Morphisms, infinite words, and symmetries

HABILITATION THESIS

Prague, 2016

Štěpán Starosta
Abstract

This habilitation thesis is a collection of articles covering various topics, published or submitted from 2013 to 2016. Most of the articles are set in Combinatorics on Words and are dealing with languages generated by morphisms that have reversal symmetry: with each element of the language, the mirror image of this element is also in the language.

The first part of the articles investigates some of the recent conjectures in Combinatorics on Words. The first conjecture is the Brlek–Reutenauer conjecture, connecting palindromic defect with factor and palindromic complexities. We solve this conjecture by giving an affirmative answer. The second conjecture is the Class P conjecture, stating that if a language generated by a morphism contains infinitely many palindromes, then the morphism belongs to a special class of morphism called class P. The third conjecture is the Zero defect conjecture, which states that if the generating morphism is primitive, then the palindromic defect of the language is zero or infinity. We give only partial answers to the last two conjectures: we deal with some specific subclasses of the morphisms in question. Namely, we give an affirmative answer for morphisms fixing 3 interval exchange transformation for Class P conjecture, and for binary and primitive marked morphisms for Zero defect conjecture.

The second part of the articles presents many new constructions of words with finite palindromic defect, also in a generalized sense. We enlarge the family of known examples of such words by following the construction of Rote words, by doing letter-to-letter projections of episturmian words, and by investigating generalized Thue–Morse words.

The third part of the thesis deals with efficient algorithmic analysis of languages generated by morphisms and leads toward two efficient algorithms: the first algorithm enumerates all primitive factors that occur in the generated language in any power; the second algorithm tests whether a morphism is circular.

The last part is constituted from two various results: the study of the Rauzy gasket, a set representing letter frequencies of all ternary episturmian words; and the study of a generalization of Markov constant motivated by the study of spectrum of a a differential operator. This part serves as an illustration of connection of Combinatorics on Words to other research domains.
Acknowledgements

The works included in this thesis were supported by the following grants: Czech Science Foundation GAČR grants 13-35273P, 13-03538S, 201/09/0584; grants MSM6840770039 and LC06002 of the Ministry of Education, Youth, and Sports of the Czech Republic; and grant of the Grant Agency of the Czech Technical University in Prague grant No. SGS11/162/OHK4/3T/14.

The experimental results were done using the open-source mathematical software SageMath [77], including PARI/GP [1] and FLINT [38].

The author also wishes to thank his co-authors for their joyful cooperation.

Copyright

Some of the works in this habilitation thesis are protected by copyright by various publishers. They are presented and reprinted in accordance with the copyright agreements with the respective publishers or with a specific licence allowing reprinting in this thesis.
List of included articles

This thesis is a collection of the following articles. This is an online version of the thesis and the articles themselves are not included due to various copyright issues. This version consists of an introduction (Chapter 1) covering the topics of these article (referred to by Roman numerals).

[I] L. Balková, E. Pelantová, and Š. Starosta, 

[II] Z. Masáková, E. Pelantová, and Š. Starosta, 

[III] S. Labbé, E. Pelantová, and Š. Starosta, 

[IV] Š. Starosta, 

[V] E. Pelantová and Š. Starosta, 

[VI] T. Jajcayová, E. Pelantová, and Š. Starosta, 

[VII] K. Klouda and Š. Starosta, 

[VIII] K. Klouda and Š. Starosta, 

[IX] P. Arnoux and Š. Starosta, 

[X] E. Pelantová, Š. Starosta, and M. Znojil, 
Contents

Abstract ................................. iii
Acknowledgements ....................... v
Copyright ................................ v
List of included articles .................. vii

1 Introduction ........................... 1
1.1 On Combinatorics on Words ................. 1
   1.1.1 Brief history and connection to other domains . 1
   1.1.2 Current challenges .......................... 4
1.2 Notation and terminology .................. 6
   1.2.1 Palindromic defect ......................... 7
   1.2.2 Infinite words with finite palindromic defect . 8
   1.2.3 Morphisms and languages they generate ... 10
1.3 Overview of results and comments .......... 12
   1.3.1 Three conjectures related to palindromes ... 12
   1.3.2 Construction of words with finite palindromic defect . 16
   1.3.3 D0L-systems and algorithms ............... 21
   1.3.4 Rauzy gasket and generalized Markov constant . 24
1.4 Contribution of the author ............... 27
References ................................ 29

2 Proof of the Brlek–Reutenauer conjecture 37

3 Exchange of three intervals: itineraries, substitutions and palindromicity 39

4 On the Zero Defect Conjecture ............... 41
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Morphic images of episturmian words having finite palindromic defect</td>
<td>43</td>
</tr>
<tr>
<td>6</td>
<td>Constructions of words rich in palindromes and pseudopalindromes</td>
<td>45</td>
</tr>
<tr>
<td>7</td>
<td>Palindromic closures using multiple antimorphisms</td>
<td>47</td>
</tr>
<tr>
<td>8</td>
<td>An algorithm for enumerating all infinite repetitions in a D0L-system</td>
<td>49</td>
</tr>
<tr>
<td>9</td>
<td>Characterization of circular D0L-systems</td>
<td>51</td>
</tr>
<tr>
<td>10</td>
<td>The Rauzy Gasket</td>
<td>53</td>
</tr>
<tr>
<td>11</td>
<td>Markov constant and quantum instabilities</td>
<td>55</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

This introduction serves as an overview of the included results accompanied with comments. A summary of used notions is part of this overview, including a brief general overview of the research domain.

1.1 On Combinatorics on Words

The research domain of most of the included articles is Combinatorics on Words which is catalogued in the Mathematics Subject Classification database under 68R15. We start by giving a brief overview of this domain.

1.1.1 Brief history and connection to other domains

The beginning of Combinatorics on Words is mostly attributed to Axel Thue and his article [78] from 1906 and 3 following articles until 1914. The reason is that he gave birth to a systematic study of objects called words: finite or infinite sequences of elements from a finite set called alphabet. The reader may refer to [9, 75, 62] for translations of Thue’s papers and comments on his results.

In 1921, Marston Morse published an article studying geodesics [54]. The article contained an overlap with Axel Thue’s study. This overlap gave rise to a name of a famous infinite word: the Thue–Morse word. The Thue–Morse word, denoted t, is an infinite word of elements from the alphabet \{0, 1\}, i.e., an infinite sequence of 0s and 1s. It can be constructed by building prefixes of t, i.e., finite sequences that form the beginning of t. We start by setting the first prefix to \(p_1 = 0\), the finite word of length 1 consisting of the letter 0. We apply the following rewriting rule to all the elements of \(p_1\): we replace 0 by
Introduction

On Combinatorics on Words

01 and 1 by 10. The obtained word is $p_2 = 01$ and it is the next constructed prefix of $t$. Repeating the procedure, we obtain

$$
p_1 = 0, \\
p_2 = 01, \\
p_3 = 0110, \\
p_4 = 01101001, \\
p_5 = 0110100110010110, \\
p_6 = 01101001100101101001011001101001.\
$$

We can observe that $p_i$ is also a prefix of $p_{i+1}$ and the length of $p_{i+1}$ is twice the length of $p_i$. Therefore, there is a unique infinite word having each $p_i$ as a prefix, i.e., the procedure to construct $t$ is unambiguous.

The constructions of the Thue–Morse word of Thue and Morse differ as they appear in different context. As already mentioned, Morse was studying geodesics. Thue was solving the following problem: does there exist an infinite word over a 2-letter alphabet which does not contain a cube? A cube is a contiguous subsequence that can be written as a repetition of 3 words. For instance, the word 011011011 is a cube since 011 is repeated 3 times. The Thue–Morse word $t$ has such a property: there are no cubes in the Thue–Morse word.

Let us note that the word $t$ is also sometimes called Prouhet–Thue–Morse since it appeared already in 1851 in [63] by Eugène Prouhet.

The next stepping stone in the history of Combinatorics on Words is the article [55] of Marston Morse and Gustav Hedlund from 1940. Their work includes the study of another famous infinite words called Sturmian words in the honour of the famous mathematician Jacques Charles François Sturm. A Sturmian word is an infinite word over a two-letter alphabet having factor complexity $n + 1$, that is, for each $n$ the number of total distinct contiguous subsequences of length $n$ found in the word is $n + 1$.

