Abstract. This work considers a natural generalization of primitivity with respect to a language. Given a language $L$, a nonempty word $w$ is said to be $L$-primitive if $w$ is not a proper power of any word in $L$. After ascertaining the number of primitive words in submonoids of a free monoid, the work proceeds to count $L$-primitive words in submonoids of a free monoid. The work also studies the distribution of $L$-primitive words in certain subsets of free monoids.

Introduction

A nonempty word which is not a power of any other word is called a primitive word. It is well known that every nonempty word can be uniquely expressed as a power of a primitive word [9]. The study of primitivity of words is often the first step towards the understanding of words and plays an important role in the theory of languages. Ito et al. have investigated the number of primitive words in the languages accepted by automata [6]. Shyr and Tseng have proved that any noncommutative submonoid of a free monoid contains infinitely many primitive words [11]. In the literature, there are various types of generalizations/extensions of the classic definition of primitive words [3, 4, 5, 7]. We propose yet another generalization of primitive word, viz. $L$-primitive word – a nonempty word that is not a proper power of any word in a given language $L$.

In this paper, we first investigate the primitive words in the submonoids of free monoids. We could ascertain that the number of primitive words in a submonoid of a free monoid is either at most one or infinity. Then, we study the distribution of $L$-primitive words in certain subsets of free monoids. In particular, we target to count the $L$-primitive words in the submonoids of free monoids.

The paper is organized as follows. In Section 1, we will present some necessary preliminaries of the paper. We introduce the concept of $L$-primitive words in Section 2 and study some basic properties. In Section 3, we count the number of $L$-primitive words in a language $L$ as well as in the submonoids of a free monoid. Finally, Section 4 concludes the paper.

1. Preliminaries

In this section, we present some basic definitions and fix our notations. For more details one may refer [1] [8] [12].

Let $A$ be a nonempty finite set called an alphabet with its elements as letters. The free monoid over $A$ is denoted by $A^*$ whose elements are called words, and $\varepsilon$ denotes the identity element of $A^*$ – the empty word. The set of all nonempty words over $A$ is denoted by $A^+$, i.e. $A^+ = A^* \setminus \{\varepsilon\}$.
A word \( u \) is said to be a prefix of a word \( v \) if there exists a word \( t \) such that \( ut = v \). A set \( X \) of words is called a prefix set if no element of \( X \) is a prefix of another word of \( X \). A power of a word \( u \) is a word of the form \( u^k \) for some \( k \in \mathbb{N} = \{0, 1, 2, \ldots \} \) – the set of natural numbers. It is convenient to set \( u^0 = \varepsilon \), for each word \( u \). If \( k \in \mathbb{N} \setminus \{0, 1\} \), we say that \( u^k \) is a proper power of \( u \). A word \( x \in A^+ \) is said to be a primitive word if it is not a proper power of another word in \( A^* \), i.e. for \( u \in A^+ \),

\[
x = u^k \implies k = 1.
\]

For a subset \( X \) of \( A^* \), we denote the set of all primitive words in \( X \) by \( X_p \). We recall the following well known property of primitive words.

**Proposition 1.1.** For every \( w \in A^+ \), there exists a unique primitive word \( u \) and a unique integer \( k \geq 1 \) such that \( w = u^k \).

The unique primitive word \( u \) obtained in Proposition 1.1 is called the primitive root of \( w \), denoted by \( \sqrt{w} \). By a language over an alphabet \( A \) is meant a subset of \( A^* \). The root of a language \( L \) (\( \subseteq A^* \)), denoted by \( \sqrt{L} \), is defined as

\[
\sqrt{L} = \{ \sqrt{w} \in A_+^* \mid w \in L \setminus \{\varepsilon\} \}.
\]

A language \( L \) is said to be commutative if \( uv = vu \), for all \( u, v \in L \). It is known that a language \( L \) is commutative if and only if there exists \( w \in A^* \) such that \( L \subset \{w\}^* \).

