\}^*$. A word w is θ-primitive if there exists no non-empty word $t \in V^+$ such that $w \in t\{t,\ \theta(t)\}^+$. The word t is called pseudo-period of w relative to θ, or simply θ-period of w if $w \in t\{t, \theta(t)\}^*$. We call a word $w \in V^+$ θ-primitive if there exists no non-empty word $t \in V^+$ such that w is a θ-power of t and $|w| > |t|$. We define the θ-primitive root of w, denoted by $\rho_\theta(w)$, as the shortest word t such that w is a θ-power of t.

Some results from [36] for a morphic involution θ which is used later, are as follows:

Corollary 2.3. *For any word $w \in V^*$ there exists a unique θ-primitive word $t \in V^*$ such that $w \in t\{t, \theta(t)\}^*$, i.e., $\rho_\theta(w) = t$.*

Corollary 2.4. *Let $u, v \in V^+$ be two words such that $u, v \in t^*$ for some $t \in Q$. Then $\rho_\theta(u) = \rho_\theta(v) = \rho_\theta(t)$.*

Corollary 2.5. *If we have two words $u, v \in V^+$ such that $u \in v\{v, \theta(v)\}^*$, then $\rho_\theta(u) = \rho_\theta(v)$.*

If every position in a string w is covered by an occurrence of a string t then we say that t covers w. For example $w = aabaaabaaabaabaa$ is covered by $t = aabaa$. If t covers w then t is both a prefix and a suffix of w. A string is *quasiperiodic* if it can be covered by a shorter string. A string is *superprimitive* if it is not quasiperiodic. If a superprimitive string t covers a string w then t is called quasiperiod of w [37, 38]. For example, let $V = \{a, b, c\}$ be the alphabet. Then the word $w = abcabcab$ is primitive but not superprimitive, covered by $abcab$, whereas $a^m b^n$ is superprimitive for $m,\ n \geq 1$.

A quasiperiod t of a string w is a unique substring of w which covers the word w, therefore t is both prefix and suffix of w [37].

2.4 L-Primitive Words

Let L be a language over an alphabet V. A word $x \in V^+$ is said to be an L-primitive word if x is not a proper power of any word in L. The set of L-primitive words over the alphabet V is denoted by $QL(V)$ or simply QL and the set of non-L-primitive words over the alphabet V is denoted by ZL.

An L-primitive word need not be primitive. For instance, let $L = \{abab\} \subseteq \{a, b\}^*$. Clearly, the word $abab$ is an L-primitive word, but not a primitive word. For $w \in V^+$, we define the set of L-primitive roots of w, denoted by $\sqrt[L]{w}$, is defined as

$$\sqrt[L]{w} = \{x \in QL \mid x^k = w,\ for\ some\ integer\ k \geq 1\}.$$

Further, for $X \subseteq V^*$, the L-primitive root of X, denoted by $\sqrt[L]{X}$, is defined as

$$\sqrt[L]{X} = \bigcup_{w\ \in\ X \setminus \{\lambda\}} \sqrt[L]{w}.$$

Some basic properties of L-primitive words are as follows.

(a) If $L = \phi$, then $QL = V^+$, the set of all nonempty words over V.

(b) If $L = V^*$, then $QL = Q_V$, the set of all primitive words over V.

Proposition 2.2 ([39])**.** *If L_1 and L_2 are two subsets of V^*, then $L_1 \subseteq L_2 \implies QL_2 \subseteq QL_1$.*

Proposition 2.3 ([39]). *Every primitive word is an L-primitive word. Hence, if $|V| \geq 2$, then $|QL| = \infty$.*

"God helps those who help themselves."

Swami Vivekananda

Chapter 3

Robustness of Primitive Words

3.1 Substitute-Robust Primitive Words

3.1.1 Symbol Substitution and Primitivity

Let V be an alphabet. For $x \in V^+$, consider the set

$$one(x) = \{x_1 b x_2 \mid x = x_1 a x_2, x_1, x_2 \in V^*, a, b \in V, a \neq b\}$$

A primitive word is Subst-robust if w remains primitive on substitution of any arbitrary symbol from the word w. In other words we say that x is subst-robust primitive word if $one(x) \subseteq Q$ [15].

The language of primitive words, Q, contains both subst-robust and non-subst-robust words of arbitrary length. For example, $w_n = aba^2b^2 \dots a^n b^n,\ n \geq 3$ are subst-robust, whereas $z_n = ab^n, n \geq 1$ are not. (Note that replacing the second occurrence of b by a in primitive word $u = bbabaa$ we get the non-primitive word $baabaa$, therefore u is not subst-robust.)

Proposition 3.1 ([15]). *If V is an alphabet containing at least three symbols, then for each 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.*

If we start with $x \in V^*$, $x \in Z$, then all substitutions in x gives a primitive word: from Corollary 2.2(b), we know that if $x = x_1 a x_2$ is not primitive then all words $x_1 b x_2, b \neq a$, are primitive. The argument holds even for $V = \{a,\ b\}$. The assertion in Proposition (3.1) does not hold true for $V = \{a, b\}$, e.g., we can not replace the second occurrence of a by b in the word $abaabb$, or the last occurrence of b by a without loosing the primitivity. However, we have the following result.

Lemma 3.1. *If $V = \{a, b\}$, then for each word $x \in V^*$, $|x| \geq 3$, and for each decomposition $x = x_1 c d x_2, x_1, x_2 \in V^*$, $c, d \in V$, at least one of the words $x_1 c' d x_2, x_1 c d' x_2$ is primitive, where $c, d, c', d' \in V$, $c' \neq c$ and $d \neq d'$.*

Proof. We prove it by contradiction. Consider a word x with $|x| \geq 4$. Then x can be written as $x = x_1 \alpha \beta x_2$ where $\alpha, \beta \in V$, $|x_1 x_2| \geq 2$. As Q is reflective, then to prove the lemma it is sufficient to prove that at least one of the $x_2 x_1 \alpha' \beta$, $x_2 x_1 \alpha \beta'$ is primitive.

Assume on the contrary, $x_2 x_1 \alpha' \beta = u^m$, $x_2 x_1 \alpha \beta' = v^n$ 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 we can assume that at least one of m and n is greater than 2 and without loss of generality we assume that $m \geq 3$, $n \geq 2$. Similarly, we cannot have $|u| = 1$. Otherwise $x_2 x_1 \alpha' \beta = u^m$ implies that $u \in \{a, b\}$. Then as $\alpha \neq \alpha'$, $\beta \neq \beta'$ then $x_2 x_1 \alpha \beta'$ is primitive which is a contradiction to the assumption. Hence we have $|u| \geq 2$.

Now, we have $|x_2 x_1| = m|u| - 2$ and $|x_2 x_1| = n|v| - 2$ which implies that

$$2|x_2 x_1| = m|u| + n|v| - 4$$

$$\implies |x_2 x_1| = \frac{m}{2}|u| + \frac{n}{2}|v| - 2$$

As $m \geq 3$, $n \geq 2$, we can write $|x_2 x_1| \geq |u| + |v| + \frac{1}{2}|u| - 2$. Since $|u| \geq 2$ we obtain $|x_2 x_1| \geq |u| + |v| - 1$. Consider the following cases.

(a) If $|x_2 x_1| = |u| + |v| - 1$ then $m = n = 2$ which leads to a contradiction.

(b) If $|x_2 x_1| > |u| + |v| - 1$ then by Theorem 2.5, there exist a word y such that $u = y^k$ and $v = y^l$ for some integers k, l. Hence $x_2 x_1 \alpha' \beta = y^{km}$ and $x_2 x_1 \alpha \beta' = y^{ln}$ which is a contradiction.

Thus at least one of the $x_2 x_1 \alpha' \beta$ and $x_2 x_1 \alpha \beta'$ is a primitive word. □

The condition $|x| \geq 3$ in the Lemma 3.1 is necessary: for $x = ab$, neither aa nor bb is primitive. Note also that ab is primitive, hence the condition of x being primitive does not help.

A subst-robust primitive word w is a primitive word which remains primitive on substitute of any arbitrary symbol from the word w. The formal definition is as follows.

Definition 3.1 (Substitute-Robust Primitive Word)**.** *A primitive word w of length n is said to be subst-robust primitive word if and only if the word*

$$pref(w,i)\ .a.\ suf(w,n-i-1)$$

is a primitive word for all $i \in \{0,1,\dots,n-1\}$ and for all $a \in V$.

For example, the words $abba$ and $a^n b^n$ for $n \geq 2$ are subst-robust primitive words.

The collection of all subst-robust primitive words over an alphabet V is denoted by Q_S. Clearly, the language of subst-robust primitive words is a subset of the set of primitive words, Q. Next lemma is a structural reformulation of definition of subst-robust primitive words.

Proposition 3.2. *A primitive word w is not subst-robust if and only if w can be expressed in the form of $u^{k_1}u_1cu_2u^{k_2}$ where $u, u_1, u_2 \in V^*$, $k_1, k_2 \geq 0$, $k_1 + k_2 \geq 1$ and $u_1bu_2 = u$, for some $c \neq b$, $c, b \in V$.*

Proof. We prove the sufficient and necessary conditions below.

($\Leftarrow$) This part is straightforward. Let us consider a word $w = u^{k_1}u_1cu_2u^{k_2}$ where $u_1bu_2 = u$ for some $b \neq c$ and $b, c \in V$. Now substitution of the letter b at place of c in w gives the exact power of u which will be a non-primitive word. Hence, w is not a subst-robust primitive word.

($\Rightarrow$) Let w be a primitive word but not subst-robust primitive word. Then there exists a decomposition $w = w_1cw_2$ for $c \in V$ such that w_1bw_2 is not a primitive word for some $b \neq c$ and $b \in V$. That is, $w_1bw_2 = u^n$ for some $u \in Q$ and $n \geq 2$. Therefore $w_1 = u^r u_1$ and $w_2 = u_2 u^s$ for $r, s \geq 0$ *and* $r + s \geq 1$ such that $u_1bu_2 = u$. Hence $w = u^r u_1cu_2u^s$.

□

We denote the set of non-subst-robust primitive words as $Q_{\overline{S}} = Q \setminus Q_S$, where '\' is the set difference operator. By Lemma 2.9, we know that the language of primitive words Q and the language of non-primitive words Z over an alphabet V are reflective. Similarly, we have the property of reflectivity for the language of subst-robust primitive words Q_S.

Lemma 3.2. *If $w \in Q_S$ then $rev(w) \in Q_S$.*

Proof. We prove this by contradiction. Let $w \in Q_S$ such that $rev(w)$ is not a subst-robust primitive word. Therefore, $rev(w) = p^r p_1ap_2p^s$ where $p \in Q$ and $p = p_1bp_2$ for some

$a \neq b$. Then the word $w = rev(p^r p_1 a p_2 p^s) = (rev(p))^s\ rev(p_2)\ \ a\ rev(p_1)\ (rev(p))^r$ and since $p = p_1\ b\ p_2$, so $rev(p) = rev(p_2)\ b\ rev(p_1)$. By Proposition 3.2, w is not a subst-robust primitive word, which is a contradiction. Therefore, if $w \in Q_S$ then $rev(w) \in Q_S$. □

Next we show that any cyclic permutation of a subst-robust primitive word is also a subst-robust primitive word.

Theorem 3.1. *Q_S is reflective.*

Proof. (By contradiction.) Let there be a word $w = xy \in Q_S$ such that $yx \notin Q_S$. Since $w \in Q_S$ hence $w \in Q$. Q is reflective (lemma 2.9), therefore $yx \in Q$ and so $yx \in Q \setminus Q_S$, that is, $yx \in Q_{\overline{S}}$. Using Proposition 3.2, we have $yx = u^r u_1 a u_2 u^s$ where $u = u_1 b u_2 \in V^*$ for some $b \in V$, $b \neq a$ and $r + s \geq 1$. We consider here two cases depending on the inclusion of the letter a either in word y or in the word x.

Case A If a is contained in y, we consider two subcases.

- **Case A.1** If $u_1 a u_2$ is contained in y then $y = u^r u_1 a u_2 u^{r'} u'_1$, and $x = u'_2 u^{s'}$ for $u = u_1 b u_2 = u'_1 u'_2$. So $xy = u'_2 u^{s'} u^r u_1 a u_2 u^{r'} u'_1$ which is not subst-robust as after substitution of a by b the new word will be $u'_2 u^{s'} u^r u_1 b u_2 u^{r'} u'_1 = (u'_2 u'_1)^{s'+r+r'+2}$ which is not a primitive word. This is a contradiction to the assumption that $w = xy \in Q_S$.
- **Case A.2** If a portion of u_2 belongs to y then $y = u^r u_1 a u'_2$, and $x = u''_2 u^s$ for $u = u_1 b u_2$ *and* $u_2 = u'_2 u''_2$. Now, $xy = u''_2 u^s u^r u_1 b u'_2$ which is not subst-robust as after substitution of a by b, and the result will be $(u''_2 u_1 b u'_2)^{s+r+1}$ a non-primitive word. This is a contradiction to the assumption that $w = xy \in Q_S$.

Case B If a belongs to x, similar subcases as in Case A are to be considered and proved.

Hence Q_S is reflective. □

Corollary 3.1. *$Q_{\overline{S}}$ is reflective.*

Proof. We prove it by contradiction. Let there be a word $w = xy \in Q_{\overline{S}}$ such that $yx \notin Q_{\overline{S}}$. By Lemma 2.9, we have that $xy \in Q$ and Q is reflective, so, $yx \in Q$. Therefore $yx \in Q \setminus Q_{\overline{S}}$, that is, $yx \in Q_S$. Since Q_S is reflective, by Theorem 3.1, we have $xy \in Q_S$, which is a contradiction. Hence $yx \in Q_{\overline{S}}$. □

Corollary 3.2. *A word w is in the set $Q_{\overline{S}}$ if and only if it is of the form $u^n u' a$ or its cyclic permutation for some $u \in Q$, $u \neq a$, $n \geq 1$ and $u = u'b$ for some $b \in V$ and $b \neq a$.*

Proof. We prove the sufficient and necessary conditions below.

(⇒) Let $w \in Q_{\overline{S}}$, then w can be written as $w = u^r u_1 a u_2 u^s$ where $u(= u_1 b u_2) \in Q$ for some $b \neq a$ and $a, b \in V$. Since $Q_{\overline{S}}$ is reflective, and $u_2 u^s u^r u_1 b = ((u_2 u_1 b)^{r+s+1}) \in Z$ and so w is a cyclic permutation of $v^{r+s} v' a$ where $v = v' b$ and $v' = u_2 u_1$.

(⇐) If a word w is a cyclic permutation of $u^n u' a$ for $n \geq 1$ then after replacing a by b it gives a cyclic permutation of u^{n+1} which is non-primitive. Since Z is reflective therefore, $w \in Q_{\overline{S}}$.

□

Observe that not all infinite subsets of Q are reflective. For example, the subset $\{a^n b^n \mid n \geq 1\}$ of Q over the alphabet $\{a, b\}$ is not reflective. We now investigate the relation between the language of non-subst-robust primitive words with the traditional languages in Chomsky hierarchy.

Theorem 3.2. *$Q_{\overline{S}}$ is not a context-free language.*

Proof. On contradiction, let us assume that $Q_{\overline{S}}$ is a context-free language. Let $p > 0$ be an integer which is the pumping length that is guaranteed to exist by the pumping lemma. Consider the string $s = a^{p+1} b^{p+1} a^{p+2} b^p$, where a and b are distinct letters from the underlying alphabet V. It is easy to see that $s \in Q_{\overline{S}}$ and $|s| \geq p$.

Hence, by Pumping Lemma for context free languages, s can be written in the form $s = uvwxy$, where u, v, w, x, and y are factors, such that $|vwx| \leq p$, $|vx| \geq 1$, and $uv^k wx^k y$ is in $Q_{\overline{S}}$ for every $k \geq 0$. By the choice of s and the fact that $|vwx| \leq p$, it is easily seen that the substring vwx can contain no more than two distinct symbols. That is, we have $uvwxy = a^{p+1} b^{p+1} a^{p+2} b^p$, $vwx \leq p$, $|vx| \geq 1$. There are four main cases to be considered. The string vwx is

(a) power of a.

(b) power of b.

(c) of the form $a^j b^k$, $j, k \geq 1$.

(d) of the form $b^j a^k$, $j, k \geq 1$.

Case (a) First we discuss pumping lemma such that vwx is a power of a. There are two possible cases.

(i) $u = a^m$, $v = a^j$, $w = a^k$, $x = a^l$, $y = a^n b^{p+1} a^{p+2} b^p$,
$j + l \geq 1$, $j + k + l \leq p$, $m + j + k + l + n = p + 1$.

In this case, pumping lemma does not satisfy for $i = 0$ as $uv^iwx^iy = a^{p'}b^{p+1}a^{p+2}b^p \notin Q_{\overline{S}}$, where $1 \leq p' (= m + k + n) \leq p$.

(ii) $u = a^{p+1}b^{p+1}a^m$, $v = a^j$, $w = a^k$, $x = a^l$, $y = a^nb^p$,
$j + l \geq 1$, $j + k + l \leq p$, $m + j + k + l + n = p + 2$.
For $i = 3$, $uv^iwx^iy = a^{p+1}b^{p+1}a^{p'}b^p \notin Q_{\overline{S}}$ as $p' (= m+3j+k+3l+n = p+2+2j+2l) > p + 3$
and therefore no replacement in uv^iwx^iy is possible to make it non-primitive, and therefore pumping lemma does not hold.

Case (b) Next we discuss the pumping lemma such that vwx is a power of b.

(i) $u = a^{p+1}b^m$, $v = b^j$, $w = b^k$, $x = b^l$, $y = b^na^{p+2}b^p$, $j + l \geq 1$, $j + k + l \leq p$, $m + j + k + l + n = p + 1$.
$uv^iwx^iy = a^{p+1}b^{p'}a^{p+2}b^p \notin Q_{\overline{S}}$ for $i = 2$ as $p' \geq p + 3$.
Clearly in this case pumping lemma does not hold.

(ii) $u = a^{p+1}b^{p+1}a^{p+2}b^m$, $v = b^j$, $w = b^k$, $x = b^l$, $y = b^n$, $j + l \geq 1$, $j + k + l \leq p$, $m + j + k + l + n = p$.
$uv^iwx^iy = a^{p+1}b^{p+1}a^{p+2}b^{p'} \notin Q_{\overline{S}}$ for $i = 4$ as $p' \geq p + 3$
that is, in this case also no replacement will give non-primitive word.

Case (c): In this case we discuss the pumping lemma so that $vwx = a^jb^k$, $j, k \geq 1$.
We have, $u = a^m$, $v = a^j$, $w = a^{k'}$, $x = a^lb^k$, $y = b^na^{p+2}b^p$.
Here $j + l + k \geq 1$, $j + k' + k + l \leq p$, $m + j + k' + l = p + 1$ and $k + n = p + 1$.
In this case, $uv^iwx^iy = a^ma^{3j}a^{k'}a^lb^ka^lb^ka^lb^kb^na^{p+2}b^p$

$= a^{p+1}a^{2j}b^ka^lb^ka^lb^{p+1}a^{p+2}b^p$.

In this case for $i = 3$, $uv^iwx^iy \notin Q_{\overline{S}}$ as $0 \leq l, k < p$.

Case (d) Similar to Case (c), we can find i in this case as well, such that pumping lemma does not hold.

Since pumping lemma does not hold in any case, therefore $Q_{\overline{S}}$ is not context-free. □

3.1.2 Recognizing Subst-Robust Primitive Words

In this section we give a linear time algorithm to recognize a subst-robust primitive word. An algorithm to test whether a given word is primitive, is based on the lemma 2.7 which state that a word w is primitive if and only if w is not an internal factor of ww.

Observe that if a word $w \in Q_{\overline{S}}$, then by Corollary 3.2, there exists a cyclic permutation of w which contains a factor of length $|w|-1$ with periodicity p which divides $|w|$ and $\frac{|w|}{p} \geq 2$. We make use of this fact in the following theorem by observing that the word ww contains a periodic substring (of length $|w|-1$) of one of the cyclic permutation of a word w.

Lemma 3.3. *Let u be a primitive word. Then u is a non-subst-robust primitive word if and only if the word uu contains a periodic word of length $|u|-1$ with periodicity p such that p divides $|u|$ and $\frac{|u|}{p} \geq 2$.*

Proof. We prove the sufficient and necessary conditions below.

(⇒) Let u be a non-subst-robust primitive word. Thus u can be written as $t^r t_1 a t_2 t^s$ where $t_1, t_2 \in V^*$, $a \in V$, $r+s \geq 1$ and $t = t_1 b t_2$ for some $a \neq b$. Thus, the word $uu = t^r t_1 a t_2 t^s t^r t_1 a t_2 t^s$ contains a factor $t_2 t^s t^r t_1$ of length $|u|-1$ which is equal to the primitive word $(t_2 t_1 b)^{r+s} t_2 t_1$.

(⇐) Let the word uu have a factor of length $|u|-1$ which is periodic with periodicity p such that $\frac{|u|}{p} \geq 2$ where u is a primitive word. Then $uu = t_1 p^r p' a t_2$, where $t_1, t_2 \in V^*$, $|p^r p'| = |u|-1$, $p = p'b \in Q$ for some $a \neq b$ and $r \geq 1$. Since $|p^r p' a| = |u|$, and therefore u is a cyclic permutation of non-subst-robust primitive word $p^r p' a$. Since $Q_{\overline{S}}$ is reflective, therefore $u \in Q_{\overline{S}}$.

□

Let u be primitive word. The following corollary claims that there are some maximal repetitions with specific periods in the word uu whose lengths are at least $|u|-1$ if $u \in Q_{\overline{S}}$.

Corollary 3.3. *Let u be a primitive word. If the word uu contains a maximal repetition of length at least $|u|-1$ with a period p where p divides $|u|$ and $p < |u|$ then u is a non-subst-robust primitive word.*

Proof. Let a maximal repetition $v^k v_1$ be a factor of uu for $v \in Q$, $k \geq 2$ and v_1 be prefix of v such that $|v_1| = |v|-1$. Since $|v|$ divides $|u|$, we have $|u| = r|v|$ for some $r \geq 2$ and $r \leq k$, that is, uu contains $v^r v_1$. Hence by Lemma 3.3, u is a non-subst-robust primitive word. □

The computation of maximal repetitions in a word can be done in linear time in terms of the length of the input word [17].

3.2 Del-Robust Primitive Words

In this section, we present another type of primitive words which remain primitive after deletion of any one symbol. Such words are called *del-robust primitive words*. On the basis of characteristics of primitive words we observe some properties of such words and relation with some formal languages. We give a linear time algorithm to verify the del-robustness of primitive words.

Definition 3.2 (Del-Robust Primitive Word)**.** *A primitive word w of length n is said to be del-robust primitive word if and only if the word*

$$pref(w,i)\ .\ suf(w,n-i-1)$$

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

For example, the words a^4b^5 and $aba^2b^2\ldots a^mb^m$ for $m \geq 2$ are del-robust primitive words.