Such finite contiguous subsequence is called a factor, thus the name factor complexity since it is one of the measures of chaos (or order) of an infinite word. The basic property of factor complexity is the following: if the factor complexity of an infinite word is bounded, then the word is (eventually) periodic. The converse is also true and we may deduce that Sturmian words are binary words having the least possible factor complexity so that they are not periodic.

Let us illustrate the notion of Sturmian words by defining the most famous word of this class: the Fibonacci word $f$. The word $f$ may be defined using the construction procedure of the Thue–Morse word $t$ but using a distinct
rewriting rule. We set \( q_1 = 0 \). To obtain \( q_2 \), we apply the rewriting rule \( 0 \mapsto 01 \) and \( 1 \mapsto 0 \) to \( q_1 \). We have \( q_2 = 01 \). We repeat the procedure using the new rewriting rule and obtain

\[
\begin{align*}
q_1 &= 0, \\
q_2 &= 01, \\
q_3 &= 010, \\
q_4 &= 01001, \\
q_5 &= 01001010, \\
q_6 &= 0100101001001.
\end{align*}
\]

As in the case of the word \( t \), each word \( q_i \) is a prefix of \( q_{i+1} \) and the length of \( q_i \) is strictly increasing. Thus, there is a unique infinite word over \{0, 1\} having each \( q_i \) as its prefix, and it is the Fibonacci word \( f \). Its name comes from the fact that we can recover the Fibonacci sequence by looking at the lengths of the prefixes \( q_i \), denoted by \( |q_i| \). We have

\[
|q_{i+1}| = |q_i| + |q_{i-1}| \quad \text{for all } i > 2.
\]

In other words, the length of a prefix \( q_{i+1} \) equals the sum of the sum of the lengths of the previous two prefixes. As the initial conditions of this recurrence are \( |q_1| = 1 \) and \( |q_2| = 2 \), we retrieve the Fibonacci sequence.

After the mentioned works, the field of Combinatorics on Words has been growing steadily. The reader may refer to [10] for an overview of early progress in the area. The steady growth of the domain is underlined by collective publications containing overview of results in Combinatorics on Words and closely related domains and various monographs.

The first item on the list of such publications is the book *Combinatorics on Words*, first published in 1983, written by a collective of authors under the pseudonym M. Lothaire [49]. Two more books by M. Lothaire were published later, *Algebraic Combinatorics on Words* in 2002 [50] and *Applied Combinatorics on Words* in 2005 [51].

The growth of Combinatorics of Words may be also seen in its increasing connection to other domains. *Substitutions in Dynamics, Arithmetics and Combinatorics* published in 2002 [33] is a basic reference for the connection of Combinatorics of Words and Symbolic Dynamics. The publication *Combinatorics, Automata, and Number Theory* of 2010 [12] contains useful results interconnecting the domain in the title of the publication. The strong connection to Automata, Theory of Codes and Formal Languages may be also observed in the following books [10, 52, 67, 68].
Introduction

Besides the mentioned domains, Combinatorics on Words finds its application in many other domains. Let us name some of them: Algebra, Logic, Music Theory, Stringology, and Biology. The reader may refer to [41] for an overview of some mentioned applications.

Let us mention, as a brief illustration of one of the connections, an attractive application in Biology, namely in DNA sequencing. One of the basic questions is the following: let \( w \) be a finite word of length \( n \) over an alphabet \( \mathcal{A} \). Determine the least number \( k \) such that the word \( w \) can be reconstructed from the knowledge of \( k \) distinct subwords of \( w \), where subword is a subsequence of \( w \). The question is already stated using the terminology of Combinatorics on Words, which serves as a theoretical basis to study problems such as DNA sequencing. In the terms of DNA sequencing, the alphabet is fixed to be the four-elements set of nucleotides, the word \( w \) is the part of the DNA sequence to be recovered, and the subwords are the results of the experiments when analysing \( w \). The reader may refer to [48] for more details on the links between Combinatorics on Words and DNA sequencing.

1.1.2 Current challenges

In this section we manifest the current state of the Combinatorics on Words by listing some of the actual challenges recognized by the international community. While listing below some of the current challenges, we omit giving specific details since the list is quite extensive and may be extended even more. In Section 1.3, we give more details for topics that are directly related to the results contained in this thesis.

We start the list by giving the reference to Ten Conferences WORDS: Open Problems and Conjectures [57] published in 2016 by Jean Néraud. As the title indicates, the author enumerates some of the questions that came up during WORDS conferences in the last ten years. The mentioned conference may be considered to be one of the most important conference series for researchers working in Combinatorics on Words.

The challenges mentioned in [57] are grouped as follows.

- **Pattern avoidance**: questions related to existence and properties of words avoiding certain patterns such as cubes in the case of the Thue–Morse word are being investigated.

- **Complexity studies**: besides factor complexity, open questions related to palindromic complexity, Abelian complexity, arithmetical complexity and other complexity measures are in the focus of researchers.
Introduction

On Combinatorics on Words

- *Factorization of words and equations*: questions concerning specific factorizations of words or equations such as $xy = yx$ with $x, y$ being finite words are within the scope of ongoing research topics.

Inspired by the overview done by Jean Néraud, we may investigate the most frequent topics in some other important conference series and regular scientific meetings focusing on Combinatorics on Words: *Mons Theoretical Computer Science Days, RuFiDiM - Russian Finnish Symposium on Discrete Mathematics, Development in Language Theory,* and *International school and conference on Combinatorics, Automata and Number Theory*. We list some other frequent topics:

- balance properties of words,
- Rauzy fractals and tilings generated by words,
- morphisms and morphic words,
- automatic sequences,
- properties of words in link with other domains such as:
  - diophantine approximation and number theory in general,
  - numeration,
  - formal language theory,
  - automata theory.

We finish this overview by mentioning the growing software support for Combinatorics on Words in the open-source computer algebra system *SageMath*. The library is being developed by some of the researchers in the domain and its possibilities range from basic support for finite or infinite words and morphisms to more advanced methods and algorithms. Many researchers use this library and advance faster in their discoveries thanks to it. As the system *SageMath* is built to represent mathematical knowledge, not just implement various algorithms, we can observe the connections of Combinatorics on Words by looking at how the library is used by other parts of the software.

Before giving more specific overview of the results included in this thesis, we summarize the used notation and terminology.

---

1The source code of the library can be explored at [https://git.sagemath.org/sage.git/tree/src/sage/combinat/words](https://git.sagemath.org/sage.git/tree/src/sage/combinat/words) (October 2016).
1.2 Notation and terminology

We mostly follow the usual terminology of Combinatorics on Words, see for instance [49]. Let $A$ be a finite set, called an alphabet. Its elements are called letters. A finite word $w$ is an element of $A^n$ for $n \in \mathbb{N}$. The length of $w$ is $n$ and is denoted $|w|$. The set of all finite words over $A$ is denoted $A^*$. An infinite word over $A$ is an infinite sequence of letters from $A$. Examples of infinite words over the alphabet $\{0,1\}$ are the Thue-Morse word $t$ and the Fibonacci word $f$ defined above. The set of all (right-)infinite words over $A$ is denoted $A^\omega$.

A finite word $w$ is a factor of a finite or infinite word $v$ if there exist words $p$ and $s$ such that $v$ is a concatenation of $p$, $w$, and $s$, denoted $v = pws$. The word $p$ is said to be a prefix and $s$ a suffix of $v$. The set of all factors of a word $u$ is the language of $u$ and is denoted $\mathcal{L}(u)$. All factors of $u$ of length $n$ are denoted by $\mathcal{L}_n(u)$.

A factor $w \in \mathcal{L}(u)$ is right special if there exist two distinct letters $a$ and $b$ such that $wa, wb \in \mathcal{L}(u)$. Analogously, it is left special if $aw, bw \in \mathcal{L}(u)$. A factor is bispecial if it is both left and right special. For a factor $w$ we define its bilateral multiplicity (or bilateral order), see [23], $m(w)$ as follows:

$$m(w) = \#\{awb \in \mathcal{L}(u) : a, b \in A\} - \#\{wb \in \mathcal{L}(u) : b \in A\} - \#\{aw \in \mathcal{L}(u) : a \in A\} + 1.$$ 

An occurrence of $w = w_0w_1 \cdots w_{n-1} \in A^n$ in a word $v = v_0v_1v_2 \cdots$ is an index $i$ such that $v_i \cdots v_{i+n-1} = w$. A factor $w$ is unioccurrent in $v$ if there is exactly one occurrence of $w$ in $v$. A complete return word of a factor $w$ (in $v$) is a factor $f$ (of $v$) containing exactly two occurrences of $w$ such that $w$ is its prefix and also its suffix.