Now, we recall some properties of numerical monoids from [10]. A numerical monoid is a submonoid of the monoid \((\mathbb{N}, +)\) whose complement in \( \mathbb{N} \) is finite. For a nonempty subset \( X \) of \( \mathbb{N} \), the submonoid of \( \mathbb{N} \) generated by \( X \) is denoted by \( \langle X \rangle \), i.e.

\[
\langle X \rangle = \{ \lambda_1 x_1 + \cdots + \lambda_n x_n \mid n, \lambda_i \in \mathbb{N}, x_i \in X, \forall i (1 \leq i \leq n) \}.
\]

**Theorem 1.2.** Let \( X \) be a nonempty subset of \( \mathbb{N} \). The submonoid \( \langle X \rangle \) is a numerical monoid if and only if \( \gcd(X) = 1 \).

**Theorem 1.3.** Every numerical monoid admits a unique finite minimal set of generators.

**Theorem 1.4.** Any nontrivial submonoid of \( \mathbb{N} \) is isomorphic to a numerical monoid.

### 2. \( L \)-primitive words

In this section, we introduce the notion of primitive words relative to a language \( L \), called \( L \)-primitive words, and obtain some properties related to \( L \)-primitive words. We prove that every primitive word is an \( L \)-primitive word so that the latter notion is a generalization of the former one. Unless it is specified otherwise, in what follows, \( L \) is an arbitrary language over \( A \).

**Definition 2.1.** A word \( x \in A^+ \) is said to be an \( L \)-primitive word if \( x \) is not a proper power of any word in \( L \), i.e., for \( u \in L \),

\[
x = u^k \implies k = 1.
\]

**Notation 2.2.** Let \( X \subseteq A^* \) and \( X^c \) denotes the complement of \( X \) in \( A^* \).

(i) The set of \( L \)-primitive words in \( X \) is denoted by \( X_{L-p} \).

(ii) The set \( (X^*)_{L-p} \) of \( L \)-primitive words in \( X^* \) is simply denoted by \( X^*_{L-p} \).

(iii) The set \( (X^c)_{L-p} \) of \( L \)-primitive words in \( X^c \) is simply denoted by \( X^c_{L-p} \).

We begin with some basic properties of \( L \)-primitive words.
Definition 2.7.

(i) If \( L = \emptyset \), then \( A_{L-p}^* = A^+ \), the set of all nonempty words over \( A \).
(ii) If \( L = A^* \), then \( A_{L-p}^* = A_p^* \), the set of all primitive words over \( A \).

Proposition 2.4. If \( L_1 \) and \( L_2 \) are two subsets of \( A^* \), then
\[
L_1 \subseteq L_2 \implies A_{L_2-p}^* \subseteq A_{L_1-p}^*.
\]
Proof. On the contrary, let us assume that \( A_{L_2-p}^* \not\subseteq A_{L_1-p}^* \). Then there exists \( w \in A_{L_2-p} \), but \( w \not\in A_{L_1-p} \). Since \( w \not\in A_{L_1-p} \), there exists \( u \in L_1 \) such that \( w = u^k \), for some \( k > 1 \). In view of hypothesis, we have \( u \in L_2 \). Consequently, \( w \not\in A_{L_2-p}^* \); a contradiction.

Remark 2.3 (ii), we have the following corollary of Proposition 2.4.

Corollary 2.5. Every primitive word is an \( L \)-primitive word. Hence, if \( |A| \geq 2 \), then \( |A_{L-p}^*| = \infty \).

Remark 2.6. 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.