The collection of all del-robust primitive words over an alphabet V is denoted by Q_D. Clearly, the language of del-robust primitive words is a subset of the set of primitive words, Q. Next we give a structural characterization of the words that are in the set Q but not in the set Q_D. The definition can be written in form of following lemma.

Proposition 3.3. *A primitive word w is not del-robust if and only if w can be expressed in the form of $u^{k_1}u_1cu_2u^{k_2}$ where $u,u_1,u_2 \in V^*$, $u_1u_2 = u$, $c \in V$, $k_1,k_2 \geq 0$ and $k_1+k_2 \geq 1$.*

Proof. We prove the sufficient and necessary conditions below.

(⇐) This part is straightforward. Let us consider a word $w = u^{k_1}u_1cu_2u^{k_2}$ where $u_1u_2 = u$ and $c \in V$. Now deletion of the letter c in w gives the exact power of u which will be a non-primitive word. Hence, w is not a del-robust primitive word.

(⇒) Let w be a primitive word but not del-robust primitive word. Then there exists a decomposition $w = w_1cw_2$ for $c \in V$ such that w_1w_2 is not a primitive word. That is, $w_1w_2 = u^n$ for some $u \in Q$ and $n \geq 2$. Therefore $w_1 = u^ru_1$ and $w_2 = u_2u^s$ for $r,s \geq 0$ *and* $r+s \geq 1$ such that $u_1u_2 = u$. Hence $w = u^ru_1cu_2u^s$.

□

Definition 3.3 (Non-Del-Robust Primitive Words)**.** *A primitive word w, is said to be non-del-robust primitive word if and only if $w \in Q$ and $w \notin Q_D$. Further, $Q_{\overline{D}} = Q \setminus Q_D$, where '\' is the set difference operator.*

Proposition 3.4. *Let $u, v \in Q$, $u^m = u_1au_2$ and $v = u_1u_2$. Then $u^mv^n \in Q_{\overline{D}}$ for $m, n \geq 2$.*

Proof. From Lemma 2.10, we know that at least one of u_1au_2 and u_1u_2 is primitive. Since $u^m = u_1au_2$ and $v = u_1u_2$, therefore $u^mv^n = u_1au_2(u_1u_2)^n$. After deletion of the letter a we will get $(u_1u_2)^{n+1}$ which is not a primitive word. However, by Lemma 3.5, u^mv^n is a primitive word for $m, n \geq 2$. Hence it is not a del-robust word, that is, $u^mv^n \in Q_{\overline{D}}$. □

Next, we discuss the reflective property for the language of del-robust primitive words Q_D.

Lemma 3.4. *If $w \in Q_D$ then $rev(w) \in Q_D$.*

Proof. We prove this by contradiction. Let $w \in Q_D$ such that $rev(w)$ is not a del-robust primitive word. Therefore, $rev(w) = p^rp_1ap_2p^s$ for some $p \in Q$ and $p = p_1p_2$. Then the word $w = rev(p^rp_1ap_2p^s) = (rev(p))^s\ rev(p_2)\ \ a\ rev(p_1)\ (rev(p))^r$ and since $p = p_1p_2$, so $rev(p) = rev(p_2)rev(p_1)$. By Proposition 3.3, w is not a del-robust primitive word, which is a contradiction. Therefore, if $w \in Q_D$ then $rev(w) \in Q_D$. □

Next we show that any cyclic permutation of a del-robust primitive word is also a del-robust primitive word.

Theorem 3.3. *Q_D is reflective.*

Proof. (By contradiction.) Let there be a word $w = xy \in Q_D$ such that $yx \notin Q_D$. Since $w \in Q_D$, hence $w \in Q$. By Lemma 2.9, we know that Q is reflective. Therefore $yx \in Q$ and so $yx \in Q \setminus Q_D$, that is, $yx \in Q_{\overline{D}}$. Using Proposition 3.3, we have $yx = u^ru_1au_2u^s$ where $u = u_1u_2 \in V^*$, $a \in V$ and $r + s \geq 1$. We consider here two cases depending on the inclusion of the letter a either in word y or in the word x.

Case A If a is contained in y, we consider two sub-cases.

Case A.1 If u_1au_2 is contained in y then $y = u^ru_1au_2u^{r'}u_1'$, and $x = u_2'u^{s'}$ for $u = u_1u_2 = u_1'u_2'$. So $xy = u_2'u^{s'}u^ru_1au_2u^{r'}u_1'$ which is not del-robust as after deletion of a the new word will be $(u_2'u_1')^{s'+r+r'+2}$ which is not a primitive word. This is a contradiction to the assumption that $w = xy \in Q_D$.

Case A.2 If a portion of u_2 belongs to y then $y = u^ru_1au_2'$, and $x = u_2''u^s$ for $u = u_1u_2$ *and* $u_2 = u_2'u_2''$. Now, $xy = u_2''u^su^ru_1au_2'$ which is not del-robust as after deletion of a, and the result will be $(u_2''u_1u_2')^{s+r+1}$ a non-primitive word. This is a contradiction to the assumption that $w = xy \in Q_D$.

Case B If a belongs to x, similar sub-cases as in Case A can be considered and proved.

Hence Q_D is reflective. □

Corollary 3.4. *$Q_{\overline{D}}$ is reflective.*

Proof. We prove it by contradiction. Let there be a word $w = xy \in Q_{\overline{D}}$ such that $yx \notin Q_{\overline{D}}$. We have that $xy \in Q$ and Q is reflective, so, $yx \in Q$. Therefore $yx \in Q \setminus Q_{\overline{D}}$, i.e. $yx \in Q_D$. Since Q_D is reflective by Theorem 3.3, we have $xy \in Q_D$, which is a contradiction. Hence $yx \in Q_{\overline{D}}$. □

Corollary 3.5. *A word w is in the set $Q_{\overline{D}}$ if and only if it is of the form $u^n a$ or its cyclic permutation for some $u \in Q$, $u \neq a$ and $n \geq 2$.*

Proof. We prove the sufficient and necessary conditions below.

(⇒) Let $w \in Q_{\overline{D}}$, then w can be written as $w = u^r u_1 a u_2 u^s$ for some $u(= u_1u_2) \in Q$ and $a \in V$. Since $Q_{\overline{D}}$ is reflective, therefore $u_2 u^s u^r u_1 a = ((u_2u_1)^{r+s+1}a)$ is also in $Q_{\overline{D}}$.

(⇐) If a word w is a cyclic permutation of $u^n a$ for $n \geq 2$ then after deletion of a it gives a cyclic permutation of u^n which is non-primitive (since Z is reflective). Therefore, $w \in Q_{\overline{D}}$.

□

We now investigate the relation between the language of non-del-robust primitive words with the traditional languages in Chomsky hierarchy.

Theorem 3.4. *$Q_{\overline{D}}$ is not a context-free language.*

Proof. Let us assume that $Q_{\overline{D}}$ is a context-free language. Let $p > 0$ be an integer which is the pumping length that is guaranteed to exist by the pumping lemma. Consider the string $s = a^{p+1}b^pa^pb^pa^pb^p$, where a and b are distinct letters from an alphabet V. It is easy to see that $s \in Q_{\overline{D}}$ and $|s| \geq p$.

Hence, by the Pumping Lemma 2.6, $s = uvwxy$, where $u, v, w, x, y \in V^*$ such that $|vwx| \leq p$, $|vx| \geq 1$, and $uv^iwx^iy \in Q_{\overline{D}}$ for every $i \geq 0$. By the choice of s and the fact that $|vwx| \leq p$, it is easily seen that the substring vwx can contain no more than two distinct symbols. That is, we have $s = uvwxy = a^{p+1}b^pa^pb^pa^pb^p$, $vwx \leq p$, $|vx| \geq 1$. There are four main cases to be considered. The string vwx is

(a) power of a.

(b) power of b.

(c) of the form $a^j b^k$, $j, k \geq 1$.

(d) of the form $b^j a^k$, $j, k \geq 1$.

Case (a) vwx is a power of a.

a(1) In this case we check for first substring of a, that is in a^{p+1}. $u = a^m$, $v = a^j$, $w = a^k$, $x = a^l$, $y = a^n b^p a^p b^p a^p b^p$, $j + l \geq 1$, $j + k + l \leq p$, $m + j + k + l + n = p + 1$.
Now, $uv^i wx^i y = a^{p'} b^p a^p b^p a^p b^p \notin Q_{\overline{D}}$ for $i = 0$ as $1 \leq p' \ (= m + k + n) \leq p$. Therefore, in this case pumping law does not hold for $Q_{\overline{D}}$. Next two cases a(2) and a(3) is to check pumping lemma in both the substring a^p at second and third occurrence in the string s.

a(2) $u = a^{p+1} b^p a^m$, $v = a^j$, $w = a^k$, $x = a^l$, $y = a^n b^p a^p b^p$, $j + l \geq 1$, $j + k + l \leq p$, $m + j + k + l + n = p$.
$uv^i wx^i y = a^{p+1} b^p a^{p'} b^p a^p b^p \notin Q_{\overline{D}}$ for $i = 0$ as $0 \leq p' \ (= m + k + n) < p$.

a(3) $u = a^{p+1} b^p a^p b^p a^m$, $v = a^j$, $w = a^k$, $x = a^l$, $y = a^n b^p$, $j + l \geq 1$, $j + k + l \leq p$, $m + j + k + l + n = p$.
$uv^i wx^i y = a^{p+1} b^p a^p b^p a^{p'} b^p \notin Q_{\overline{D}}$ for $i = 0$ as $0 \leq p' \ (= m + k + n) < p$.

Case (b) vwx is a power of b. We check for the partition of s such that $vwx = b^n$ in *b(1)*, *b(2)* and *b(3)* cases.

b(1) $u = a^{p+1} b^m$, $v = b^j$, $w = b^k$, $x = b^l$, $y = b^n a^p b^p a^p b^p$, $j + l \geq 1$, $j + k + l \leq p$, $m + j + k + l + n = p$.
$uv^i wx^i y = a^{p+1} b^{p'} a^p b^p a^p b^p \notin Q_{\overline{D}}$ for $i = 0$ as $0 \leq p' \ (= m + k + n) < p$.

b(2) $u = a^{p+1} b^p a^p b^m$, $v = b^j$, $w = b^k$, $x = b^l$, $y = b^n a^p b^p$, $j + l \geq 1$, $j + k + l \leq p$, $m + j + k + l + n = p$.
$uv^i wx^i y = a^{p+1} b^p a^p b^{p'} a^p b^p \notin Q_{\overline{D}}$ for $i = 0$ as $0 \leq p' \ (= m + k + n) < p$.

b(3) $u = a^{p+1} b^p a^p b^p a^p b^m$, $v = b^j$, $w = b^k$, $x = b^l$, $y = b^n$, $j + l \geq 1$, $j + k + l \leq p$, $m + j + k + l + n = p$.
$uv^i wx^i y = a^{p+1} b^p a^p b^p a^p b^{p'} \notin Q_{\overline{D}}$ for $i = 0$ as $0 \leq p' \ (= m + k + n) < p$.

Case (c): In case (c), we discuss for the partition os s such that $vwx = a^j b^k$, $j, k \geq 1$. In this case there are nine cases based upon the division of v, w, and k in $a^j b^k$ and position of vwx in s.

c(1) $u = a^m$, $v = a^j$, $w = a^{k'}$, $x = a^l b^k$, $y = b^n a^p b^p a^p b^p$, $j + l + k \geq 1$, $j + k' + k + l \leq p$, $m + j + k' + l = p + 1$ and $k + n = p$.
$uv^i wx^i y = a^{p_1} b^{p_2} a^p b^p a^p b^p \notin Q_{\overline{D}}$ for $i = 0$ as $0 \leq p_1 \ (= m + k') \leq p$ and $p_2 \ (= n) < p$.

c(2) $u = a^{p+1}b^pa^m$, $v = a^j$, $w = a^{k'}$, $x = a^lb^k$, $y = b^na^pb^p$, $j + l + k \geq 1$, $j + k' + k + l \leq p$, $m + j + k' + l = p$ and $k + n = p$.
$uv^iwx^iy = a^{p+1}b^pa^{p_1}b^{p_2}a^pb^p \notin Q_{\overline{D}}$ for $i = 0$ as $0 \leq p_1\ (= m+k') < p$ and $p_2\ (= n) < p$.

c(3) $u = a^{p+1}b^pa^pb^pa^m$, $v = a^j$, $w = a^{k'}$, $x = a^lb^k$, $y = b^n$, $j + l + k \geq 1$, $j + k' + k + l \leq p$, $m + j + k' + l = p$ and $k + n = p$.
$uv^iwx^iy = a^{p+1}b^pa^pb^pa^{p_1}b^{p_2} \notin Q_{\overline{D}}$ for $i = 0$ as $0 \leq p_1\ (= m+k') < p$ and $p_2\ (= n) < p$.

c(4) $u = a^m$, $v = a^j$, $w = a^{k'}b^l$, $x = b^{l'}$, $y = b^na^pb^pa^pb^p$, $j + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j + k' = p + 1$ and $l + l' + n = p$.
$uv^iwx^iy = a^{p_1}b^{p_2}a^pb^pa^pb^p$ for $i = 0$, where $p_1 = m + k'$ and $p_2 = l + n$. Since $j + l' \geq 1$ therefore either ($p_1 \leq p$ and $p_2 \leq p$) or ($p_1 \leq p + 1$ and $p_2 \leq p - 1$). In both the cases $uv^iwx^iy = a^{p_1}b^{p_2}a^pb^pa^pb^p \notin Q_{\overline{D}}$.

c(5) $u = a^{p+1}b^pa^m$, $v = a^j$, $w = a^{k'}b^l$, $x = b^{l'}$, $y = b^na^pb^p$, $j + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j + k' = p$ and $l + l' + n = p$.
$uv^iwx^iy = a^{p+1}b^pa^{p_1}b^{p_2}a^pb^p$ for $i = 0$, where $p_1 = m + k'$ and $p_2 = l + n$. Since $j + l' \geq 1$ therefore either ($p_1 < p$ and $p_2 \leq p$ or ($p_1 \leq p$ and $p_2 < p$). In both the cases $uv^iwx^iy = a^{p_1}b^{p_2}a^pb^pa^pb^p \notin Q_{\overline{D}}$.

c(6) $u = a^{p+1}b^pa^pb^pa^m$, $v = a^j$, $w = a^{k'}b^l$, $x = b^{l'}$, $y = b^n$, $j + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j + k' = p$ and $l + l' + n = p$. Similar to the case c(5).

c(7) $u = a^m$, $v = a^jb^{k'}$, $w = b^l$, $x = b^{l'}$, $y = b^na^pb^pa^pb^p$, $j + k' + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j = p + 1$ and $k' + l + l' + n = p$.
$uv^iwx^iy = a^{p_1}b^{p_2}a^pb^pa^pb^p \notin Q_{\overline{D}}$ for $i = 0$ as $p_1(= m) \leq p$ and $p_2\ (= l + n) \leq p$.

c(8) $u = a^{p+1}b^pa^m$, $v = a^jb^{k'}$, $w = b^l$, $x = b^{l'}$, $y = b^na^pb^p$, $j + k' + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j = p$ and $k' + l + l' + n = p$.
$uv^iwx^iy = a^{p+1}b^pa^{p_1}b^{p_2}a^pb^p \notin Q_{\overline{D}}$ for $i = 0$ as $p_1(= m) < p$ and $p_2\ (= l + n) \leq p$.

c(9) $u = a^{p+1}b^pa^pb^pa^m$, $v = a^jb^{k'}$, $w = b^l$, $x = b^{l'}$, $y = b^n$, $j + k' + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j = p$ and $k' + l + l' + n = p$.
$uv^iwx^iy = a^{p+1}b^pa^pb^pa^{p_1}b^{p_2} \notin Q_{\overline{D}}$ for $i = 0$ as $p_1(= m) < p$ and $p_2\ (= l + n) \leq p$.

Case (d) Next we discuss for the division of $s = uvwxy$ such that $vwx = b^ja^k$, $j, k \geq 1$.

d(1) $u = a^{p+1}b^m$, $v = b^j$, $w = b^{k'}$, $x = b^la^{l'}$, $y = a^nb^pa^pb^p$, $j + l + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j + k' + l = p$ and $l' + n = p$.
$uv^iwx^iy = a^{p+1}b^{p_1}a^{p_2}b^pa^pb^p \notin Q_{\overline{D}}$ for $i = 0$ as $p_1\ (= m + k') \leq p$ and $p_2(= n) < p$

d(2) $u = a^{p+1}b^pa^pb^m$, $v = b^j$, $w = b^{k'}$, $x = b^la^{l'}$, $y = a^nb^p$, $j + l + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j + k' + l = p$ and $l' + n = p$. Similar to case d(1).

d(3) $u = a^{p+1}b^m$, $v = b^j$, $w = b^{k'}a^l$, $x = a^{l'}$, $y = a^n b^p a^p b^p$, $j + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j + k' = p$ and $l + l' + n = p$.
$uv^iwx^iy = a^{p+1}b^{p_1}a^{p_2}b^pa^pb^p$ for $i = 0$, where $p_1 = m + k'$ and $p_2 = l + n$. Since $j + l' \geq 1$ therefore either ($p_1 < p$ and $p_2 \leq p$ or ($p_1 \leq p$ and $p_2 < p$). In both the cases $uv^iwx^iy = a^{p_1}b^{p_2}a^pb^pa^pb^p \notin Q_{\overline{D}}$.

d(4) $u = a^{p+1}b^pa^pb^m$, $v = b^j$, $w = b^{k'}a^l$, $x = a^{l'}$, $y = a^nb^p$, $j + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j + k' = p$ and $l + l' + n = p$. Similar to case d(3).

d(5) $u = a^{p+1}b^m$, $v = b^ja^{k'}$, $w = a^l$, $x = a^{l'}$, $y = a^nb^pa^pb^p$, $j + k' + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j = p$ and $k' + l + l' + n = p$
$uv^iwx^iy = a^{p+1}b^{p_1}a^{p_2}b^pa^pb^p \notin Q_{\overline{D}}$ for $i = 0$ as $p_1\ (= m) < p$ and $p_2 = (l + n) \leq p$.

d(6) $u = a^{p+1}b^pa^pb^m$, $v = b^ja^{k'}$, $w = a^l$, $x = a^{l'}$, $y = a^nb^p$, $j + l' \geq 1$, $j + k' + l + l' \leq p$, $m + j = p$ and $k' + l + l' + n = p$. Similar to case d(5).

In any of the above cases Pumping Lemma does not hold, therefore the assumption that $Q_{\overline{D}}$ is context-free must be false. □

3.2.1 Recognizing Del-Robust Primitive Words

In this section we give a linear time algorithm to recognize a del-robust primitive word. An existing algorithm to test whether a given word is primitive, is based on the idea that a word w is primitive if and only if w is not an internal factor of its square ww, that is, $ww = xwy$ implies that either $x = \lambda$ or $y = \lambda$ [1].

Observe that if a word $w \in Q_{\overline{D}}$, then by Corollary 3.5 there exists a cyclic permutation of w which contains a non-primitive factor of length $|w| - 1$. We make use of this fact in the following theorem by observing that the word ww consists of all the cyclic permutation of a word w.

Theorem 3.5. *Let u be a primitive word. Then u is a non-del-robust primitive word if and only if the word uu contains at least one non-primitive word of length $|u| - 1$.*

Proof. We prove the sufficient and necessary conditions below.

(⇒) Let u be a non-del-robust primitive word. Thus u can be written as $t^rt_1at_2t^s$ for some primitive word $t = t_1t_2$ where $t_1, t_2 \in V^*$, $a \in V$ and $r + s \geq 1$. Thus, the word $uu = t^rt_1at_2t^st^rt_1at_2t^s$ contains a factor $t_2t^st^rt_1$ of length $|u| - 1$ which is equal to the non-primitive word $(t_2t_1)^{r+s+1}$.

(⇐) Let the word uu have a non-primitive factor of length $|u| - 1$ where u is a primitive word. Then $uu = t_1 p^r t_2$, where $t_1, t_2 \in V^*$, $|p^r| = |u| - 1$, $p \in Q$ and $r \geq 2$. Here we have two cases to consider viz. either the word p^r is entirely contained in the word u or some segment of the word p^r is contained in the word u.

Case A Let p^r be entirely in u. Then u is not del-robust as u can be either ap^r or $p^r a$ for some $a \in V$.

Case B Let some portion of p^r be contained in u. Since $uu = t_1 p^r t_2$ and Z is reflective (by Lemma 2.9), we have $t_2 t_1 p^r = u'u'$, where u' is cyclic permutation of u. Here p^r is entirely in u' which is the Case (A). Therefore $u' = ap^r$, that is, u' is a non-del-robust word. Thus u, which is nothing but a cyclic permutation of u', is also a non-del-robust word.

□

Recall the definition of maximal repetitions given in Section 2.1. Let u be primitive word. The following lemma claims that there are some maximal repetitions with specific periods in the word uu whose lengths are at least $|u| - 1$ if $u \in Q_{\overline{D}}$.

Corollary 3.6. *Let u be a primitive word. If the word uu contains a maximal repetition of length at least $|u| - 1$ with a period p where p divides $|u| - 1$ and $p < |u| - 1$ then u is a non-del-robust primitive word.*

Proof. Let a maximal repetition $v^k v_1$ be a factor of uu for $v \in Q$, $k \geq 2$ and v_1 be prefix of v. Since $|v|$ divides $|u| - 1$, we have $|u| - 1 = r|v|$ for some $r \geq 2$ and $r \leq k$, that is, uu contains v^r. Hence by the Theorem 3.5, u is a non-del-robust primitive word. □

The computation of maximal repetitions in a word can be done in linear time in terms of the length of the input word [17]. Next we present a linear time algorithm to test del-robustness of a primitive word by using the algorithm, FINDMAXIMALREPETITIONS, that finds maximal repetitions and testing primitivity in linear time for a given word.