We say that an infinite word $u$ is recurrent if every its factor has infinitely many occurrences in $u$. The word $u$ is uniformly recurrent if it is recurrent and every its factor has a finite number of complete return words in $u$. In other words, the gaps between successive occurrences of a factor are bounded.

The reversal or mirror mapping assigns to a word $w \in A^*$ the word $R(w)$ with the letters reversed, i.e.,

$$R(w) = w_{n-1}w_{n-2} \cdots w_1w_0 \quad \text{where} \quad w = w_0w_1 \cdots w_{n-1} \in A^n.$$

A word is palindrome if $w = R(w)$. We say that a language $\mathcal{L} \subset A^*$ is closed under reversal if for all $w \in \mathcal{L}$ we have $R(w) \in \mathcal{L}$.

Given an infinite word $u$, its factor complexity $\mathcal{C}_u(n)$ is the count of its factors of length $n$:

$$\mathcal{C}_u(n) = \#\mathcal{L}_n(u) \quad \text{for all} \quad n \in \mathbb{N}.$$
With \( \text{Pal}(u) \) being the set of all palindromic factors of the infinite word \( u \), the \textit{palindromic complexity} \( \mathcal{P}_u(n) \) of \( u \) is given by
\[
\mathcal{P}_u(n) = \#(\mathcal{L}_n(u) \cap \text{Pal}(u)) \quad \text{for all } n \in \mathbb{N}.
\]
We omit the subscript \( u \) if there is no confusion.

### 1.2.1 Palindromic defect

One property of words which is in the research focus is related to palindromic factors of a finite or infinite word. In [27], given a finite word \( w \in A^* \), the authors investigate the set of all its palindromic factors, denoted \( \text{Pal}(w) \), and give the following upper bound on \( \#\text{Pal}(w) \):
\[
\#\text{Pal}(w) \leq |w| + 1.
\] (1)

Note that the \textit{empty word} \( \varepsilon \), the unique word of length 0, is an element of \( \text{Pal}(w) \) for all \( w \).

For instance, we have \( \text{Pal}(011) = \{\varepsilon, 0, 1, 11\} \). Thus, for the word 011, the upper bound on \( \#\text{Pal}(011) \) is attained.

The difference of the upper bound and the actual number of palindromic factors is the \textit{palindromic defect} of \( w \), see [17]. It is denoted \( D(w) \). We have
\[
D(w) = |w| + 1 - \#\text{Pal}(w).
\]

A basic property of the palindromic defect is that \( D(w) \geq 0 \) for any \( w \) and \( D(v) \leq D(w) \) for any factor \( v \) of \( w \).

The properties of palindromic defect allow for a natural extension to infinite words:
\[
D(u) = \sup\{D(w) : w \in \mathcal{L}(u)\}.
\]

One can say that it measures the number of “missing” palindromic factors in the given word.

There exist words that have palindromic defect 0. The famous Sturmian words are an example. Such words are also called \textit{rich} or \textit{full}, and are being investigated as they possess some notable properties. The first interesting property is the existence of the following theorem listing many known characterizations of words with zero palindromic defect (provided the language in question is closed under reversal). For a palindrome \( w \in \mathcal{L}(u) \), we set \( \text{Pext}(w) = \{awa \in \mathcal{L}(u) : a \in A\} \).

**Theorem 1.** For an infinite word \( u \) with language closed under reversal the following statements are equivalent:
1. \( D(u) = 0 \) (i.e., \( u \) is rich) \((27)\);

2. any complete return word of any palindromic factor of \( u \) is a palindrome \((27)\);

3. for any factor \( w \) of \( u \), every factor of \( u \) that contains \( w \) only as its prefix and \( R(w) \) only as its suffix is a palindrome \((24)\);

4. the longest palindromic suffix of any factor \( w \in L(u) \) is unioccurent in \( w \) \((27, 34)\);

5. for each \( n \) the following equality holds
   \[ C(n + 1) - C(n) + 2 = P(n) + P(n + 1) \]
   \((20)\);

6. any bispecial factor \( w \in L(u) \) satisfies
   \[ m(w) = \begin{cases} 0 & \text{if } w \neq R(w); \\ \#Pext(w) - 1 & \text{otherwise}. \end{cases} \]
   \((7)\).

### 1.2.2 Infinite words with finite palindromic defect

A significant part of this thesis is connected to infinite words having finite palindromic defect. We give a list of their characterizations, which can be seen as a generalization of Theorem 1.

**Theorem 2.** For an infinite word \( u \) with language closed under reversal the following statements are equivalent:

1. \( D(u) \) is finite \((17)\);

2. there exists an integer \( P \) such that any prefix of \( u \) longer than \( P \) has a unioccurrent longest palindromic suffix \((27, 8)\);

3. there exists an integer \( N \) such that for any palindromic factor of \( u \) having length at least \( N \), every its complete return word is a palindrome \((8, 60)\);

4. there exists an integer \( N \) such that for any factor \( w \) of \( u \) having length at least \( N \), every factor of \( u \) that contains \( w \) only as its prefix and \( R(w) \) only as its suffix is a palindrome \((8, 60)\);
5. There exists an integer \( N \) such that for each \( n \geq N \) the following equality holds

\[
C(n+1) - C(n) + 2 = \mathcal{P}(n) + \mathcal{P}(n+1)
\]

(\([\text{I}]\)).

6. There exists an integer \( N \) such that any bispecial factor \( w \in \mathcal{L}(u) \) of length at least \( N \) satisfies

\[
m(w) = \begin{cases} 
0 & \text{if } w \neq R(w); \\
\#\text{Ext}(w) - 1 & \text{otherwise};
\end{cases}
\]

and

\[
C(N+1) - C(N) + 2 = \mathcal{P}(N) + \mathcal{P}(N+1)
\]

(\([\text{II}]\)).

Note that a generalization of property \([\text{I}]\) of Theorem \([\text{I}]\) is not included in the last theorem. It may be included only if we add the assumption of uniform recurrence. In other words, if \( u \) is uniformly recurrent and has finite palindromic defect, then there exists an integer \( T \) such that any factor of \( u \) longer than \( T \) has a unioccurrent longest palindromic suffix. See [61, Theorem 35] for a proof\(^2\). The converse follows directly from Theorem 2, item 2.

To see that uniform recurrence is indeed required, consider the following example. Let \( p = 1231321 \). We have \( D(p) = 1 \). Set \( w_0 = p \) and \( w_i = w_{i-1}0^i w_{i-1} \) for all \( i > 0 \) where \( 0^i \) is the word consisting of the letter 0 repeated \( i \) times. Let \( w \) be the infinite word having \( w_i \) as its prefix for all \( i \). We have

\[
w = \overbrace{pp}^{w_2} \overbrace{pp000p000p00p000p}^{w_1} \overbrace{pp000p00p00p00p000p} \ldots
\]

It can be verified that \( D(w) \) is finite:

**Proposition 3.** The word \( w \) satisfies \( D(w) = D(p) = 1 \).

**Proof.** We obtain \( D(p) = 1 \) by direct calculation.

By [4, Corollary 3], the palindromic defect of \( w \) is equal to the number of its prefixes such that their longest palindromic suffix is not unioccurrent. We show for all \( i \) that each prefix of \( w_i \) longer than \( |p| \) has a unioccurrent longest

\(^2\)The proof in the mentioned reference is given in a more general context of more possible symmetries than just the reversal symmetry.
palindromic suffix. We proceed by induction on $i$. First note that for all $i$, the word $w_i$ is a palindrome and it contains exactly two occurrences of $w_{i-1}$.

The word $w_0 = p$ has no prefix longer than $|p|$ thus the claim is true.

Let $i > 0$. Assume that for $w_{i-1}$, each its prefix longer than $|p|$ has a unoccurent longest palindromic suffix.

