Factor frequencies of reversal closed languages

Ľubomíra Balková

3rd year of PGS, email: `l.balkova@centrum.cz`
Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, CTU
advisor: Doc. Zuzana Masáková, Ph.D., Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, CTU

Abstract. We study infinite words over a finite alphabet. In particular, we focus on frequencies of factors (subwords) of infinite words whose language is reversal closed, i.e. u contains with each factor also its mirror image. Crucial is the notion of Rauzy graphs associated with the infinite word. Investigation of symmetries of the reduced Rauzy graph Γ_n, $n \in \mathbb{N}$, allows us to determine a good and easily calculable upper bound on the number of different factor frequencies.

Abstrakt. Studujeme nekonečná slova nad konečnou abecedou. Speciálně se zaměřujeme na frekvence faktoru nekonečných slov, jejichž jazyk je uzavřen na reverzi, tj. s každým slovem obsahuje také slovo, které získáme, když přečteme dané slovo pozpátku. Klíčovým pojmem je Rauzyho graf přiřazený nekonečnému slovu. Zkoumání symetrií redukovaného Rauzyho grafu Γ_n, $n \in \mathbb{N}$, nám umožní nalézt dobrý a snadno vypočitatelný horní odhad na počet ruzných frekvencí faktoru nekonečného slova.

1 Introduction

Everybody who is about to study a foreign language is interested in word frequencies of this language. The reason is simple. If you start, there is no point in beginning with low-frequency words provided your aim is to manage everyday communication. Word frequencies are in focus of designers of internet search engines, but also of the one who wants to raise the visit rate of his web page. There exist so-called "stoplists" which provide frequencies of most often used words. For instance, just three words *I, and, the* account for ten percent of all words in printed English. This is "easy" to calculate. Prepare a sheet of paper, go through all printed matters in English, for each word you read, put a black tally on the sheet, and each time you see *I* or *and* or *the*, put a red tally on the sheet. At the end, divide the number of red tallies by the number of black tallies and you should obtain approximately $0,1$. In the Czech language, similar role is played by words *a, v, se, na, je, že, o* which take about 9 percent of a written text. In this paper, our point of view will not be linguistic (statistic), we will instead move to the domain of Combinatorics on Words and Graph Theory. We will turn our attention to factor frequencies in infinite words, so the number of occurrences of a factor will be possibly infinite and the definition of factor frequency will have to be generalized. We will show how to find a good upper bound on the number of different factor frequencies in infinite words which contain with every factor also its mirror image. Let us also mention that we have studied factor frequencies in several classes of infinite words (to be found in the thesis) and the results confirm accuracy of the obtained upper bound.

Having introduced notation and basic definitions, we will first recall well-known relations holding for frequencies of edges and vertices in Rauzy graphs (Kirchhoff's law). Afterwards, we will introduce a useful tool- *reduced Rauzy graph*. With this in hand, one can easily deduce the upper bound derived by Boshernitzan (Theorem 1). Knowing that for any infinite reversal closed word u, the mirror map does not change factor frequencies will allow us to improve essentially the upper bound in case of words whose language is reversal closed (Theorem 2).

2 Preliminaries

First, let us recall our "vocabulary" which will be used throughout this paper. An *alphabet* $\mathcal{A}$ is a finite set of symbols called *letters*. A concatenation of letters is a *word*. *Length* of a word w is the number of letters contained in w and is denoted $|w|$. We will also deal with right-sided infinite words $u = u_0u_1u_2....$ A finite word w is called a *factor* of the word u (finite or infinite) if there exist a finite word $w^{(1)}$ and a word $w^{(2)}$ (finite or infinite) such that $u = w^{(1)}ww^{(2)}$. An infinite word u is said to be *recurrent* if each of its factors occur infinitely many times in u and u is *uniformly recurrent* if for any $n \in \mathbb{N}$ there exists an $R(n) \in \mathbb{N}$ such that any factor of u of length $R(n)$ contains all factors of length n. An infinite word u is said to be *eventually periodic* if there exist finite words v, w such that $u = vw^{\omega}$, where w^{ω} means that w is repeated infinitely many times. A word which is not eventually periodic is called *aperiodic*. *Language* $\mathcal{L}(u)$ of an infinite word u is the set of all factors of u. A language $\mathcal{L}(u)$ is reversal closed, if for every factor $w = w_0w_1 \dots w_n$, where $w_i \in \mathcal{A}$, $i \in \{0, \dots, n\}$, also its mirror image $\overline{w} = w_n \dots w_1w_0$ belongs to $\mathcal{L}(u)$. We denote by $\mathcal{L}_n(u)$ the set of factors of length n of the infinite word u. Then, we can define *complexity function* (or *complexity*) $C_u : \mathbb{N} \to \mathbb{N}$ which associates to every n the number of different factors of length n of the infinite word u, i.e.