Theorem 3.6 (Correctness of the ISDELROBUST Algorithm)**.** *Let u be a word. The Algorithm 1 returns True if and only if u is a del-robust primitive word.*

Proof. The correctness of the algorithm follows from the Corollary 3.6 which is used in the step 8 of the algorithm. □

Theorem 3.7 (Complexity of the ISDELROBUST Algorithm)**.** *The time complexity of the Algorithm 1 for an input word with length n is $O(n)$.*

Algorithm 1 DEL-ROBUST PRIMITIVE WORD

Input: A finite word u

Output: "True" if u is a del-robust primitive word, else "False"

```
1: procedure IsDelRobust
2:     Let v ← uu.
3:     S ← FindMaximalRepetitions(v)          ▷ S is a set of pairs of period and length.
4:     for all (p_i, l_i) ∈ S do
5:         if |u| mod p_i = 0 and p_i < |u| then                   ▷ Testing primitivity.
6:             return False                               ▷ The word u is not primitive.
7:         end if
8:         if p_i < |u| − 1 and (|u| − 1) mod p_i = 0 and l_i ≥ |u| − 1 then
9:             return False (Corollary 3.6)              ▷ The word u is not del-robust.
10:        end if
11:        return True                                       ▷ The word u is del-robust.
12:    end for
13: end procedure
```

Proof. In step (3), the set of pairs of periods and corresponding lengths of maximal repetitions of uu can also be computed in linear time [17]. The number of pairs returned in step (3) is bounded by $O(n)$. Thus, step (4) - (9) takes linear time to find those periods which are mentioned in Corollary 3.6. Therefore the total time taken by the algorithm to test del-robustness of a word of length n is $O(n)$. □

3.2.2 Counting Del-Robust Primitive Words

In this section we give a lower bound on number of n-length del-robust primitive words. Let V be an alphabet and $Z(n) = V^n \setminus Q(n)$ be the set of n-length non-primitive words. Given a word $w \in Z(n-1)$ and a symbol $a \in V$, the number of the words that are obtained by inserting a in w is equal to

$$|\{w_1 a w_2 \mid w = w_1 w_2, w_1, w_2 \in V^*\}| = n - |w|_a.$$

(For example if $w = w_1.a.a.w_2$ then insertion of a immediately before aa or in between aa or after aa gives the same word, that is, insertion at two positions is not required which is same as $|aa|$. Similarly, we can prove it for $w = w_1 a w_2 a w_2$.)

Now for a given word w the number of all words that can be obtained by inserting any one symbol from V is given by

$$
\begin{aligned}
& |\{w_1 a w_2 \mid w = w_1 w_2,\ w_1, w_2 \in V^*, a \in V\}| \\
= \quad & \sum_{a \in V} (n - |w|_a) = n|V| - \sum_{a \in V} |w|_a = n|V| - (n-1) = n|V| - n + 1
\end{aligned}
$$

We know from Lemma 2.10 that a non-primitive word w either remains non-primitive after inserting a symbol a if $w = a^{n-1}$ or become non-del-robust primitive word. Therefore, the number of non-del-robust primitive words of length n, $Q_{\overline{D}}(n)$, is the difference between the number of all words obtained by inserting a symbol in the words from set $Z(n-1)$ and the number of elements in set $\{a^n \mid a \in V\}$. We can find an upper bound on number of non-del-robust primitive words of length n as follows.

$$
\begin{aligned}
|Q_{\overline{D}}(n)| \quad & = |\{w_1 a w_2 \mid w = w_1 w_2 \in Z(n-1), w_1, w_2 \in V^*, a \in V\}| - |V| \\
& \leq \sum_{a \in V} \sum_{w \in Z(n-1)} |\{w_1 a w_2 \mid w = w_1 w_2\}| - |V| \\
& \leq \sum_{w \in Z_{n-1}} (n|V| - n + 1) - |V| \\
& \leq (n|V| - n + 1)|Z(n-1)| - |V|
\end{aligned}
$$

From the Proposition 3.6, we have the number of primitive words of length n that is $|Q(n)|$. Since $Z(n) = V^n \setminus Q(n)$ we have $|Z(n)| = |V^n| - |Q(n)|$ and number of del-robust-primitive words of length n over alphabet V is $|Q_D(n)| = |Q(n)| - |Q_{\overline{D}}(n)|$.

3.3 Ins-Robust Primitive Words

Definition 3.4 (Ins-Robust Primitive Word). *A primitive word w of length n is said to be ins-robust primitive word if the word*

$$
pref(w, i) \ . \ a \ . \ suf(w, n - i)
$$

is a primitive word for all $i \in \{0, 1, \ldots, n\}$ where $a \in V$.

There are infinitely many primitive words which are ins-robust. For example, the words $a^n b^n c^n$ for $n \geq 1$ are ins-robust primitive words. We denote the set of all ins-robust primitive words over an alphabet V by Q_I. Clearly the language of ins-robust primitive words is a subset of the set of primitive words, that is, $Q_I \subset Q$.

Following theorem is a reformulation of the definition of ins-robust primitive words.

Theorem 3.8. *A primitive word w is not ins-robust if and only if w can be expressed in the form of $u^r u_1 u_2 u^s$ where $u = u_1 c u_2 \in Q$, u_1, $u_2 \in V^*$, for some $c \in V$, r, $s \geq 0$ and $r + s \geq 1$.*

Proof. We prove the sufficient and necessary conditions below.

($\Leftarrow$) This part is straightforward. Let us consider a word $w = u^{k_1} u_1 u_2 u^{k_2}$ where $u_1 c u_2 = u$ for some $c \in V$. The word w is primitive by Lemma 2.10(b). Now insertion of the letter c in w (between u_1 and u_2) gives the exact power of u which become a non-primitive word. Hence, w is not an ins-robust primitive word.

($\Rightarrow$) Let w be a primitive word but not ins-robust. Then there exists a decomposition $w = w_1 w_2$ such that $w_1 c w_2$ is not a primitive word for some letter $c \in V$. That is, $w_1 c w_2 = p^n$ for some $p \in Q$ and $n \geq 2$. Therefore $w_1 = p^r p_1$ and $w_2 = p_2 p^s$ for $r, s \geq 0$ *and* $r + s \geq 1$ such that $p_1 c p_2 = p$. Hence $w = p^r p_1 p_2 p^s$.

$\square$

Definition 3.5 (Non-Ins-Robust Primitive Words)**.** *A primitive word w is said to be non-ins-robust if $w \in Q$ but $w \notin Q_I$. We denote the set of all non-ins-robust primitive words as $Q_{\overline{I}}$. So, $Q_{\overline{I}} = Q \setminus Q_I$, where '$\setminus$' is the set difference operator.*

The next theorem is about an equation in words and identifies a sufficient condition under which three words are power of a common word.

Theorem 3.9 ([11])**.** *If $u^m v^n = w^k \neq \lambda$ for words u, v, $w \in V^*$ and natural numbers m, n, $k \geq 2$, then u, v and w are powers of a common word.*

The following lemma is a consequence of the Theorem 3.9 which states that a word obtained by concatenating powers of two distinct primitive words is also primitive.

Lemma 3.5 ([33])**.** *If $p, q \in Q$ with $p \neq q$ then $p^i q^j \in Q$ for all i, $j \geq 2$.*

Proposition 3.5. *If $u, v \in Q$, $u^m = u_1 u_2$ and $v = u_1 c u_2$ for some $c \in V$ then $u^m v^n \in Q_{\overline{I}}$ for $m, n \geq 2$.*

Proof. From Lemma 2.10(b), we know that at least one of $u_1 u_2$ and $u_1 c u_2$ is primitive. Since $u^m = u_1 u_2$ for $m \geq 2$ and $v = u_1 c u_2$, therefore v is primitive and so is $u^m v^n = u_1 u_2 (u_1 c u_2)^n$. After insertion of the letter c we will get $(u_1 c u_2)^{n+1}$ which is not a primitive word. However, by Lemma 3.5, $u^m v^n$ is a primitive word for $m, n \geq 2$. Hence it is not a ins-robust word, that is, $u^m v^n \in Q_{\overline{I}}$. □

As mentioned earlier, if a word w is primitive then $rev(w)$ is also primitive. We prove this for ins-robust primitive word.

Lemma 3.6. *If $w \in Q_I$ then $rev(w) \in Q_I$.*

Proof. Assume that for a word $w \in Q_I$, $rev(w)$ is not a ins-robust primitive word. i.e. $rev(w) = p^r p_1 p_2 p^s$ where $p = p_1 c p_2 \in Q$ for some $c \in V$. Then the word $w = rev(rev(w)) = rev(p^r p_1 p_2 p^s) = (rev(p))^s \, rev(p_2) \; rev(p_1) \, (rev(p))^r$ and $p = p_1 c p_2$, $rev(p) = rev(p_2) \, c \, rev(p_1)$. By Theorem 3.8, w is not a ins-robust primitive word, which is a contradiction. Therefore, if $w \in Q_I$ then $rev(w) \in Q_I$. □

Next, we show that the language of ins-robust primitive words, Q_I, is reflective.

Theorem 3.10. *Q_I is reflective.*

Proof. Let there be a word $w = xy \in Q_I$ such that $yx \notin Q_I$. Since $w \in Q_I$, hence $w \in Q$. By Lemma 2.9, we know that Q is reflective. Therefore $yx \in Q$ and so $yx \in Q \setminus Q_I$, i.e. $yx \in Q_{\overline{I}}$. Using Theorem 3.8, we have $yx = u^r u_1 u_2 u^s$ where $u = u_1 c u_2 \in V^*$ for some $c \in V$ and $r + s \geq 1$. There are three possibilities which are as follows.

Case A If $y = u^{r_1} u'$, $x = u'' u^{r_2} u_1 u_2 u^s$ where $u = u' u''$ and $r_1 + r_2 + 1 = r$.

In this case $xy = u'' u^{r_2} u_1 u_2 u^s u^{r_1} u' = (u''u')^{r_2} u'' u_1 u_2 u' (u''u')^{s+r_1}$.

Since $u = u_1 c u_2$, therefore $u'' u_1 c u_2 u' = u'' u u' = (u''u')^2$.

Therefore $(u''u')^{r_2} u'' u_1 c u_2 u' (u''u')^{s+r_1} = (u''u')^{s+r+1}$, that is, $xy \in Q_{\overline{I}}$, which is a contradiction.

Case B $y = u^r u'$, $x = u'' u^s$ where $u'u'' = u_1 u_2$

Case B.1 If $u' = u'_1$ and $u'' = u''_1 u_2$ where $u'_1 u''_1 = u_1$.

Since $u = u_1 c u_2 = u'_1 u''_1 c u_2$.

In this case $xy = u'' u^s u^r u' = u''_1 u_2 u^s u^r u'_1$.

Now $u''_1 c u_2 u^s u^r u'_1 = (u''_1 c u_2 u'_1)^{r+s+1}$.

Therefore $xy \in Q_{\overline{I}}$, a contradiction.

Case B.2 If $u' = u_1 u'_2$ and $u'' = u''_2$ where $u'_2 u''_2 = u_2$. This is similar to Case B.1.

Case C If $y = u^r u_1 u_2 u^{s_1} u'$, $x = u'' u^{s_2}$ where $u = u'u''$. This case is similar to the Case A.

Hence Q_I is reflective. $\square$

Corollary 3.7. *$Q_{\overline{I}}$ is reflective.*

Proof. We prove it by contradiction. Let there be a word $w = xy \in Q_{\overline{I}}$ such that $yx \notin Q_{\overline{I}}$. We have $xy \in Q$ and Q is reflective, so $yx \in Q$ by Lemma 2.9. Therefore $yx \in Q \setminus Q_{\overline{I}}$, i.e. $yx \in Q_I$. But Q_I is reflective by Theorem 3.10, we have $xy \in Q_I$, which is a contradiction. Hence $yx \in Q_{\overline{I}}$. $\square$

Theorem 3.11. *A word w is in the set $Q_{\overline{I}}$ if and only if it is of the form $u^n u'$ or its cyclic permutation for some $u \in Q$, $u = u'a$, $a \in V$ and $n \geq 1$.*

Proof. We prove the sufficient and necessary conditions below.

(⇒) Let $w \in Q_{\overline{I}}$, then w can be written as $w = u^r u_1 u_2 u^s$ for some u $(= u_1 a u_2) \in Q$ and $a \in V$. Since $Q_{\overline{I}}$ is reflective, therefore $u_2 u^s u^r u_1 = (u_2 u_1 a)^{r+s} u_2 u_1$ is also in $Q_{\overline{I}}$.

(⇐) If a word w is a cyclic permutation of $u^n u'$ for $n \geq 1$ and $u = u'a$ then after insertion of a symbol a, it gives a cyclic permutation of u^{n+1} which is non-primitive (since Z is reflective). Therefore, $w \in Q_{\overline{I}}$.

$\square$

We observe that a word w is periodic with minimum period p (≥ 2) divides $|w| + 1$ and $p \leq |w|$ then w is non-ins-robust primitive word. Since Q_I is reflective, therefore any cyclic permutation of w is also non-ins-robust primitive word. We know by Theorem 2.7 that cyclic permutation of a primitive word is also primitive , so the cyclic permutation of an ins-robust primitive word is primitive. In next result, we show that it remains ins-robust too.

Corollary 3.8. *Cyclic permutation of a ins-robust primitive word is ins-robust.*

Proof. Let $w \in Q_I$. Then cyclic permutation of w will be yx for some partition $w = xy$. Since Q_I is reflective. Therefore yx is also ins-robust primitive word. This proves that any cyclic permutation of an ins-robust primitive word is ins-robust. $\square$

3.3.1 Ins-Robust Primitive Words and Density

It is easy to see that Q is right dense [41]. In the following theorem we discuss the denseness of language of non-ins-robust primitive words $Q_{\overline{I}}$.

Theorem 3.12. *Let $w \in V^*$ be a word. If $|w| = n$ and $wa^n \in Q_{\overline{I}}$ where $w \notin a^*$ and $n \geq 1$, then there exists words $u, u_1, u_2 \in V^*$ such that $wa^n = u^2 u_1 u_2$ and $u = u_1 b u_2$ for some $b \neq a$.*

Proof. Let $wa^n \in Q_{\overline{I}}$.

If $wa^n = u^r u_1 u_2 u^s$, where $u = u_1 b u_2$, for some $b \in V$. We claim that $r = 2$ *and* $s = 0$.

Case A. First we prove that $s \ngeq 1$.
If $s \geq 1$ then $|u| \leq n+1$. We have two cases depending on the length of u.

Case A(i). In this case we prove that $|u| \leq n$. On contrary if $|u| = n+1$ that the possibility can be $wa^n = u_1 u_2 u$. But then $u = ba^n$ for some $b \neq a$. Since $|u| = n+1$ and so $|u_1 u_2| = n$. $|wa^n| = 2n+1$, which is a contradiction as $|w| = n$ and so $|wa^n| = 2n$.

Case A(ii). If $s \geq 1$ and $|u| \leq n$, then $u = a^r$, where $r = |u|$, and therefore $u_1 u_2 = a^{r-1}$. $wa^n = a^{2n} \notin Q_{\overline{I}}$ which leads to a contradiction.

Therefore $s = 0$. Hence $wa^n = u^r u_1 u_2$. Next we prove that $r = 2$ is only possibility.

Case B. $r = 1$. This case is not possible. Because in this case $wa^n = uu_1u_2$, $|wa^n| = 2n$ which implies $|u| = (2n+1)/2$ a non-integral value.

Case C. If $r \geq 2$. In this case we prove that $r \geq 3$ is not possible.

Let $r \geq 3$ then $|wa^n| = |u^r u_1 u_2| = ((r+1)|u| - 1)$. In this case $|u| = \frac{2n+1}{r+1} * 2 \leq n$. Therefore $u = a^{k+1}$ but then $wa^n \notin Q_{\overline{I}}$. Hence $r \geq 3$ is also not possible.

Thus the only possibility is $r = 2$, $|wa^n| = u^2 u_1 u_2$. Since $u_1 u_2 = a^k$ and $wa^n \in Q_{\overline{I}}$ therefore $w, u \notin a^*$, and so $u = a^{k_1} b a^{k_2}$ where $k_1 + k_2 = k$, $b \neq a$ and $k_2 \geq k_1 + 2$. ☐

Lemma 3.7. *Let V be an alphabet, $w \in V^*$, $|w| = n$ and $a \in V$. If $wa^n \in Q_{\overline{I}}$ then for $b \neq a$, $wb^n \in Q_I$.*

Proof. Let $wa^n \in Q_{\overline{I}}$. Then, by Theorem 3.12 we have,

$wa^n = u^2 u_1 u_2 \;\; and \;\; u_1 u_2 = a^k.$

$u = a^{k_1} b a^{k_2} \;\; where \;\; a \neq b.$

Let wc^n be also in $Q_{\overline{I}}$ for some $c \neq a$. Then, by Theorem 3.12 we have,

$wc^n = v^2 v_1 v_2 \;\; and \;\; v_1 v_2 = c^k.$

$v = c^{k'_1} d c^{k'_2} \;\; where \;\; c \neq d.$

But since $|u| = |v|$ and $w = uu' = vv'$ where $u' = uu_1u_2$ and $v' = vv_1v_2$, therefore $u = v$, that is, $a^{k_1} b a^{k_2} = c^{k'_1} d c^{k'_2}$.
If $k_1 < k'_1$ then $a = b = c = d$, which is a contradiction. Alternatively, if $k_1 = k'_1$ then $a = c$ which is again a contradiction. Therefore $wc^n \in Q_I$. ☐

Theorem 3.13. *The language Q_I is dense over the alphabet V.*

Proof. Consider a word w. We only need to consider the case when $w \notin Q_I$, that is, $w \in V^* \setminus Q_I$. By Lemma 3.7, there exists $b \in V$ such that $wb^n \in Q_I$, where $n = |w|$. Hence Q_I is dense over V. □

3.3.2 Relation of Q_I with Other Formal Languages

We now investigate the relation between the language of ins-robust primitive words with the traditional languages in Chomsky hierarchy. We prove that the language of ins-robust primitive words over an alphabet is not regular and also show that the language of non-ins-robust primitive words is not context-free. For completeness, we recall the pumping lemma for regular languages and pumpimg lemma for context-free languages which will be used to show that Q_I is not regular and $Q_{\overline{I}}$ is not context-free respectively.

Let us recall a result which will be used in proving that the language of ins-robust primitive words is not regular.

Lemma 3.8 ([42]). *For any fixed integer k, there exist a positive integer m such that the system of equations $(k-j)x_j + j = m$, $j = 0, 1, 2, \ldots, k-1$ has a nontrivial solution with appropriate positive integers $x_1, x_2, \ldots, x_j > 1$.*

Theorem 3.14. *Q_I is not regular.*

Proof. Let us suppose that the language of ins-robust primitive words Q_I is regular. Then there exist a natural number $n > 0$ depending upon the number of states of finite automaton for Q_I.

Consider the word $w = a^n ba^m b, m > n+1$ and $m \neq 2n$. Note that w is an ins-robust primitive word over V, where $|V| \geq 2$ and $a \neq b$. Since $w \in Q_I$ and $|w| \geq n$, then it must satisfy the other conditions of pumping Lemma for regular languages. So there exist a decomposition of w into x, y and z such that $w = xyz$, $|y| > 0$ and $xy^iz \in Q_I$ for all $i \geq 0$.

Let $x = a^k$, $y = a^{(n-j)}$, $z = a^{j-k}ba^mb$. Now choose $i = x_j$ and since we know by Lemma 3.8 that for every $j \in \{0, 1, \ldots, n-1\}$, there exists a positive integer $x_j > 1$ such that $xy^{x_j}z = a^k a^{(n-j)x_j} a^{j-k} ba^m b = a^{(n-j)x_j+j} ba^m b = a^m ba^m b = (a^m b)^2 \notin Q_I$ which is a contradiction. Hence the language of ins-robust primitive words Q_I is not regular.

□

Theorem 3.15. *$Q_{\overline{I}}$ is not a context-free language for a binary alphabet.*

Proof. Let $V = \{a,\ b\}$ be an alphabet. By contradiction, let us assume that $Q_{\overline{I}}$ is a context-free language. Let $p > 0$ be an integer which is the pumping length for the language $Q_{\overline{I}}$.

Consider the string $s = a^{p+1}b^{p+1}a^{p+1}b^p$, where $a,\ b \in V$ are distinct. It is easy to see that $s \in Q_{\overline{I}}$ and $|s| \geq p$.

Hence, by the Pumping Lemma 2.6, s can be written as $uvwxy$, where $u, v, w, x,$ and y are factors of s, such that $|vwx| \leq p$, $|vx| \geq 1$, and $uv^iwx^iy \in Q_{\overline{I}}$, $\forall i \geq 0$. By the choice of s and the fact that $|vwx| \leq p$, we have one of the following possibilities for vwx:

(a) $vwx = a^j$ for some $1 \leq j \leq p$.

(b) $vwx = a^jb^k$ for some j and k with $j + k \leq p$ and $j, k \geq 1$.

(c) $vwx = b^j$ for some $1 \leq j \leq p$.

(d) $vwx = b^ja^k$ for some $j, k \geq 1$ with $j + k \leq p$.

In Case (a), since $vwx = a^j$, therefore $vx = a^t$ for some $t \geq 1$ and hence $uv^iwx^iy = a^{p-t+1}b^{p+1}a^{p+1}b^p \notin Q_{\overline{I}}$ for $i = 0$.

Case (b) can have several subcases.

(i) $v = a^{j_1},\ w = a^{j_2},\ x = a^{j_3}b^k$ where $j_1 + j_2 + j_3 + k \leq p$ and $j_1 + j_3 + k \geq 1$.

If vwx is in the prefix substring string $a^{p+1}b^{p+1}$, then

$uv^4wx^4y = a^{p+1-j_1-j_2-j_3}\ a^{4j_1}a^{j_2}a^{j_3}b^ka^{j_3}b^ka^{j_3}b^k\ a^{j_3}b^kb^{p+1-k}\ a^{p+1}b^p = a^{p+1}a^{3j_1}b^ka^{j_3}b^k$ $a^{j_3}b^ka^{j_3}b^{p+1}a^{p+1}b^p \notin Q_{\overline{I}}$ for $i = 4$ as $0 \leq j_1, j_3, k \leq p - 1$ and $k \geq 1$ so insertion of a or b at any place can not make it non-primitive.

Similarly, we can show for the occurrence in suffix substring $a^{p+1}b^p$.

(ii) $v = a^{j_1},\ w = a^{j_2}b^{k_1},\ x = b^{k_2}$ where $j_1 + k_2 \geq 1$, $j_1 + j_2 + k_1 + k_2 \leq p$ and $j_1, j_2, k_1, k_2 \geq 0$.

$uv^4wx^4y = a^{p+1+3j_1}b^{p+1+3k_2}\ a^{p+1}b^p \notin Q_{\overline{I}}$ for $i = 4$ because atleast j_1 or k_2 must be greater than or equal to 1 and less than or equal to p.

(iii) $v = a^jb^{k_1},\ w = b^{k_2},\ x = b^{k_3}$.

Case (b) (iii) is similar to case b(i).

Case (c) is similar to case (a) and case (d) is similar to case (b). Therefore, our initial assumption that $Q_{\overline{I}}$ is context-free, must be false. □

Next we prove that the language of non-ins-robust primitive words is not context-free in general.

Lemma 3.9. *The language $Q_{\overline{I}}$ is not context-free over an alphabet V where V has at least two distinct letters.*

Proof. The proof of Theorem 3.15 can be generalized to arbitrary alphabet V having at least two letters. The set of all words over alphabet having greater than two distinct letters also contains the words with two letters. If $Q_{\overline{I}}$ is assumed to be a CFL over V where $|V| \geq 3$, then we can choose words of the form used in Theorem 3.15 and obtain a contradiction. Hence the language of non-ins-robust primitive words $Q_{\overline{I}}$ is not context-free over V where $|V| \geq 2$. □

3.3.3 Counting Ins-Robust Primitive Words

In this section we give a lower bound on number of n-length ins-robust primitive words. Let V be an alphabet and $Z(k) = V^k \setminus Q$ be the set of n-length non-primitive words.