The prefixes of $w_i$ of length less than $|w_{i-1}| + 1$ satisfy the claim by the induction hypothesis. The prefix of $w_i$ of the form $w_{i-1}0^k$ with $0 < k < i$ has its longest palindromic suffix equal to $0^k w_{i-2} 0^k$. As there are exactly two occurrences of $w_{i-2}$ in $w_{i-1}$ and the other occurrence is as a prefix, this longest palindromic suffix is unoccurent. The prefix of $w_i$ of the form $w_{i-1}0^i s$ where $s$ is a prefix of $w_{i-1}$ has its longest palindromic suffix equal to $R(s)0^i s$ since $w_{i-1}$ is a palindrome. As $0^i$ has exactly one occurrence in $w_i$, the longest palindromic suffix $R(s)0^i s$ is unoccurent.

As $w_i$ is a prefix of $w$ for all $i$, we have $D(w) = D(p) = 1$.

To see that the word $w$ is the counterexample we are looking for, consider for each integer $k$ the word $0^k 1231$ which is a factor of $w$. Its longest palindromic suffix is 1 and it is not unoccurent.

### 1.2.3 Morphisms and languages they generate

Morphisms are an important tool to generate infinite words and their languages. A morphism $\varphi$ is a mapping $A^* \rightarrow B^*$ where $A$ and $B$ are alphabets such that $\forall v, w$ we have $\varphi(vw) = \varphi(v) \varphi(w)$ (it is a homomorphism of monoids $A^*$ and $B^*$). Its action is extended to $A^\mathbb{N}$: if $u = u_0 u_1 u_2 \ldots \in A^\mathbb{N}$ with $u_i \in A$, then

$$\varphi(u) = \varphi(u_0) \varphi(u_1) \varphi(u_2) \ldots \in B^\mathbb{N}.$$  

If $\varphi$ is an endomorphism of $A^*$, we may find its fixed point, i.e., a word $u$ such that

$$\varphi(u) = u.$$  

We are interested mainly in the case of $u$ being infinite. A morphism $\varphi : A^* \rightarrow A^*$ is primitive if for each $a, b \in A$ there exists an integer $k$ such that $b$ occurs in $\varphi^k(a)$. It is well known that fixed points of primitive morphisms are uniformly recurrent.

Two morphisms $\varphi, \psi : A^* \rightarrow B^*$ are conjugate if there exists a word $w \in B^*$ such that

$$\forall a \in A, \varphi(a)w = w\psi(a) \quad \text{or} \quad \forall a \in A, w\varphi(a) = \psi(a)w.$$  

If $\varphi$ is primitive, then the languages of fixed points of $\varphi$ and $\psi$ are the same.
A morphism \( \psi : A^* \to B^* \) is of class \( P \) if \( \psi(a) = pp_a \) for all \( a \in A \) where \( p \) and \( p_a \) are both palindromes (possibly empty). A morphism \( \varphi \) is of class \( P' \) if it is conjugate to a morphism of class \( P \).

A morphism is uniform if the lengths of images of letters are the same.

The following examples illustrate the last few notions.

Example 4. Let \( \varphi : \{a, b\}^* \to \{a, b\}^* \) be determined by

\[
\begin{align*}
    a & \mapsto abab, \\
    b & \mapsto aab.
\end{align*}
\]

The fixed point of \( \varphi \) is

\[
    u = \lim_{k \to +\infty} \varphi^k(a) = abab aab abab abab \ldots
\]

As \( \varphi \) is primitive, the word \( u \) is uniformly recurrent. The morphism \( \varphi \) is of class \( P' \) since it is conjugate to \( \psi \) given by

\[
\begin{align*}
    \psi : & \\
    a & \mapsto abab, \\
    b & \mapsto aba.
\end{align*}
\]

Indeed, we have \( ab\varphi(a) = \psi(a)ab \) and \( ab\varphi(b) = \psi(b)ab \). To see that \( \psi \) is of class \( P \), i.e., it is of the form \( a \mapsto pp_a \) and \( b \mapsto pp_b \), it suffices to set \( p = aba \), \( p_a = b \) and \( p_b = \varepsilon \). The fixed point of \( \psi \) is

\[
    v = \lim_{k \to +\infty} \psi^k(a) = abab aba abab abab \ldots
\]

We have \( L(u) = L(v) \).

Since \( |\varphi(a)| \neq |\varphi(b)| \), the morphism \( \varphi \) is not uniform.

Example 5. The two already mentioned famous examples of infinite words, the Thue-Morse word \( t \) and the Fibonacci word \( f \), are both fixed points of a morphism.

The word \( t \) is fixed by the morphism \( \varphi_{TM} \) determined by \( \varphi_{TM}(0) = 01 \) and \( \varphi_{TM}(1) = 10 \). Note that this uniform morphism in fact has two fixed points, one being the other one after replacing 0 with 1 and 1 with 0. The word \( t \) as given above is the fixed points starting in 0.

The word \( f \) is fixed by the morphism \( \varphi_F \) defined by \( \varphi_F(0) = 01 \) and \( \varphi_F(1) = 0 \).

An (infinite) fixed point of a morphism of class \( P' \) clearly contains infinitely many palindromes which is one motivation for this notion. Class \( P \) is introduced in [39] in the context of discrete Schrödinger operators.
1.3 Overview of results and comments

We give an overview of the results of the research articles which are part of this thesis. The overview is divided into the following groups:

1. Three conjectures related to palindromes
2. Construction of words with finite palindromic defect
3. D0L-systems and algorithms
4. Rauzy gasket and generalized Markov constant

1.3.1 Three conjectures related to palindromes

This section contains results on recent conjectures from Combinatorics on Words dealing with general properties of words related to palindromes and palindromic defect.

Brlek–Reutenauer conjecture ([I])

The first conjecture relates the palindromic defect of an infinite word with its factor and palindromic complexities.

Conjecture 1 (Brlek–Reutenauer conjecture [I]). Let $u$ be an infinite word with language closed under reversal. We have

$$2D(u) = \sum_{n=0}^{+\infty} (C_u(n+1) - C_u(n) + 2 - P_u(n+1) - P_u(n)).$$

The authors of the conjecture proved already in [I] that the conjecture is true for periodic infinite words. The article [20] gives a positive answer for words having palindromic defect zero (it is a consequence of already mentioned Theorem [I], item [5]). The article [I] completes the study by giving an affirmative answer to Conjecture [I]. The proof is done by showing the two following theorems.

Theorem 6. If $u$ is an infinite word with language closed under reversal such that both $D(u)$ and $\sum_{n=0}^{+\infty} (C_u(n+1) + C_u(n) + 2 - P_u(n+1) - P_u(n))$ are finite, then

$$2D(u) = \sum_{n=0}^{+\infty} (C_u(n+1) + C_u(n) + 2 - P_u(n+1) - P_u(n)).$$
Theorem 7. If \( u \) is an infinite word with language closed under reversal, then \( D(u) < +\infty \) if and only if
\[
\sum_{n=0}^{+\infty} (C_u(n+1) + C_u(n) + 2 - \mathcal{P}_u(n+1) - \mathcal{P}_u(n)) < +\infty.
\]

The relation of palindromic defect and factor and palindromic complexity given by the statement of Conjecture 1 is not very practical to compute the palindromic defect of an infinite word. However, it gives more insight into words with finite palindromic defect. The methods used to prove the conjecture resulted in giving the characterization of Theorem 2 and subsequently also the characterization of Theorem 6. The article [III] exploits these results in the proof of its main theorem (see below).

Class \( P \) conjecture ([III])

The second conjecture is the following.

Conjecture 2 (Class \( P \) conjecture [39]). Let \( u \) be a fixed point of a primitive morphism \( \varphi \) containing infinitely many palindromic factors. There exists a morphism of class \( P' \) such that its fixed point has the same language as \( u \).

The original statement of the conjecture in [39] is ambiguous and allows for more interpretations, see also [II] or [37]. The above given statement of Conjecture 2 follows from two results. First, for binary alphabet the question is solved in [76]: if a fixed point of a primitive morphism \( \varphi \) over a binary alphabet contains infinitely many palindromes, then \( \varphi \) or \( \varphi^2 \) is of class \( P' \). Second, in [46], the author shows that if we restrict ourselves just to infinite words, not more general languages of fixed points, the answer is negative: there exists a word \( w \) over ternary alphabet which is a fixed point of a primitive morphism, containing infinitely many palindromic factors, and not being fixed by any morphism of class \( P' \). However, the authors of [37] note that the language of the word \( w \) may indeed be generated by a morphism of class \( P \).