Definition 2.7. For \( w \in A^+ \), we define the set of \( L \)-primitive roots of \( w \), denoted by \( \sqrt[\mathbb{L}]{w} \), is defined as
\[
\sqrt[\mathbb{L}]{w} = \{ x \in A_{L-p}^* \mid x^k = w, \text{ for some } k \geq 1 \}.
\]
Further, for \( X \subseteq A^* \), the \( L \)-primitive root of \( X \), denoted by \( \sqrt[\mathbb{L}]{X} \), is defined as
\[
\sqrt[\mathbb{L}]{X} = \bigcup_{w \in X \setminus \{\varepsilon\}} \sqrt[\mathbb{L}]{w}.
\]

Remark 2.8. The primitive root of a nonempty word is an \( L \)-primitive root of the word. Thus, if \( w \neq \varepsilon \), then \( \sqrt[\mathbb{L}]{w} \neq \emptyset \).

3. \( L \)-PRIMITIVE WORDS IN VARIOUS SUBSETS

This section is divided into two subsections. We investigate \( L \)-primitive words in some sets related to \( L \) itself in Subsection 3.1. Then, we carry on the investigations on submonoids in Subsection 3.2.

3.1. \( L \)-PRIMITIVE WORDS IN \( L \). In this subsection, we make an attempt to investigate \( L \)-primitive words in \( L \) and also in \( L^\ell \). In this connection, we provide some sufficient conditions and characterizations. In fact, we give a relation between \( L \)-primitive words and \( L \)-primitive roots in \( L \).

Theorem 3.1. If \( \varepsilon \not\in L \), \( L \neq \emptyset \) if and only if \( L_{L-p} \neq \emptyset \).
Proof. Let us assume that \( L \neq \emptyset \) and choose \( w \in L \). If \( w \in L_{L-p} \), then we are through. Otherwise, there exists \( u \in L \) such that \( w = u^k \), for some \( k > 1 \). Clearly, \( |u| < |w| \). If \( u \in L_{L-p} \), then we are through. Otherwise, we continue to choose shorter words in \( L \) whose power is \( u \). But this process terminates at a finite stage and eventually we get a word \( x \in L_{L-p} \) and \( w = x^m \), for some \( m > 1 \). Hence, \( L_{L-p} \neq \emptyset \). The converse is straightforward.

Theorem 3.2. If \( L \subseteq A^+ \) is a prefix set, then \( L = L_{L-p} \).
Remark 3.3. The converse of Theorem 3.2 is not necessarily true. For instance, let \( L = \{a, ab\} \subseteq \{a, b\}^+ \). Clearly, \( L = L_{L^p} \), but \( L \) is not a prefix set.

It is clear that \( L_{L^p} \subseteq \sqrt{\mathcal{T}} \). Now, we explore the possibilities so that \( L_{L^p} = \sqrt{\mathcal{T}} \).

For this, we need the notion of power of a subset of \( A^* \) introduced by Calbrix and Nivat (cf. [2]). The power of a subset \( X \) of \( A^* \), denoted by \( \text{pow}(X) \), is defined as

\[
\text{pow}(X) = \{x^k \mid x \in X \text{ and } k \geq 1\}.
\]

Remark 3.4. Clearly, \( \text{pow}(A^*_{L^p}) = A^+ \).

Theorem 3.5.

(i) \( \sqrt{\mathcal{T}} \subseteq L \iff L_{L^p} = \sqrt{\mathcal{T}} \).

(ii) \( L^c = \text{pow}(L^c) \Rightarrow L_{L^p} = \sqrt{\mathcal{T}} \).

(iii) \( L \subseteq A^+ \Rightarrow L_{L^p} = \sqrt{\mathcal{T}} = L \).

Proof. We first note that \( L_{L^p} \subseteq \sqrt{\mathcal{T}} \).

(i) \((\Leftarrow)\) Since \( L_{L^p} \subseteq L \), from the hypothesis, we have \( \sqrt{\mathcal{T}} \subseteq L \).

\((\Rightarrow)\) Let \( x \in \sqrt{\mathcal{T}} \); then \( x \) is \( L \)-primitive word. Also, from the hypothesis, we have \( x \in L \). Thus, \( x \in L_{L^p} \). Hence, we have the part (i).

