RAIRO-Theor. Inf. Appl. 56 (2022) 4
https://doi.org/10.1051/ita/2022004

RAIRO - Theoretical Informatics and Applications
www.rairo-ita.org

ON PRIMITIVE WORDS WITH NON-PRIMITIVE PRODUCT

Othman Echi*, Adel Khalfallah and Dhaker Kroumi

Abstract. Let $\mathcal{A}$ be an alphabet of size $n \geq 2$. Our goal in this paper is to give a complete description of primitive words $p \neq q$ over $\mathcal{A}$ such that pq is non-primitive. As an application, we will count the cardinality of the set $\mathcal{E}(l, \mathcal{A})$ of all couples (p, q) of distinct primitive words such that $|p| = |q| = l$ and pq is non-primitive, where l is a positive integer. Then we give a combinatorial formula for the cardinality $\varepsilon(n, l)$ of this set. The density in $\{(p, q) : p, q$ are distinct primitive words and $|p| = |q| = l\}$ of the set $\mathcal{E}(l, \mathcal{A})$ is also discussed.

Mathematics Subject Classification. 68R15, 68Q45.

Received August 16, 2021. Accepted February 1, 2022.

1. Introduction

Combinatorics on words is an area of research focusing on combinatorial properties of words applied to formal languages. It plays an important role in several mathematical research areas as well as theoretical computer science (see [2, 6, 9, 21]).

This paper deals with primitive words over a nontrivial alphabet $\mathcal{A}$ (having at least two letters). The empty word over $\mathcal{A}$ will be denoted by ε, $\mathcal{A}^*$ is the set of all words over $\mathcal{A}$ and $\mathcal{A}^+$ is the set of all nonempty words over $\mathcal{A}$. A primitive word over $\mathcal{A}$, is a nonempty word u which is not a proper power of another word (*i.e.*, if $u = v^m$ then $m = 1$). We denote by $\mathbf{Q}(\mathcal{A})$ the set of all primitive words over $\mathcal{A}$ and $\mathbf{Q}_l(\mathcal{A})$ the set of all primitive words of length l. Primitive and non-primitive words play a crucial role in algebraic coding theory and the theory of formal languages (see for instance Lothaire [15] and Shyr [21]).

Whether $\mathbf{Q}(\mathcal{A})$ is a context-free language or not is a well-known long-standing open problem posed by Dömsi *et al.* [4, 8]. This problem was the origin of most of the combinatorial studies of primitive words.

In [18], Reis-Shyr have proved that every non-empty word which is not a power of a letter is a product of two primitive words. So one may think that $\mathbf{Q}(\mathcal{A})$ is "very large" in some sense; in fact the natural density of the language of primitive words is 1 (see [19]).

One of the well-known classical results about primitive words is Shyr-Yu Theorem [20]. Letting $p^+ := \{p^n : n$ is a positive integer$\}$, the theorem states that if $p \neq q$ are primitive words, then the language p^+q^+ contains at most one non-primitive word of the form pq^m or p^mq (since p^nq^m is primitive for all $n, m \geq 2$ [16]).

In [20] Shyr-Yu gave a necessary condition when the product pq^m is a k-power of a primitive word (*i.e.*, $pq^m \in \mathbf{Q}^{(k)}(\mathcal{A})$), with $m, k \geq 2$. It is also worth noting that no information about the case $m = 1$ has been provided, as far as we know.

Keywords and phrases: Primitive words, primitive root, Möbius inversion formula.

Department of Mathematics, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia.

* Corresponding author: othechi@yahoo.com

© The authors. Published by EDP Sciences, 2022

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In [10], the author gave necessary and sufficient conditions to get $pq^m \in \mathbf{Q}^{(k)}(\mathcal{A})$, for $m \geq 1$ and $k \geq 2$, but the expression of p was not explicit.

The aim of this paper is to give a complete description of primitive words $p \neq q$ such that pq is non-primitive (*cf.*, Thm. 3.1). If, in addition $|p| = |q|$, the number of such couples (p, q) is computed and their density is discussed.

2. Preliminaries

For any integer $k \geq 1$, we denote by $\mathbf{Q}^{(k)}(\mathcal{A}) = \{p^k : p \in \mathbf{Q}(\mathcal{A})\}$. For two words u and v we say that u is a prefix (resp., a suffix) of v if there exist a word x (resp., a word y) such that $v = ux$ (resp., $v = yu$). Here, we present some well known results which we will need in the sequel. In the remainder, $\mathcal{A}$ denotes an alphabet of size $n \geq 2$.

Lemma 2.1 ([16]). *Let u, v be nonempty words over $\mathcal{A}$, then the following properties hold.*

1. *$uv = vu$ if and only if there exists a word w such that $u, v \in w^+$.*
2. *There exist a unique primitive word $\sqrt{u}$ (called* the primitive root *of u) and a unique positive integer $\mathbf{e}$ (called* the exponent *of u) such that $u = \sqrt{u}^{\mathbf{e}}$.*

The following lemma, which is the main ingredient to solve most problems on primitivity, is due to Fine-Wilf [11].

Lemma 2.2 (Fine-Wilf Theorem). *Let $u, v \in \mathcal{A}^+$. Then the following statements are equivalent.*

(i) There exists a word w such that $u, v \in w^+$.
(ii) There exist $i, j > 0$ so that u^i and v^j have a common prefix (suffix) of length $|u| + |v| - \gcd(|u|, |v|)$.

