Complexity of cutting words on regular tilings

Pascal Hubert *and Laurent Vuillon †

Abstract: *We show that the complexity of a cutting word u in a regular tiling by a polyomino Q is equal to $P_n(u) = (p+q-1)n+1$ for all $n \geq 0$, where $P_n(u)$ counts the number of distinct factors of length n in the infinite word u and where the boundary of Q is constructed by $2p$ horizontal and $2q$ vertical unit segments.*

Keywords: Cutting words, regular tilings, complexity function, flow on the torus, combinatorics on words.

1 Introduction

Cutting words on regular tilings are natural coding of geometrical objects appearing in billiard theory [1, 7, 11], combinatorics on words [5, 8], dynamical systems [6] and optimization theory [9]. The simplest case is given by a half-line with irrational slope in a unit square grid. This geometrical object is coded in order to find an infinite word u called *cutting word on the square grid*. The plane is equipped with the unit grid, and, travelling along the half-line, one meets the vertical and the horizontal segments of the grid (we assume the half-line avoids the vertices; in any case it can happen at most once). Encode the former event by a and the latter by b. Thus to the half-line an infinite word u of a's and b's corresponds, which we call the cutting word of the half-line.

This construction leads to the well known Sturmian words [4, 5, 8]. Sturmian words are infinite words over a two-letter alphabet that have exactly $n+1$ factors of length n for each $n \geq 0$. These infinite words have many equivalent characterizations. In particular, Sturmian words are cutting words on

*Institut de Mathématiques de Luminy, CNRS UPR 9016, Université Aix-Marseille II, Campus de Luminy, Case 907 13288 Marseille Cedex 9, France, hubert@iml.univ-mrs.fr

†Laboratoire de Mathématiques, CNRS UMR 5127, Université de Savoie, 73376, le Bourget-du-Lac, France, Laurent.Vuillon@univ-savoie.fr

a square grid with irrational slope and also are coding of billiard words on a square table with irrational direction. On a square table, we consider a starting point and an irrational direction; then a point moves under the table with constant speed. When it reaches the boundary, the angle of incidence is equal to the angle of reflection. In order to have an infinite word, we code the trajectory on the square billiard so that opposite edges are coded by the same letter. In fact, for cutting words on square grid and for coding of square billiards the two constructions give Sturmian words. More generally, we can define cutting words on tilings and coding of polygonal billiard. In some cases, the two constructions coincide in particular for so called tiling billiards; that is square billiards, equilateral triangle billiards and isoceles right triangle billiards (see [7, 10]). In the majority of cases, the notions of cutting words on tilings and polygonal billiard words are different.

In this article, we investigate cutting lines on a regular tiling by computing the complexity of the infinite cutting word u. We find the complexity of cutting words for each regular tiling by translations of one polyomino. Our methods are based on combinatorics on words and also on geometrical notions such as flow on a torus and computation of the Euler characteristic. The last part deals with the link between cutting words, Sturmian words and automata.

2 Cutting words and complexity

A *polyomino* or tile is a simply connected union of unit squares, that is a union of squares forming a 4-connected figure without any holes. Let Q be a polyomino. A *tiling by translations* of Q is a covering of the whole plane using translates of Q, such that the interiors of the tiles do not overlap. A *reduced word* is a word on the free group on $\{a, b\}$ (with $\bar{a}$ is the inverse of a and $\bar{b}$ is the inverse of b) where all cancellations are done (namely each occurrence of $a\bar{a}$, $\bar{a}a$, $b\bar{b}$ and $\bar{b}b$ is replaced by the empty word). Let C be the *boundary word* of the polyomino Q that is a reduced word in the free group on $\{a, b\}$, where a represents a right step, b a up step, $\bar{a}$ a left step and $\bar{b}$ a bottom step, that codes the boundary of the polyomino Q in the following way. Starting from a point on the boundary of the polyomino Q, the boundary word C is the concatenation of labels of boundary unit segments read in trigonometric order. The starting point is not meaningful. Thus the boundary word C is a cyclic word.

The bar operation can be extended to reduced words by $\overline{w} = \overline{w_1 w_2 \cdots w_n} = \overline{w}_n \overline{w}_{n-1} \cdots \overline{w}_1$ where $w_i \in \{a, b, \bar{a}, \bar{b}\}$. And of course we have that $\bar{\bar{a}} = a$ and

$\bar{\bar{b}} = b$.