$$C_u(n) = \#\mathcal{L}_n(u).$$

Let us mention that if there exists $n \in \mathbb{N}$ such that $C_u(n) \leq n$, then the infinite word u is eventually periodic. In other words, aperiodic words has complexity $C(n) \geq n + 1$ for all $n \in \mathbb{N}$. Aperiodic words with the lowest possible complexity are called *Sturmian*. Similarly, let us denote by $\mathcal{P}al_n(u)$ the set of palindromes of length n contained in u and let us define *palindromic complexity* $P_u : \mathbb{N} \to \mathbb{N}$ which associates to every n the number of different palindromes of length n of the infinite word u. We recall that *palindrome* is a word which is equal to its mirror image. We say that $a \in \mathcal{A}$ is *right extension* of a factor $w \in \mathcal{L}(u)$ if wa is also a factor of u. We denote by $Rext(w)$ the set of all right extensions of w in u, i.e. $Rext(w) = \{a \in \mathcal{A} \mid wa \in \mathcal{L}(u)\}$. If $\#Rext(w) \geq 2$, then the factor w is called *right special* (RS for short). Analogously, we define *left extensions, $Lext(w)$, left special factor* (LS for short). Moreover, we say that a factor w is *bispecial* (BS for short) if w is LS and RS. With this in hand, we can give a formula for the *first difference of complexity* $\Delta C_u(n) = C_u(n+1) - C_u(n)$. We leave the proof as an easy exercise.

$$\Delta C_u(n) = \sum_{w \in \mathcal{L}_n(u)} \big(\#Rext(w) - 1\big) = \sum_{w \in \mathcal{L}_n(u)} \big(\#Lext(w) - 1\big), \quad n \in \mathbb{N}. \tag{1}$$

To have everything prepared for the deduction of an improved upper bound on the number of different frequencies, it remains to define *Rauzy graph*, and, of course, *factor frequency.*

Definition 1. *Rauzy graph Γ_n of an infinite word u (of order n) is a directed graph whose set of vertices is $\mathcal{L}_n(u)$ and set of edges is $\mathcal{L}_{n+1}(u)$. Let $w_0, w_1, \ldots, w_n$ be letters in $\mathcal{A}$ and let $e = w_0 w_1 \ldots w_{n-1} w_n$ be an edge of Γ_n, then e starts in the vertex $w = w_0 w_1 \ldots w_{n-1}$ and ends in the vertex $v = w_1 \ldots w_{n-1} w_n$.*

Definition 2. *Let w be a factor of an infinite word u over a finite alphabet $\mathcal{A}$, then (factor) frequency of w (in u) is defined as*

$$\rho(w) = \lim_{|v|\to\infty, v\in\mathcal{L}(u)} \frac{\#\{\textit{occurrences of } w \textit{ in } v\}}{|v|}$$

if the limit exists.

3 Upper bound on the number of factor frequencies

In the sequel, let us suppose that **frequencies of all factors of $\mathcal{L}(u)$ exist**. It is not difficult to see that the frequency of a vertex w in Γ_n is equal to the sum of frequencies of the edges starting in w, or, by symmetry, the sum of frequencies of the edges ending in w. Let us formalize this observation and leave its proof as a simple exercise.

Lemma 1 (Kirchhoff's law)**.** *Let w be a factor of u, then*

$$\rho(w) = \sum_{a\in Lext(w)} \rho(aw) = \sum_{a\in Rext(w)} \rho(wa).$$

Consequently, if a factor $w \in \mathcal{L}(u)$ is neither LS nor RS, then both the frequency of the unique edge starting in w and the frequency of the unique edge ending in w is equal to $\rho(w)$. Formally rewritten, this observation has the following reading.