An affirmative answer to Conjecture 2 would provide a strong tool to investigate fixed points of primitive morphisms that contain infinitely many palindromic factors. However, at this moment, only partial answers are known: as already mentioned, the binary case is solved ([76]); for larger alphabets an affirmative answer is provided only for some special classes of morphisms ([II] and [III]).

The affirmative answer in [III] is for the case of \( \varphi \) fixing a word coding a non-degenerate exchange of 3 intervals. Moreover, the result states that \( \varphi \) itself or \( \varphi^2 \) is in class \( P' \).
Words coding exchange of $k$ intervals form a well-known class of words over ternary alphabet. Such words may be defined as follows. Let $J$ be a left-closed right-open interval. Consider a partition $J = J_0 \cup \cdots \cup J_{k-1}$ of $J$ into a disjoint union of left-closed right-open subintervals such that $\forall x \in J_i, \forall y \in J_{i+1}, x < y$. A bijection $T : J \to J$ is an exchange of $k$ intervals with permutation $\pi$ if there exist numbers $c_0, \ldots, c_{k-1}$ such that for $0 \leq i < k$ one has

$$T(x) = x + c_i \text{ for } x \in J_i,$$

where $\pi$ is a permutation of $\{0, 1, \ldots, k-1\}$ determining the order of intervals $T(J_i)$, i.e., such that $\pi(i) < \pi(j)$ implies $\forall x \in T(J_i), \forall y \in T(J_j), x < y$. If $\pi$ is the permutation $i \mapsto k - i - 1$, then $T$ is called a symmetric interval exchange transformation. The orbit of a given point $\rho \in J$ is the infinite sequence $\rho, T(\rho), T^2(\rho), T^3(\rho), \ldots$. It can be coded by an infinite word $u_\rho = u_0 u_1 u_2 \ldots$ over the alphabet $\{0, 1, \ldots, k-1\}$ as follows:

$$u_n = X \text{ if } T^n(\rho) \in J_X \text{ for } X \in \{0, 1, \ldots, k-1\}.$$

If for every $n \in \mathbb{N}$ we have $C_{u_\rho}(n) = (k - 1)n + 1$, then the transformation $T$ and the word $u_\rho$ are said to be non-degenerate.

The class of words coding symmetric interval exchange transformations are well-studied (see [31, 32, 30]). In particular, they have zero palindromic defect and their language is closed under reversal. It follows that they contain infinitely many palindromic factors. In [33], besides giving an affirmative answer to the Class P conjecture in the case of 3 intervals, we enlarge the knowledge of languages of words coding symmetric 3 interval exchange transformation by giving a detailed description of the return times to a subinterval and the corresponding itineraries. This description is then refined with respect to the reversal symmetry present in the language. The partial solution to the Class P conjecture is based on these results and known results on a relation with Sturmian words. The exploited relation with Sturmian words is also the reason why this result is only for the ternary case, not for a general coding of $k$-interval exchange transformation.

**Zero defect conjecture ([33])**

The last conjecture is the following.

**Conjecture 3** (Zero defect conjecture [15]). Let $u$ be an aperiodic fixed point of a primitive morphism having its language closed under reversal. We have $D(u) = 0$ or $D(u) = +\infty$.  

---

**Introduction**

Overview of results and comments
The Thue–Morse word $t$ and the Fibonacci word $f$ are examples of aperiodic fixed points of a primitive morphism (see Example 5) having their language closed under reversal. We have $D(f) = 0$ (in fact, it is true for all Sturmian words, as already mentioned). On the other hand, $D(t) = +\infty$.

In 2012, Bucci and Vaslet [22] found a counterexample to this conjecture on a ternary alphabet. They showed that the fixed point of the primitive morphism determined by 

$$a \mapsto aabcaca, b \mapsto aa, c \mapsto a$$

has finite positive palindromic defect and is not periodic. In [23], the author also gives a counterexample. Thus, the current statement of the conjecture is not true. However, there still might some refinement of the current statement that is valid as there are many witnesses and the found counterexamples seem to have some specific properties. For instance, the mentioned counterexample of [22] is not injective. Indeed, in [III] we prove that the conjecture is true for a special class of morphisms. A morphism $\varphi$ is marked if there exists two morphisms $\varphi_1$ and $\varphi_2$, both being conjugate to $\varphi$, such that

$$\{\text{last letter of } \varphi_1(a) : a \in A\} = \{\text{first letter of } \varphi_2(a) : a \in A\} = A.$$ 

In other words, the set of the last letters of the images of letters by $\varphi_1$ is the whole alphabet $A$ and the set of the first letters of the images of letters by $\varphi_2$ is also the whole alphabet $A$.

For instance, $\varphi = \varphi_{TM} : 0 \mapsto 01, 1 \mapsto 10$ is marked (here $\varphi = \varphi_1 = \varphi_2$). For $\varphi = \varphi_F : 0 \mapsto 01, 1 \mapsto 0$ we have $\varphi = \varphi_1$ and $\varphi_2 : 0 \mapsto 10, 1 \mapsto 0$. Thus, $\varphi_F$ is also marked. In fact, any non-trivial morphism on binary alphabet is marked.

In [III] we show the following theorem:

**Theorem 8.** Let $\varphi$ be a primitive marked morphism and let $u$ be its fixed point with finite palindromic defect. If all complete return words of all letters in $u$ are palindromes or there exists a conjugate of $\varphi$ distinct from $\varphi$ itself, then $D(u) = 0$.

The proof is based on the characterization of Theorem 2. The characterization is expressed using the notion of an extension graph of a factor, which is used to describe possible extensions of factors in a language. This notion is also used in [I4] to study languages having some specific properties of extension graphs of its elements. In [I3], this notion is further studied and the authors make a connection to words with finite palindromic defect.

Let us comment also on the assumptions of the last theorem. The case which does not satisfy the assumptions, i.e., the case of primitive marked
morphisms such that the morphism is only conjugate to itself and its fixed point $u$ contains a non-palindromic complete return word to a letter, remains an open question. The proof in [III] is not applicable to this case and probably it is not easily extendable to it.

In [III], the special case of binary alphabet is also solved. In this case, the requirements on the morphism may be dropped:

**Theorem 9.** If $u \in A^N$ is a fixed point of a primitive morphism over binary alphabet and $D(u) < +\infty$, then $D(u) = 0$ or $u$ is periodic.

### 1.3.2 Construction of words with finite palindromic defect ([VI, V, IV])

Many examples and properties are known for words with zero finite palindromic defect, i.e., rich or full words. Besides the articles mentioned in connection with Theorem 1 ([27, 34, 20, 7]), let us mention some other relevant results:

- In [21], the authors give another characterization of words with zero palindromic defect.
- In [19], the relation of words with zero palindromic defect to so-called periodic-like words is exhibited.
- Links to another class of words, trapezoidal words, are shown in [26].
- The number of all words with zero palindromic defect of a given length and other properties are investigated in [80, 35].
- As already mentioned, words coding symmetric interval exchange transformations have zero palindromic defect by [9] and Theorem 1 item 5.
- In [16], the authors show that words coding rotation on the unit circle with respect to partition consisting of two intervals have zero palindromic defect.
- In [66], the author show a connection of words having zero palindromic defect with Burrows-Wheeler transform.

The mentioned results give rise to many examples of words with zero palindromic defect. On the other hand, there are not many specific classes of words with finite but non-zero palindromic defect. In [8], the following relation of words with finite palindromic defect to words with zero palindromic is given:
**Theorem 10.** If $u \in A^\mathbb{N}$ is a uniformly recurrent infinite word such that $D(u) < +\infty$, then there exist a morphism $\varphi : B^* \rightarrow A^*$ and an infinite uniformly recurrent word $v \in B^\mathbb{N}$ such that $u = \varphi(v)$ and $D(v) = 0$.

This theorem gives insight on the relation of words with finite and zero palindromic defect, but it cannot be used directly to construct words of finite palindromic defect. Although the morphism $\varphi$ is of a very special class, so-called class $P_{ret}$ introduced in [8], the idea cannot be reversed as there exists a word with zero palindromic defect and a morphism of class $P_{ret}$ such that the image of this word by the morphism has infinite palindromic defect [8, Proposition 5.7].