Lemma 2.3 ([14, Corollary 4],Shyr-Yu). *If $uq^m = g^k$ for some $m, k \geq 1$, $u \in \mathcal{A}^+$, and $g, q \in \mathbf{Q}(\mathcal{A})$, with $u \notin q^+$, then $g \neq q$ and $|g| > |q^{m-1}|$.*

The following lemma is a classical result in combinatorics on words, originally due to Lyndon-Schützenberger [16] (see also [1, 15, 17]). This lemma will play a crucial role in the proof of our main result (see Thm. 3.1).

Lemma 2.4 ([16]). *Let t, v be two distinct nonempty words over $\mathcal{A}$ such that $tu = uv$. Then there exists a unique pair of words (p, q) and a unique positive integer m such that pq is primitive, $t = (pq)^m$, $v = (qp)^m$, and $u = (pq)^j p$, for some integer $j \geq 0$.*

Lemma 2.5 ([20]). *Let $p \neq q \in \mathbf{Q}(\mathcal{A})$. Then the following properties hold.*

1. *The language p^+q^+ contains at most one non-primitive word.*
2. *If $|p| = |q|$, then $p^+q^+ \setminus \{pq\} \subseteq \mathbf{Q}(\mathcal{A})$.*

It is worth noting that an alternative proof of the previous result has been given in [7].

Studying local distribution of non-primitive words, Shyr-Tu have established the following result.

Lemma 2.6 ([22]). *Let $u \in \mathcal{A}^+$ and each $x \neq y \in \mathcal{A}^+$ such that $|x| = |y| \leq \frac{|u|}{2}$ and $x \neq y$, then either ux or uy is a primitive word.*

The following is a direct consequence of Theorems 13, 14 in [5].

Lemma 2.7 (Prefix-Suffix). *Let $q \in \mathbf{Q}(\mathcal{A})$, $x \in \mathcal{A}^+$, with $x \neq q$. Then the following properties hold.*

1. *If x is a prefix of q, then $q^k x$ is primitive for all $k \geq 2$.*
2. *If x is a suffix of q, then xq^k is primitive for all $k \geq 2$.*

Lemma 2.8 ([3]). *Let $p \neq q \in \mathbf{Q}(\mathcal{A})$ such that $|p| = r|q|$, for some integer $r \geq 2$. Then the following properties hold.*

1. pq^m *is primitive, for all* $m \geq r$.
2. $p^m q$ *is primitive for all* $m \geq 2$ *(even for* $r = 1$*).*

3. Primitive words With non-primitive product

Note that some information about non-primitive products of the form pq^m, for $m \geq 2$ was provided in [20]. But once again, it was not an "if and only if result", and the case $m = 1$ was not discussed. The following result gives a complete description of primitive words p, q such that pq is not primitive.

Theorem 3.1. *Let* p *and* q *be two distinct primitive words and* $k \geq 2$ *be a given integer. Then the following statements are equivalent.*

1. $pq \in \mathbf{Q}^{(k)}(\mathcal{A})$.
2. *One of the following statements holds.*
   (a) $p = (xq)^{k-1}x$, *with* $x \in \mathcal{A}^+$ *and* $xq \in \mathbf{Q}(\mathcal{A})$.
   (b) *There exist an integer* $1 \leq s \leq k-1$, *and* $\alpha, \beta \in \mathcal{A}^+$ *such that* $\alpha\beta \in \mathbf{Q}(\mathcal{A})$, $q = (\alpha\beta)^s\alpha$ *and* $p = (\beta\alpha)^{k-s-1}\beta$.

Proof. $(\mathbf{1}) \Longrightarrow (\mathbf{2})$. Assume that $pq \in \mathbf{Q}^{(k)}(\mathcal{A})$, so that $pq = g^k$ for some $g \in \mathbf{Q}(\mathcal{A})$. Clearly $|g| \neq |q|$, otherwise we have $g = q$, which implies that $p = g^{k-1}$. As p and g are primitive words, we get $p = g$ and $k - 1 = 1$. This contradicts the fact that $p \neq q$. Hence, we consider two mutually exclusive cases.

- *Case 1* $|g| > |q|$: By $pq = g^k$, we deduce that $g = xq$ for some $x \in \mathcal{A}^+$. Then, we have $p = g^{k-1}x = (xq)^{k-1}x$.

- *Case 2* $|g| < |q|$: As $pq = g^k$, there exists $y \in \mathcal{A}^+$ such that $q = yg$. Hence $py = g^{k-1}$. Of course $|y| \neq |g|$, otherwise we get $y = g$ and then $q = g^2$, a contradiction. Thus, two mutually exclusive subcases are to be considered: $|g| > |y|$ and $|g| < |y|$.

Sub-case 2-1 $|g| > |y|$: From the equality $py = g^{k-1}$, we deduce that there exists $y_1 \in \mathcal{A}^+$ such that $g = y_1 y$, which yields $p = g^{k-2}y_1$. Let us rewrite $q = yg = yy_1y = uv = tu$, where $u = y$, $v = g$ and $t = yy_1$. We have $v \neq t$, otherwise $g = yy_1 = y_1y$, which contradicts the fact that g is primitive. By Lemma 2.4, there exist two words α and β over $\mathcal{A}$ with $\alpha\beta \in \mathbf{Q}(\mathcal{A})$ and two integers $s \geq 1$ and $j \geq 0$ such that $yy_1 = (\alpha\beta)^s$, $g = (\beta\alpha)^s$ and $y = (\alpha\beta)^j\alpha$. As g and $\beta\alpha$ are primitive, we get $s = 1$, that is $g = \beta\alpha$, which yields $yy_1 = \alpha\beta = (\alpha\beta)^j\alpha y_1$. This implies that $j = 0$, and consequently we have $\beta = y_1$ and $\alpha = y$. We conclude that $q = \alpha\beta\alpha$ and $p = g^{k-2}y_1 = (\beta\alpha)^{k-2}\beta$, so that p and q have the desired form $q = (\alpha\beta)^s\alpha$ and $p = (\beta\alpha)^{k-s-1}\beta$, with $s = 1$.