Corollary 1. *Let w be a factor of u which is neither LS nor RS. Let us denote by a the only left extension of w and by b its only right extension. Then, $\rho(w) = \rho(aw) = \rho(wb)$.*

We can label every edge e in the Rauzy graph Γ_n of u by $\rho(e)$. Then the number of different frequencies of factors in $\mathcal{L}_{n+1}(u)$ corresponds to the number of different edge labels in Γ_n. For a factor $w \in \mathcal{L}_n(u)$ which is neither LS nor RS, it is thus evident that the unique edge ending in w has the same label $\rho(w)$ as the unique edge starting in w. Consequently, if we are interested in the number of different edge labels, we can remove the vertex w from the graph and replace the incoming and outgoing edge with a new edge keeping the label $\rho(w)$. Repeating this procedure, we obtain the so-called reduced Rauzy graph, which has obviously the same set of edge labels. Let us give precise definitions.

Definition 3. *Let Γ_n be the Rauzy graph of order n of an infinite word u. A directed path $w^{(0)} w^{(1)} \ldots w^{(m)}$ in Γ_n such that its initial vertex $w^{(0)}$ is LS or RS, its final vertex $w^{(m)}$ is also LS or RS, and the other vertices are neither LS nor RS factors is called simple. We define label of the simple path as the label of any edge of this path.*

Definition 4. *Reduced Rauzy graph $\tilde{\Gamma}_n$ of u (of order n) is a directed graph whose set of vertices is formed by LS and RS factors of $\mathcal{L}_n(u)$ and whose set of edges is given in the following way. Vertices w and v are connected with an edge e if there exists in Γ_n a simple path starting in w and ending in v. We assign to such an edge e the label of the corresponding simple path.*

The number of different edge labels in the reduced Rauzy graph $\tilde{\Gamma}_n$ is clearly less or equal to the number of edges in $\tilde{\Gamma}_n$. Let us thus calculate the number of edges in $\tilde{\Gamma}_n$ in order to get an upper bound on the number of frequencies of factors in $\mathcal{L}_{n+1}(u)$. For every RS factor $w \in \mathcal{L}_n(u)$, it holds that $\#Rext(w)$ edges begin in w, and for every LS factor $v \in \mathcal{L}_n(u)$ which is not RS, only one edge begins in v, thus we get the following relation

$$\#\{e|\ e \text{ edge in } \tilde{\Gamma}_n\} = \sum_{w \text{ RS}} \#Rext(w) + \sum_{v \text{ LS not RS}} 1. \tag{2}$$

Using Equation 1, we deduce that

$$\#\{e|\ e \text{ edge in } \tilde{\Gamma}_n\} = \Delta C(n) + \sum_{v \text{ RS}} 1 + \sum_{v \text{ LS not RS}} 1. \tag{3}$$

The following result initially proved by Boshernitzan in [3] follows immediately.

Theorem 1 (Boshernitzan). *Let u be an infinite word such that for every factor $w \in \mathcal{L}(u)$, the frequency $\rho(w)$ exists. Then for every $n \in \mathbb{N}$, it holds*

$$\#\{\rho(e) \bigm| e \in \mathcal{L}_{n+1}(u)\} \leq 3\Delta C(n).$$

This upper bound can be lowered for an infinite word u whose language $\mathcal{L}(u)$ is reversal closed. In this case, each factor of u has the same frequency as its mirror image.

Lemma 2. *Let u be an infinite word whose language $\mathcal{L}(u)$ is reversal closed and such that for each factor $w \in \mathcal{L}(u)$, the frequency $\rho(w)$ exists. Then $\rho(w) = \rho(\overline{w})$ holds for each factor w of $\mathcal{L}(u)$.*

Proof. Take an arbitrary factor $w \in \mathcal{L}(u)$ and let $(v^{(n)})_{n=1}^{\infty}$ be any sequence of a strictly growing length in $\mathcal{L}(u)$. Since the frequency of w exists, we can write

$$\rho(w) = \lim_{n\to\infty} \frac{\#\{\text{occurrences of } w \text{ in } v^{(n)}\}}{|v^{(n)}|}.$$

As $\mathcal{L}(u)$ is reversal closed, we get

$$\#\{\text{occurrences of } w \text{ in } v^{(n)}\} = \#\{\text{occurrences of } \overline{w} \text{ in } \overline{v^{(n)}}\}.$$

Using $|v^{(n)}| = |\overline{v^{(n)}}|$, we can then rewrite $\rho(w)$ as follows

$$\rho(w) = \lim_{n\to\infty} \frac{\#\{\text{occurrences of } \overline{w} \text{ in } \overline{v^{(n)}}\}}{|\overline{v^{(n)}}|} = \rho(\overline{w}).$$

The last equality holds thanks to the assumption that frequencies of all factors exist. $\square$

We have now everything prepared for an improvement of the upper bound on the number of edge labels in $\tilde{\Gamma}_n$, or, equivalently, on the number of different factor frequencies of $\mathcal{L}_{n+1}(u)$ of an infinite word u whose language is reversal closed. The following lemma will play an essential role in this improvement.