(ii) Let us assume that \( x \in \sqrt{\mathcal{T}} \setminus L_{L^p} \). Since \( x \) is an \( L \)-primitive word and \( x \notin L_{L^p} \), we have \( x \notin L \). Then, from the hypothesis, we have \( x^k \in L^c \), for all \( k \geq 1 \). But, since \( x \in \sqrt{\mathcal{T}} \), we have \( x \in \sqrt{w} \), for some \( w \in L \). That is, there is a number \( t \geq 1 \), such that \( x^t = \sqrt{w} \in L \); a contradiction. Hence, \( \sqrt{\mathcal{T}} = L_{L^p} \).

(iii) Clearly, \( L_{L^p} \subseteq L \). Let \( x \in L \); from the hypothesis, we have \( x \in A^+ \). By Corollary 2.3, since every primitive word is an \( L \)-primitive word, we have \( x \in L_{L^p} \). Thus, \( L = L_{L^p} \).

It is clear that for \( w \in L^c \), we have \( \sqrt{w} = \{w\} \). Since \( L \subseteq A^+ \), we have

\[
\sqrt{\mathcal{T}} = \bigcup_{w \in L} \sqrt{w} = \bigcup_{w \in L} \{w\} = L.
\]

Hence, if \( L \subseteq A^+ \), we have \( L_{L^p} = \sqrt{\mathcal{T}} = L \).

Corollary 3.6. \( L = \sqrt{\mathcal{T}} \iff L_{L^p} = \sqrt{\mathcal{T}} = L \).

Remark 3.7. The converse of Theorem 3.5(ii) is not necessarily true. For instance, consider \( L = \{a, b, ab\} \subseteq \{a, b\}^+ \). Observe that \( L_{L^p} = \sqrt{\mathcal{T}} = \{a, b\} \). Clearly, since \( a^2 \in L^c \), we have \( a^6 \in \text{pow}(L^c) \); but, \( a^6 \notin L^c \). Hence, \( L^c \neq \text{pow}(L^c) \).

Theorem 3.8. \( L = \text{pow}(L) \iff L_{L^p} = L^c \).

Proof.

\((\Rightarrow)\) Clearly, \( L_{L^p} \subseteq L^c \). Let \( x \in L^c \), but \( x \notin L_{L^p} \). There exists a word \( y \in L \) such that \( x = y^k \), for some \( k > 1 \). Since \( y \in L \), we have \( y^k \in \text{pow}(L) \). It follows that \( x \in \text{pow}(L) \). But, \( L = \text{pow}(L) \), we have \( x \in L \). This is a contradiction.
Theorem 3.9. Let $x \in \text{pow}(L)$, but $x \notin L$. There exists a word $y \in L$ such that $x = y^k$, for some $k > 1$. Since, $x \notin L$, we have $x \in L^*$. But, $L_{L-p}^* = L^*$, it follows that $x$ is an $L$-primitive word; which is a contradiction.

\[\square\]

3.2. $L$-primitive words in submonoids. In this subsection, for comparison, we first investigate the number of primitive words in the submonoids of a free monoid. Further, we study the $L$-primitive words in the submonoids of a free monoid. We count the $L$-primitive words in a submonoid of the free monoid over a unary alphabet. In this case, when $L$ is finite, we prove that a submonoid has either at most one or infinitely many $L$-primitive words. Finally, we leave certain remarks on estimating the number of $L$-primitive words over an arbitrary alphabet. We require the following theorem.

**Theorem 3.9** ([11]). Let $H$ be a submonoid of $A^*$. $H$ is noncommutative if and only if $|H_p| = \infty$.

Now, we observe that a submonoid of $A^*$ contains either at most one primitive word or infinitely many primitive words.