A theorem of Beauquier and Nivat [3] states:

Theorem 1 (Beauquier, Nivat) *A polyomino Q tiles the plane by translations if and only if the boundary word C is equal up to a cyclic permutation of the symbols to $xyz\bar{x}\bar{y}\bar{z}$ or $xy\bar{x}\bar{y}$ where $x, y, z \in \{a, b, \bar{a}, \bar{b}\}^*$.*

If the boundary word is equal to $xyz\bar{x}\bar{y}\bar{z}$ (resp. $xy\bar{x}\bar{y}$) such a polyomino is called pseudo-hexagon (resp. pseudo-square). For example, the polyomino on the left in Figure 1 is a pseudo-hexagon and the boundary word is equal to $xyz\bar{x}\bar{y}\bar{z} = a\ ab\ \bar{a}b\ \bar{a}\ \bar{b}\bar{a}\ \bar{b}a$ (where $x = a, y = ab, z = \bar{a}b$). More geometrically, the boundary of Q is given by a pseudo-hexagon that has six sides in correspondence two-by-two by one of the following translations by the vectors t_1, t_2 or t_3 where t_1, t_2 are noncollinear vectors (see Figure 1)); by construction $t_3 = t_1 + t_2$ and t_1 (resp. t_2, t_3) maps x (resp. y, z) to $\bar{x}$ (resp. $\bar{y}, \bar{z}$). In case of a pseudo-square, four sides are in correspondence two-by-two by one of the following translations by the vectors t_1, t_2 where t_1, t_2 are noncollinear vectors.

Given a tiling P by a polyomino Q, a vector v is a translation vector of P if there exist tiles Q_1 and Q_2 in P such that Q_2 is the translation of Q_1 by v. A tiling P of the plane by one polyomino is called *regular* if the vectors of translations form the lattice R (for each vector of translation v, there exist a and b such that $v = at_1 + bt_2, (a, b) \in \mathbb{Z}^2$ where t_1 and t_2 are the translation vectors constructed by the previous mapping of the sides of Q). A subset S is a *fundamental domain* for the lattice R if $\forall (x, y) \in \mathbb{R}^2$, it exists a unique $v \in R$ such that $(x, y) + v \in S$. The polyomino Q is a fundamental domain for the lattice R (see Figure 2a for a regular tiling and 2b for a nonregular tiling). In combinatorial terms, the tiling is regular if each copy of the polyomino has the same factorization of its boundary word, that is each tile of a regular tiling is surrounded in the same way by copies of the tile (see [3]). For example, in Figure 1 the boundary word of the right polyomino is equal to either $\bar{b}a\ aba\ b\ \bar{a}b\ \bar{a}\bar{b}\bar{a}\ \bar{b}$ or $\bar{b}a\ a\ bab\ \bar{a}b\ \bar{a}\ \bar{b}\bar{a}\bar{b}$ and both factorizations leads to different computations of t_1, t_2 and t_3 the translation vectors and then to different regular tilings.

We suppose that the boundary of the polyomino Q is built with $2p$ unit horizontal segments and $2q$ unit vertical segments. We code two segments identified by one translation t_1, t_2 or t_3 by the same letter (note that now we just code a unit segment by a letter identifying a_i and $\overline{a}_i$, b_i and $\overline{b}_i$). Thus the polyomino Q and the tiling P are coded by $p+q$ letters on the alphabet $A =$

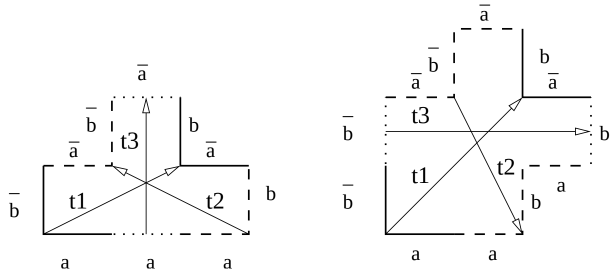

Figure 1: Polyomino and translations.

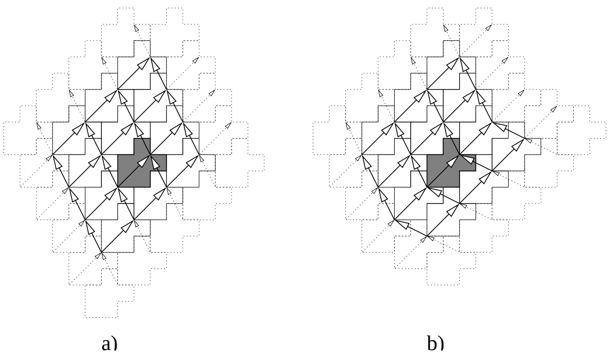

Figure 2: Regular and nonregular tilings.