We have the following result that gives the number of the primitive words of length m.

Proposition 3.6 ([43]). *Let $m \in N$ and $m = {m_1}^{r_1}{m_2}^{r_2}\ldots{m_t}^{r_t}$ be the factorization of m, where all $m_i, 1 \leq i \leq t$, are prime and $m_i \neq m_j$ for $i \neq j$, then the number of primitive words of length m is equal to*

$$|V|^m - \sum_{1 \leq i \leq t} |V|^{\frac{m}{m_i}} + \sum_{1 \leq i \leq j \leq t} |V|^{\frac{m}{m_i m_j}}$$

$$- \sum_{1 \leq i \leq j \leq k \leq t} |V|^{\frac{m}{m_i m_j m_k}} + \cdots + (-1)^{t-1}|V|^{\frac{m}{m_1 m_2 \cdots m_t}}$$

We observe that the deletion of a symbol from a n-length non-primitive word gives a maximum of $(n-1)$-different non-ins-robust primitive words when the word is of type $a_1a_2\ldots a_n$ such that $a_i \neq a_{i+1}$ for $1 \leq i \leq n-1$ and minimum it can be $zero$ if the word is of type $a^r,\ r > 2,\ a \in V$. Given a word $w \in Z(n)$. The number of words that can be obtained by deleting a symbol from w is

$$0 \leq |\{w_1w_2 \mid w_1aw_2 = w,\ w_1, w_2 \in V^*, a \in V\}| \leq n.$$

We know from Lemma 2.10 that a non-primitive word w remains non-primitive after deleting a symbol a if $w = a^n$ and $n \geq 3$. Otherwise a non-ins-robust primitive word.

$$Q_{\overline{I}}(n) = \{w_1w_2 \mid w_1aw_2 \in Z_{n+1}, a \in V, w_1, w_2 \in V^*\}$$

Therefore, the number of non-ins-robust primitive words of length n, $Q_{\overline{I}}(n)$, is the difference between the number of all words obtained by deleting a symbol from the words of set

$$Z(n+1) \setminus V^{n+1} \ where\ V^{n+1} = \{a^{n+1} \mid a \in V\}\ for\ n \geq 2$$

We can find an upper bound on number of non-ins-robust primitive words of length $n \geq 2$ as follows.

$$|Q_{\overline{I}}(n)| = |\{w_1w_2 \mid w = w_1bw_2 \in Z(n+1) \setminus V^{n+1}, w_1, w_2 \in V^*, b \in V\}|$$

$$|Q_{\overline{I}}(n)| = |\{w_1w_2 \mid w = w_1aw_2 \in Z(n+1), w_1, w_2 \in V^*, a \in V\}| - (n+1).|V|$$

$$\leq (n+1).(|Z(n+1|) - |V|).$$

From the Proposition 3.6, we know the number of primitive words of fixed length. Thus the number of ins-robust-primitive words of length n, $Q_I(n)$, over an alphabet V is equal to $|Q_n| - |Q_{\overline{I}}(n)|$.

3.3.4 Recognizing Ins-Robust Primitive Words

In this section, we give a linear time algorithm to determine if a given primitive word w is ins-robust. We design the algorithm that exploits the property of the structure of ins-robust primitive words. We state some simple observations before presenting the algorithm. The following theorem is based on the structure of ins-robust primitive word.

Theorem 3.16. *Let u be a primitive word. Then u will be **non-ins-robust** primitive word iff uu contains at least one periodic word of length $|u|$ with period p such that p divides of length $|u|+1$ and $p \leq |u|$.*

Proof. ($\Rightarrow$) If u is a non-ins-robust primitive word, then u can be written as $t^rt_1t_2t^s$ for some primitive word t, $r+s \geq 1$ and $t = t_1at_2$ for some symbol $a \in V$ where $t_1, t_2 \in V^*$. $uu = t^rt_1t_2t^st^rt_1t_2t^s$, This word contains a subword $t_2t^st^rt_1$ of length $|u|$ that is $(t_2t_1a)^{r+s}t_2t_1$ which is a periodic word with period $|t_2t_1a| = |t|$ which divides $|u|+1$.
($\Leftarrow$) Let uu has a periodic substring of length $|u|$ with period p ($p/|u|+1$ and $p \leq |u|$) where u is primitive word. Then $uu = t_1x^rx't_2$, where $t_1, t_2 \in V^*$, $|x^rx'| = |u|$, $x \in Q$, $r \geq 1$ and $x = x'a$ for some $a \in V$. $|t_1t_2| = |u|$. Here we have two cases, either x^rx' entirely contained in u or some portion of x^rx' contained in u.
Case A. Let x^rx' entirely in u. Then u is not ins-robust as $u = x^rx'$.
Case B. Let some portion of x^rx' contained in u. Since $uu = t_1x^rx't_2$, and Z is reflective, therefore $t_2t_1x^rx' = u'u'$, where u' is cyclic permutation of u. Hence, $u' = x^rx'$, is non-ins-robust. Since Q_I is reflective, therefore u is also non-ins-robust.

□

Corollary 3.9. *Let u be a primitive word. Then u will be **non-ins-robust** primitive word if* *and only if there exists a cyclic permutation of u, say u', which is a periodic with period p such that p divides $|u|+1$ and $p \leq |u|$.*

Proof. The proof follows from Theorem 3.16.

□

Next we present a linear time algorithm to test ins-robustness of a primitive word by using the existing algorithm for finding maximal repetitions in linear time. For more details on the maximal repetition, see section 4 [17].

Algorithm 2 INS-ROBUST PRIMITIVE WORD

Input: A finite word u

Output: "True" if u is a ins-robust primitive word, else "False"

1: **procedure** ISINSROBUST
2:     Let $v \leftarrow uu$.
3:     $S \leftarrow$ FindMaximalRepetitions(v) ▷ S is a set of pairs of period and length.
4:     **for all** $(p_i, l_i) \in S$ **do**
5:         **if** $|u| \mod p_i = 0$ **and** $p_i < |u|$ **then** ▷ Testing primitivity.
6:             Return False ▷ The word u is not primitive.
7:         **end if**
8:         **if thenp**$_i < |u|$ **and** $(|u|+1) \mod p_i = 0$ **and** $l_i \geq |u|$
9:             Return False (Corollary 3.6) ▷ The word u is not ins-robust.
10:         **end if**
11:         Return True ▷ The word u is ins-robust.
12:     **end for**
13: **end procedure**

Theorem 3.17. *Let w be a word given as input to Algorithm 2. The algorithm returns true if and only if the word w is ins-robust.*

Proof. In step (3), the algorithm finds the maximal repetitions with their periods. Since $Q_{\overline{I}}$ is closed under reflectivity, therefore uu has all the cyclic permutations of u. There is a periodic word $x^r x'$, a permutation of u such that $x = x'a$ for some $a \in V$. Therefore uu also has this periodic word which is proved in Theorem 3.16. That is for a non-ins-robust primitive word u, uu contains a periodic word of length at least $|u|$ with a period p such that p divides $(|u|+1)$ and $p < |u|$. This is explained in Step (8) where u is a primitive word Step (6). Otherwise u is ins-robust primitive word. □

Theorem 3.18. *The property of being ins-robust primitive is testable on a word of length n in $O(n)$ time.*

Proof. The Step (2) in Algorithm 2 has $O(1)$ running time. In Step (3) maximal repetition algorithm is computed using algorithm given in section 4 [17] is used which has linear time complexity. Now from Step (4) to Step (9), the complexity depends on the cardinality of S, which is less than n. Hence it also has linear time complexity. Therefore by Theorem 3.16 testing ins-robustness for primitive word can be done in linear time. □

3.4 Exchange-Robust Primitive Words

We consider a new formal language class known as exchange-robust primitive words in which exchanging two different consecutive symbols in a primitive word preserve primitivity.

Definition 3.6 (Exchange-Robust Primitive Words)**.** *A primitive word $w = a_1a_2 \cdots a_{i+1}a_{i+2} \cdots a_n$ of length n is said to be exchange-robust if and only if*

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

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

Observe that if a primitive word is exchange robust then it must remain primitive on exchange of any two consecutive symbols. We denote by Q_X the set of all primitive words which are exchange-robust over an alphabet V. Clearly, the set of all exchange-robust primitive words is a subset of Q. There are infinitely many primitive words which are exchange-robust. For example, $a^nb^{2n}a^n,\ n \geq 2$ is exchange-robust. We exchange two consecutive symbols a and b if $a,\ b \in V$ and $a \neq b$.

Our next result is concerned about the exchange of two different symbols at consecutive places in a nonprimitive word. We prove that the new word which we obtain by exchanging two different and consecutive symbols at any position in a nonprimitive word results in a primitive word.

Lemma 3.10. *Let w be a word with $|alph(w)| \geq 2$. If $w = x_1abx_2 \in Z$ then $x_1bax_2 \in Q$.*

Proof. We prove it by contradiction. Since $w \in Z$, 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 $m_1 + m_2 + 1 \geq 2$. Assume on the contrary that $w' = u^{m_1}u_1bau_2u^{m_2} \notin Q$. As the languages Q and Z are reflective, then it is enough to consider $abu_2u^{m_2}u^{m_1}u_1$. Suppose $abu_2u^{m_2}u^{m_1}u_1 = v^m$ and $bau_2u^{m_2}u^{m_1}u_1 = y^n,\ n \geq 2$. Let p be the maximal common suffix of v^m and y^n. 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 $m = n = 2$; otherwise we have a contradiction.

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,

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

Since $|v| \geq 2$, we obtain that $|p| \geq |y| + |v| - 1$. Hence by Fine and Wilf's theorem, 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 $w' = u^{m_1}u_1bau_2u^{m_2} \in Q$. □

Next we study the primitive words in which exchange of two different and consecutive symbols result in a nonprimitive word.

Definition 3.7 (Non-exchange-robust Primitive Words)**.** *A primitive word is said to be non-exchange-robust if and only if exchange of two different symbols at some consecutive positions results a nonprimitive word.*

We call this set of words as non-exchange-robust primitive words. We denote the set of non-exchange-robust primitive words over the alphabet V as $Q_{\overline{X}}$. By definition, we have $Q_{\overline{X}} \cup Q_X = Q$.

3.4.1 Structural Characterization of Exchange-Robust Primitive Words

We give the structural characterization of non-exchange-robust primitive words.

Theorem 3.19. *A primitive word w is non-exchange-robust if and only if w is a primitive word of the form $u^{k_1}u_1abu_2u^{k_2},\ a,\ b \in V,\ a \neq b,\ k_1 + k_2 \geq 0$ such that $u_1bau_2 = u^m$ for some $m \geq 2$.*

Proof. (⇒) Let w be a primitive word. Suppose $w = v_1xyv_2 = u^{k_1}u_1abu_2u^{k_2}$ where $a \neq b$ such that $v_1 = u^{k_1}u_1,\ v_2 = u_2u^{k_2}$. If we exchange x and y, we get $w' = v_1yxv_2 = u^{k_1}u_1bau_2u^{k_2}$ such that $u_1bau_2 = u^m$ for $m \geq 2$. Hence $w' = u^k,\ k \geq 2$ where $k_1+m+k_2 = k$ and thus w is not an exchange-robust primitive word.

(⇐) Let $w \in Q$ which is not an exchange-robust word. Then there exists at least one consecutive positions where exchanging them makes the word nonprimitive. The word w can be written as either v_1abv_2 where $v_1,\ v_2 \in V^*$ and $a,\ b \in V$. Let $w' = v_1bav_2 \in Z$ that is $w' = v_1bav_2 = u^m$ for $m \geq 2$. Now $v_1 = u^iu_1$ and $v_2 = u_2u^j$ for $i,\ j \geq 0$. Combining both we have $v_1bav_2 = u^iu_1bau_2u^j$ where $u_1bau_2 = u^k$ for $k \geq 2$. □

$Q_{\overline{X}} = Q \setminus Q_X$ where '\' is the set minus operator. There are finite length as well as arbitrary length primitive words which are non-exchange-robust; for example, $abba$ and $(ab)^nba(ab)^n$ for $n \geq 1$.

Unlike the languages of del-robust and ins-robust primitive words which are closed under the cyclic permutation [44], the set of $Q_{\overline{X}}$ is not closed under the cyclic permutation. For example, consider the word $abbabbbab \in Q_{\overline{X}}$. One of the cyclic permutation of the word is $ababbabbb$, which is exchange robust. Hence the language of $Q_{\overline{X}}$ is not closed under cyclic permutation.

Before we prove the denseness of the language of non-exchange-robust primitive words, we prove the following result which we require to prove the denseness of $Q_{\overline{X}}$.

Lemma 3.11. *The language $Q_{\overline{X}}$ is dense over the alphabet V.*

Proof. Let w be a word. We consider two different possibilities depending upon whether w is a primitive word or a non-primitive word.

Case (A) Suppose w is a primitive word. If $|w| = 1$, then there exist $a \in V$ such that $w \neq a$ and $waaw \in Q_{\overline{X}}$. Suppose $|w| \geq 2$. We can express $w = w_1abw_2$ where $w_1, w_2 \in V^*$ and $a \neq b$. Then we can choose $x = w_1baw_2$ and $z = \lambda$ so that $xwz \in Q_{\overline{X}}$.

Case (B) If w is a non-primitive word. Suppose $w = a^n$ for some $a \in V$, $n \geq 2$ and $|w| = n$. We can choose $x = \lambda$ and $z = ba^{n-2}b$. Then we have $xwz = a^nba^{n-2}b \in Q$ and also it is non-exchange robust. Suppose $w = u^m$ for $m \geq 2$ and $|alph(w)| \geq 2$. As $|alph(w)| \geq 2$ then $|u| \geq 2$. Suppose $u = u_1abu_2$. If we choose $x = \lambda$ and $z = u_1bau_2$ then $xwz \in Q$ and $xwz \in Q_{\overline{X}}$. Hence $Q_{\overline{X}}$ is dense over V. □

3.4.2 Context-freeness of $Q_{\overline{X}}$

In this section we prove that the language of non-exchange-robust primitive words is not context-free over a given alphabet. In our proof, we use the classic Ogden's lemma, the fact that intersection of a CFL and a regular language is also context-free and we also use the fact that the family of context-free languages are closed under gsm-mapping [45].

Lemma 3.12. *(Ogden's lemma [46]) For each context-free grammar $G = (V, \Sigma, P, S)$ there is an integer k such that any word w in $L(G)$, if any k or more distinct positions in are designated as distinguished, then there is some A in $V \setminus \Sigma$ and there are words u, v, x, y and z in Σ^* such that:*

(a) $S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz = w$.

(b) x *contains at least one of the distinguished positions.*

(c) *Either* u *and* v *both contain distinguished positions, or* y *and* z *both contain distinguished positions.*

(d) vxy *contains at most* k *distinguished positions.*

Theorem 3.20. *The language of non-exchange robust words is not context-free over the alphabet* $V = \{a,\ b\}$.

Proof. Consider the regular language $R = ba^+ba^+ba^+ba^+$. We obtain a new language L by intersecting $Q_{\overline{X}}$ and R as $Q_{\overline{X}} \cap R = L$ where

$$L = \{ba^{n_1}ba^{n_2}ba^{n_3}ba^{n_4} \mid n_1,\ n_2,\ n_3,\ n_4 \geq 1,\ (|n_1 - n_3| \leq 1,\ |n_2 - n_4| \leq 1,\\ |(n_1 + n_2) - (n_3 + n_4)| = 0 \ or\ 2)\ and\ (n_1 \neq n_3\ or\ n_2 \neq n_4)\} \quad (3.1)$$

We claim that $Q_{\overline{X}} \cap R = L$.

We prove it in both directions. The inclusion $Q_{\overline{X}} \cap R \supseteq L$ is easy to observe. For the converse, let us take a word $w = ba^{n_1}ba^{n_2}ba^{n_3}ba^{n_4} \in Q_{\overline{X}} \cap R$. As $w \in Q_{\overline{X}}$, 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 (a) $aba^{n_1-1}ba^{n_2}ba^{n_3}ba^{n_4}$

Case (b) $ba^{n_1-1}ba^{n_2+1}ba^{n_3}ba^{n_4}$

Case (c) $ba^{n_1+1}ba^{n_2-1}ba^{n_3}ba^{n_4}$

Case (d) $ba^{n_1}ba^{n_2-1}ba^{n_3+1}ba^{n_4}$

Case (e) $ba^{n_1}ba^{n_2+1}ba^{n_3-1}ba^{n_4}$

Case (f) $ba^{n_1}ba^{n_2}ba^{n_3-1}ba^{n_4+1}$

Case (g) $ba^{n_1}ba^{n_2}ba^{n_3+1}ba^{n_4-1}$

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

(i) $n_1 \neq n_3\ or\ n_2 \neq n_4$ (otherwise $ba^{n_1}ba^{n_2}ba^{n_1}ba^{n_2} \notin Q$)

(ii) $|n_1 - n_3| \leq 1,\ |n_2 - n_4| \leq 1,\ |(n_1 + n_2) - (n_3 + n_4)| = 0\ or\ 2$ (otherwise the word $w' \in Q_X$

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

A CFL is closed under gsm mapping [47]. Using a sequential transducer (a gsm), the language $Q_{\overline{X}} \cap R$ can be translated into a new language

$$L' = \{a^{n_1}b^{n_2}c^{n_3}d^{n_4} \mid n_1,\ n_2,\ n_3,\ n_4 \geq 1,\ |n_1 - n_3| \leq 1,\ |n_2 - n_4| \leq 1,$$
$$|(n_1 + n_2) - (n_3 + n_4)| = 0 \text{ or } 2 \text{ and } (n_1 \neq n_3 \text{ or } n_2 \neq n_4)\} \quad (3.2)$$

We have to prove that L' is not a context-free language. Assume by 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, 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 and (iii) $uv^ixy^iz \in L'$ 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 $w = uvxyz$ such that all the conditions of Ogden's lemma satisfy.

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

Case (a) 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 $|p_2 - p_4| \geq 2$.

Case (b) 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. Since the family of context-free languages is closed under sequential transducers and intersection with regular languages [47], we conclude that $Q_{\overline{X}}$ is not context-free. □

3.5 Conclusions

We have investigated four different types of point mutation operations on primitive words. We have studied to preserve the primitivity by substitute a symbol by another symbol, deletion or insertion a symbol and exchanging two consecutive symbols. The structural characterization of each of the class of primitive words have been discussed and also some important combinatorial properties related to each of the class have been identified. It has been proved that the languages of non-del-robust, non-ins-robust and non-exchange-robust primitive words are not context-free. We have also proved that Q_D, Q_S and Q_I are reflective. We have linear time algorithms to recognize the del-robust and ins-robust primitive words, but for exchange robust this problem is still open.

We summarize the results as follows.

	Q_S	Q_D	Q_I	Q_X
Definition	Elements remain primitive after a symbol substitution	Elements remain primitive after a symbol deletion	Elements remain primitive after a symbol insertion	Elements remain primitive after an exchange of distinct consecutive symbols
Reversibility	Reversible	Reversible	Reversible	Reversible
Reflectivity	Reflective	Reflective	Reflective	Not Reflective
Context-Freeness	$Q_{\overline{S}}$ is not CFL	$Q_{\overline{D}}$ is not CFL	$Q_{\overline{I}}$ is not CFL	$Q_{\overline{X}}$ is not CFL
Algorithm for recognition	Result for Linear time Algorithm	Linear time Algorithm	Linear time Algorithm	No linear time algorithm is known

“Once you start working on something, don’t be afraid of failure and don’t abandon it. People who work sincerely are the happiest.”

- Chanakya

Chapter 4

Robustness of L-Primitive Words

4.1 L-Primitive Words

The primitive words has been studied in [48, 20, 49, 36, 50], which is generated by letters of alphabet V such that it is not proper power of $x \in V^*$. In this chapter, we deal with a language of primitive words with respect to a language $L \subseteq V^*$, called L-primitive words, that is, if x is L-primitive then x is not a power of any word $y \in L$.

Definition 4.1. *[39] Let L be a language over an alphabet V. A word $x \in V^+$ is said to be an L-primitive word if x is not a proper power of any word in L, that is,*

$$x = u^k \text{ for } u \in L, \implies k = 1.$$

Let $X \subseteq V^*$ and X^c denotes the complement of X in V^*. The set of L-primitive words over an alphabet V is denoted by $QL(V)$ or simply QL and the set of non-L-primitive words over an alphabet V is denoted by ZL.

A word over an alphabet has unique primitive root but it can have more than one L-primitive roots. For example if $L = \{aa, aaa\}$, then $aaaaaa$ has only one primitive root, which is a, whereas there are two L-primitive roots which are aa and aaa.

Proposition 4.1. *[39] If L_1 and L_2 are two subsets of V^*, then*

$$L_1 \subseteq L_2 \implies QL_2 \subseteq QL_1$$

.

The Proposition 4.1 is proved for two languages such that one is subset of other. In next proposition we prove above result for independent languages.

Proposition 4.2. *Let $L_k = \{u^{2^k}\}$ for the natural number k where u is a primitive word. Then $QL_i \subseteq QL_j$ for $i \leq j$.*

Proof. For $L_k = \{u^{2^k}\}$, $ZL_k = \{u^{2^k.i} \mid i \geq 2\}$. Therefore $ZL_j \subseteq ZL_k$ for $j \geq k$. Therefore $QL_k \subseteq QL_j$ for $k \leq j$. □

Proposition 4.3. *Let for $i \geq 1$, $L_{ik} = \{u^{i^k}\}$ where $k \geq 0$ and u is a primitive word. Then $QL_{ik} \subseteq QL_{ij}$ for $k \leq j$.*

Proof. Proof is similar to that of Proposition 4.2. □

The language of primitive words, Q, is reflective but the language of L-primitive words, QL, need not be reflective. Consider for example, a language L that contains ab but not ba. Then $baba \in QL$ as it is not proper power of any word contained in L but $abab \notin QL$.

Lemma 4.1. *Let L be a language. Then QL is reflective if and only if ZL is reflective.*

Proof. If part: Since QL is reflective, we have $vu \in QL$ for all $w' = u'v' \in QL$. On contrary, let ZL is not reflective, then there exists a word $w = uv \in ZL$ such that $vu \in QL$ but then $uv \in QL$, which is contradiction.

Proof of *only if part* is similar to *if part*. □

Lemma 4.2. *Let L be a language. Then L is not reflective if $u_2u_1 \in ZL$ for some $u_1u_2 \in QL$.*

Proof. Since $u_2u_1 \in ZL$, there exists $v \in L$ such that $u_2u_1 = v^k$ for some $k \geq 2$. Therefore $u_1u_2 = v'^k$ for some v', cyclic permutation of v. But since $u_1u_2 \in QL$, $v' \notin L$. Therefore L is not reflective. □

Lemma 4.3. *Let L be a language. Then if L is reflective then QL and ZL are also reflective.*

Proof. Suppose L is reflective. Then for all $w = w_1w_2 \in L$, $w_2w_1 \in L$, and suppose for contradiction a partition of a word $v = v_1v_2 \in QL$, $v_2v_1 \notin QL$. Therefore $v_2v_1 \in ZL$ which implies that there exists $u \in L$ such that $v_2v_1 = u^k$ for some $k \geq 2$. Therefore $v_2 = u^{k_1}u'$ and $v_1 = u''u^{k_1}$ where $u'u'' = u$. Also $v_1v_2 = u''u^{k_1}u^{k_1}u' = (u''u')^k$. Since $u''u' \in L$, we have $v_1v_2 \in ZL$, which is contradiction. □

Converse of Lemma 4.3 need not be true. For example, if $L = \{abc, abcabc, cab, bca\}$, then for all $uv \in QL$, $vu \in QL$, that is, QL is reflective even though $bcabca, cabcab \notin L$.