The general goal of the 3 articles of this section is to enlarge the family of known examples of words with finite palindromic defect and investigate their properties, and also to broaden the family of known words having finite generalized palindromic defect introduced in [60] and studied in [61] (see below for a definition).

In [IV], we focus on episturmian words, see [27]. A word is episturmian if its language is closed under reversal and for each $n$ there is at most one right special factor of length $n$. Aperiodic binary episturmian words are exactly the Sturmian words. Episturmian words have all zero palindromic defect.

The first results of [IV] is that that the image of an episturmian word by a morphism of class $P_{ret}$ has always finite palindromic defect. This result is used to show the second result:

**Theorem 11.** Let $u$ be an episturmian word over a ternary alphabet $A$. Let $B$ be an alphabet and $\pi : A^* \rightarrow B^*$ be a letter-to-letter morphism. We have $D(\pi(u)) = 0$.

In other words, any letter-to-letter image of a ternary episturmian word has zero palindromic defect.

To illustrate this theorem, let us consider the so-called Tribonacci word which is the fixed point of the morphism $\varphi : 0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 0$. For $u = \varphi(u)$ we have

$$u = 0102010010201010201001\ldots$$

Let $\pi : \{0, 1, 2\} \rightarrow \{0, 1\}$ be a letter-to-letter morphism determined by $\pi(0) = 0, \pi(1) = 1$ and $\pi(2) = 1$. The last theorem then implies that the word

$$\pi(u) = 010100101010101001\ldots$$
Introduction

Overview of results and comments

Both main mentioned results of [IV] exploit many properties of episturmian words and reveal more of the ingenious structure they possess. Theorem [11] is shown only for ternary episturmian words; however, computational evidence suggests that we may drop the requirement on the size of the alphabet. Unfortunately, the small size of the alphabet is crucial in the presented proof in [IV] and it does not allow a simple extension to even 4-letter alphabet.

The article [IV] investigates also the generalized palindromic defect, as do the articles [VI, V]. We introduce here the notion for the case of binary alphabet \{0, 1\}. Let \(E : \{0, 1\}^* \to \{0, 1\}^*\) be an antimorphism, i.e., \(\forall v, w \in \{0, 1\}^*\) we have \(E(vw) = E(w)E(v)\), that is given by \(E(0) = 1\) and \(E(1) = 0\). The mapping \(E\) exchanges the two letters and reverses the order of letters in a word as the mapping \(R\) does. For instance, we have \(E(011) = E(1)E(1)E(0) = 001\). This mapping can be viewed as a generalization of the concept of the mirror mapping \(R\). Its fixed points are called pseudopalindromes, antipalindromes, or \(E\)-palindromes (studied for instance in [15, 35]). The \(E\)-palindromic defect measuring the number of missing \(E\)-palindromes can be defined analogously to the classical (\(R\)-)palindromic defect, see [73].

A further generalization of these notions is given in [60] where the so-called \(H\)-palindromic defect is introduced with \(H = \{E, R, ER, Id\}\). This value considers classical palindromes and \(E\)-palindromes at the same time and measures the difference of their actual count and their maximal count. To count the palindromes and \(E\)-palindromes at the same time we consider, given a word \(f\), the set \([f] = \{\mu(f) : \mu \in H\}\) and for a factor \(w\) we count the following number: \(\tau(w) = \#\{[f] : f\ \text{factor of } w, f = R(f) \text{ or } f = E(f)\}\). In other word, the number \(\tau(w)\) counts the number of sets \([f]\) where \(f\) is a palindrome or an \(E\)-palindrome, thus counting both at the same time. The maximal count, for the case of \(H\), is given by \(\tau(w) \leq |w| + 1\). Finally, the \(H\)-palindromic defect is given by the difference of this upper bound \(|w| + 1\) and \(\tau(w)\). The basic property of \(H\)-palindromic defect is also its nonnegativity and its interpretation is analogous to classical palindromic defect: it measures how many palindromes and \(E\)-palindromes, counted at the same time using the classes \([f]\), are missing to attain the maximum. It is also extended to infinite words in a similar way.

The Thue–Morse word contains infinitely many palindromes and also infinitely many \(E\)-palindromes. Moreover, it has infinite palindromic defect and \(E\)-palindromic defect, and its \(H\)-palindromic defect is zero. To demonstrate it, let us take the prefix \(w = 0110100110\) of the Thue–Morse word. The count
of palindromes and $E$-palindromes is
\[
\tau(w) = \# \left\{ [\varepsilon], [0], [11], [01], [101], [0110], [1010], [0011], \\
[110100], [100110], [01101001] \right\} = 11 = |w| + 1.
\]

In [V] we show that the mapping $S$ defined on $\{0, 1\}^*$ by $S(u_0u_1u_2\ldots) = v_1v_2v_3\ldots$ with $v_i = u_{i-1} + u_i \mod 2$ can be used to construct new classes of words with zero or finite palindromic defect and words with zero or finite $H$-palindromic defects. The article is inspired by the results of [69] stating that a word $u$ is a complementary-symmetric Rote word if and only if $S(u)$ is a Sturmian word, and by the results of [16] claiming that every complementary-symmetric Rote word has zero palindromic defect. A complementary-symmetric Rote word may be defined as a coding of an irrational rotation of the unit circle with respect to the partition into two intervals of same length.

Namely, we show the following theorems.

**Theorem 12.** Let $u \in \{0, 1\}^N$ be an infinite word having its language closed under all elements of $H = \{Id, E, R, ER\}$. The word $u$ has zero $H$-palindromic defect (resp. finite $H$-palindromic defect) if and only if $S(u)$ has zero palindromic defect (resp. finite palindromic defect).

A corollary of the last theorem is that complementary-symmetric Rote words have zero $H$-palindromic defect since their languages are indeed closed under all elements of $H$. As already mentioned, they have also zero palindromic defect. Thus, they have finite $H$-palindromic and palindromic defect at the same time.

**Theorem 13.** Let $u \in \{0, 1\}^N$ be a uniformly recurrent word. If $u$ has finite palindromic defect, then the word $S(u)$ has finite palindromic defect.

The last theorem gives a procedure constructing possibly many words with finite palindromic defect. For example, if $u$ is a Sturmian word, then $S^k(u)$ has finite palindromic defect for all $k > 0$. In this case, computer experiments suggest that $S^k(u)$ is not Sturmian for all $k$ and its palindromic defect is not zero.

In [IV], we combine Theorems 11 and 12 in order to exhibit another new class of words with zero $H$-palindromic defect. Namely, we show that applying an operation inverse to $S$ to a non-trivial letter-to-letter projection of a ternary Arnoux–Rauzy word, we obtain a new class of words with zero $H$-palindromic defect.
In [VI], we investigate infinite words which are constructed using iterated palindromic closure operators with respect to $R$ and $E$. Such words are in general called \textit{generalized pseudostandard words}. The operator constructs successively palindromic and $E$-palindromic prefixes of an infinite word. We focus on so-called generalized Thue-Morse words which are a generalization of the Thue-Morse word $t$ for larger alphabets: given two integers $b$ and $m$ such that $b > 1$ and $m > 1$, the generalized Thue-Morse word $t_{b,m}$ is defined over $\{0, \ldots, m - 1\}$ by $t_{b,m} = (s_b(n) \mod m)_{n=0}^{\infty}$ where the number $s_b(n)$ denotes the digit sum of the expansion of the number $n$ in the base $b$. The classical Thue-Morse word $t$ equals $t_{2,2}$. These words are studied for instance in [79, 74, 2].

The article [VI] is mainly motivated by [25] where the authors show that the Thue-Morse word is a generalized pseudostandard word. Its construction using this method is governed by two directive sequences $01^\omega$ and $(RE)^\omega$, where the superscript $\omega$ denotes infinite repetition of the word, i.e., $01^\omega = 01111\ldots$. The first prefix is constructed by taking the first element of the first sequence, i.e., 0, and creating its $R$-palindromic closure, i.e., finding the shortest $R$-palindrome such that 0 is its prefix. The $R$-palindromic closure of a word $w$ is denoted $w^R$. We have the first prefix to be $f_1 = 0^R = 0$. To obtain the next prefix, we take the second element of the first directive sequence, append it to $f_1$ and find its $E$-palindromic closure since $E$ is the second element of the second directive sequence. Thus we have $f_2 = (f_11)^E = (01)^E = 01$. We continue this process to obtain more prefixes of the Thue-Morse word, taking again at step $i$ the $i$th elements of both directive sequences:

$$
\begin{align*}
f_3 &= (f_21)^R = (011)^R = 0110, \\
f_4 &= (f_31)^E = (01101)^E = 01101001, \\
f_5 &= (f_41)^R = (011010011)^R = 0110100110010110.
\end{align*}
$$

In [VI], we investigate the generalized Thue-Morse words and find a characterization of those that can be constructed using the same procedure.