Sub-case 2-2 $|g| < |y|$: Let us perform the Euclidean division of $|y|$ by $|g|$: $|y| = r|g| + h$ with $0 \leq h < |g|$. Suppose that $h = 0$, then from $py = g^{k-1}$, we get $y = g^r$, so that $q = yg = g^{r+1}$. This contradicts the fact that q is primitive. Hence, we have $h \neq 0$.

From $py = g^{k-1}$, we deduce that $y = y_1 g^r$, for some $y_1 \in \mathcal{A}^+$, with $|y_1| = h < |g|$ and $py_1 = g^{k-r-1}$. But as $|y_1| < |g|$, $g = g_1 y_1$ for some $g_1 \in \mathcal{A}^+$. Consequently, we get $p = g^{k-r-2}g_1$. Now, let us write $y = y_1 g^r = y_1 g^{r-1} g_1 y_1$. Letting $u = y_1$, $v = g^r$ and $t = y_1 g^{r-1} g_1$, we have $uv = tu$. Of course $v \neq t$, otherwise we have $g^r = y_1 g^{r-1} g_1$, which yields $(g_1 y_1)^r = y_1 (g_1 y_1)^{r-1} g_1 = (y_1 g_1)^r$. Since $g_1 y_1$ and $y_1 g_1$ are primitive, we deduce that $g = g_1 y_1 = y_1 g_1$, contradicting the fact that g is primitive. Therefore, by Lemma 2.4, there exist $\alpha, \beta \in \mathcal{A}^*$, $s \geq 1$ and $j \geq 0$ integers such that $\alpha\beta \in \mathbf{Q}(\mathcal{A})$, $t = y_1 g^{r-1} g_1 = (\alpha\beta)^s$, $v = g^r = (\beta\alpha)^s$ and $u = y_1 = (\alpha\beta)^j\alpha$. As g and $\beta\alpha$ are primitive, we deduce that $g = \beta\alpha$ and $r = s$. This leads to $(\alpha\beta)^j\alpha(\beta\alpha)^{s-1}g_1 = (\alpha\beta)^s$ and consequently we have $j = 0$, which yields $y_1 = \alpha$ and $g_1 = \beta$.

We conclude that $q = yg = y_1 g^r g = \alpha(\beta\alpha)^{s+1} = (\alpha\beta)^{s+1}\alpha$ and $p = g^{k-r-2}g_1 = (\beta\alpha)^{k-s-2}\beta = (\beta\alpha)^{k-s-2}\beta$.

Taking into consideration the fact that p and q are primitive words, in both subcases we have necessarily $\alpha, \beta \in \mathcal{A}^+$.

$(\mathbf{2}) \Longrightarrow (\mathbf{1})$. Straightforward.

□

Remark 3.2. According to Lemma 2.1 [16], in the previous theorem, the primitive root and the exponent of pq may be identified according to Conditions $(2) - a)$ and $(2) - b)$ of the theorem. Letting $l := |p| + |q|$, the following properties hold.

1. Under Assumption $(2)-a)$ of Theorem 3.1, $\sqrt{pq}=xq$ and the exponent of pq is given by $k=\frac{l}{|x|+|q|}$.
2. Under Assumption $(2)-b)$ of Theorem 3.1, $\sqrt{pq}=\beta\alpha$ and $k=\frac{l}{|\alpha|+|\beta|}$.

So, we obtain the following inequalities:

$$1<\frac{l}{\delta}\leq k\leq\frac{l}{d}<l,$$

where d (resp., δ) is the smallest (resp., greatest) divisor of l in the interval $(1,l)$.

The following proposition shows that every $k\in(1,l)$ is reached as an exponent of some pq, with $l=|p|+|q|$.

Proposition 3.3. *Let l be a composite positive integer and k be a proper divisor of l (i.e., $k\in(1,l)$). Then there exist two primitive words p,q over $\mathcal{A}$ such that pq is not primitive, $|p|+|q|=l$ and k is the exponent of pq.*

Proof. Let u be a primitive word with length $\frac{l}{k}$, and $x\in\mathcal{A},u_1\in\mathcal{A}^+$ such that $u=xu_1$. We let $p=x$, $q=u_1u^{k-1}$. We will show that q is primitive.

- If $k\geq 3$, then, as u_1 is a suffix of u, by the Prefix-Suffix Lemma 2.7, $q=u_1u^{k-1}$ is primitive.
- Now, if $k=2$, then $q=u_1u=u_1xu_1$ is conjugate to u_1^2x. So, according to Lemma 2.8 (second assertion), u_1^2x is primitive.