Lemma 4.4. *Let L be a language over an alphabet V. Then if $vu \in QL$ for all $u, v \in V^*$ such that $uv \in QL$ then $vu \in L$ for all $u, v \in V^*$ such that $uv \in L$ and uv is L-primitive.*

Proof. For contradiction, let a word $w = uv \in L \cap QL$ such that $vu \notin L$. Since QL is reflective, we have $vu \in QL$. Since QL is reflective, ZL is also reflective and $(uv)^k \in ZL$ for all $k \geq 2$ implies that $(vu)^k \in ZL$. Also we have $vu \in QL$ therefore $vu \in L$, which is a contradiction. □

Theorem 4.1. *Let L be a language over an alphabet V. QL is reflective if and only if $vu \in L$ for all $u, v \in V^*$ such that $uv \in L$ and uv is L-primitive.*

Proof. Only if part: Let $L \cap QL$ be reflective but on contrary QL is not, then there exist $w \in QL$ such $w = uv$ for some $u,\ v \in V^+$ such that $vu \in ZL$. Therefore $vu = t^k$ for some $k \geq 2$ and $t \in L \cap QL$. Since Q is reflective, we have $uv = t'^k$ for some $t' \in V^*$ such that t' is cyclic permutation of t. Since $L \cap QL$ is reflective, we have $t' \in L \cap QL$. Therefore $uv \in ZL$, which is a contradiction.

If part: This part is the same as Lemma 4.4. □

The language of primitive words, Q, is closed under reverse operation on words but the language of L-primitive words, QL, need not be closed under reverse operation. For example, if $L = \{ab, baba\}$, then $baba \in QL$ but $abab \notin QL$.

Lemma 4.5. *Let L be a language over an alphabet V. Then if $rev(w) \in L$ for all $w \in L$ then $rev(w) \in QL$ for all $w \in QL$.*

Proof. Let for a word $v \in QL$, $rev(v) \notin QL$, then there exists a word $w \in L$ such that $rev(v) = w^k$ for some $k \geq 2$. Since for all $w \in L$, $rev(w) \in L$, $v = rev(rev(v)) = rev(w^k) = (rev(w))^k \notin QL$, which is a contradiction. This proves the result. □

But converse of the above statement need not be true. For example, consider $L = \{ab, abab, ba\}$. Then $rev(w) \in QL$ for all $w \in QL$, but $rev(abab) \notin L$.

Lemma 4.6. *Let L be a language over an alphabet V. Then if $rev(w) \in QL$ for all $w \in QL$ then $rev(v) \in L$ for all $v \in L \cap QL$.*

Proof. For $w \in QL$, $rev(w) \in QL$. For contradiction, let a word $v \in L$, $v \in QL$, $v^2 \in L$, $rev(v^2) = (rev(v))^2 \notin L$ and also $rev(v) \notin L$. Since $(rev(v))^2 \notin L$ therefore $(rev(v))^2 \in QL$ therefore by assumption we have, $rev((rev(v))^2) = v^2 \in QL$ which is a contradiction as $v \in QL$. Therefore $rev(v) \in L$. □

Theorem 4.2. *Let L be a language over an alphabet V. $rev(w) \in QL$ for all $w \in QL$ if and only if $rev(v) \in L$ for all $v \in L \cap QL$.*

Proof. This result follow from the Lemma 4.5 and Lemma 4.6. □

Lemma 4.7. *Let L be a language over an alphabet V. For $w \in QL$, $rev(w) \in QL$ if and only if $rev(\sqrt[L]{v}) \in L$ for all $v \in L$.*

Proof. If part: Let for any $w \in QL$, $rev(w) \in QL$. Therefore we have $rev(w) \in \overline{QL}$ for all $w \in \overline{QL}$.

If $w \in L$, then $w = u^k$ for some $k \geq 1$ and $u \in L$. $rev(w) = (rev(u))^k$ where $rev(u) \in L$. Let reverse of L-primitive root of w is not in L for a word $w \in L$.

Only if Part: If L-primitive root of $rev(w)$ is in L for all $w \in L$. and let $w \in QL$ but $rev(w) \notin QL$, then $rev(w) = v^k$ for some $v \in L \cap QL$ and $k \geq 2$. $w = rev(rev(w)) = rev(v^k) = (rev(v))^k$. But since $v \in L$ which is L-primitive root of $rev(w)$, therefore $rev(v) \in L$. Hence $w \notin QL$, which is a contradiction. □

Lemma 4.8. *$Q \subseteq QL$, for any language $L \subseteq V^*$, where Q is the language of primitive words and QL is set of L-primitive words.*

Proof. Let $w \in Q$ but $w \notin QL$, then $w = u^k$ for some $u \in L$ and $k \geq 2$. Since $L \subseteq V^*$, therefore $w \notin Q$, which is contradiction. □

As we know that at least one of w and wa is primitive for $w \notin a^*$ (lemma (2.10)), this is also true in case of L-primitive. For every word $u \in V^+$ and every symbols $a, b \in V$, $a \neq b$, at least one of the words $ua,\ ub$ is primitive as well as L-primitive. This result has several consequences, proving in some sense that "there are very many L-primitive words".

Corollary 4.1. *(a) For every word $u \in V^*$, at most one of the words ua with $a \in V$, is not L-primitive.*

(b) For every words $u_1, u_2 \in V^*$, at most one of the words u_1au_2, with $a \in V$, is not L-primitive.

This corollary claims that the language QL is right 1-dense and therefore right k-dense for every k.

Lemma 4.9. *Let L be a language over an alphabet V. Then if $Q \not\subseteq L$ then there exist a word $u \in Q$ and an integer $k \geq 2$ such that $u^k \in QL$.*

Proof. If $u \in Q$ but $u \notin L$ and let for some $k \geq 2$, u^k in ZL then there exist a minimum of all $i \geq 2$ such that i divides k and $u^i \in L$. If that minimum is j then $u^j \in QL$ as it is not a proper power of any element in L. □

Theorem 4.3. *Let L be a language over an alphabet V. $QL = Q$ if and only if $Q \subseteq L$.*

Proof. **If part:** Let $QL = Q$ but $Q \nsubseteq L$, then there exists $u \in Q$ which is not in L, therefore there exists $k \geq 2$ such that $u^k \in QL$. But $Q = QL$, $u^k \in Q$ which is a contradiction.

Only if Part: Let $Q \subseteq L$, we have $Q \subseteq QL$. Let $QL \nsubseteq Q$, then there exists $u \in QL$, such that $u = x^k$ for some $x \in Q$ and $k \geq 2$. But since $Q \subseteq L$, we have $x \in L$. Therefore $x^k \in ZL$, which is a contradiction. Hence $Q = QL$. □

Corollary 4.2. *Let L be a language over an alphabet V. Then if $L = Q$ then $QL = Q$ and equivalently $ZL = Z$.*

Proof. Follows from Theorem 4.3. □

Corollary 4.3. *Let L be a language over an alphabet V. $QL \cap Z$ is empty if and only if $Q \subseteq L$.*

Proof. Follows from Theorem 4.3. □

L^n is defined as concatenation of $L^{n-1}.L$ where $n \geq 2$ and $L^1 = L$.

Theorem 4.4. *For a language L, $ZL = \bigcup_{n \geq 2} L^n$ if and only if there exists a unique L-primitive word $u \in L$ such that and for any $x \in L$, $x = u^k$ for some $k \geq 1$.*

Proof. This is obvious that to equate ZL and union of L^n, the elements of L should be powers of a unique primitive word. Now remaining part is proved below.

If part: On contradiction let $ZL = \bigcup_{n \geq 2} L^n$ and let $x_1, x_2 \in L$ such that $x_1 = u^{r_1 \cdot s_1}$ and $x_2 = u^{r_2 \cdot s_2}$ for two L-primitive words u^{r_1} and u^{r_2}, where $(r_1, r_2) = 1$. Then $u^{r_1 + r_2} \in L^2$ but not in ZL, which is contradiction.

Only if part: Let $u \in L$ be a unique L-primitive word such that and for any $x \in L$, $x = u^k$ for some $k \geq 1$. Then for any $w \in L^n$, $w = u^{k_1}.u^{k_2} \ldots u^{k_n} = u^{k_1 + k_2 \ldots + k_n} \in ZL$. Therefore, for every $n \geq 2$, $L^n \subseteq ZL$ and so $\bigcup_{n \geq 2} L^n \subseteq ZL$.
Next, we have to prove that $ZL \subseteq \bigcup_{n \geq 2} L^n$. Let $w \in ZL$, then $w = u^k$ for $k \geq 2$ as every element of L can be represented as power of $u \in L$. Therefore, $w \in L^k$. Similarly we can show for all elements of ZL. Therefore $ZL \subseteq \bigcup_{n \geq 2} L^n$. □

4.2 Other Formal Languages and L-Primitive Words

We know that the language of primitive words, Q, is not regular [51]. This is still an open problem that whether Q is context-free language or not [45, 52]. But we know that primitive words can be identified with 2DPDA [13]. In this section we identify relation of other formal languages with the language of L-primitive words. The question arises that whether nature of QL depends on the nature of language L. We identify conditions under which language of L-primitive words is regular or context-free. We discuss some results related to this.

Theorem 4.5. *Let L be a language on an alphabet V. Then the language of L-primitive words is regular if any of the following conditions holds.*

(a) If L is finite.

(b) If $L = \{w^n \mid n \geq 1\}$ for some word $w \in V^*$ of finite length.

(c) If $L = \bigcup_{w \in \{w_1, w_2, \ldots w_m\} \subset V^*} \{w^n \mid n \geq 1\}$ for a finite number m where $|w_i|$ is finite for $1 \leq i \leq m$.

Proof. **Case (a):** This case is related to finite language L. $ZL = \bigcup_{u \in L} \{u^n \mid n \geq 2\}$. L is regular, Therefore $\{u^n \mid n \geq 2\} = u^* \setminus \{u\}$ is also regular language. Since the regular languages are closed under finite union, we have ZL is regular. The regular languages are closed under set complement, therefore $QL (= V^* \setminus ZL)$ is regular.

The next cases are for the infinite language L.

Case (b): Since $L = \{u^n \mid n \geq 1\}$ for some $u \in V^*$, we have $ZL = \{u^n \mid n \geq 2\}$ is regular language. Therefore QL is regular.

Case (c): If $L = \bigcup_{w \in \{w_1, w_1, \ldots w_n\} \subset V^*} \{w^n \mid n \geq 1\}$ then $ZL = \bigcup_{u \in \{w_1, w_1, \ldots w_n\} \subset V^*} \{u^n \mid n \geq 2\}$ is regular language as the regular languages are closed under finite union. Since the regular languages are closed under set complement, we have QL is regular. □

There are examples for non-trivial regular languages for which ZL is not regular. Let $L = aaa(aa)^*$ then L is regular. Consider $ZL = \{a^n \mid n > 3 \; and \; n \neq 2^k \; where \; k \in N\}$. However, ZL is not even context free.

There exists an infinite non-regular language L such that language of L-primitive words is regular. For example, $L = \{a^k \mid k \; is \; prime\}$ is not regular. Now, $ZL = \{a^n \mid n \geq 2\}$. Clearly, ZL is a regular language. Therefore QL is also regular.

Theorem 4.6. *Let L be a regular language. Then ZL is regular if and only if $ZL \setminus L$ is a regular language and primitive root of $ZL \setminus L$ is a finite language.*

Proof. If part: The class of regular languages is closed under union and taking the difference of two sets, therefore if $ZL \setminus L$ is a regular language then so is $(ZL \setminus L) \cup L = ZL$.

Only if part: If ZL is regular then so is $ZL \setminus L$. Let there be an n-state minimal deterministic automaton which accept $ZL \setminus L$. Now suppose that root of $ZL \setminus L$, represented by R, is infinite. Then there is a word $w \in ZL \setminus L$ such that root of w is greater than n. Now according to pumping lemma for regular languages there exists an integer n for $w = xyz$ such that $|y| \geq 1$ and $|xy| \leq n$ and $xy^iz \in ZL \setminus L$ for all $i \geq 0$, so xy^iz is non-L-primitive and so non-primitive for all i. That is for $i, j \geq 1$ with $i < j$ such that both xy^iz and xy^jz are non-primitive. Since language of non-primitive words is reflective, zxy^i and zxy^j are non-primitive. By Theorem 2.4, roots of zxy^i and zxy^j are same and equal to root ot y. Hence root of xyz is same as root of y and so less than n, which is a contradiction to the assumption that root of w is greater than n.

Therefore primitive root of $ZL \setminus L$ must be finite. □

Corollary 4.4. *For a context free language L, the language ZL need not be a context-free language.*

Proof. Let $L = \{a^n b^n \mid a \neq b,\ n \geq 1\}$. L is a context-free language. Corresponding to it $ZL = \{(a^n b^n)^k \mid a \neq b,\ n \geq 1\ k \geq 2\}$. Consider the string $s = a^{p+1}b^{p+1}a^{p+1}b^{p+1}$, where $a,\ b \in V$ and $a \neq b$. It is easy to see that $s \in ZL$ and $|s| \geq p$. We prove it by using pumping lemma for CFL.

Hence, s can be written as $uvwxy$, where u, v, w, x, and y are factors, such that $|vwx| \leq p$, $|vx| \geq 1$, and $uv^iwx^iy \in Q_{\overline{I}}$ for every $i \geq 0$. By the choice of s and the fact that $|vwx| \leq p$, we have one of the following possibilities for vwx:

Case (a) $vwx = a^j$ for some $j \leq p$.

Case (b) $vwx = a^j b^k$ for some j and k with $j + k \leq p$.

Case (c) $vwx = b^j$ for some $j \leq p$.

Case (d) $vwx = b^j a^k$ for some j and k with $j + k \leq p$.

In Case (a), since $vwx = a^j$, therefore $vx = a^t$ for some $t \geq 1$ and hence $uv^iwx^iy = a^{p-t+1}b^{p+1}a^{p+1}b^{p+1} \notin ZL$ for $i = 0$.

Similarly, we can obtain contradiction in Case (c).

Case (b) has several subcases.

(i) $v = a^{j_1},\ w = a^{j_2},\ x = a^{j_3}b^k$.

(ii) $v = a^{j_1},\ w = a^{j_2}b^{k_1},\ x = b^{k_2}$.

(iii) $v = a^j b^{k_1},\ w = b^{k_2},\ x = b^{k_3}$.

In Case b(i), Case b(ii) and Case b(iii) if we take $i = 0$, $uv^i wx^i y \notin ZL$.

Similarly, we can obtain contradiction in Case (d).

Therefore ZL is not a context-free language. □

4.3 Ins-Robust L-Primitive Words

Definition 4.2. *An L-primitive word w of length n is said to be ins-robust L-primitive word if the word*

$$pref(w, i)\ .\ a\ .\ suf(w, n - i)$$

is an L-primitive word for all $i \in \{0, 1, \dots, n\}$ where $a \in V$.

We denote the language of ins-robust L-primitive words as QL_I and language of non-ins-robust L-primitive words as $QL_{\overline{I}}$. QL_I which is a subset of the set of primitive words, QL.

A language QL_I need not closed under reflective property.

For instance a language L which contains ab but not the word ba. Then $baba \in QL_I$ as it is not proper power of any word contained in L and also not power of any word of L after insertion of any symbol in this word, but $abab \notin QL_I$ as it is not even in QL.

Similarly, we can have $w \in QL_I$ such that $rev(w)$ need not be in QL_I.

Lemma 4.10. *QL_I is reflective if and only if for all $w = uv \in L \cap QL$, $vu \in L$.*

Proof. Proof is similar to Theorem 4.1. □

We know that for $w \in Q_I$, $rev(w) \in Q_I$ by Lemma (3.6), but for $w \in QL_I$, $rev(w)$ need not be in QL_I.

Theorem 4.7. *Let L be a language over an alphabet V. For $w \in QL_{\overline{I}}$, $rev(w) \in QL_{\overline{I}}$ if and only if for $u \in L \cap QL$, $rev(u) \in L$.*

Proof. Let $w \in QL_{\overline{I}}$, $u,\ rev(u) \in L \cap QL$ and $rev(u) \in L$ but $rev(w) \in QL_{\overline{I}}$. Since $w \in QL_{\overline{I}}$, therefore $w = u^r u_1 u_2 u^r$, $rev(w) = (rev(u))^s rev(u_2) rev(u_1)\ (rev(u))^r$ for some $u \in L$ and $u = u_1 a u_2$ for some $a \in V$. It follows that $rev(w) \in QL_{\overline{I}}$.

Conversely let for all $w,\ rev(w) \in QL_{\overline{I}}$ and $u \in L \cap QL$ but $rev(u) \notin L$. Therefore $w = u^r u_1 u_2 u^r$, $rev(w) = (rev(u))^s rev(u_2) rev(u_1) (rev(u))^r \in QL_{\overline{I}}$ for some $u \in L \cap QL$. The following two cases may arises.

Case (A) If $rev(u) = x^k$ for $k \geq 2$ where $x \in L \cap QL$, then $(rev(u))^s rev(u_2)\, rev(u_1)(rev(u))^r x \in QL_{\overline{I}}$. But then $w = rev(x)(rev(x)^k)^r rev(u_1)\ rev(u_2)(rev(x)^k)^r \in QL_{\overline{I}}$ which implies $rev(x) \in L$ which is contradicting to $u = (rev(x))^k \in L \cap QL$.

Case (B) If $rev(w) = (rev(u))^s rev(u_2) rev(u_1)(rev(u))^r = x^m x_1 x_2 x^n$ for $x = (rev(u))^k$ and $k \geq 2$. The proof is similar to case (A). □

Theorem 4.8. *Let L be a language over an alphabet V. For $w \in QL_I$, $rev(w) \in QL_I$ if and only if for $u \in L \cap QL$, $rev(u) \in L$.*

Proof. Proof is similar to Theorem 4.7. □

Lemma 4.11. *Let L be a language over an alphabet V. $Q_I \subseteq QL_I$, for any language $L \subseteq V^*$, where Q_I is the language of ins-robust primitive words and QL_I is set of ins-robust L-primitive words.*

Proof. Let $w \in Q_I$ but $w \notin QL_I$. Then for some partition of w, say w_1w_2, and $a \in V$, $w_1 a w_2 = u^k$ for some $u \in L$ and $k \geq 2$. Since $L \subseteq V^*$, therefore $w \notin Q_I$, which is a contradiction. □

Proposition 4.4. *Let L and M be two subsets of V^*. Then $L \subseteq M \implies QM_I \subseteq QL_I$.*

Proof. On the contrary, let us assume that $QM_I \nsubseteq QL_I$. Then there exists $w \in QM_I$, but $w \notin QL_I$. Therefore there exists partition of $w = w_1w_2$ such that $w_1 a w_2 = u^k$ for some $u \in L$ and $k > 1$. Since $L \subseteq M$, we have $u \in M$. Consequently, $w \neq QM_I$ a contradiction. □

Remark 4.1. *Clearly, an ins-robust L-primitive word need not be ins-robust primitive. For instance, let $L = \{abb\} \subseteq \{a,b\}^*$. The word abb is an ins-robust L-primitive word, but not an ins-robust primitive word.*

4.4 Del-Robust L-Primitive Words

Definition 4.3. *A L-primitive word w of length n is said to be del-robust primitive word if and only if the word*

$$pref(w,i)\ .\ suf(w, n-i-1) \neq u^k,\ k \geq 2,\ u \in L,\ \ i \in \{0, 1, \ldots, n-1\}\ .$$

For example, if $L = \{(ab)^n \mid n \geq 1, a, b \in V\}$, then the words $b(abb)^k, k \geq 1$ and $a^m b^n$ for $m, n \geq 2$ are del-robust L-primitive words, whereas $b(abb)^k, k \geq 1$ are not in Q_D. We denote the language of del-Robust L-primitive words as QL_D and language of non-del-robust L-primitive words as $QL_{\overline{D}}$.

Lemma 4.12. *Let L be a language over an alphabet V. $Q_D \subseteq QL_D$, for any language $L \subseteq V^*$, where Q_D is the language of del-robust primitive words and QL_D is set of del-robust L-primitive words.*

Proof. Let $w \in Q_D$ but $w \notin QL_D$, then for some partition of w (say w_1aw_2), $w_1w_2 = u^k$ for some $u \in L$ and $k \geq 2$. Since $L \subseteq V^*$, therefore $w \notin Q_D$, which is contradiction. □

A language QL_D need not closed under reflective property. For example, let a language L contains ab but not ba. Then $baba \in QL_D$ as it is not proper power of any word contained in L and also not proper power of any word of L after deletion of any symbol from this word, but $abab \notin QL_D$ as it is not even in QL. For $w \in QL_D$, $rev(w)$ need not be in QL_D. Similarly, $QL_{\overline{D}}$ need not closed under reflective property.

Lemma 4.13. *Let L be a language over an alphabet V. QL_D is reflective if and only if for all $w = uv \in L \cap QL$, $vu \in L$.*

Proof. Proof of this lemma is similar to that of Theorem 4.1. □

From corollary 3.5, we know that a word w is in the set $Q_{\overline{D}}$ if and only if it is of the form $u^n a$ or its cyclic permutation for some $u \in Q$, $u \neq a$ and $n \geq 2$. But since $QL_{\overline{D}}$ is not reflective, so this result may not be true for such words.

Proposition 4.5. *If L and M are two subsets of V^*, then $L \subseteq M \implies QM_D \subseteq QL_D$.*

Proof. On the contrary, let us assume that $QM_D \nsubseteq QL_D$. Then there exists $w \in QM_D$, but $w \notin QL_D$. Therefore there exists partition of $w = w_1aw_2$ such that $w_1w_2 = u^k$ for some $u \in L$ and $k > 1$. By hypothesis, we have $u \in M$. Consequently, $w \neq QM_D$ a contradiction. □

Remark 4.2. *A del-robust L-primitive word need not be del-robust primitive. For instance, let $L = \{(abb)^n \mid n \geq 1\} \subseteq \{a, b\}^*$. Clearly, the word abb is a del-robust L-primitive word, but not a del-robust primitive word.*

It is easy to prove that an L-primitive word w is not del-robust if and only if w can be expressed in the form of $u^{k_1}u_1cu_2u^{k_2}$ where $u_1,\ u_2 \in V^*$, $u_1u_2 = u \in L$, $c \in V$, $k_1, k_2 \geq 0$ and $k_1 + k_2 \geq 1$.