Lemma 3. *Let u be an infinite word whose language $\mathcal{L}(u)$ is reversal closed and such that for each factor $w \in \mathcal{L}(u)$, the frequency $\rho(w)$ exists. Then for every $n \in \mathbb{N}$, we have*

$$\#\{\rho(e)|e \in \mathcal{L}_{n+1}\} \leq \frac{1}{2}\left(P(n) \ + \ P(n+1) \ + \ \Delta C(n) \ -\sum_{w \textit{ BS in } \mathcal{L}_n} 1 \ - \sum_{w \textit{ BS in } \mathcal{P}al_n} 1\right) + \sum_{w \textit{ RS in } \mathcal{L}_n} 1.$$

Proof. Let Γ_n be the Rauzy graph of u of order n. Let us define a mapping f which to every vertex $w \in \mathcal{L}_n(u)$ associates the vertex $\overline{w}$, to every edge $e \in \mathcal{L}_{n+1}(u)$ associates the edge $\overline{e}$, and to every path $w^{(0)}w^{(1)}\dots w^{(m)}$ in Γ_n associates the path $\overline{w^{(m)}}\dots\overline{w^{(1)}}\ \overline{w^{(0)}}$. Then, clearly, $f^2 = Id$ and thanks to the closeness of $\mathcal{L}(u)$ under reversal, f maps the Rauzy graph Γ_n onto itself, in fact, f is an automorphism of Γ_n. Let us replace the Rauzy graph Γ_n by the reduced Rauzy graph $\tilde{\Gamma}_n$. We know already that the set of edge labels of $\tilde{\Gamma}_n$ is equal to the set of edge labels of Γ_n. Let us denote by A the number of edges e in $\tilde{\Gamma}_n$ such that e is mapped by f onto itself and by B the number of edges e in $\tilde{\Gamma}_n$ such that e is not mapped by f onto itself, then clearly, $\#\{e|\ e \text{ edge in } \tilde{\Gamma}_n\} = A + B$. To be more precise, if e is an edge in $\tilde{\Gamma}_n$ corresponding to the simple path $w^{(0)}w^{(1)}\dots w^{(m)}$ in Γ_n, then $f(e)$ is the edge in $\tilde{\Gamma}_n$ corresponding to the simple path $f(w^{(0)}w^{(1)}\dots w^{(m)}) = \overline{w^{(m)}}\dots\overline{w^{(1)}}\ \overline{w^{(0)}}$. Consequently, if e is mapped by f onto itself, then the corresponding simple path $w^{(0)}w^{(1)}\dots w^{(m)}$ satisfies that its central vertex $w^{(\frac{m}{2})}$ is a palindrome (for m even) or its central edge going from $w^{(\frac{m-1}{2})}$ to $w^{(\frac{m+1}{2})}$ is a palindrome (for m odd). On the other hand, every palindrome of length $n+1$ forms the central edge of a simple path in Γ_n which is mapped by f onto itself and every palindrome of length n forms either the central vertex of a simple path which is mapped by f on itself or is BS and thus a vertex in Γ_n. Therefore,

$$A = P(n) + P(n+1) - \#\{w \in \mathcal{L}_n | w \text{ BS in } \mathcal{P}al_n\}. \tag{4}$$

We subtract the number of palindromic BS factors of $\mathcal{L}_n(u)$ since they form vertices, not edges in $\tilde{\Gamma}_n$. Now, let us turn our attention to edges e which are not mapped by f onto themselves. If e is an edge in $\tilde{\Gamma}_n$ going from a vertex w to v, where $f(e) \neq e$, then there exists an edge e' in $\tilde{\Gamma}_n$ going from $\overline{v}$ to $\overline{w}$ with $e' \neq f(e')$, namely $e' = f(e)$. However, e and e' have the same label. (If e corresponds to the simple path $w^{(0)}w^{(1)}\dots w^{(m)}$ in Γ_n, then e' corresponds to the simple path $\overline{w^{(m)}}\dots\overline{w^{(1)}}\ \overline{w^{(0)}}$ in Γ_n. Lemma 2 implies that the label of these simple paths is the same.) These considerations lead to the following estimate

$$\#\{\rho(e)|\ e \in \mathcal{L}_{n+1}(u)\} \leq A + \frac{1}{2}B = \frac{1}{2}A + \frac{1}{2}(A+B) \tag{5}$$