In both cases, p,q are primitive, $pq=u^k\in\mathbf{Q}^{(k)}(\mathcal{A})$ and $|p|+|q|=l$. □

4. Primitive words $p\neq q$ such that $|p|=|q|$ and pq is non-primitive

Let $p\neq q$ be primitive words. Then following Shyr-Yu 2.5, or Lentin-Schützenberger [14], we know that the language p^+q^+ contains at most one non-primitive word.

A well known class of couples (p,q) of distinct primitive words such that $p^+q^+\setminus\{pq\}\subseteq\mathbf{Q}(\mathcal{A})$, is

$$\{(p,q)\in\mathbf{Q}(\mathcal{A})\times\mathbf{Q}(\mathcal{A})\colon p\neq q,|p|=|q|\}.$$

The aim of this section is to shed light on this class.

Proposition 4.1. *Let p and q be two distinct primitive words. Then, the following statements are equivalent.*

1. *$|p|=|q|$ and $pq\notin\mathbf{Q}(\mathcal{A})$.*
2. *There exist two words $\alpha,\beta\in\mathcal{A}^+$ such that $|\alpha|=|\beta|$, $\alpha\beta\in\mathbf{Q}(\mathcal{A})$, $q=(\alpha\beta)^s\alpha$ and $p=(\beta\alpha)^s\beta$ for some integer $s\geq 1$.*

Proof. $(\mathbf{1})\Longrightarrow(\mathbf{2})$. First, we will show the direct implication. Assume that pq is not primitive. Then, there exists an integer $k\geq 2$ such that $pq\in\mathbf{Q}^{(k)}(\mathcal{A})$. Using Theorem 3.1, there exist an integer $s\geq 1$ and two nonempty words α and β such that $\alpha\beta\in\mathbf{Q}(\mathcal{A})$, $q=(\alpha\beta)^s\alpha$ and $p=(\beta\alpha)^{k-s-1}\beta$. On the other hand, we have $|p|=|q|$, from which we get $(s+1)|\alpha|+s|\beta|=(k-s)|\beta|+(k-s-1)|\alpha|$ and then $(k-2s)(|\alpha|+|\beta|)=2|\alpha|$. Necessarily we have $k-2s=1$, otherwise $k-2s\geq 2$, and consequently we get

$$2|\alpha|=(k-2s)(|\alpha|+|\beta|)\geq 2(|\alpha|+|\beta|)>2|\alpha|,$$

a contradiction. We conclude that $k=2s+1$. It follows that $q=(\alpha\beta)^s\alpha$ and $p=(\beta\alpha)^s\beta$.

$(\mathbf{2})\Longrightarrow(\mathbf{1})$. We finish the proof by showing the reverse implication. Assume that $q=(\alpha\beta)^s\alpha$ and $p=(\beta\alpha)^s\beta$, with $s\geq 1$, $|\alpha|=|\beta|$ and $\alpha\beta\in\mathbf{Q}(\mathcal{A})$. As a result, we have $|p|=(2s+1)|\alpha|=|q|$ and $pq=(\beta\alpha)^{2s+1}\in\mathbf{Q}^{(2s+1)}(\mathcal{A})$. □

In the remainder of this section, we will count the couples (p, q) of distinct primitive words of the same length such that pq is not primitive. For a given positive integer l, define the set

$$\mathcal{E}(\mathcal{A}, l) = \left\{(p, q) \in \mathbf{Q}(\mathcal{A})^2 \colon p \neq q,\ |p| = |q| = l \text{ and } pq \notin \mathbf{Q}(\mathcal{A})\right\}.$$

The cardinalities of $\mathcal{E}(\mathcal{A}, l)$ and $\mathbf{Q}_l(\mathcal{A})$ will be denoted by $\varepsilon(n, l)$ and $\pi_n(l)$, respectively:

$$\varepsilon(n, l) := |\mathcal{E}(\mathcal{A}, l)| \text{ and } \pi_n(l) := |\mathbf{Q}_l(\mathcal{A})|.$$

Finally, define

$$\Lambda^1(l) := \{d \in \mathbb{N} \colon d|l, d \not\equiv 0 \pmod 2, d \geq 3\}.$$

Theorem 4.2. *We have*

$$\varepsilon(n, l) = \sum_{d \in \Lambda^1(l)} \pi_n\left(\frac{2l}{d}\right).$$

Proof. The proof is based on a bijection between $\mathcal{E}(\mathcal{A}, l)$ and the disjoint union $\bigcup_{d\in\Lambda^1(l)} \mathbf{Q}_{\frac{2l}{d}}(\mathcal{A})$.

Define the map $\varphi \colon \bigcup_{d\in\Lambda^1(l)} \mathbf{Q}_{\frac{2l}{d}}(\mathcal{A}) \longrightarrow \mathcal{E}(\mathcal{A}, l)$ assigning to a $u \in \mathbf{Q}_{\frac{2l}{d}}(\mathcal{A})$ the couple (p, q) such that $p = (\beta\alpha)^s\beta$ and $q = (\alpha\beta)^s\alpha$, where $|\alpha| = |\beta| = \frac{l}{d}$, $u = \alpha\beta$ and $s = \frac{d-1}{2}$.