Theorem 4.9. *Let L be a language over an alphabet V. $QL_D = Q_D$ if and only if $Q \subseteq L$.*

Proof. If $Q \subseteq L$ then it is easy to prove that $QL_D = Q_D$. Conversely, if $QL_D = Q_D$ and $Q \nsubseteq L$, then there exists a primitive word u such that $u \notin L$. We have $u^k a \in QL_D$ but $u^k a \notin Q_D$, which is a contradiction. □

4.5 Exchange Robust L-Primitive Words

Definition 4.4. *An L-primitive word w of length n (≥ 2) is said to be exchange robust primitive word if and only if the word*

$$pref(w,i)\ w_{i+1}w_i\ suf(w,n-i-2) \neq u^k,\ k \geq 2,\ u \in L,\ \ i \in \{0,1,\dots,n-2\}$$

.

For example, if $L = \{ab\}$, then the word $baaaba$ is exchange robust L-primitive words, whereas $baaaba$ is not in Q_X. We denote the language of Exchange Robust L-primitive words as QL_X and language of non-exchange robust L-primitive words as $QL_{\overline{X}}$.

Lemma 4.14. *$Q_X \subseteq QL_X$, for any language $L \subseteq V^*$, where Q_X is the language of exchange robust primitive words and QL_X is set of exchange robust L-primitive words.*

Proof. Let $w \in Q_X$ but $w \notin QL_X$, then for some partition of w (say w_1abw_2), $w_1baw_2 = u^k$ for some $u \in L$ and $k \geq 2$. Since $L \subseteq V^*$, therefore $w \notin Q_X$, which is contradiction. □

Lemma 4.15. *Let L be a language over an alphabet V. If $L \subseteq V$ then $QL_X = QL$.*

Proof. To prove this, we have to prove that $QL_{\overline{X}} (= QL \setminus QL_X)$ is empty for $L \subseteq V$. Let $QL_{\overline{X}}$ be not empty, then there exist $w_1abw_2 \in QL$ such that $b \neq a$ and $w_1baw_2 \in ZL$. Since $L \subseteq V$, we have $ZL\ \{a^k \mid a \in L\}$. Therefore $b = a$, which is a contradiction. Therefore $QL_{\overline{X}}$ is empty and so $QL_X = QL$. □

Theorem 4.10. *QL_X is reflective if and only if $L \subseteq V$.*

Proof. To prove this we prove that QL_X is not closed under reflective property if and only if $L \nsubseteq V$.

Proof of *if part* is similar to Lemma 4.15. If $L \subseteq V$ then QL_X is reflective.

Only if: QL_X is reflective, hence $uv \in QL_X$ implies $vu \in QL_X$. Let $L \nsubseteq V$ (i.e. there exists a word $w \in L$ such that $|alph(w)| \geq 2$) then $QL_{\overline{X}}$ is not empty and so there exists $u^r u_1 bau_2 u^s \in QL_{\overline{X}}$ for some $u = u_1abu_2 \in L$. But $au_2u^su^ru_1b \notin QL_{\overline{X}}$, in fact $au_2u^su^ru_1b \in QL_X$, which is a contradiction, as QL_X is reflective. Therefore $L \subseteq V$. □

Lemma 4.16. *Let L be a language over an alphabet V. If $rev(u) \in L$ for $u \in L$ then $rev(w) \in QL_X$ for $w \in QL_X$.*

Proof. Let $rev(u) \in L$ for all $u \in L$ but $rev(w) \notin QL_X$ for some $w \in QL_X$. Since $rev(w) \notin QL_X$, we have $rev(w) \in ZL$ or $rev(w) \in QL_{\overline{X}}$.

Case (A) If $rev(w) \in ZL$, then $w \in ZL$, which is a contradiction.

Case (B) If $rev(w) \in QL_{\overline{X}}$, $rev(w) = u^r u_1 bau_2 u^s$ for some $u = u_1abu_2 \in L$ and $a \neq b$. Therefore $w = rev(rev(w)) = (rev(u))^s \; rev(u_2) \; ab \; rev(u_1) \;\; (rev(u))^r \in QL_{\overline{X}}$ as $rev(u) = rev(u_2) \; ab \; rev(u_1) \in L$ which is contradiction.

Therefore $rev(w) \in QL_X$. □

Theorem 4.11. *For $w \in QL_X$, $rev(w) \in QL_X$ if and only if for $u \in L \cap QL$, $rev(u) \in L$.*

Proof. Proof of *if part* is similar to that of Lemma 4.16.

For other part, let $w \in QL_X$ implies $rev(w) \in QL_X$ but for some $u \in L \cap QL$, $rev(u) \notin L$. Then there exists $r, s \geq 0, \; r + s \geq 1$ such that $(rev(u))^r \; rev(u_2) \; ab \; rev(u_1) \; (rev(u))^s \in QL_X$ where $u = u_1 \; ab \; u_2$.

$rev((rev(u))^r \; rev(u_2) \; ba \; rev(u_1) \; (rev(u))^s) = u^s \; u_1 \; ba \; u_2 \; u^r$ is in $QL_{\overline{X}}$, which contradict the assumption that QL_X is closed under reverse operation. □

Proposition 4.6. *If L and M are two subsets of V^*, then $L \subseteq M \implies QM_X \subseteq QL_X$.*

Proof. On the contrary, let us assume that $QM_X \nsubseteq QL_X$. Then there exists $w \in QM_X$, but $w \notin QL_X$. Therefore there exists partition of a word $w = w_1abw_2$ such that $w_1baw_2 = u^k$ for some $u \in L$ and $k > 1$. By hypothesis, we have $u \in M$. Consequently, $w \neq QM_X$ a contradiction. □

Remark 4.3. *An exchange robust L-primitive word need not be exchange robust primitive. For instance, let $L = \{abba\} \subseteq \{a, b\}^*$. Clearly, the word $abba$ is an exchange robust L-primitive word, but not exchange robust primitive word.*

Theorem 4.12. *An L-primitive word w is non-exchange robust if and only if w can be expressed in the form of $u^{k_1}u_1bau_2u^{k_2}$ where $u_1, \; u_2 \in V^*$, $u_1abu_2 = u^2 \in L$, $c \in V$, $k_1, k_2 \geq 0$ and $k_1 + k_2 \geq 0$.*

Proof. We prove the necessary and sufficient conditions are as follows:
($\Rightarrow$) Let w be a primitive word. Suppose $w = v_1xyv_2 = u^{k_1}u_1abu_2u^{k_2}$ where $a \neq b$ such that $v_1 = u^{k_1}u_1, \; v_2 = u_2u^{k_2}$. First we consider that x and y are not hole. If we exchange

x and y, we get $w' = v_1yxv_2 = u^{k_1}u_1bau_2u^{k_2}$ such that $u_1bau_2 = u^m$ for $m \geq 2$. Hence $w' = u^k,\ k \geq 2$ where $k_1+m+k_2 = k$ and thus w is not an exchange-robust primitive word.

($\Leftarrow$) Let $w \in Q$ which is not an exchange-robust word. Then there exists at least one consecutive positions where exchanging them makes the word nonprimitive. w can be written as either v_1abv_2 where $v_1,\ v_2 \in V^*$ and $a,\ b \in V$. Let $w' = v_1bav_2 \in Z$ that is $w' = v_1bav_2 = u^m$ for $m \geq 2$. Now $v_1 = u^iu_1$ and $v_2 = u_2u^j$ for $i,\ j \geq 0$. Combining both we have $v_1bav_2 = u^iu_1bau_2u^j$ where $u_1bau_2 = u^k$ for $k \geq 2$. □

4.6 Conclusions

In this chapter, we have discussed a special type of words which are primitive with respect to a language L, called L-primitive words. We have characterize them and identified several properties. We have also defined the robustness of these words. We have identified the conditions to show the reflectivity of QL. Various robustness of L-primitive words and their properties are also discussed in this chapter.

"The fragrance of flowers spread only in the direction of wind. But the goodness of a person spreads in all direction."

- Chanakya: Indian teacher, philosopher, economist, jurist and royal advisor

Chapter 5

Pseudo Quasiperiodic Words

A *morphism* $h : U^* \to V^*$ (or $h : U^+ \to V^+$) is a mapping which satisfies: $h(ww') = h(w)h(w')$ for all $w, w' \in U^*$, where U and V are alphabets. In particular, if h is morphism then $h(\lambda) = \lambda$ and h is completely specified by the words $h(a)$ with $a \in V$. A morphism h is λ-free if $h(a) \neq \lambda$ for all $a \in V$. A morphism h is called injective if and only if, for all $v, w \in V^*$, $h(v) = h(w)$ implies $v = w$. A morphism h is periodic if $\exists z$ such that $h(a) \in z^*$, for all $a \in V$. For a morphism h, if $|h(a)| = |h(b)|$ for all $a, b \in V$ then h is called uniform morphism. h is prefix (resp. suffix) if none of the words in $h(V)$ is a prefix (resp. suffix) of another and if h is injective then h is called code [53].

The notion of a morphism is very important in combinatorics of words. A mapping $\theta : V^* \to V^*$ is called a *morphic involution* of V^* if $\theta(xy) = \theta(x)\theta(y)$ for any $x, y \in V^*$ (morphism), and θ^2 is equal to the identity (involution). Throughout this chapter, θ denotes an morphic involution.

5.1 Robustness of θ-Primitive Words

A word $w \in V^*$ is a *pseudo-power* of a non-empty word $t \in V^+$ relative to θ, or simply θ-power of t, if $w \in t\{t, \theta(t)\}^*$. Conversely, t is called *pseudo-period* of w relative to θ, or simply θ-period of w if $w \in t\{t, \theta(t)\}^*$. A word w is θ-primitive if there exists no non-empty word $t \in V^+$ such that w is a θ-period of t and $|w| > |t|$ [36]. We represent language of θ-primitive words as Q_θ.

For example, let $V = \{a, b, c\}$ be the alphabet and $\theta : V^* \to V^*$ be a morphic involution define as

$$\theta(a) = b, \theta(b) = a \text{ and } \theta(c) = c.$$

Then the word $w = abcbac$ is primitive but not θ-primitive, and θ-period of w is abc. $a^m b^n$

is not θ-primitive but $a^m c^n$ is θ-primitive for $m,\ n \geq 1$.

Next we discuss some robustness of θ-primitive words.

5.1.1 Ins-Robustness of θ-Primitive Words

For an involution morphism θ on the alphabet V, a θ-primitive word w of length n is said to be ins-robust θ-primitive word if the word $\mathit{pref}(w,i)\ .\ a\ .\ \mathit{suf}(w,n-i)$ is a θ-primitive word for all $i \in \{0,1,\ldots,n\}$ where $a \in V$.

We denote the set of all ins-robust θ-primitive words over an alphabet V by $Q_{\theta I}$. A θ-primitive word w is said to be non-ins-robust if w is θ-primitive but $w \notin Q_{\theta I}$. The set of non-ins-robust θ-primitive words is denoted by $Q_{\theta \bar{I}}$. Clearly the language of ins-robust θ-primitive words is a subset of the language of θ-primitive words.

Lemma 5.1. *Let $\theta : V^* \to V^*$ be a morphic involution. A θ-primitive word w is non ins-robust if and only if w can be expressed in the form of $u_1u_2\ldots u_i\ldots u_k$ where $u_j \in \{u,\theta(u)\}$ for $1 \leq j(\neq i)\ \leq k$ and $u_i = u_{i_1}u_{i_2}$ such that $u_{i_1}cu_{i_2} \in \{u,\theta(u)\}$ for some $c \in V$, $u_{i_1}, u_{i_2} \in V^*$ and $u \in Q$.*

Proof. Only if part: Let us consider a word $w = u_1u_2\ldots u_{i_1}u_{i_2}\ldots u_k$ where $u_j \in \{u,\theta(u)\}$ for $1 \leq j(\neq i) \leq k$ and $u_{i_1}cu_{i_2} \in \{u,\theta(u)\}$ for some $c \in V$. Now insertion of the letter c in w (between u_{i_1} and u_{i_2}) gives the exact θ-power of u. Hence it is a non-θ-primitive word after insertion of c at some position of w. Therefore, w is non-ins-robust θ-primitive word.

If part: Let w be a θ-primitive word but not ins-robust. Then there exists a decomposition $w = w_1w_2$ such that w_1cw_2 is not a θ-primitive word for some letter $c \in V$. Hence, w_1cw_2 is θ-power of word p for some $p \in Q$. Therefore $w_1 = up_1$ and $w_2 = p_2v$ such that $p_1cp_2 \in \{p,\theta(p)\}$ and uv is θ-power of p . □

Corollary 5.1. *If $w \in Q_{\theta I}$ then $rev(w) \in Q_{\theta I}$ for an involution morphism θ.*

Proof. Let $w \in Q_{\theta I}$, but $rev(w)$ is not ins-robust, i.e., $rev(w) = u_1u_2\ldots u_i\ldots u_k$ where $u_j \in \{u,\theta(u)\}$ for $1 \leq j(\neq i) \leq k$, $u_{i_1}cu_{i_2} \in \{u,\theta(u)\}$ for some $c \in V$ and $u_i = u_{i_1}u_{i_2}$. Then the word $w = rev(rev(w)) = rev(u_1u_2\ldots u_i\ldots u_k) = rev(u_k)\ldots rev(u_i)\ldots rev(u_1)$. $rev(u_j) \in \{rev(u),\theta(rev(u))\}$ for $1 \leq j(\neq i) \leq k$, and $rev(u_i) = rev(u_{i_2})rev(u_{i_1})$, $rev(u_{i_2})$ $c\ rev(u_{i_1}) \in \{rev(u),\theta(rev(u))\}$. Since after insertion of c, the word become θ-power of $rev(u)$. Hence w is not an ins-robust θ-primitive word, which is a contradiction. Therefore, if $w \in Q_{\theta I}$ then $rev(w) \in Q_{\theta I}$. □

For a morphic involution $\theta : V^* \to V^*$, the language of ins-robust θ-primitive words need not be reflective. For example, let $\theta(a) = b$, $\theta(b) = a$ and $\theta(c) = c$. $bbbcabacbaabca \in Q_{\theta I}$ but $bcabacbaabcabb \notin Q_{\theta I}$.

We can easily prove that for a morphic involution $\theta : V^* \to V^*$, the language of θ-primitive words is reflective if and only if θ is identity function.

Theorem 3.16 need not be true for non-ins-robust pseudo-primitive words. For example, if $u = abcbacab \in Q_{\theta\bar{I}}$ where $\theta(a) = b$, $\theta(b) = a$ and $\theta(c) = c$. There is no non-θ-primitive word of length $|u| - 1$ in uu. In next lemma we discuss the condition on the words so that the theorem holds on non-ins-robust pseudo-primitive words.

Lemma 5.2. *Let $u = u_1u_2$ be a non-ins-robust θ-primitive word such that $u_1au_2 = v\theta(v)$ for some $a \in V$, $u_1, u_2 \in V^*$ and $v \in V^+$. For a morphic involution θ over alphabet V, the word uu contains at least one θ-periodic word of length $|u|$ with θ-period p such that p divides of length $|u| + 1$ and $p \leq |u|$.*

Proof. If $u = u_1u_2$ such that $u_1au_2 = v\theta(v)$ for some $a \in V$ and $v \in V^+$, then there are two cases, either u_1a is prefix of v or au_2 is suffix of $\theta(v)$.

Case(A). If u_1a is prefix of v, i.e., $v = u_1av'$ such that $u = u_1v'\theta(u_1\ a\ v')$, then $uu = u_1v'\theta(u_1av')\ u_1v'\theta(u_1av')$. Here, $\theta(v')u_1v'\theta(u_1)\theta(a)$ is a non-ins-robust θ-primitive word of length $|u|$ and $\theta(\theta(v')u_1) = v'\theta(u_1)$.

Case(B). Similarly we can prove for the case au_2 is suffix of $\theta(v)$. □

5.1.2 Context-freeness of $Q_{\theta I}$

For a morphic involution, we prove that the language of ins-robust θ-primitive words over an alphabet V ($|V| \geq 3$) is not regular and also show that the language of non-ins-robust θ-primitive words is not context-free.

Theorem 5.1. *$Q_{\theta I}$ is not regular for an involution morphism θ.*

Proof. Suppose that the language $Q_{\theta I}$ over an alphabet $V = \{a,\ b,\ c\}$ is regular for an involution morphism θ such that $\theta(a) = a$, $\theta(b) = c$ and $\theta(c) = b$. Then there exist a natural number $n > 0$ depending upon the number of states of finite automaton for $Q_{\theta I}$. Consider the word $w = a^nca^mb, m > n + 1$ and $m \neq 2n$. Note that $w \in Q_{\theta I}$, where $|V| \geq 3$ and a, b and c are distinct symbols. Since $w \in Q_{\theta I}$ and $|w| \geq n$, then it must satisfy the other conditions of pumping Lemma for regular languages. So there exist a decomposition of w into x, y and z such that $w = xyz, |y| > 0$ and $xy^iz \in Q_{\theta I}$ for all $i \geq 0$.

Let $x = a^k,\ y = a^{(n-j)},\ z = a^{j-k}ca^mb$. Now choose $i = x_j$ and since we know by Lemma 3.8 that for every $j \in \{0, 1, \ldots, n-1\}$, there exists a positive integer $x_j > 1$ such that

$xy^{x_j}z = a^k a^{(n-j)x_j} a^{j-k} ca^m b = a^{(n-j)x_j+j} ca^m b = a^m ca^m b = (a^m c)\theta(a^m c) \notin Q_{\theta I}$ which is a contradiction. Hence the language of ins-robust θ-primitive words $Q_{\theta I}$ is not regular. $\square$

We know that, $Q_{\overline{I}}$ is not a context-free language for alphabet V such that $|V| \geq 2$ by Lemma 3.9. But $Q_{\theta\overline{I}}$ is regular for alphabet V such that $|V| = 2$ if $\theta \neq id_V$, where id_V is identity mapping over V. In the next theorem, we discuss for $|V| \geq 3$.

Theorem 5.2. *Let θ be an involution morphism. $Q_{\theta\overline{I}}$ is not a context-free language for alphabet V such that $|V| \geq 3$.*

Proof. Let $V = \{a,\ b,\ c\}$ be an alphabet. Assume that for an involution morphism θ, $Q_{\theta\overline{I}}$ is context-free language such that $\theta(a) = a$, $\theta(b) = c$ and $\theta(c) = b$. Let $p > 0$ be an integer which is the pumping length for the language $Q_{\theta\overline{I}}$. Consider the string $s = a^{p+1}c^{p+1}a^{p+1}b^p$, where $a,\ b,\ c \in V$ are distinct. It is easy to see that $s \in Q_{\theta\overline{I}}$ and $|s| \geq p$.

Hence, by the Pumping Lemma 2.6, s can be written in the form $s = uvwxy$, where u, v, w, x, and y are factors, such that $|vwx| \leq p$, $|vx| \geq 1$, and uv^iwx^iy is in $Q_{\theta\overline{I}}$ for every integer $i \geq 0$. By the choice of s and the fact that $|vwx| \leq p$, we have one of the following possibilities for vwx:

(a) $vwx = a^j$ for some $j \leq p$.

(b) $vwx = a^jc^k$ for some j and k with $j + k \leq p$.

(c) $vwx = c^j$ for some $j \leq p$.

(d) $vwx = c^ja^k$ for some j and k with $j + k \leq p$.

(e) $vwx = a^jb^k$ for some j and k with $j + k \leq p$.

(f) $vwx = b^j$ for some $j \leq p$.

In Case (a), since $vwx = a^j$, therefore $vx = a^t$ for some $t \geq 1$ and hence $uv^iwx^iy = a^{p-t+1}c^{p+1}a^{p+1}b^p \notin Q_{\theta\overline{I}}$ for $i = 0$.

Case (b) can have several subcases.

(i) $v = a^{j_1},\ w = a^{j_2},\ x = a^{j_3}c^k$.

(ii) $v = a^{j_1},\ w = a^{j_2}c^{k_1},\ x = c^{k_2}$.

(iii) $v = a^jc^{k_1},\ w = c^{k_2},\ x = c^{k_3}$.

In Case (1), Case (2) and Case (3) if we take $i = 4$, $uv^iwx^iy \neq Q_{\theta\bar{I}}$.

Similarly, we can obtain contradiction in rest of the case (5.4), case (5.4), case (5.1.2) and case (5.1.2) by choosing a suitable i.

Therefore, our initial assumption that $Q_{\theta\bar{I}}$ is context-free, must be false. □

5.1.3 Other Robustness of θ-Primitive Words

A θ-primitive word w of length n is said to be del-robust θ-primitive word if and only if the word $pref(w,i)$. $suf(w,n-i-1)$ is a θ-primitive word for any $i \in \{0,1,\ldots,n-1\}$.

Theorem 5.3. *A θ-primitive word w is not del-robust if and only if w can be expressed in the form of $w_1u_1cu_2w_2$ where $w_1, w_2, u_1, u_2 \in V^*$, $u_1u_2 = u$, $w_1, w_2 \in \{u, \theta(u)\}^*$ but $w_1w_2 \in \{u, \theta(u)\}^+$ and $c \in V$.*

Proof. We prove the sufficient and necessary conditions below.

($\Leftarrow$) Let us consider a word $w = w_1u_1cu_2w_2$ where $w_1, w_2, u_1, u_2 \in V^*$, $u_1u_2 = u$, $w_1, w_2 \in \{u, \theta(u)\}^*$ but $w_1w_2 \in \{u, \theta(u)\}^+$ and $c \in V$. Now deletion of the letter c in w gives the exact θ-power of u which is not a θ-primitive word. Hence, w is not a del-robust θ-primitive word.

($\Rightarrow$) Let w be a θ-primitive word but not del-robust. Then there exists a decomposition $w = w_1cw_2$ for $c \in V$ such that w_1w_2 is not a θ-primitive word. That is, $w_1w_2 \in p\{p, \theta(p)\}^+$ for some $p \in Q$. Therefore $w_1 = w_1'p_1$ and $w_2 = p_2w_2$ such that $p_1p_2 \in \{p, \theta(p)\}$ and $w_1', w_2' \in \{p, \theta(p)\}^*$ such that $w_1'w_2' \in \{p, \theta(p)\}^+$. Hence proved.

□

Lemma 5.3. *If w is del-robust θ-primitive then $rev(w)$ is also del-robust θ-primitive.*

Proof. We prove this by contradiction. Let w is del-robust θ-primitive but $rev(w)$ is not del-robust. Therefore, $rev(w) = w_1u_1cu_2w_2$ where $w_1, w_2, u_1, u_2 \in V^*$, $u_1u_2 = u$, $w_1, w_1 \in \{u, \theta(u)\}^*$ but $w_1w_2 \in \{u, \theta(u)\}^+$ and $c \in V$. Then the word $w = rev(w_1u_1cu_2w_2) = rev(w_2)\ rev(u_2)\ \ c\ rev(u_1)\ rev(w_1)$ where $rev(w_1), rev(w_2) \in \{rev(u), \theta(rev(u))\}^*$ but $rev(w_1)rev(w_2) \in \{rev(u), \theta(rev(u))\}^+$ and since $u = u_1u_2$, so $rev(u) = rev(u_2)rev(u_1)$. By Theorem 5.3, w is not a del-robust θ-primitive word, which is a contradiction. Therefore, $rev(w)$ is also del-robust θ-primitive. □

Cyclic permutation of a del-robust θ-primitive word need not be del-robust. For example $\theta : V^* \to V^*$ such that $\theta(a) = b$, $\theta(b) = a$ and $\theta(c) = c$, then $aacbabc$ is del-robust θ-primitive word but $acbabca$ is not.