Rewriting Equation (3), we obtain

$$A + B = \Delta C(n) + 2\sum_{w \text{ RS in } \mathcal{L}_n} 1 - \sum_{w \text{ BS in } \mathcal{L}_n} 1.$$

The statement follows then using Equation (4). $\square$

Theorem 2. *Let u be an infinite word whose language $\mathcal{L}(u)$ is reversal closed and such that for every factor $w \in \mathcal{L}(u)$, the frequency $\rho(w)$ exists. Then for every $n \in \mathbb{N}$, we have*

$$\#\{\rho(e)|e \in \mathcal{L}_{n+1}\} \quad \leq \quad 2\Delta C(n)+1-\frac{1}{2}\left(\sum_{w \ BS \ in \ \mathcal{P}al_n} 1+\sum_{w \ BS \ in \ \mathcal{L}_n} 1\right) \quad \leq \quad 2\Delta C(n)+1.$$

The equality $\#\{\rho(e)|e \in \mathcal{L}_{n+1}(u)\} = 2\Delta C(n)+1$ holds for all sufficiently large n if and only if u is periodic.

Remark 1. *To approve that the estimate from Theorem 2 cannot be easily lowered keeping its general validity, let us demonstrate that it is reached for all lengths $n \in \mathbb{N}$ in the case of Sturmian words. Thanks to [4], we know that every Sturmian word is reversal closed and all BS factors are palindromes. Moreover, since $\Delta C(n) = 1$ for all $n \in \mathbb{N}$, the upper bound on the number of different frequencies can be simplified as follows*

$$\#\{\rho(e)|e \in \mathcal{L}_{n+1}(u)\} \quad \leq \quad 3 \; - \sum_{w \ BS \ in \ \mathcal{L}_n} 1.$$

To see that the upper bound is reached, it suffices to recall the result of Berthé in [2]

$$\#\{\rho(e)|e \in \mathcal{L}_{n+1}(u)\} = \begin{cases} 2 & \text{if } n \text{ is the length of a BS factor,} \\ 3 & \text{otherwise.} \end{cases}$$

Proof of Theorem 2. It has been shown in [1] that

$$P(n)+P(n+1) \leq \Delta C(n)+2 \quad \text{for every } n \in \mathbb{N}. \tag{6}$$

The term $\sum_{w \text{ RS in } \mathcal{L}_n} 1$ can be bounded by $\sum_{w \text{ RS in } \mathcal{L}_n} (\#Rext(w)-1) = \Delta C(n)$. Applying these bounds on the result of Lemma 3, we obtain

$$\#\{\rho(e)|e \in \mathcal{L}_{n+1}\} \leq 2\Delta C(n)+1-\frac{1}{2}\left(\sum_{w \text{ BS in } \mathcal{P}al_n} 1+\sum_{w \text{ BS in } \mathcal{L}_n} 1\right).$$

Let us turn our attention to eventually periodic words. Since $\mathcal{L}(u)$ is reversal closed, it follows immediately that u is recurrent. If u is eventually periodic and recurrent, then u is known to be periodic. Thus, there exists a minimal period K such that $u = z^\omega$, where $|z| = K$. Then, $C(n) = K$ for every $n \geq K$ and every factor of length n occurs with frequency $\frac{1}{K}$. Thus, $\#\{\rho(e)| \ e \text{ edge in } \Gamma_n\} = 2\Delta C(n)+1 = 1$ for $n \geq K$. If u is aperiodic, then $\Delta C(n) \geq 1$ together with the fact that every LS factor is prefix of a BS factor implies that for every $N \in \mathbb{N}$, there exists a BS factor in $\mathcal{L}(u)$ of length $n \geq N$, hence $\#\{\rho(e)| \ e \text{ edge in } \Gamma_n\} \leq 2\Delta C(n)+1-\frac{1}{2}\left(\sum_{w \text{ BS in } \mathcal{P}al_n} 1+\sum_{w \text{ BS in } \mathcal{L}_n} 1\right) < 2\Delta C(n)+1$. □