\textbf{Theorem 14.} The generalized Thue-Morse word $t_{b,m}$ is a generalized pseudostandard word if and only if $b \leq m$ or $b - 1 = 0 \mod m$.

In [74] it is shown that all the generalized Thue-Morse words have zero $I_2(m)$-palindromic defect for some group $I_2(m)$ in the sense of the mentioned $H$-palindromic richness (see [60, 61] for an exact definition of this notion).
Introduction

The group $I_2(m)$ is isomorphic to the dihedral group of order $2m$. A section of [I] is inspired by this fact and investigates the generalized palindromic defect of the words $S^i(t_{b,m})$ where the mapping $S$ is generalized to the alphabet $\{0, 1, \ldots, m-1\}$ as follows: for every word $w = w_0 \cdots w_n$ with $w_i \in \{0, 1, \ldots, m-1\}$ we set

$$S(w_0w_1\cdots w_n) = v_1\cdots v_n,$$

where $v_i = (w_{i-1} + w_i) \mod m$ for every $i \in \{1, \ldots, n\}$.

The main result of the last part of [V] is the following theorem.

**Theorem 15.** Let $m, b \in \mathbb{Z}$ such that $m \geq 3$ and $b \geq 3$. There exists a group $I'_2(m)$ with $I'_2(m) = I_2(m)$ if $m$ is odd, and $I'_2(m)$ being isomorphic to $I_2(m)$ if $m$ is even, such that

1. The word $S(t_{b,m})$ has finite $I'_2(m)$-palindromic defect;

2. if $m$ or $b$ is odd, the word $S(t_{b,m})$ has zero $I'_2(m)$-palindromic defect.

1.3.3 D0L-systems and algorithms ([VII, VIII])

The name “D0L-system” stands for deterministic context-independent Lindenmayer system. It is defined as a triple $G = (A, \varphi, w)$ with $A$ an alphabet, $\varphi$ an endomorphism of $A^*$, and $w \in A^*$. The word $w$ is the axiom of $G$. The language of $G$ is the set $\{\varphi^i(w) : i \leq 0\}$. The set of all factors of the elements of the language of $G$ is denoted $\mathcal{F}(G)$. See more in [70, 71] on D0L-systems and more general concept of L-systems which were introduced by Aristid Lindenmayer to model the plant growth.

Our motivation to study this notion stems from the fact that it generalizes the notion of the language of a fixed point of a morphism. Indeed, by taking the axiom to be the first letter of a fixed point of morphism we can recover the language of the fixed point by considering the set $\mathcal{F}(G)$ of the corresponding D0L-system $G$.

For instance, consider the morphism $\varphi_{TM}$ fixing the Thue–Morse word $t$ (see Example 5) and set $G_w = (\{0, 1\}, \varphi_{TM}, w)$ for $w \in \{0, 1\}^*$. We have $\mathcal{F}(G_0) = \mathcal{L}(t)$. As $\varphi_{TM}$ is primitive, all its fixed points have the same language, thus $\mathcal{F}(G_1) = \mathcal{L}(t)$. In general, if $w \in \mathcal{L}(t)$, we have $\mathcal{F}(G_w) = \mathcal{L}(t)$.

Further motivation for research in this domain comes from many open questions concerning efficient analysis of a given D0L-system. Let us list 3 of these open questions concerning a given D0L-system $G$:

---

3By efficient analysis we mean an algorithm that in general does not need to enumerate factors of some length as this is feasible only for small lengths even if we use for instance bounds on linear recurrence constant for primitive morphisms.

---
Question 1: Find an efficient algorithm to generate all elements of \( \mathcal{F}(G) \) of given length (and thus calculate the factor complexity of the generated language).

Question 2: Detect symmetries in \( \mathcal{F}(G) \) (by a symmetry we mean for instance the closedness of \( \mathcal{F}(G) \) under some involutive antimorphism).

Question 3: Test efficiently whether the morphism of \( G \) is injective on \( \mathcal{F}(G) \), which is the following condition: \( \forall w, v \in \mathcal{F}(G) \) we have \( w \neq v \Rightarrow \varphi(w) \neq \varphi(v) \).

To complete the overview, let us also give an example of an answered question. The problem whether \( \mathcal{F}(G) \) is periodic and ultimately periodic is treated and solved in [36, 59, 47] and [72] the binary case is fully described. The question of possible order of magnitude of factor complexities of \( \mathcal{F}(G) \) is also solved, see [58].

It seems that a helpful property of a D0L-system \( G \) that would be needed to answer at least Question 1 is the synchronizing delay. Simply put, knowing the synchronizing delay of a system, one can find preimages of elements that are longer than certain bound. Moreover, these preimages are unambiguous except for some prefix and suffix of bounded length of the word in question. Knowing the unique preimages gives much insight into the language. D0L-systems admitting a finite synchronizing delay are called circular. The formal definition follows.

**Definition 16.** Let \( G = (\mathcal{A}, \varphi, w) \) be a D0L-system, \( \varphi(a) \neq \varepsilon \) for all \( a \in \mathcal{A} \) and \( u \in \mathcal{F}(G) \).

A triplet \((p, v, s)\) where \( p, s \in \mathcal{A}^* \) and \( v = v_1 \cdots v_n \in \mathcal{F}(G) \) with \( n > 0 \) is an interpretation of the word \( u \) if \( \varphi(v) = pus \).

Let \((p, v, s)\) and \((p', v', s')\) be two interpretations of a non-empty word \( u \in \mathcal{F}(G) \) with \( v = v_1 \cdots v_n \in \mathcal{A}^n \), \( v' = v'_1 \cdots v'_m \in \mathcal{A}^m \) and \( u = u_1 \cdots u_\ell \in \mathcal{A}^\ell \).

We say that \( G \) is circular with synchronization delay \( D > 0 \) if whenever we have

\[
|\varphi(v_1 \cdots v_i)| - |p| > D \quad \text{and} \quad |\varphi(v_{i+1} \cdots v_n)| - |s| > D
\]

for some \( i \) such that \( 1 \leq i \leq n \), then there exists \( j \) such that \( 1 \leq j \leq m \) and

\[
|\varphi(v_1 \cdots v_{i-1})| - |p| = |\varphi(v'_1 \cdots v'_{j-1})| - |p'|
\]

and \( v_i = v'_j \).
See Section 3 of [VIII] for an overview concerning the notion of synchronizing delay and circularity. This notion is tightly connected to recognizability, see [50], where the author also shows that a primitive morphism is circular.

If we restrict ourselves to circular morphisms, then using [42] we may design an efficient algorithm to answer Question 1 of enumerating all factors. However, to do this, the synchronizing delay is still required.

The task to efficiently determine the synchronizing delay is also an unsolved question. It is known for binary alphabets and uniform morphisms, see [43] and the references therein. For larger alphabets and primitive morphisms, the recent result [28] follows the work [50] and unveils an explicit upper bound. However, this upper bound is far from optimal and it cannot be used in a practical computation.

Our works [VII, VIII] aim towards this partial goal in the analysis of $\mathcal{F}(G)$ for a given D0L-system $G$.

In [VII] we show the following theorem. A non-empty word $w$ is primitive if it is not a non-trivial (integer) power of another word, i.e., $w = v^k$ implies $k = 1$.

**Theorem 17.** Let $G = (A, \varphi, w)$ be a D0L-system. There exists a constant $M$, a finite set of primitive words $P$ such that $v^k \in \mathcal{F}(G)$ for every $k \geq M$ and $v \in P$.

Moreover, using the known results on periodicity of D0L-systems of [47], we design a simple algorithm finding all primitive words $v$ such that $v^k \in \mathcal{F}(G)$ for all $k$. The algorithm may be outlined as follows. A letter $a \in A$ is unbounded or growing if $|\varphi^k(a)| \to +\infty$ as $k \to +\infty$.