Let us show that φ is well defined, that is p and q are primitive, $|p| = |q| = l$ and pq is not primitive. Indeed, if $s \geq 2$, then by the Prefix-Suffix Lemma 2.7, $p = (\alpha\beta)^s\alpha$ and $q = (\beta\alpha)^s\beta$ are primitive words. If $s = 1$, then as $\alpha\beta\alpha$ is conjugate to $\beta\alpha^2$ and $|\alpha| = |\beta|$, $p = \beta\alpha^2$ is a primitive word, by Shyr-Yu [20]. Using an analogues argument, we show that $q = \beta\alpha\beta$ is primitive. As a result, p and q are primitive for any $s \geq 1$. Summarizing the properties of p and q as: $p = (\beta\alpha)^s\beta$ and $q = (\alpha\beta)^s\alpha$ are primitive, where $|\alpha| = |\beta|$ and $\alpha\beta \in \mathbf{Q}(\mathcal{A})$. By Proposition 4.1, we have $|p| = |q|$ and $pq \notin \mathbf{Q}(\mathcal{A})$, which is equivalent to $(p, q) \in \mathcal{E}(\mathcal{A}, l)$.

φ is one-to-one. Let $u_1 = \alpha_1\beta_1 \in \mathbf{Q}_{\frac{2l}{d_1}}(\mathcal{A})$ and $u_2 = \alpha_2\beta_2 \in \mathbf{Q}_{\frac{2l}{d_2}}(\mathcal{A})$ be such that $\varphi(u_1) = \varphi(u_2)$, hence $(\alpha_1\beta_1)^{s_1}\alpha_1 = (\alpha_2\beta_2)^{s_2}\alpha_2$ and $(\beta_1\alpha_1)^{s_1}\beta_1 = (\beta_2\alpha_2)^{s_2}\beta_2$, where $s_i = \frac{d_i-1}{2}$ for $i = 1, 2$. Then, we get $(\alpha_1\beta_1)^{2s_1+1} = (\alpha_2\beta_2)^{2s_2+1}$. On the other hand, as $\alpha_1\beta_1$ and $\alpha_2\beta_2$ are primitive words, we have $u_1 = \alpha_1\beta_1 = \alpha_2\beta_2 = u_2$.

Note that by the previous proposition, φ is onto. This completes the proof.

□

Before providing a combinatorial formula for $\varepsilon(n, l)$, let us first recall the Möbius inversion formula. Define the Möbius function

$$\mu(n) = \begin{cases} 1 & \text{if} \quad n = 1, \\ 0 & \text{if} \quad n \text{ is a multiple of the square of a prime number}, \\ (-1)^k & \text{if} \quad n \text{ is a squarefree with } k \text{ prime factors}. \end{cases}$$

Clearly μ is a multiplicative function (*i.e.*, $\mu(mn) = \mu(m)\mu(n)$, for any relatively prime numbers m, n). Let $f, g \colon \mathbb{N} \longrightarrow \mathbb{C}$ be arithmetic functions, then the following property holds.

Möbius inversion formula [12]

Let n be a positive integer. Then, we have the following equivalence:

$$g(m) = \sum_{d|m} f(d) \text{ for all } m|n \iff f(m) = \sum_{d|m} \mu(d) g\left(\frac{m}{d}\right) \text{ for all } m|n.$$

Let us fix some notations. For a prime number $\mathbf{r}$ and $l = \mathbf{r}^m l_1$ such that $\mathbf{r}$ does not divide l_1, we denote by

$$\tau(n, \mathbf{r}, l) = n^{pl} - n^l - \pi_n(\mathbf{r}l),$$

$$\Delta(l, \mathbf{r}) = \{\mathbf{r}\} \cup \mathbf{pf}(l_1),$$

$$\Gamma(l, \mathbf{r}) \colon = \{L \subseteq \Delta(l, \mathbf{r}) \colon \emptyset \neq L \mathrm{and} \mathrm{L} \neq \{\mathbf{r}\}\},$$

$$\mathbf{p}(L) \colon = \prod_{x \in L} x,$$

where $\mathbf{pf}(l_1)$ denotes the set of all prime factors of l_1.

The next result gives the exact expression of $\varepsilon(n, l)$ in a combinatorial form.

Theorem 4.3. *Suppose that $l = 2^m l_1$ such that l_1 is odd. Then, we have*

$$\varepsilon(n, l) = \tau(n, 2, l) = \sum_{L \in \Gamma(l, 2)} (-1)^{|L|+1} n^{\frac{2l}{\mathbf{p}(L)}}.$$

The proof relies on two lemmata.

Lemma 4.4. *If l is a positive integer such that $\mathbf{r}$ does not divide l and m is a nonnegative integer, then we have*

$$\pi_n(\mathbf{r}^{m+1} l) = \sum_{d|l} \mu(d) \left(n^{\frac{\mathbf{r}^{m+1} l}{d}} - n^{\frac{\mathbf{r}^m l}{d}} \right).$$

Proof. Clearly, we have

$$\begin{aligned} \pi_n(\mathbf{r}^{m+1} l) &= \sum_{d | \mathbf{r}^{m+1} l} \mu(d) n^{\frac{\mathbf{r}^{m+1} l}{d}} \\ &= \sum_{d|l} \mu(d) n^{\frac{\mathbf{r}^{m+1} l}{d}} + \sum_{d|l} \mu(\mathbf{r}d) n^{\frac{\mathbf{r}^{m+1} l}{\mathbf{r}d}} \\ &= \sum_{d|l} \mu(d) n^{\frac{\mathbf{r}^{m+1} l}{d}} - \sum_{d|l} \mu(d) n^{\frac{\mathbf{r}^m l}{d}}. \end{aligned}$$