$\{a_1, a_2, \cdots, a_p, b_1, b_2, \cdots, b_q\}$ where the a_i's code unit horizontal segments and the b_i's code unit vertical segments (see Figures 3 and 4).

Let D be a half-line with irrational slope α that contains no points with both integral coordinates. Let M_0 be the origin of D. We define the cutting word by the sequence of segments met by the line from M_0 to infinity. By this construction, we obtain an infinite word $u_{M_0,\alpha} = u_0u_1\cdots$ on the alphabet A (see Figure 4).

A factor of an infinite word $u_0u_1\cdots$ is a word w such that $w = u_iu_{i+1}\cdots u_j$ for some i, j with $i \leq j$. Consider now the set $L(n, u)$ of all factors of length n of the infinite word u, defined by $L(n, u) = \{w | u_iu_{i+1}\cdots u_{i+n-1} = w, \text{ for some } i \in \mathbb{N}\}$. The complexity function (sometimes called subword complexity function) computes the number of distinct factors of length n in u that is $P_n(u) = \text{Card } L(n, u)$.

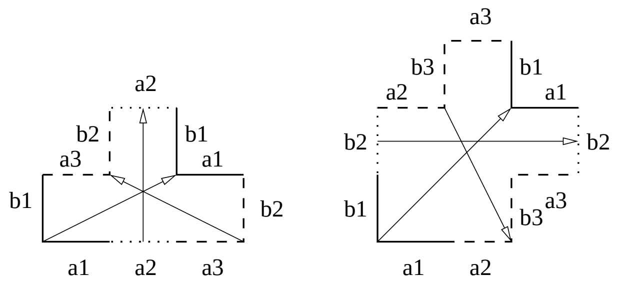

Figure 3: Coding of polyominoes.

Theorem 2 *The complexity $P_n(u)$ of the cutting word is given by*

$$P_n(u) = (p + q - 1)n + 1, \text{ for all } n \geq 0.$$

We prove Theorem 2 in the next section. The proof is divided in four lemmas. A special case of this theorem is the classical result for cutting words on a regular tiling with a unit square [8, 5]. We just construct in this case the well-known Sturmian words with complexity $P_n(u) = n + 1$. As an example of our result, by Theorem 2 the complexity function for cutting words in the regular tiling of Figure 4 is $P_1(u) = \text{Card } \{a_1, a_2, a_3, b_1, b_2\} = 5$ (the cardinality of the alphabet), $P_2(u) = 9$ (the number of factors of length two of u) and for all n, $P_n(u) = 4n+1$ independently of the origin point M_0 and of the irrational slope α.

In the next section, we will see that cutting words with same irrational slope α for a given regular tiling P have the same infinite set of finite factors. That is, all cutting words with the same irrational slope α have the same language (the language of u is $L(u) = \cup_n L(n, u)$) independently of the origin of the half-line.

3 Flow on tori

Let R be a sub-lattice of $\mathbb{Z}^2$, Π^2 be the two-dimensional torus given by the quotient $\mathbb{R}^2/R$ and let π be the canonical projection on the torus $\pi : \mathbb{R}^2 \to \mathbb{R}^2/R = \Pi^2$.

Let ϕ_t be the flow of constant slope α on the torus Π^2. This means that each point is moved on the torus along a line of slope α with a constant

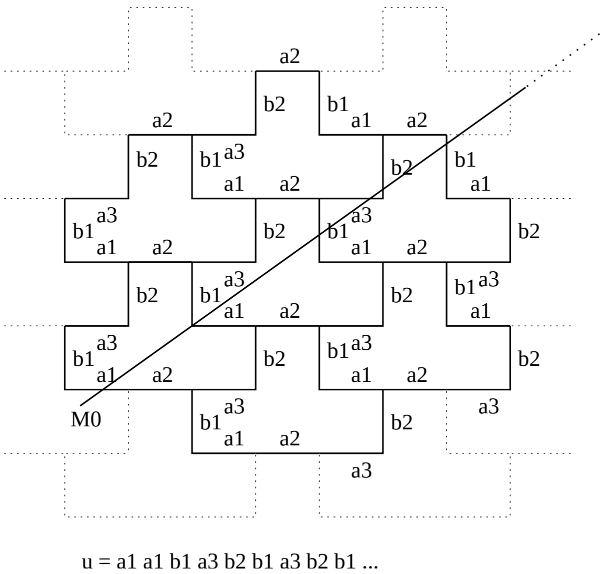

Figure 4: Cutting word.

speed t. In particular, the image of a point (x, y) by the flow ϕ_t is

$$\phi_t(x, y) = \left(\frac{t}{\sqrt{1+\alpha^2}} + x, \frac{t\alpha}{\sqrt{1+\alpha^2}} + y \right) \mod R.$$

A flow is called *minimal* on Π^2 if the trajectory of every point is dense on the torus Π^2.