1. Construct an injective simplification of $\varphi$ (injective simplification is an injective morphism that may be used to generate the same language, see [29]).
2. Find all unbounded letters (done by inspecting the incidence matrix of the morphism).
3. Construct two finite directed labeled graphs describe relation between unbounded letters (done by investigating occurrences of unbounded letters in letter images of unbounded letters).
4. Loop over all cycles in the constructed graphs and look for specific labels that indicate one kind of primitive word we are looking for.
5. Test all growing letters if they are first letters of purely periodic fixed points of the morphism (using [47]). If a purely periodic fixed point
is found, then the second kind of primitive word we are looking for is found.

6. If the original morphism was simplified, an inverse operation needs to be applied to the list of discovered primitive words.

7. Output the list of all found primitive words and their conjugates.

In [VIII], we give the following characterization of circularity of a D0L-system. We say that a system \( G = (A, \varphi, w) \) is \textit{unboundedly repetitive} if there exists \( v \in \mathcal{F}(G) \) such that \( v^k \in \mathcal{F}(G) \) for all \( k \) and \( v \) contains at least one unbounded letter.

**Theorem 18.** An D0L-system \( G = (A, \varphi, w) \) injective on \( \mathcal{F}(G) \) is not circular if and only if it is unboundedly repetitive.

To obtain this result we prove that every non-circular D0L-system contains arbitrarily long words of the form \( v^k \), using the results of [VII]. The last theorem is important as it may be used to design an efficient algorithm to test circularity, provided that the tested system is injective on its language. An efficient test of injectivity on a language is still an open question (Question 3), thus for the moment, the test may be done using a stronger condition of general injectivity that is easy to test.

Both algorithms presented in this section are efficient and present a step forward in solving Question 1 which, at least by algorithms enumerating factors, may then be used to solve the other two questions and other possible similar challenges.

### 1.3.4 Rauzy gasket and generalized Markov constant ([IX, XI])

This section contains a short overview of two articles included in this thesis. Each article deals with a problem with no direct relation to the above problems. They may be perceived as witnesses of the mentioned strong connection of Combinatorics on Words to other research domains.

In [IX] we deal with ternary episturmian words (see the definition in Section 1.3.2). We study the set of all triples of letters frequencies of all episturmian words. Given an infinite word \( u \), the frequency of a letter \( a \) is given by the limit

\[
\lim_{n \to +\infty} \frac{\text{number of occurrences of the letter } a \text{ in the prefix of } u \text{ of length } n}{n},
\]
if it exists. The letter frequencies of episturmian words exist.

Let us give an example, the Tribonacci word already mentioned above is a ternary episturmian word. It is the fixed point of the morphism \( \varphi : 0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 0 \). For \( u = \varphi(u) \) we have

\[
u = 0102010010201010201001\]

and the frequencies of 0, 1, and 2 are \( \zeta, \zeta^2, \) and \( \zeta^3 \) respectively with \( \zeta + \zeta^2 + \zeta^3 = 1 \), see [64].

The question to study the set of all letter frequencies of all ternary episturmian words is motivated by the study of systems generating such words in \( [3, 65] \). The set in question is called “the Rauzy gasket” to honour Gérard Rauzy and due to its nature (namely its relation to the Sierpinski and Apollonian gaskets). Thus, a point of the Rauzy gasket is the point \( (\zeta, \zeta^2, \zeta^3) \) of the letter frequencies of the Tribonacci word. Figure 1 shows an approximate shape of the Rauzy gasket.

The main result of \( [IX] \) is a proof of the fact that the Rauzy gasket is of Lebesgue measure 0 (in \( R^2 \)). The presented proof is done by finding a relation to known results concerning the so-called Fully subtractive algorithm, see \([53, 44, 11]\).

In \( [4] \), the authors continue the study of the Rauzy gasket by showing that its Hausdorff dimension is less than 2.

In \( [X] \) we investigate the set

\[
S(\alpha) = \text{set of all accumulation points of } \left\{ m^2 \left( \frac{k}{m} - \alpha \right) : k, m \in \mathbb{Z} \right\}
\]

where \( \alpha \in \mathbb{R} \). The motivation stems from spectral properties of a differential operator connected to properties of metamaterials. From another point of view, the value \( \inf |S(\alpha)| \) is well-known in the literature and sometimes it is called the Markov constant of \( \alpha \). See for instance \([24]\).

The notion of Markov constant is connected to a famous theorem of Hurwitz: for every irrational number \( \xi \) there exist infinitely many relatively prime integers \( m \) and \( n \) such that

\[
\left| \frac{m}{n} - \xi \right| < \frac{1}{\sqrt{5}n^2}
\]

and the bound \( \frac{1}{\sqrt{5}} \) is optimal. The optimality is for \( \xi \) equal to the golden ratio (and other numbers having the tail of their continued fraction expansion equals to an infinite sequence of 1s). Thus, the Markov constant of the golden ratio \( \frac{1+\sqrt{5}}{2} \) is \( \frac{1}{\sqrt{5}} \) which is the maximum value of a Markov constant of any number.
Figure 1: An approximation of the Rauzy gasket (a superset of the Rauzy gasket). The Rauzy gasket is a set contained in the convex set \( \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 1, x, y, z \geq 0\} \). The plane containing this set is depicted.
In [X] we use methods and results from Number Theory and Combinatorics on Words in order to describe the set \( S(\alpha) \). Besides some basic properties of the set \( S(\alpha) \) such as the closedness under multiplication by \( z^2 \) for every \( z \in \mathbb{Z} \) and its connection to known results on Markov constants and continued fractions, we exhibit different behaviour of the set for different parameter \( \alpha \). We show that there exist irrational numbers \( \alpha_1, \alpha_2 \) and \( \alpha_3 \) such that

1. \( S(\alpha_1) = \mathbb{R} \);
2. \( S(\alpha_2) = (-\infty, -\varepsilon] \cup [\varepsilon, +\infty) \), where \( \varepsilon = \frac{\sqrt{2}}{8} \approx 0.18 \);
3. the Hausdorff dimension of \( S(\alpha_3) \cap (-\frac{1}{2}, \frac{1}{2}) \) is positive but less than 1.

### 1.4 Contribution of the author

Since this thesis contains also co-authored articles, we briefly comment on the author’s contribution. For all the co-authored articles, it is difficult to determine exactly the author’s total contribution. The creation process of these articles consisted of a collective work towards the solution of the given question, either in person or by the means of electronic communication, followed by a collective creation of the article itself. The following specific contributions may be separated:

- In [I I I, V, VI], the author is responsible for doing all the supporting computer experiments.
- In [X], the author’s contribution is in Sections 2–5.
References


<table>
<thead>
<tr>
<th>References</th>
<th>References</th>
</tr>
</thead>
</table>


References


<table>
<thead>
<tr>
<th>References</th>
<th>References</th>
</tr>
</thead>
</table>

Proof of the Brlek–Reutenauer conjecture

Authors: L. Balková, E. Pelantová, and Š. Starosta
Cited as: [I]
Chapter 3

Exchange of three intervals: itineraries, substitutions and palindromicity

Authors: Z. Masáková, E. Pelantová, and Š. Starosta
Cited as: [II]
Chapter 4

On the Zero Defect Conjecture

Authors: S. Labbé, E. Pelantová, and Š. Starosta
Cited as: [III]
Morphic images of episturmian words having finite palindromic defect

Authors: Š. Starosta
Cited as: [IV]
Chapter 6

Constructions of words rich in palindromes and pseudopalindromes

Authors: E. Pelantová and Š. Starosta
Cited as: [V]
Chapter 7

Palindromic closures using multiple antimorphisms

Authors: T. Jajcayová, E. Pelantová, and Š. Starosta
Cited as: [VI]
Chapter 8

An algorithm for enumerating all infinite repetitions in a D0L-system

Authors: K. Klouda and Š. Starosta
Cited as: [VII]
Chapter 9

Characterization of circular D0L-systems

Authors: K. Klouda and Š. Starosta
Cited as: [VIII]
Chapter 10

The Rauzy Gasket

Authors: P. Arnoux and Š. Starosta

Cited as: [IX]
Chapter 11

Markov constant and quantum instabilities

Authors: E. Pelantová, Š. Starosta, and M. Znojil
Cited as: [X]