□

Lemma 4.5. *Let l be a positive integer. Then we have*

$$\tau(n, \mathbf{r}, l) = \sum_{L \in \Gamma(l, \mathbf{r})} (-1)^{|L|+1} n^{\frac{\mathbf{r}l}{\mathbf{p}(L)}}.i$$

Proof. From Lemma 4.4, we get

$$\begin{aligned}\tau(n,\mathbf{r},l) &= n^{\mathbf{r}l} - n^l - \pi_n(\mathbf{r}l) = n^{\mathbf{r}l} - n^l - \sum_{d|l}\mu(d)\left(n^{\frac{\mathbf{r}l}{d}} - n^{\frac{l}{d}}\right)\\ &= -\sum_{d|l,d\neq 1}\mu(d)\left(n^{\frac{\mathbf{r}l}{d}} - n^{\frac{l}{d}}\right) = -\sum_{d|l,d\neq 1}\mu(d)n^{\frac{\mathbf{r}l}{d}} + \sum_{d|l,d\neq 1}\mu(d)n^{\frac{\mathbf{r}l}{\mathbf{r}d}}\\ &= \sum_{\emptyset\neq L\subseteq\Delta(l,\mathbf{r}),\mathbf{r}\notin L}(-1)^{|L|+1}n^{\frac{\mathbf{r}l}{\mathbf{P}(L)}} + \sum_{\emptyset\neq L\subseteq\Delta(l,\mathbf{r}),\mathbf{r}\in L,L\neq\{\mathbf{r}\}}(-1)^{|L|+1}n^{\frac{\mathbf{r}l}{\mathbf{P}(L)}}\\ &= \sum_{L\in\Gamma(l,\mathbf{r})}(-1)^{|L|+1}n^{\frac{\mathbf{r}l}{\mathbf{P}(L)}}.\end{aligned}$$

□

Proof of Theorem 4.3. By Proposition 4.1, we have

$$\begin{aligned}\varepsilon(n,l) &= \sum_{d\in\Lambda^1(l)}\pi_n\left(\frac{2l}{d}\right) = \sum_{d|l_1,d\neq 1}\pi_n\left(\frac{2l}{d}\right)\\ &= \sum_{d|l_1}\pi_n\left(\frac{2^{m+1}l_1}{d}\right) - \pi_n\left(2^{m+1}l_1\right),\end{aligned} \tag{4.1}$$

so that

$$\varepsilon(n,2^m l_1) + \pi_n\left(2^{m+1}l_1\right) = \sum_{d|l_1}\pi_n\left(\frac{2^{m+1}l_1}{d}\right). \tag{4.2}$$

On the other hand and according to Lemma 4.4, we have

$$\pi_n(2^{m+1}l_1) = \sum_{d|l_1}\mu(d)\left(n^{\frac{2^{m+1}l_1}{d}} - n^{\frac{2^m l_1}{d}}\right). \tag{4.3}$$

According to Möbius inversion formula, we get

$$n^{2^{m+1}l_1} - n^{2^m l_1} = \sum_{d|l_1}\pi_n\left(\frac{2^{m+1}l_1}{d}\right). \tag{4.4}$$

Combining equations (4.2) and (4.4), we obtain

$$\varepsilon(n,l) = n^{2l} - n^l - \pi_n(2l) = \tau(n,2,l).$$

The second form of $\varepsilon(n,l)$ in Theorem 4.3 is a direct consequence of Lemma 4.5. □

Example 4.6. From the previous theorem, we have the following two examples.

1. $\varepsilon(n,2^m) = 0$.

2. Let p be an odd prime number and $k \geq 1$ be an integer. Note that $\Delta(p^k, 2) = \{2, p\}$, hence

$$\varepsilon(n, p^k) = (-1)^{|L_1|+1} n^{\frac{2p^k}{\mathbf{p}(L_1)}} + (-1)^{|L_2|+1} n^{\frac{2p^k}{\mathbf{p}(L_2)}},$$

where $L_1 = \{p\}$ and $L_2 = \{2, p\}$. Therefore, we get $\varepsilon(n, p^k) = n^{2p^{k-1}} - n^{p^{k-1}}$.

The next results shows how fast $\varepsilon(n, l)$ grows when l is large enough.

Proposition 4.7. *We have* $\varepsilon(n, l) = \mathbf{O}(n^{\frac{2l}{3}})$, *as* $l \to \infty$.

For the proof, we need a lemma.

Lemma 4.8. *Let* i, l *be two integers such that* $i \geq 1$ *and* $l \geq 4$. *Then, we have*

$$\pi_n(l) \leq \frac{\pi_n(l+2i)}{(n^2-1)^i}.$$

Proof. We have $\pi_n(l) \leq \frac{\pi_n(l+2)}{n^2-1}$. Indeed, by lemma 2.6, for a given $u \in \mathcal{A}^+$ with $|u| = l$ and each $x, y \in \mathcal{A}^+$ with $|x| = |y| = 2$ and $x \neq y$, we have either ux or uy is a primitive word. It follows that $(n^2-1)\pi_n(l) \leq \pi_n(l+2)$. By induction on i, we complete the proof. □