Lemma 1 *The flow ϕ_t of constant irrational slope α is minimal on Π^2.*

Proof This is a general result of dynamical systems theory. The minimality is implied by the irrationality of α (see [11] chapter 3 page 58–59 for example). ∎

Definition 1 *A point M in the boundary of the polyomino Q is called in-going (resp. out-going) point if a nonempty segment of line with slope α contained in the interior of Q ends (resp. begins) in M (see Figure 5).*

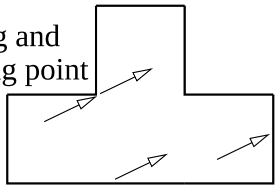

Figure 5: In-going and out-going points.

Remark 1 *We extend the notion of coding to half-lines with irrational slope α that contain a point with both integer coordinates. Such a half-line defines two cutting words which only differ by two consecutive letters. We arbitrarily choose one of them.*

In the next lemma, we just consider trajectories beginning at out-going points.

Lemma 2 *The complexity $P_n(u)$ is the number of distinct prefixes of length n of cutting words starting at out-going points.*

Proof This lemma is a consequence of Lemma 1. As the trajectory of M_0 by the flow ϕ_t is dense on Π^2 all finite prefixes of cutting words from out-going points of Q appear in u. ∎

Lemma 3 *Let k be the number of vertices of Q in Π^2 ($k =$ Card $\{\pi(v),$ v vertices of $Q\}$). Then*

$$k = p + q - 1.$$

Proof The Euler characteristic of a torus is 0 and $\chi(\Pi^2) = v - e + f$ where v is the number of vertices of Q in Π^2, e is the number of edges in Π^2 and f the number of faces in Π^2.

Since k is the number of vertices of Q in Π^2, $v = k$. The boundary of the polyomino Q has $2(p + q)$ segments in the plane. In Π^2, the segments are pairwise identified, that is $e = p + q$. As Q is simply connected then it follows that $f = 1$. Thus we have $k = p + q - 1$. ∎

Definition 2 *The geometric cylinder of order* n *of a word* $d = d_0 d_1 \cdots d_{n-1}$ *is the set of out-going points such that* $d_0 d_1 \cdots d_{n-1}$ *codes the first* n *intersections of the cutting line of these out-going points and this set is denoted* C_d. *If* a *is a letter,* C_a *is called the out-going segment associated with* a.

In the following an interval is a connected subset of an out-going segment.

Lemma 4 *Geometric cylinders are intervals and*

$$P_{n+1}(u) - P_n(u) = p + q - 1.$$

Proof By induction on n.

Basis case: For $n = 0$, $P_0(u) = 1$. Indeed the unique word of length 0 is the empty word. For $n = 1$, $P_1(u) = p+q$. Furthermore, geometric cylinders of length 1 are intervals. Indeed they are exactly the $p+q$ out-going segments of the polyomino Q.

Induction: We suppose that geometric cylinders of length n are intervals. Let T be the first return map of the flow ϕ_t on the boundary δQ of Q. A word w of length n has j followers $(wv_1, wv_2, \cdots, wv_j)$ if

$$T^{-n-1}(v_1), T^{-n-1}(v_2), \cdots, T^{-n-1}(v_{j-1})$$

are elements of C_w where the v_i's are distinct vertices of Q in Π^2. Indeed, if $C_w = [a, b]$ then the intervals $[a, T^{-n-1}(v_1)]$, $[T^{-n-1}(v_1), T^{-n-1}(v_2)]$, $\cdots$ $[T^{-n-1}(v_{j-1}), b]$ are associated with the cylinders $C_{wv_1}, C_{wv_2}, \cdots, C_{wv_j}$ with $v_i \neq v_j$ if $i \neq j$. In addition to that, they are non empty intervals because the irrationality of α implies $T^{-i}(v_k) \neq T^{-j}(v_\ell)$ if $(i, k) \neq (j, \ell)$. Furthermore, all these geometric cylinders are distinct because their last letters are different (the segments of the polyomino Q are coded by different letters). We have just shown that cylinders of length $n+1$ are intervals. Thus the complexity is increasing by the number of vertices $\pi(v)$ on the torus (i.e., $P_{n+1}(u) - P_n(u) = k = p + q - 1$). ∎

Lemmas 2, 3 and 4 give the Theorem 2.