**Theorem 3.10.** Let $H$ be a submonoid of $A^*$; then either $|H_p| \leq 1$ or $|H_p| = \infty$.

**Proof.** If $H = \{\varepsilon\}$, then $|H_p| = 0$. Let us assume that $H \neq \{\varepsilon\}$. If $H$ is noncommutative, then by Theorem 3.9, we have $|H_p| = \infty$. Otherwise, we have $H \subseteq \{w\}^*$, for some word $w \in A^+$. Without loss of generality, assume that $w \in A_p^*$. Thus, according to $w \in H$ or not, we have $|H_p| = 1$ or 0. \[\square\]

**Corollary 3.11.** If $H$ is a nontrivial submonoid of $A^*$, then either $|\sqrt{H}| = 1$ or $\infty$.

Let $A = \{a\}$ be a unary alphabet. It is known that $A^*$ is isomorphic to the additive monoid of natural numbers $(\mathbb{N}, +)$ under the isomorphism given by $a^k \mapsto k$. Thus, each word $a^k$ of $A^*$ is characterized by its length $k \in \mathbb{N}$. Hence, we count the $L$-primitive words in the submonoids of $\mathbb{N}$, instead of $A^*$. In what follows, $H$ is a nontrivial submonoid of $\mathbb{N}$ and $L$ is a nonempty subset of $\mathbb{N}$. Now, we count the number of $L$-primitive words in $H$. We begin with the following remark.

**Remark 3.12.** If $1 \in L$, then according to $1 \in H$ or not, we have $|H_{L-p}| = 1$ or $|H_{L-p}| = 0$, respectively.

Let us assume that $1 \notin L$. In view of Theorem 1.4 and Theorem 1.3, let $y$ be the finite minimal generating set of $H$.

**Theorem 3.13.** If $\gcd(Y) = 1$, then $|H_{L-p}| = \infty$.

**Proof.** If $\gcd(Y) = 1$, by Theorem 1.2, the submonoid $H$ is a numerical monoid so that $|\mathbb{N} \setminus H| < \infty$. Thus, $H$ contains infinitely many prime numbers. Since $1 \notin L$, every prime number is $L$-primitive. Hence, $|H_{L-p}| = \infty$. \[\square\]

**Theorem 3.14.** If $L$ is a finite set and $\gcd(Y) > 1$, then $|H_{L-p}| \leq 1$ or $|H_{L-p}| = \infty$.

**Proof.** We first assume that $l \nmid d$, for all $l \in L$ and claim that $|H_{L-p}| = \infty$. Let $\gcd(Y) = d$. Since $d \neq 1$, by Proposition 1.2, the submonoid $H$ is not a numerical monoid. We define the function

$$f : H \rightarrow \mathbb{N} \text{ by } hf = \frac{h}{d},$$
Clearly, \( f \) is a monomorphism and therefore the image of \( f \), \( \text{Im}(f) \), is isomorphic to \( H \). By Theorem 1[2], the submonoid \( \text{Im}(f) \) is a numerical monoid.

Clearly, \( \text{Im}(f) \) has infinitely many prime numbers. Let \( p \in \text{Im}(f) \) be a prime number such that \( p > \max(L) \), then \( pd \in H \). By Euclid’s lemma, \( l \nmid pd \), for all \( l \in L \). Since \( \text{Im}(f) \) has infinitely many such prime numbers, we have \( |H_{L-p}| = \infty \).

Now, we assume that \( l \mid d \), for some \( l \in L \). Here, we determine \( |H_{L-p}| \) with respect to \( d \in L \) or not. If \( d \notin L \), then clearly \( |H_{L-p}| = 0 \). If \( d \in L \), we consider the cases \( d \in H \) or not. If \( d \notin H \), then clearly \( |H_{L-p}| = 0 \). In case \( d \in H \), if there is an \( l'(\neq d) \) which divides \( d \), then \( |H_{L-p}| = 0 \); otherwise \( |H_{L-p}| = 1 \). \( \square \)

**Remark 3.15.** If \( L \) is an infinite subset of \( \mathbb{N} \), then \( |H_{L-p}| \) need not satisfy the Theorem 3[4]. For instance, let \( H \) be the submonoid of \( \mathbb{N} \) generated by the set \( \{4, 6\} \) and \( L = \{4\} \cup \{\mathbb{P} \setminus \{2, 5\}\} \), where \( \mathbb{P} \) is the set of all prime numbers in \( \mathbb{N} \). We observe that \( H_{L-p} = \{4, 10\} \) and so \( |H_{L-p}| = 2 \). Similarly, if \( L = \{4\} \cup \{\mathbb{P} \setminus \{2, 5, 7\}\} \), then \( H_{L-p} = \{4, 10, 14\} \) and so \( |H_{L-p}| = 3 \).