Proof of Proposition 4.7. Let l be a positive integer that is not a power of 2. By Theorem 4.2, we have

$$\varepsilon(n, l) = \sum_{d \in \Lambda^1(l)} \pi_n\left(\frac{2l}{d}\right) \leq \sum_{1 \leq i \leq \frac{l}{d_1}} \pi_n(2i), \tag{4.5}$$

where d_1 is the smallest divisor of l greater than or equal to 3. On the other hand, using Lemma 4.8, we get

$$\pi_n(2i) \leq \frac{1}{(n^2-1)^{\frac{l}{d_1}-i}} \pi_n\left(\frac{2l}{d_1}\right) \leq \frac{n^{\frac{2l}{d_1}}}{(n^2-1)^{\frac{l}{d_1}-i}},$$

for every integer $i \geq 2$, so that

$$\sum_{2 \leq i \leq \frac{l}{d_1}} \pi_n(2i) \leq \sum_{1 \leq i \leq \frac{l}{d_1}} \frac{n^{\frac{2l}{d_1}}}{(n^2-1)^{\frac{l}{d_1}-i}} \leq \sum_{k=0}^{\infty} \frac{n^{\frac{2l}{d_1}}}{(n^2-1)^k} = n^{\frac{2l}{d_1}} \frac{n^2-1}{n^2-2} \leq n^{\frac{2l}{3}} \frac{n^2-1}{n^2-2}. \tag{4.6}$$

Combining (4.5) and (4.6) completes the proof. □

Now, we discuss the asymptotic behavior of $\varepsilon(n, l_k)$ for some specific sequences $(l_k, k \in \mathbb{N})$.

Proposition 4.9. *Let* p *be a given odd prime number. If* $(l_k, k \in \mathbb{N})$ *is an increasing sequence of elements of the set* $\Psi_p = \{l \in \mathbb{N}\colon\ p \text{ is the smallest prime factor of } l\}$, *then*

$$\varepsilon(n, l_k) \sim n^{\frac{2l_k}{p}} \text{ as } k \to \infty.$$

First, let us show a technical lemma.

Lemma 4.10. *We have* $\pi_n(l) \sim n^l$, *as* l *goes to* ∞.

Proof. As $\mathbf{Q}_l(\mathcal{A}) \subseteq \mathcal{A}^l$, we have $\pi_n(l) \leq n^l$. From Shyr-Tu [20], we deduce that if $x \neq y \in \mathcal{A}$, then given $p \in \mathbf{Q}_l(\mathcal{A})$ (with $l \geq 2$), either px or py is primitive. This leads to the inequality $(n-1)\pi_n(l) \leq \pi_n(l+1)$. In particular, $\pi_n(l)$ is increasing with respect to l. Hence, as $n^l = \sum_{d|l} \pi_n(d)$, we get

$$n^l - \pi_n(l) = \sum_{d|l, d\neq l} \pi_n(d) \leq \sum_{d|l, d\neq l} \pi_n(\delta) \leq l\pi_n(\delta) \leq ln^{\delta},$$

with δ is the largest divisor of l not equal to l. Writing $l = \delta k$, for some $k \geq 2$, we obtain $\delta = \frac{l}{k} \leq \frac{l}{2}$. Therefore, we have $n^l - \pi_n(l) \leq ln^{\delta} \leq ln^{\frac{l}{2}}$. It follows that

$$n^l - ln^{\frac{l}{2}} \leq \pi_n(l) \leq n^l.$$

This is equivalent to

$$1 - \frac{l}{n^{\frac{l}{2}}} \leq \frac{\pi_n(l)}{n^l} \leq 1,$$

which completes the proof. □

Proof of 4.9. Let us rewrite

$$\varepsilon(n, l_k) = \sum_{d\in\Lambda^1(l_k)} \pi_n\left(\frac{2l_k}{d}\right) = \pi_n\left(\frac{2l_k}{p}\right) + \sum_{d\in\Lambda^1(l_k)\setminus\{p\}} \pi_n\left(\frac{2l_k}{d}\right). \tag{4.7}$$

If d_2 denotes the smallest element of $\Lambda^1(l_k) \setminus \{p\}$, then necessarily $d_2 \geq p+2$. So, we have

$$\sum_{d\in\Lambda^1(l_k)\setminus\{p\}} \pi_n\left(\frac{2l_k}{d}\right) \leq \sum_{d\in\Lambda^1(l_k)\setminus\{p\}} \pi_n\left(\frac{2l_k}{d_2}\right) \leq l_k\pi_n\left(\frac{2l_k}{d_2}\right) \leq l_k n^{\frac{2l_k}{d_2}}. \tag{4.8}$$

Since $l_k n^{\frac{2l_k}{d_2}} = o\left(n^{\frac{2l_k}{p}}\right)$, we obtain

$$\sum_{d\in\Lambda^1(l_k)\setminus\{p\}} \pi_n\left(\frac{2l_k}{d}\right) = o\left(n^{\frac{2l_k}{p}}\right). \tag{4.9}$$

On the other hand, using lemma (4.10), we get

$$\pi_n\left(\frac{2l_k}{p}\right) \sim n^{\frac{2l_k}{p}}. \tag{4.10}$$

Combining (4.9) and (4.10) in (4.7) completes the proof. □

It is worth noting that in [13] Horváth has observed that almost all words are primitive, so there is a very little chance to "go out from $\mathbf{Q}(\mathcal{A})$" even by means of strong combinatorial manipulations. Recently, Ryoma [19] has showed that the natural density of the language of primitive words over $\mathcal{A}$ is 1 (this is equivalent to Lem. 4.10).