5.2 θ-Superprimitive Words

A word $w \in V^+$ is θ-primitive if there exists no non-empty word $t \in V^+$ such that w is a θ-power of t and $|w| > |t|$. The θ-primitive root of w, denoted by $\rho_\theta(w)$, is the shortest word t such that w is a θ-power of t.

A string w covers another string z if for every $i \in \{1, \ldots, |z|\}$ there exists a $j \in \{1, \ldots, |w|\}$ such that there is an occurrence of w starting at position $i - j + 1$ in string z. A string z is quasiperiodic if z is covered by $w \neq z$, and the ordered sequence of all occurrences of w in z is called the w-cover of z. A string z is superprimitive if it is not quasiperiodic.

If a string u is simultaneously a θ-prefix and a θ-suffix of string x then u is a θ-border of x. The longest nontrivial θ-border of x is denoted by $T_\theta(x)$. By convention, we refer to $T_\theta(x)$ as the θ-border of x and to other as pseudo-border of x. Let $t_{\theta x}$ be a θ-quasiperiod of x. A word u is called θ-cover of a word w if w can be written as concatenation or superposition u or $\theta(u)$ or both.

Definition 5.1. *A word, w, is θ-quasiperiodic if there exist a word x such that x is θ-cover of w and $|x| < |w|$. A word is θ-superprimitive if it is not θ-quasiperiodic.*

Example Let $\theta : \{a, b, c, d\}^* \to \{a, b, c, d\}^*$ be a morphic involution defined by $\theta(a) = c$, $\theta(c) = a$, $\theta(b) = d$, and $\theta(d) = b$. Then the word $w = adcbb$ is θ-superprimitive, while its conjugate $w' = badcb$ is not. $abcbabcabccba$ is θ-quasiperiodic for $\theta(a) = c$, $\theta(b) = b$, $\theta(c) = a$ and its θ-cover is abc.

Lemma 5.4. *A θ-superprimitive word is superprimitive.*

Proof. Suppose that w is a θ-superprimitive word but not superprimitive. Then there exists some $t \in L$ which is cover of w and $|t| < |w|$, therefore t is also a θ-cover for w, which is a contradiction. □

Converse need not be true. Let $\theta(a) = b$, $\theta(b) = a$ and $\theta(c) = c$. Here θ is an involution as $\theta(\theta(a)) = a$. $u = acbca$ is superprimitive but not a θ-superprimitive as acb is a θ-cover of u.

Lemma 5.5. *The θ-cover of a word is θ-superprimitive.*

Proof. Let $w \in V^+$ and t be its θ-cover. Suppose, now that t is not θ-superprimitive. Then there exists a word $s \in V^*$ such that s is a θ-cover of t and $|s| < |t|$. Since $\theta(t)$ is covered by either s or $\theta(s)$. Thus, s is a θ-cover of w. However, this contradicts t being the θ-cover of w because $|s| < |t|$. □

Alternate Proof Let θ-cover of a word be not θ-superprimitive, then the θ-cover (say u) must have a proper cover say u' such that $|u'| < |u|$ and so u' can cover entire u, which is a contradiction, as u is the smallest word which is θ-cover of the word. Hence u is θ-superprimitive.

Corollary 5.2. *The θ-cover of a word is superprimitive and so it is primitive.*

Proof. The proof follows from Lemmas 5.4 and 5.5. □

A quasiperiodic word need not be reflective. For example, $ababa$ is quasiperiodic, but $babaa$ is not quasiperiodic that is superprimitive. A θ-quasiperiodic word need not be reflective. For example, $abcbabc$ is θ-quasiperiodic for $\theta : V^* \to V^*$ $\theta(a) = c$, $\theta(b) = b$, $\theta(c) = a$, but $cabcbab$ is not θ-quasiperiodic that is θ-superprimitive.

Lemma 5.6. *If y is a θ-border of x and $|y| \geq |t_{\theta x}|$ for a θ-border $t_{\theta x}$ of x, then $t_{\theta x}$ is θ-cover of y.*

Proof. Since $|y| \geq |t_{\theta x}|$ and $t_{\theta x}$ is a θ-border of x, then s is also a border of y. We distinguish two cases:

Case A $|y| \leq 2|t_{\theta x}|$. Then, every symbol of y is θ-covered by at least one of the two occurrences of $t_{\theta x}$ or $\theta(t_{\theta x})$, that start at positions 1 and $|y| - |t_{\theta x}| + 1$ of y, respectively.

Case B $|y| > 2|t_{\theta x}|$. Then, there exists some string u such that $y = t_1 u t_2$, where $t_1, t_2 \in \{t_{\theta x}, \theta(t_{\theta x})\}$. However, since $t_{\theta x}$ θ-covers x, we know that every symbol in u is θ-covered by an occurrence of $t_{\theta x}$ or $\theta(t_{\theta x})$. Therefore, $t_{\theta x}$ is θ-cover of y. □

Corollary 5.3. *If y is a θ-border of x and $|y| \geq |t_{\theta x}|$, then, for any θ-quasiperiod $t_{\theta y}$ of y, $t_{\theta y} \in \{t_{\theta x}, \theta(t_{\theta x})\}$.*

Proof. Assume first that $|t_{\theta y}| \geq |t_{\theta x}|$. Since $t_{\theta y}$ is a θ-border of y and y is a θ-border of x, then $t_{\theta y}$, is a θ-border of x. If $|t_{\theta y}| \geq |t_{\theta x}|$, we have by Lemma 5.6 that $t_{\theta x}$ θ-covers $t_{\theta y}$. Now if $|t_{\theta y}| \leq |t_{\theta x}|$, then $t_{\theta x}$ is a θ-border of y and $|t_{\theta x}| \geq |t_{\theta y}|$, therefore by Lemma 5.6, $t_{\theta y}$ θ-covers $t_{\theta x}$. In either case, one has that $t_{\theta y} \in \{t_{\theta x},\ \theta(t_{\theta x})\}$, as $t_{\theta x}$ and $t_{\theta y}$ are both θ-superprimitive by definition of θ-quasiperiod. □

Lemma 5.7. *Any word x has at most two θ-quasiperiods (if one is denoted by $Q_\theta(x)$ then other is $\theta(Q_\theta(x))$).*

Proof. Assume that x has any third distinct θ-quasiperiod, denoted as $t_{\theta x}$ and assume to fix the ideas that $|Q_\theta(x)| > |t_{\theta x}|$. Let y be a border of x such that $|y| > |Q_\theta(x)|$. If $t_{\theta y}$ is θ-cover of y, then by Corollary 5.3, we have that $t_{\theta y} \in \{Q_\theta(x), \theta(Q_\theta(x))\}$ and $t_{\theta y} \in \{t_{\theta x}, \theta(t_{\theta x})\}$. Therefore, $t_{\theta x} \in \{Q_\theta(x), \theta(Q_\theta(x))\}$. □

Lemma 5.8. *If $T_\theta(x)$ is the θ-border of x and $Q_\theta(T_\theta(x)) = Q_\theta(x)$.*

Proof. Since $T_\theta(x)$ is the θ-border of x, then $|T_\theta(x)|$ is maximum among all nontrivial psuedo-borders with respect to θ and in particular it is not shorter than $Q_\theta(x)$. The claim then follows from Corollary 5.3. □

5.2.1 Pseudo L-Primitive Words

For a morphic involution $\theta : V^* \to V^*$ and a language L, we call a word $w \in V^+$, θ_L-primitive if there exists no non-empty word $t \in L$ such that w is a θ-power of t and $|w| > |t|$. We define the θ_L-primitive root (in short, θ_L-root) of w, denoted by $\rho_{\theta L}(w)$, as the shortest word $t \in L$ such that w is a θ-power of t and there is no word $x \in L$ which is θ_L-root of t and $|x| < |t|$.

We represent the set of θ_L-primitive words as QL_θ and set of non-θ_L-primitive words as ZL_θ. This is obvious that $QL_\theta \subseteq QL$ ($ZL \subseteq ZL_\theta$).

For example, for $L = \{ab, abab, ba\}$, $\theta : V^* \to V^*$, a morphism involution, such that $\theta(a) = b$ and $\theta(b) = a$, $abba$ is L-primitive, but not θ_L-primitive, as $abba = ab\theta(ab)$.

Proposition 5.1. *Let f be a λ-free morphism of $L \subseteq V^*$ with $|V| \geq 2$. Then $f(QL) \cap QL$ is infinite if f is injective,*

Proof. The words $a^n b^n$ are L-primitive for all $n \geq 2$. Let $u = a^m b^m$ with $m \geq 2$ and suppose that $f(u)$ is not L-primitive. Let $f(a) = p^r$, $f(b) = q^s$ where $p, q \in Q$. Then $f(u) = p^{rm} q^{sm}$ with $rm, sm \geq 2$. Since $f(u)$ is not L-primitive as it is not primitive which is only possible if $p = q$. Hence $f(ab) = f(ba)$, a contradiction. □

Lemma 5.9. *Let L be a language over an alphabet V and $\theta : V^* \to V^*$, a morphic involution on V^*. Then the θ_L-primitive root of a word is θ_L-primitive.*

Proof. Let $w \in V^+$ and $t = \rho_{\theta L}(w)$ be its θ_L-primitive root, that is, w is a θ-power of $t \in L$. Suppose, now that t is not θ_L-primitive. Then there exists a word $s \in L$ such that t is a θ-power of s and $|s| < |t|$. Since t is a θ-power of s, thus, w is a θ-power of s, which contradicts t being the θ_L-primitive root of w because $|s| < |t|$ and $s \in L$. $\square$

Lemma 5.10. *Let L be a language over an alphabet V and $\theta : V^* \to V^*$, a morphic involution on V^*. Then a θ_L-primitive word is L-primitive.*

Proof. Suppose that w is a θ_L-primitive word but not L-primitive. Then there exists some $t \in L$ such that $w = t^n$ with $n \geq 2$ and $|t| < |w|$, therefore w is also a θ-power of t, which is a contradiction. $\square$

Converse need not be true. Since θ is not the identity function, there exists a word $u \in L$ such that $\theta(u) \neq u$. Then, if we take $w = u^3\theta(u)$, then w is not θ_L-primitive, but w is L-primitive as if w and t^k have common prefix of length $|t| + |u| - 1$ for some $t \in L$ then by Fine and Wilf Theorem [22], t and u have same root and so it is for $\theta(u)$, which implies that $u = \theta(u)$, contradiction.

The θ_L-primitive root of a word need not be θ-primitive and so need not to be primitive. Let L $\{bb, cc\}$ such that $\theta(b) = c$ and $\theta(c) = b$. For a word $w = bbccbbbbcc$, θ_L-primitive roots is bb or cc which are neither θ-primitive nor primitive. The θ-primitive roots of w are b and c.

The class of θ_L-primitive words is not necessarily closed under circular permutations. For example, Let $\theta : \{a, b, c, d\}^* \to \{a, b\}^*$ be a morphic involution such that $\theta(a) = b$, $\theta(b) = a$, $\theta(c) = d$, $\theta(d) = c$ and a language $L = \{ab, cd\}$. $w = babcdcdaba$ is θ_L-primitive but $abcdcdabab$ is not. Similarily we can show that the class of θ-superprimitive words is not necessarily closed under circular permutations.

For a morphism θ, a language L is called θ-closed if for every $u \in L$, $\theta(u) \in L$.

5.3 Robustness of Primitive Morphism

A morphism $f : V^* \to V^*$ is k**-primitive** if for all $x \in Q$ and $|x| \leq k$, $f(x) \in Q$, where $k \geq 1$. The morphism f is **primitive** if it is k-primitive for all $k \geq 1$. A morphism f is called **uniform** if $|f(a)| = |f(b)|$ for all $a, b \in V$ and $a \neq b$. A morphism f is called 1-*uniform* if $|f(a)| = 1$ for all $a \in V$. A word v is *morphically primitive* if, for every word w with $|w| \leq |v|$, there do not exist morphisms $h, h' : V^* \to V^*$ satisfying $h(v) = w$ and $h'(w) = v$, and we call v *morphically imprimitive* if it is not morphically primitive.

Definition 5.2. *A morphism $f : V^* \to V^*$ is k-del-robust-primitive if for all $x \in Q_D$ and $|x| \leq k$, $f(x) \in Q_D$. The morphism f is del-robust-primitive if it is k-del-robust-primitive for all $k \geq 1$.*

Example. Define a morphism $f : V^* \to V^*$, such that $f(a) = ab$ and $f(b) = a$, f is primitive morphism but not del-robust-primitive morphism, as, $ab \in Q_D$ but $f(ab) = aba \notin Q_D$.

Theorem 5.4. *A* 1-$uniform$ *primitive morphism is del-robust-primitive morphism.*

Proof. Since f is 1-$uniform$ primitive morphism, therefore $f(a) \neq f(b)$ for $a \neq b$. $|f(a)| = 1$ for every $a \in V$. If $w \in Q_D$ then w can not be written as either u^r, $r \geq 2$ or $u^r u_1 a u_2 u^s$, $r + s \geq 1$, $r, s \geq 0$ where $u, u_1, u_2 \in V^*$ and since f is primitive morphism, so $f(w)$ can neither be written as $f(u)^r$ nor $f(u)^r f(u_1) f(a) f(u_2) f(u)^s$ and so $f(w) \in Q_D$. Therefore f is del-robust-primitive morphism. $\square$

Let $f : V^* \to V^*$ be a morphism. Denote by f_Q the set of all the primitive words $u \in Q$ such that $f(u) \in Q$ and f_Z the set of all the primitive words $u \in Q$ such that $f(u) \in Z$ [15].

Definition 5.3. *Let f be a morphism of V^*. Denote by f_D the set of all the del-robust primitive words $u \in Q_D$ such that $f(u) \in Q_D$ and by f'_D the set of all the del-robust primitive words $u \in Q_D$ such that $f(u) \notin Q_D$, i.e. $f(u) = p^n$ or $f(u) = p^r p_1 a p_2 p^s$, $p \in Q$, $p_1, p_2 \in V^*$, $a \in V$, $p_1 p_2 = p$, $r, s \geq 0$, $n \geq 2$ and $r + s \geq 1$.*

Lemma 5.11. *[15] Let f be a morphism of V^*. Then*

(a) The languages f_Q and f_Z are reflective.

(b) If f is injective, then $u, v \in f_Z$, $u \neq v$ imply $uv \notin f_Z$ and $uv \in Q$.

Proposition 5.2. *Let f be a morphism of V^*. Then the languages f_D and f'_D are reflective.*

Proof. If $f_D = \phi$, this is immediate. Suppose $f_D \neq \phi$ and let $uv \in f_D$. Then $uv \in Q_D$ and $f(u)f(v) = f(uv) \in Q_D$. Since Q_D is reflective and f is morphism, then $vu \in Q_D$ and $f(vu) = f(v)f(u) \in Q_D$. Therefore, $vu \in f_D$. If $f'_D = \phi$, this is immediate. Suppose $f'_D \neq \phi$ and let $uv \in f'_D$. Then $uv \in Q$ and $f(u)f(v) = f(uv) \in \overline{Q_D}(= V^* \setminus Q_D)$. Since Q_D and $\overline{Q_D}$ are reflective, then $vu \in Q_D$ and $f(vu) = f(v)f(u) \in \overline{Q_D}$. Therefore, $vu \in f'_D$. $\square$

Denote by f_d the set of all the primitive words $u \in Q$ such that $f(u) \in Q_D$ and by f'_d the set of all the primitive words $u \in Q$ such that $f(u) \notin Q_D$.

Proposition 5.3. *Let f be a morphism of V^*. Then the languages f_d and f'_d are reflective.*

Proof. If $f_d = \phi$, this is immediate. Suppose $f_d \neq \phi$ and let $uv \in f_d$. Then $uv \in Q$ and $f(u)f(v) = f(uv) \in Q_D$. Since Q and Q_D are reflective, then $vu \in Q$ and $f(vu) = f(v)f(u) \in Q_D$. Therefore, $vu \in f_d$. If $f'_d = \phi$, this is immediate. Suppose $f'_d \neq \phi$ and let $uv \in f'_d$. Then $uv \in Q$ and $f(u)f(v) = f(uv) \in \overline{Q_D}(= V^* \setminus Q_D)$. Since Q and $\overline{Q_D}$ are reflective, then $vu \in Q$ and $f(vu) = f(v)f(u) \in \overline{Q_D}$. Therefore, $vu \in f'_d$. □

An injective morphism may not be del-robust. For example, define f on $V = \{a, b\}$, s.t., $f(a) = b$ and $f(b) = aba$. f is injective morphism, but $f(ab) = baba \notin Q_D$.

If f is a morphism of V^*, the word u is said to be f**-reductible** if $f(u) = p^m$, $p \in Q$, $m \geq 2$. Since Q, Q_D, $Q_{\overline{D}}$ and Z are reflective, then uv is f-reductible if and only if vu is f-reductible.

Proposition 5.4. *[15] Let f be an injective morphism of V^*.*

(a) If $u, v \in V^+$ with $uv \neq vu$ are f-reductible, then uv is not f-reductible.

(b) If uv is f-reductible and $uv \neq vu$, then either u or v is not f-reductible.

A word $u \in V^+$ is said to be **universally primitive** or simply u-primitive if for every injective morphism f of V^*, the word $f(u)$ is primitive [15]. Hence a u-primitive word is a word that is not f-reductible for every injective morphism of V^*. Let Q_U denote be the set of all the u-primitive words of V^*. Clearly $Q_U \subseteq Q$ Since $V \cap Q_U = \phi$, the inclusion is strict.

Definition 5.4. *A word $u \in V^+$ is said to be **universally del-robust primitive** or simply ud-primitive if for every injective morphism f of V^*, the word $f(u)$ is del-robust primitive. Let Q_{UD} denote be the set of all the ud-primitive words of V^*. Clearly $Q_{UD} \subseteq Q$. Since $V \cap Q_{UD} = \phi$, the inclusion is strict.*

Proposition 5.5. *[15] Let $w = u^m v^n$ with $u, v \in V^+$ and $m, n \geq 2$. Then the following properties are equivalent:*

(a) w is u-primitive,

(b) u and v have different roots,

(c) $uv \neq vu$.

Proposition 5.6. *Any λ-free morphism injective morphism on the set $\{a^n b^n b^n a^n \mid n \geq 2\}$ is a subset of Q_D.*

Proof. Since f is injective morphism, therefore $f(uv) \neq f(vu)$ for $uv \neq vu$, $u, v \in V^*$ and $u \neq v$ and $f(a^n b^n)$ is primitive. Let there be an injective morphism f such that $f(a^n b^n b^n a^n) \notin Q_D$ for some $n \geq 2$. Then $f(a^n b^n b^n a^n) \in Z$ or $f(a^n b^n b^n a^n) \in Q_{\overline{D}}$, which is not possible as f is λ-free injective. □

Proposition 5.7. *If a word w is ud-primitive then w is u-primitive.*

Proof. If the word w is ud-primitive then $f(w) \in Q_D$ for every injective morphism f, and so $f(w) \in Q$. Therefore w is u-primitive. □

The converse need not be true as $f(w) = (aba)^2abab \in Q \setminus Q_D$.

Proposition 5.8. *Let $w = u^m v^n$ with $u, v \in Q$, $u \neq v$ and $m, n \geq 2$. If w is ud-primitive then $w \in Q_D$ and u and v have different roots and $uv \neq vu$.*

Proof. Since $w = u^m v^n \in Q$ and for every injective morphism $f(w) \in Q_D$, and so $w \in Q_D$. Since w is ud-primitive and so it is u-primitive. Therefore by proposition 5.5, u and v have different roots and $uv \neq vu$. □

The set $\{a^n b^n b^n a^n \mid n \geq 2,\ a, b \in V\}$ contains only ud-primitive words and hence Q_{UD} is infinite.

5.4 Robustness of Abelian Primitive Words

Let $V = \{a_1, \ldots, a_n\}$ be an alphabet. The Parikh vector of a word $w \in V^*$ is $\psi(w) = (|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_n})$. For the alphabet $V = \{a, b\}$, we assume $a < b$. Thus, for example $\psi(abbaabb) = (3, 4)$. A word w is a n-th Abelian power if $x = y_1 y_2 \ldots y_n$ for some $y_1, y_2, \ldots, y_n \in V^*$ such that for all $2 \leq i \leq n$, $\psi(y_i) = \psi(y_1)$.

A word w is Abelian primitive (or A-primitive, for short) if w is not a k-th Abelian power for every $k \geq 2$. For an alphabet V, the set of all A-primitive words $w \in V^*$ is denoted by $AQ(V)$ or simply AQ if V is understood.

Definition 5.5 (Substitute-Robust Abelian Primitive Word). *A primitive word w of length n is said to be subst-robust Abelian-primitive word (or subst-robust A-primitive word) if and only if the word*

$$pref(w, i)\ .a.\ suf(w, n - i - 1)$$

is an A-primitive word for all $i \in \{0, 1, \ldots, n-1\}$ and for all $a \in V$.

For example, the word $abbababaa$ is not a subst-robust A-primitive word and the words $a^n b^n$ for $n \geq 2$ are subst-robust A-primitive words.

The collection of all subst-robust A-primitive words over an alphabet V is denoted by AQ_S.

Lemma 5.12. *If $w \in AQ_S$ then $rev(w) \in AQ_S$.*

Proof. Let $w \in AQ_S$ such that $rev(w)$ is not a subst-robust A-primitive word. Therefore, $rev(w) = p_1.p_2 \ldots p_i^{'} a p_i^{''} \ldots p_k$ where $\psi(p_i) = \psi(p)$ for some $p \in Q$ and $p = p_1 b p_2$ for some $b \neq a$. Then the word $w = rev(p_1.p_2 \ldots p_i^{'} a p_i^{''} \ldots p_k) = rev(p_k)\ rev(p_i^{''})\ \ a\ rev(p_i^{'})\ rev(p_1)$ and $\psi(p_1\ b\ p_2) = \psi(p) = \psi(rev(p_2)\ b\ rev(p_1))$. By Proposition 3.2, w is not a subst-robust A-primitive word, which is a contradiction. Therefore, if $w \in AQ_S$ then $rev(w) \in Q_S$. □

Definition 5.6. *A word w is del-robust Abelian primitive (or DA-primitive, for short) if the primitive word w can not be written as $uu_1u_2 \ldots u_i^{\prime} a u_i" \ldots u_n$ for some $u, u_i \in V^*$, $u_i = u_i^{\prime} u_i"$ for some $1 \leq i \leq n$ and $a \in V$ where $\psi(u_i) = \psi(u)$ for all $1 \leq i \leq n$.*

We denote the set of del-robust abelian primitive word as AQ_D.