In the following, we make certain remarks on the number of \( L \)-primitive words in the submonoids of a free monoid over an alphabet of size at least two. First observe that if the submonoid \( H \) is \( \{\varepsilon\} \), then \( |H_{L-p}| = 0 \). If \( H \neq \{\varepsilon\} \), then by Corollary 2[5] we have the following remark.

**Remark 3.16.** If \( H \) is a noncommutative submonoid of \( A^* \), where \( |A| \geq 2 \), then \( |H_{L-p}| = \infty \).

### 4. Conclusion

Motivated by the work of Ito et al. in [6], we have considered a study on the number of primitive words in the languages of semi-flower automata (SFA). SFA precisely accept finitely generated submonoids of free monoids [12]. We extended the study to submonoids of free monoids and observed that the number is either at most one or infinite. Further, we have initiated a study on the number of \( L \)-primitive words in submonoids of free monoids. If \( L \) is a finite, we have counted the number of \( L \)-primitive words in the submonoids of a free monoid over a unary alphabet. When \( L \) is infinite, the problem appears to be more complicated and a systematic study in this regard is necessary. In case the alphabet size is at least two, we could remark only on the number of \( L \)-primitive words in noncommutative submonoids. One can consider the problem in commutative submonoids.

### References

[1] J. Berstel and D. Perrin. *Theory of codes*, volume 117 of *Pure and Applied Mathematics*. Academic Press Inc., 1985.

[2] H. Calbrix and M. Nivat. Prefix and period languages of rational \( \omega \)-languages. In *Developments in language theory, II (Magdeburg, 1995)*, pages 341–349. World Sci. Publ., River Edge, NJ, 1996.

[3] E. Czeizler, L. Kari, and S. Seki. On a special class of primitive words. *Theoret. Comput. Sci.*, 411(3):617–630, 2010.

[4] M. Domaratzki. *Trajectory-based operations*. PhD thesis, Queen’s University, Canada, 2004.

[5] H. K. Hsiao, C. C. Huang, and S. S. Yu. Word operation closure and primitivity of languages. *J. UCS*, 8(2):243–256 (electronic), 2002. Advances and trends in automata and formal languages.

[6] M. Ito, M. Katsura, H. J. Shyr, and S. S. Yu. Automata accepting primitive words. *Semigroup Forum*, 37(1):45–52, 1988.

[7] L. Kari and G. Thierrin. Word insertions and primitivity. *Util. Math.*, 53:49–61, 1998.
[8] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Co., Reading, Mass., 1983.

[9] R. C. Lyndon and M. P. Schützenberger. The equation $a^M = b^N c^P$ in a free group. *Michigan Math. J.*, 9:289–298, 1962.

[10] J. C. Rosales and P. A. García-Sánchez. *Numerical semigroups*, volume 20 of *Developments in Mathematics*. Springer, New York, 2009.

[11] H. J. Shyr and D. C. Tseng. Some properties of dense languages. *Soochow J. Math.*, 10:127–131, 1984.

[12] S. N. Singh. *Semi-Flower Automata*. PhD thesis, IIT Guwahati, India, 2012. URL: http://www.iitg.ac.in/kvk/papers/sns-phd-thesis.pdf

Department of Mathematics, Central University of Bihar, Patna, India
E-mail address: shubh@cub.ac.in

Department of Mathematics, IIT Guwahati, Guwahati, India
E-mail address: kvk@iitg.ac.in