Another point of interest is

$$\mathbf{P}(n,l) = \frac{|\mathcal{E}(\mathcal{A},l)|}{|\mathcal{C}(\mathcal{A},l)|} = \frac{\varepsilon(n,l)}{\pi_n(l)(\pi_n(l)-1)},$$

the probability for a random chosen couple (p,q) of distinct primitive words with $|p| = |q| = l$ to satisfy $p^+q^+ \not\subseteq \mathbf{Q}(\mathcal{A})$, where

$$\mathcal{C}(\mathcal{A},l) = \left\{(p,q) \in \mathbf{Q}(\mathcal{A})^2 \colon p \neq q,\ |p| = |q| = l\right\}.$$

We close the paper by discussing the asymptotic behavior of $\mathbf{P}(n,l)$ as n or l goes to ∞.

Proposition 4.11. *We have*

1. $\mathbf{P}(n,l) \sim \frac{1}{n^{2l\left(1-\frac{1}{p_1}\right)}}$, *as n goes to infinity, where p_1 is the smallest odd prime factor of l (assuming l not a power of 2).*
2. $\mathbf{P}(n,l) = \mathbf{O}\left(n^{-\frac{4l}{3}}\right)$, *as l goes to ∞.*

Remark 4.12. From Proposition 4.11, one may conclude that if the alphabet size is large enough, then almost all couples (p,q) of distinct primitive words of the same given length l satisfy $p^+q^+ \subset \mathbf{Q}(\mathcal{A})$. A similar conclusion holds if we assume that the alphabet size is fixed and l is large enough.

References

[1] G. Castiglione, G. Fici and A. Restivo, Primitive sets of words. *Theoret. Comput. Sci.* **866** (2021) 25–36.
[2] C. Choffrut and J. Karhumaki, Combinatorics of words, in Vol. 1 of *Handbook of Formal Languages*. Springer-Verlag, Berlin, Heidelberg (1997) 329–438.
[3] C. Chunhua, Y. Shuang and D. Yang, Some kinds of primitive and non-primitive words. *Acta Inform.* **51** (2014) 339–346.
[4] P. Dömösi, S. Horváth and M. Ito, Formal languages and primitive words. *Publ. Math. Debrecen* **42** (1993) 315–321.
[5] P. Dömösi and G. Horváth, Alternative proof of the Lyndon-Schützenberger theorem. *Theoret. Comput. Sci.* **366** (2006) 194–198.
[6] P. Dömösi and G. Horvath, The language of primitive words is not regular: two simple proofs. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* **87** (2005) 191–197.
[7] P. Dömösi, G. Horváth and L. Vuillon, On the Shyr-Yu theorem. *Theo. Comput. Sci.* **410** (2009) 4874–4877.
[8] P. Dömös and M. Ito, Context-free languages and primitive words. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ (2015).
[9] P. Dömösi, M. Ito and S. Marcus, Marcus contextual languages consisting of primitive words. *Discrete Math.* **308** (2008) 4877–4881.
[10] O. Echi, Non-primitive words of the form pq^m. *RAIRO-Theor. Inf. Appl.* **51** (2017) 135–139.
[11] N.J. Fine and H.S. Wilf, Uniqueness theorems for periodic functions. *Proc. Am. Math. Soc.* **16** (1965) 109–114.
[12] G.H. Hardy and E.M. Wright, An introduction to the theory of numbers, 6th edn. Oxford University Press, Oxford (2008).
[13] S. Horváth, Strong interchangeability and nonlinearity of primitive words, in: Algebraic Methods in Language Processing. Univ. of Twente, Enschede, The Netherlands (1995) 173–178.
[14] A. Lentin and M.P. Schützenberger, A combinatorial problem in the theory of free monoids. In: Combinatorial Mathematics and its Applications. Proc. Conf., Univ. North Carolina, Chapel Hill, N.C. (1967) 128–144.
[15] M. Lothaire, Combinatorics on Words. Addison-Wesley (1983).
[16] R.C. Lyndon and M.P. Schützenberger, The equation $a^M = b^N c^P$ in a free group. *Mich. Math. J.* **9** (1962) 289–298.
[17] A. Restivo, On a question of McNaughton and Papert. *Inf. Control* **25** (1974) 93–101.
[18] C. Reis and H.J. Shyr, Some properties of disjunctive languages on a free monoid. Inf. *Control* **37** (1978) 334–344.
[19] R. Sińya, Asymptotic Approximation by Regular Languages. In: theory and practice of computer science, SOFSEM 2021, *LNCS* vol 12607 (2021) 74–88.
[20] H.J. Shyr and S.S. Yu, Non-primitive words in the language p^+q^+. *Soochow J. Math.* **20** (1994) 535–546.
[21] H.J. Shyr, Free monoids and languages. 2nd edition. Lecture notes, Hon Min Book Co., Taichung (1991).

[22] H.J. Shyr and F.K. Tu, Local distribution of non-primitive words. In: Ordered structures and algebra of computer languages. World Scientific, River Edge, NJ (1993) 202–217.

Subscribe to Open (S2O)
A fair and sustainable open access model

This journal is currently published in open access under a Subscribe-to-Open model (S2O). S2O is a transformative model that aims to move subscription journals to open access. Open access is the free, immediate, online availability of research articles combined with the rights to use these articles fully in the digital environment. We are thankful to our subscribers and sponsors for making it possible to publish this journal in open access, free of charge for authors.

Please help to maintain this journal in open access!

Check that your library subscribes to the journal, or make a personal donation to the S2O programme, by contacting subscribers@edpsciences.org

More information, including a list of sponsors and a financial transparency report, available at: https://www.edpsciences.org/en/maths-s2o-programme