For completeness' sake, let us mention another proof which will not use Equation (6), nevertheless, similar ideas as those ones occurring in [1] will be present. Going through this second version of the proof, it can be in particular noticed that Theorem 2 does not require uniform recurrence of the infinite word u. We will keep notation from Proof of Lemma 3 and we will make use of a partial result rewritten in a different way this time:

$$\#\{\rho(e)|e \in \mathcal{L}_{n+1}(u)\} \le A + \frac{1}{2}B = (A+B) - \frac{1}{2}B. \tag{7}$$

We want to find a lower bound on B, i.e. on the number of edges in $\tilde{\Gamma}_n$ which are not mapped by f on themselves. $\tilde{\Gamma}_n$ contains the following disjoint subgraphs (whose union comprises all vertices of $\tilde{\Gamma}_n$) of three types:

1. subgraphs containing two vertices w and $\overline{w}$, where w is RS not LS, and all edges connecting them mutually
2. subgraphs containing two vertices w and $\overline{w}$, where w is non-palindromic BS, and all edges connecting them mutually (attention! number of subgraphs of this type is just $\frac{1}{2}\#\{w \in \mathcal{L}_n(u) \mid w \text{ non-palindromic BS}\}$)
3. subgraphs containing one vertex w, where w is a palindromic BS, and eventually edges-loops starting and ending in w

Clearly, all edges in $\tilde{\Gamma}_n$ which are mapped by f on themselves are contained in the above subgraphs. Since (reduced) Rauzy graphs of infinite words are connected, each subgraph is connected with an edge to the union of the remaining subgraphs. Moreover, since the language $\mathcal{L}(u)$ is reversal closed, if an edge e starts in a subgraph Γ and ends in a subgraph Γ', then the edge $f(e)$ starts in Γ' and ends in Γ. It follows that B is greater or equal to $2\times$ the minimal number of edges which can ensure connection of the disjoint subgraphs of the graph:

$$B \ge 2 \times \text{number of subgraphs} - 2 = 2\sum_{w \text{ RS in } \mathcal{L}_n} 1 + \sum_{w \text{ BS in } \mathcal{P}al_n} 1 - \sum_{w \text{ BS in } \mathcal{L}_n} 1 - 2. \tag{8}$$

Implanting in Equation (7) the just deduced lower bound on B together with the expression of $A+B$ derived in Proof of Lemma 3

$$A + B = \Delta C(n) + 2\sum_{w \text{ RS in } \mathcal{L}_n} 1 - \sum_{w \text{ BS in } \mathcal{L}_n} 1,$$

and with the fact that $\sum_{w \text{ RS in } \mathcal{L}_n} 1$ can be bounded by $\sum_{w \text{ RS in } \mathcal{L}_n}(\#Rext(w) - 1) = \Delta C(n)$, we have proved the upper bound from Theorem 2

$$\#\{\rho(e)|\ e \text{ edge in } \Gamma_n\} \le 2\Delta C(n) + 1 - \frac{1}{2}\left(\sum_{w \text{ BS in } \mathcal{P}al_n} 1 + \sum_{w \text{ BS in } \mathcal{L}_n} 1\right).$$

To conclude, let us throw in that we have studied frequencies of infinite words associated with β-integers for β being a quadratic non-simple Parry number, thus defined over a two-letter alphabet, and we have learned that the upper bound from Theorem 2 is either reached (for most of the lengths) or is only by 1 greater than the real number of factor frequencies of a given length. Another example of an infinite word, even over a k letter alphabet, where the upper bound is reached for all lengths, is the k-interval exchange word. (Description of frequencies has been recently given by Ferenczi [5].)

References

[1] P. Baláži, Z. Masáková, E. Pelantová. *Factor versus palindromic complexity of uniformly recurrent infinite words*. Theoret. Comput. Sci. vol. **380**, issue **3** (2007), 266–275.

[2] V. Berthé. *Fréquences des facteurs des suites sturmiennes*. Theoret. Comput. Sci. **165** (1996), 295–309.

[3] M. Boshernitzan. *A condition for minimal interval exchange maps to be uniquely ergodic*. Duke Math. **52** (1985), 723–752.

[4] X. Droubay, J. Justin, G. Pirillo. *Episturmian words and some constructions of de Luca and Rauzy*. Theoret. Comput. Sci. **255** (2001), 539–553.

[5] S. Ferenczi, L. Zamboni. *Combinatorial structure of symmetric k-interval exchange transformation*. Preprint available at http://iml.univ-mrs.fr/$\sim$ferenczi/