The word $w = aabbab$ is A-primitive and DA-primitive as well, while $u = aabbabbab$ is not a DA-primitive word, as $u = xy_1by_2$ where $x = aabb$,$y_1 = a$, $y_2 = bab$, $y = y_1y_2 = abab$ and $\psi(x) = \psi(y) = (2, 2)$.

We know that the language of del-robust primitive words Q_D over an alphabet V is reflective by Theorem 3.3. Similarly, we have the property of reflectivity for the language of del-robust abelian primitive words AQ_D.

Lemma 5.13. *If $w \in AQ_D$ then $rev(w) \in AQ_D$.*

Proof. We prove this by contradiction. Let $w \in AQ_D$ such that $rev(w)$ is not a del-robust abelian primitive word. Therefore, $rev(w) = p_1.p_2 \ldots p_i^{'} a p_i^{''} \ldots p_k$ for $p_l \in Q$, $1 \leq l \leq k$ and $p_i = p_i^{'} p_i^{''}$ such that such that for all $1 \leq i, j \leq k$, $\psi(p_i) = \psi(p_j)$. Then the word $w = rev(p_1.p_2 \ldots p_i^{'} a p_i^{''} \ldots p_k) = (rev(p_k)) \ldots\ rev(p_i^{''})\ \ a\ rev(p_i^{'})\ rev(p_1)$ and since $p_i = p_i^{'} p_i^{''}$, so $rev(p) = rev(p_i^{''}) rev(p_i^{'})$. Therefore w is not a del-robust abelian primitive word, which is a contradiction. Therefore, if $w \in AQ_D$ then $rev(w) \in AQ_D$. □

Corollary 5.4. *If $w \in AQ$ then $rev(w) \in AQ$.*

Proof. Similar to lemma 5.13. □

The language AQ is not reflective. For example, $aabb \in AQ$ but $abba \in \overline{AQ}$. AQ_D is not reflective. $aabbb \in AQ_D$ but $baabb \in \overline{AQ_D}$.

Theorem 5.5. *$AQ_{\overline{D}}$ is not a context-free language.*

Proof. By contradiction, let us assume that $AQ_{\overline{D}}$ is not a context-free language. Let $p > 0$ be an integer which is the pumping length for the language $AQ_{\overline{D}}$. Consider the string $s = a^p b^p c^p b^p c^p a^{p+1}$, where $a,\ b \in V$ are distinct. It is easy to see that $s \in AQ_{\overline{D}}$ and $|s| \geq p$.

Hence, by the Pumping Lemma 2.6, s can be written in the form $s = uvwxy$, where u, v, w, x, and y are factors, such that $|vwx| \leq p$, $|vx| \geq 1$, and uv^iwx^iy is in $AQ_{\overline{D}}$ for every integer $i \geq 0$. By the choice of s and the fact that $|vwx| \leq p$, we have one of the following possibilities for vwx:

(a) $vwx = a^j$ for some $j \leq p$.

(b) $vwx = a^jb^k$ or b^jc^k or c^jb^k for some j and k with $j + k \leq p$.

(c) $vwx = b^j$ for some $j \leq p$.

(d) $vwx = c^ja^k$ for some j and k with $j + k \leq p$.

In Case (a), since $vwx = a^j$, therefore $vx = a^t$ for some $t \geq 1$ and hence $uv^iwx^iy = a^{p-t}b^pc^pb^pa^{p+1}c^p \notin AQ_{\overline{D}}$ for $i = 3$.

Case (b) can have several subcases. We prove it for $vwx = a^jb^k$. For b^jc^k, proof will be similar.

(i) $v = a^{j_1},\ w = a^{j_2},\ x = a^{j_3}b^k$.

(ii) $v = a^{j_1},\ w = a^{j_2}b^{k_1},\ x = b^{k_2}$.

(iii) $v = a^jb^{k_1},\ w = b^{k_2},\ x = b^{k_3}$.

In Case (a), Case (b) and Case (c) if we take $i = 4$, $uv^iwx^iy \neq AQ_{\overline{D}}$.

Similarly, we can obtain contradiction in Case (c) and Case (d) by choosing a suitable i. Therefore, our initial assumption that $AQ_{\overline{D}}$ is context-free, must be false. □

Theorem 5.6. *AQ_D is not regular.*

Proof. Let us suppose that the language AQ_D is regular. Then there exist a natural number $n > 0$ depending upon the number of states of finite automaton for AQ_D.

Consider the word $w = a^nba^mb, n > m + 2$. Note that $w \in AQ_D$. Since $|w| \geq n$, then it must satisfy the other conditions of pumping Lemma for regular languages. So there exist a decomposition of w into x, y and z such that $w = xyz, |y| > 0$ and $xy^iz \in AQ_D$ for all $i \geq 0$.

Let $x = a^k,\ y = a^{(n-j)},\ z = a^{j-k}ba^mb$. Now choose $i = x_j$ and since we know by Lemma 3.8 that for every $j \in \{0, 1, \ldots, n-1\}$, there exists a positive integer $x_j > 1$ such that $xy^{x_j}z = a^ka^{(n-j)x_j}a^{j-k}ba^mb = a^{(n-j)x_j+j}ba^mb = a^mba^mb = (a^mb)^2 \notin AQ_D$ which is a contradiction. Hence AQ_D is not regular. □

We know that the set AQ is not context-free [54]. Next we show that AQ_D over an alphabet V ($|V| \geq 3$) is not context-free.

Lemma 5.14. *For a prime $p \geq 2$, the word $x = aabbc(ab)^{p-2}$ is del-robust A-primitive.*

Proof. $|x| = 2p + 1$ for $x \in M$. If x is not del-robust A-primitive, then one of three cases occurs:
(a) $\psi(x) = \psi(u).2p + \{0, 0, 1\}$ for some *letter* u,
(b) $x = u_1 u_2 c u_3 \dots u_p$ for words $u_1, \dots, u_p$ of length two such that $\psi(u_i) = \psi(u_j)$, $1 \leq i, j \leq p$ and $i \neq j$
or
(c) Otherwise.

The case (a) cannot occur because x contains occurrences of a, b and c. In case (b), since we would have $u_1 = aa$ and $u_2 = bb$ so this case is also not possible.

Thus, we must have that $x = v_1 v_2$ for $|v_1| = p + 1$ and $|v_2| = p$. If $p = 2$ or 3, then x is del-robust. If $p > 3$ and $v_1 = aabbc(ab)^{(p-5)/2}a$ which has Parikh vector $((p-5)/2 + 3, (p-5)/2 + 2, 1)$, and $v_2 = b(ab)^{(p-1)/2}$ which has Parikh vector $((p-1)/2, (p-1)/2 + 1, 0)$. We can see that the number of occurrences of a in v_1 is even, while in v_2 it is odd or vice versa. Therefore $aabb(ab)^{p-2}$ is A-primitive and so $aabbc(ab)^{p-2}$ is del-robust A-primitive. ☐

Lemma 5.15. $AQ_D \cap aabbc(ab)^* = \{aabbc(ab)^{p-2} \mid p \ is \ prime\}$.

Proof. Let $M = AQ_D \cap aabbc(ab)^*$. The if part is immediate from Lemma 5.14. For the only if part, let $x \in M$. Then $|x| = 2n$ for some $n \geq 2$. Suppose, on contrary x is not of the form $aabbc(ab)^{p-2}$ for some prime p. Then we must have that n is not prime. Let q be a prime factor of n and note that $x = (aabbc(ab)^{q-2}).((ab)^q)^{n/q-1}$ and that all factors of length $2q$ have q occurrences of a, q occurrences of b and one symbol c. Further, $aabb(ab)^{q-2}$ is an A-primitive root after deletion of a symbol c from x by Lemma 5.14. Thus, x is not del robust A-primitive, which is a contradiction. ☐

We can now show that the set of all del-robust A-primitive words is not context-free.

Theorem 5.7. *Let V be an alphabet such that $|V| \geq 3$. The set AQ_D over V is not context-free.*

Proof. We prove that M is not context-free. Let $M' = h^{-1}((aabb)^{-1}M)$ where $h : \{a, c\}^* \to \{a, b, c\}^*$ is the morphism $h(a) = ab$ and $h(c) = c$. Then $M' = \{ca^{p-2} \mid p \ is \ prime\}$. As the context-free languages are closed under quotient by regular sets and inverse homomorphism, M' is context-free if M is context-free. But as M' is unary after deletion of c from each element of M'. Since M' is not regular, by the pumping lemma. Thus, M and so AQ_D are not context-free. ☐

5.5 Conclusions

In this chapter, We have discussed the characterizations of pseudo-superprimitive words and pseudo-L-primitive words and identified several properties. We have investigated the robustness of primitive morphism and some results on universally primitive words. We have discussed the robustness of abelian primitive words and proved that the language of del-robust abelian primitive words is not context free.

Yatra Naryastu Pujyante Ramante Tatra Devata,
Yatraitaastu Na Pujyante Sarvaastatrafalaah Kriyaah.
(Manusmriti Verse 3.56)

Meaning: "Where Women Are Honored , Divinity Blossoms There; And Where they Are Dishonored , All Action Remains Unfruitful."

Chapter 6

Conclusion and Future Directions

The main motivation of this thesis is to advance our understanding of various primitive words and their robustness with respect to various operations. By providing approaches for robustness on L-primitive words, pseudo-primitive words, superprimitive words and pseudo-quasiperiodic words, we can hopefully gain more insight into the general problem of primitive words. Each of the chapters leave scope for future directions. These are some of the obvious steps to take towards advancing the current state-of-the-art.

1) It is shown that Q_D is reflective. It is also proved that the language of non-del-robust primitive words $Q_{\overline{D}}$ is not context-free. We have also presented a linear time algorithm to test if a given word is del-robust primitive. Finally, we have given a lower bound on the number of del-robust primitive words of a given length There are several interesting questions that remain unanswered about del-robust primitive words. Some of them that we plan to explore in immediate future are:
(i) Is Q_D^i for $i \geq 2$ regular? It is known that Q^i for $i \geq 2$ is regular [50] where Q is the set of primitive words.
(ii) It is known that the language of primitive words Q is accepted by 2DPDA [13]. Is the language of del-robust primitive words Q_D accepted by 2-way deterministic context-free?
(iii) Is the language of del-robust primitive words Q_D deterministic context-free? We believe that the properties we have identified for del-robust primitive words will be helpful in answering these questions.

2) We have characterized ins-robust primitive words and identified several properties and proved that the language of ins-robust primitive words Q_I is not regular. We also proved that the language of non-ins-robust primitive words $Q_{\overline{I}}$ is not context-free. We identified that Q_I is dense over an alphabet V. We have also presented a linear time

algorithm to test if a given word is ins-robust primitive. Finally, we have given a lower bound on the number of ins-robust primitive words of a given length.

There are several interesting questions that remain unanswered about ins-robust primitive words. Some of them that we plan to explore in immediate future are as follows. Is Q_I^i for $i \geq 2$ regular? It is known that Q^i for $i \geq 2$ is regular [55]. We conjecture that the the language of ins-robust primitive words Q_I is accepted by 2DPDA and indexed grammar [56, 57]. We also conjecture that the language of ins-robust primitive words Q_I is not a deterministic context-free language. We believe that the properties we have identified for ins-robust primitive words will be helpful in answering these questions.

3) It has been proved that the languages of non-exchange-robust primitive words are not context-free. We mention some of the interesting questions that are still unanswered. (1) Is the language Q_X context-free? (2) One can consider to exchange two symbols at any positions and preserve primitivity. It is an open problem to get a linear time algorithm to recognize exchange-robust primitive word.

4) It is proved that the language Q_S is reflective and $Q_{\overline{S}}$ is not a context-free language.

   There are several interesting questions that remain unanswered about subst-robust primitive words. Some of them that we plan to explore in immediate future are: (i) Is Q_S^i for $i \geq 2$ regular? (ii) Is the language of subst-robust primitive words Q_S deterministic context-free?

5) A word is L-primitive if it is not a proper power of a shorter word from the language L. It is shown that the exchange of any two consecutive distinct symbols in a non-L-primitive word w, $alph(w) \geq 2$, make it L-primitive word. If $w = x_1abx_2 \in ZL$ then $x_1bax_2 \in QL$. It also shown that the language $Q_{\overline{X}}$ need not be dense over the alphabet V. It is proved that the language of non-exchange-robust L-primitive words may be context-free for some language L and the language of exchange robust primitive words QL_X is accepted by a 2DPDA.

6) A special type of Primitive words (pseudo-superprimitive words) are defined which is based on pseudo-primitivity and superprimitivity of words. There are still to discuss the robustness of languages of pseudo-periodic, quasiperidic and pseudo-superprimitive words. There is future scope to discuss the robustness of pseudo-primitive words for the other morphisms viz. antimorphic involution, morphism without involution etc.

   We have characterized ins-robust pseudo-primitive words and identified several properties. and proved that the language of ins-robust primitive words $Q_{\theta I}$ is not context-free. We have introduced some new terms say, pseudo L-primitive word and pseudo-superprimitive words and identified some properties. Finally, we have discussed robustness for morphism.

We mention some of the interesting questions that are still unanswered. (i) Are the languages of subst-robust abelian primitive words AQ_S, ins-robust abelian primitive words AQ_I and exchange-robust abelian primitive words AQ_X context-free? (ii) Is there a linear time algorithm to find the θ-quasiperiod of a string?

Bibliography

[1] M. Lothaire. *Combinatorics on words*. Cambridge University Press, 1997.

[2] Monsieur Lothaire. *Algebraic combinatorics on words*. Number 90. Cambridge University Press, 2002.

[3] M. Lothaire. *Applied combinatorics on words*, volume 105. Cambridge University Press, 2005.

[4] Michael A. Harrison. *Introduction to Formal Language Theory*. Addison-Wesley Longman Publishing Co., Inc., 1978.

[5] Jean Berstel and Dominique Perrin. *Theory of Codes*. Academic Press, 1985.

[6] Dan Gusfield. *Algorithms on Strings, Trees and Sequences: Computer Science and computational biology*. Cambridge University Press, 1997.

[7] Lila Kari and Kalpana Mahalingam. Watson–crick palindromes in dna computing. *Natural Computing*, 9(2):297–316, June 2010.

[8] Donald E. Knuth, James H. Morris, JR., and Vaughan R. Pratt. Fast pattern matching in strings. *SIAM Journal on Computing*, 6(2):323–350, 1977.

[9] Maxime Crochemore and Wojciech Rytter. *Jewels of Stringology: Text Algorithms*. World Scientific, 2003.

[10] Pál Dömösi, Sándor Horváth, Masami Ito, László Kászonyi, and Masashi Katsura. Formal languages consisting of primitive words. In *Fundamentals of Computation Theory*, pages 194–203. Springer, 1993.

[11] Roger C. Lyndon and Marcel Paul Schützenberger. The equation $a^M = b^N c^P$ in a free group. *The Michigan Mathematical Journal*, 9:289–298, 1962.

[12] Johanna Nichols and David A. Peterson. The amerind personal pronouns. *Language*, pages 336–371, 1996.

[13] Holger Petersen. The ambiguity of primitive words. *STACS 94*, pages 679–690, 1994.

[14] M. Ito, M. Katsura, H.J. Shyr, and S.S. Yu. Automata accepting primitive words. In *Semigroup Forum*, volume 37, pages 45–52. Springer, 1988.

[15] Gheorghe Păun, Nicolae Santean, Gabriel Thierrin, and Sheng Yu. On the robustness of primitive words. *Discrete Applied Mathematics*, 117(1):239–252, 2002.

[16] H.J. Shyr and S.S. Yu. Non-primitive words in the language p^+q^+. *Soochow Journal of Mathematics*, 20(4):535–546, 1994.

[17] Roman Kolpakov and Gregory Kucherov. Finding maximal repetitions in a word in linear time. In *40th Annual Symposium on Foundations of Computer Science*, pages 596–604. IEEE, 1999.

[18] Roman Kolpakov and Gregory Kucherov. On maximal repetitions in words. In *Fundamentals of Computation Theory*, pages 374–385. Springer, 1999.

[19] J. Shallit. *A second course in formal languages and automata theory*. Cambridge University Press, 2008.

[20] Christian Choffrut and Juhani Karhumäki. Combinatorics of words. In *Handbook of formal languages*, pages 329–438. Springer, 1997.

[21] H.J. Shyr. Free monoids and languages, hon min book co. *Taichung, Taiwan*, 1991.

[22] Nathan J. Fine and Herbert S. Wilf. Uniqueness theorems for periodic functions. *Proceedings of the American Mathematical Society*, 16(1):109–114, 1965.

[23] DS Blank, MS Cohen, M Coltheart, J Diederich, BM Garner, RW Gayler, CL Giles, L Goldfarb, M Hadeishi, B Hazlehurst, et al. Connectionist symbol processing: Dead or alive? *Neural Computing Surveys*, 2:1–40, 1999.

[24] Stephen Cole Kleene. Representation of events in nerve nets and finite automata. Technical report, RAND PROJECT AIR FORCE SANTA MONICA CA, 1951.

[25] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[26] Carlos Martin-Vide, Alexandru Mateescu, and Victor Mitrana. Parallel finite automata systems communicating by states. *International Journal of Foundations of Computer Science*, 13(05):733–749, 2002.

[27] Zoltán Ésik. A note on isomorphic simulation of automata by networks of two-state automata. *Discrete Applied Mathematics*, 30(1):77–82, 1991.

[28] Ferenc Gécseg. *Products of automata*, volume 7. Springer Science & Business Media, 2012.

[29] Jeffrey Jaffe. A necessary and sufficient pumping lemma for regular languages. *ACM SIGACT News*, 10(2):48–49, 1978.

[30] Hans-Jörg Kreowski. A pumping lemma for context-free graph languages. In *International Workshop on Graph Grammars and Their Application to Computer Science*, pages 270–283. Springer, 1978.

[31] David S Wise. A strong pumping lemma for context-free languages. *Theoretical Computer Science*, 3(3):359–369, 1976.

[32] Marcel Paul Schützenberger. On context-free languages and push-down automata. *Information and control*, 6(3):246–264, 1963.

[33] Gerhard Lischke. Primitive words and roots of words. *arXiv preprint arXiv:1104.4427*, 2011.

[34] Huei-Jan Shyr and Gabriel Thierrin. Disjunctive languages and codes. In *International Conference on Fundamentals of Computation Theory*, pages 171–176. Springer, 1977.

[35] Ito Masami and Domosi Pal. *Context-free Languages And Primitive Words*. World Scientific, 2014.

[36] Elena Czeizler, Lila Kari, and Shinnosuke Seki. On a special class of primitive words. *Theoretical Computer Science*, 411(3):617–630, 2010.

[37] Gerth Stølting Brodal and Christian NS Pedersen. Finding maximal quasiperiodicities in strings. In *Annual Symposium on Combinatorial Pattern Matching*, pages 397–411. Springer, 2000.

[38] Mireille Régnier and Laurent Mouchard. Periods and quasiperiods characterization. In *Annual Symposium on Combinatorial Pattern Matching*, pages 388–396. Springer, 2000.

[39] Shubh Narayan Singh and K.V. Krishna. L-primitive words in submonoids of a free monoid. In *Recent Advances in Mathematics, Statistics and Computer Science*, pages 322–330. World Scientific, 2016.

[40] Francine Blanchet-Sadri. Primitive partial words. *Discrete Applied Mathematics*, 148(3):195–213, 2005.

[41] H.J. Shyr and Din-Chang Tseng. Some properties of dense languages. *Soochow J. Math*, 10:127–131, 1984.

[42] Pál Dömösi and Géza Horváth. The language of primitive words is not regular: two simple proofs. *Bulletin of European Association for Theoretical Computer Science*, 87:191–194, 2005.

[43] Wang Lijun. Count of primitive words. *Applied Mathematics-A Journal of Chinese Universities*, 16(3):339–344, 2001.

[44] Amit Kumar Srivastava, Ananda Chandra Nayak, and Kalpesh Kapoor. On del-robust primitive words. *Discrete Applied Mathematics*, 2016.

[45] Pál Dömösi and Masami Ito. *Context-free languages and primitive words*. World Scientific, 2014.

[46] William Ogden. A helpful result for proving inherent ambiguity. *Mathematical systems theory*, 2(3):191–194, Sep 1968.

[47] Seymour Ginsburg and Sheila Greibach. Deterministic context free languages. In *Switching Circuit Theory and Logical Design, 1965. SWCT 1965. Sixth Annual Symposium on*, pages 203–220. IEEE, 1965.

[48] Gheorghe Paun and Gabriel Thierrin. Morphisms and primitivity. *Bulletin of the European Association for Theoretical Computer Science*, 61:85–88, 1997.

[49] Cao Chunhua, Yang Shuang, and Yang Di. Some kinds of primitive and non-primitive words. *Acta Informatica*, pages 1–8, 2014.

[50] Masami Ito Pál Dömösi. *Primitive Words in Languages*, chapter 9, pages 267–308. World Scientific, 2014.

[51] Pál Domosi and Géza Horváth. The language of primitive words is not regular: two simple proofs. *BULLETIN-EUROPEAN ASSOCIATION FOR THEORETICAL COMPUTER SCIENCE*, 87:191, 2005.

[52] P. Dömösi, S. Horváth, and M. Ito. On the connection between formal languages and primitive words. 1991.

[53] Juhani Karhumäki. Combinatorics of words.

[54] Michael Domaratzki and Narad Rampersad. Abelian primitive words. *International Journal of Foundations of Computer Science*, 23(05):1021–1033, 2012.

[55] C.M. Reis and H.J. Shyr. Some properties of disjunctive languages on a free monoid. *Information and Control*, 37(3):334 – 344, 1978.

[56] Zvi Galil. Two way deterministic pushdown automaton languages and some open problems in the theory of computation. In *Switching and Automata Theory, 1974., IEEE Conference Record of 15th Annual Symposium on*, pages 170–177. IEEE, 1974.

[57] Alfred V. Aho. Indexed grammars–an extension of context-free grammars. *Journal of the ACM (JACM)*, 15(4):647–671, 1968.

Publications

Journals

- **Amit Kumar Srivastava**, Ananda Chandra Nayak, Kalpesh Kapoor: On Del-Robust Primitive Words. in *Discrete Applied Mathematics* 206(2016): 115-121.

Conferences

- Ananda Chandra Nayak, **Amit Kumar Srivastava**: On Del-Robust Primitive Partial Words with One Hole. In 10^{th} *International Conference on Language and Automata Theory and Applications*, volume 9618 of Lecture Notes in Computer Science, pages 233-244. Springer International Publishing, 2016.
- Ananda Chandra Nayak, **Amit Kumar Srivastava**, Kalpesh Kapoor: On Exchange-Robust and Subst-Robust Primitive Partial Words. On 17^{th} *Italian Conferences on Theoretical Computer Science*, volume 1720 of *CEUR workshop proceedings*, pages 190-202, 2016.