4 Related constructions

The number of vertices $\pi(v)$ on Π^2 is the number of in-going vertices of Q (and also the number of out-going vertices of Q). Indeed, two in-going vertices correspond to two distinct points in Π^2 because the codings of the

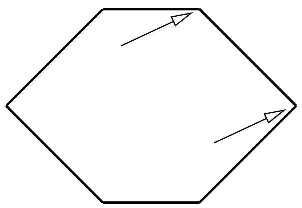

Figure 6: In-going vertices on an hexagon.

nearest segments are different. Conversely, each vertex of Π^2 is associated with at least one in-going vertex because a trajectory of the flow ϕ_t arriving at each point of the torus always exists.

Now we present a more general setting of the theorem:

Theorem 3 *Let Q be a simply connected polygon that tiles the plane by translations. If the slope of cutting lines and the slopes of the edges of Q are linearly independent over the rational numbers then the complexity $P_n(u)$ of cutting words on this tiling is given by $P_n(u) = kn + 1, \forall n \geq 0$, where k is the number of in-going vertices of Q.*

Proof Same structure as the preceding proof. We just point out the crucial fact that the complexity is increasing by the number of vertices on Π^2. That is by the number of in-going vertices of Q. ∎

Example: the complexity of the cutting word on a tiling by hexagon is equal to $P_n(u) = 2n + 1$ because the number of in-going vertices is two (see Figure 6).

5 Cutting words and automata

The goal of this section is to construct a cutting word for a given regular tiling by Q using the information of the related Sturmian word S (the related Sturmian word S is the Sturmian word of starting point M_0 and slope α on the square grid). More precisely, for fixed M_0 and α, we construct the cutting words $u_{M_0,\alpha}$ on the regular tiling by a polyomino Q from the associated Sturmian word of slope α and starting point M_0.

The boundary of the polyomino Q is built with $2p$ unit horizontal segments and $2q$ unit vertical segments. As in section 2, the boundary is coded by the $p+q$ letters of the alphabet $A = \{a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q\}$.

Now, the unit segments in the interior of Q are coded in an extra alphabet $E = \{c_1, c_2, \ldots, c_r, d_1, d_2, \ldots, d_s\}$ where c_i's code distinct unit horizontal segments and d_i's code distinct unit vertical segments (see Figure 7).

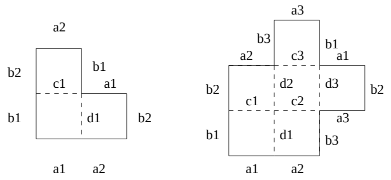

Figure 7: Polyominoes and codings.

By this coding, the boundary of each unit square of the polyomino Q (and the tiling) is coded by the alphabet $A \cup E$. Let CS ($CS_{M_0,\alpha}$) be the cutting sequence on the alphabet $A \cup E$ and S be the corresponding Sturmian word.

This section explains that CS is obtained from S by a deterministic finite automaton. Then one gets $u_{M_0,\alpha}$ from CS by a projection.

In this construction, we assume that the slope α is positive. Given a unit square in the tiling, when the coding of one out-going segment is determined then there are only two possibilities for the coding of the in-going segments of the same square. Observe that one of these segments is horizontal and the other one is vertical. This remark leads to a deterministic finite automaton $\mathcal{A}$ that codes the Sturmian words by a cutting sequence CS on the square lattice using the alphabet $A \cup E$. The states of $\mathcal{A}$ are the letters of $A \cup E$ and the input alphabet is $\{a, b\}$. Given a state s in $A \cup E$ and a label ℓ in $\{a, b\}$, then the next state is $\mathcal{A}(s, \ell) = s'$.

For example, Figure 8 displays the associated automaton for the triomino (Polyomino on the left of Figure 7).

Now suppose that M_0 is an out-going point of Q. Then the first letter of the cutting word CS is in the alphabet A. The next letter of CS is determined by $CS(1), S(2)$ and the automaton $\mathcal{A}$. $CS(1)$ encodes an out-going segment and determines a unique unit square on Q and $S(2)$ gives a way to find the letter $CS(2)$; indeed

$$CS(2) = \mathcal{A}(CS(1), S(2)).$$

More generally $CS(i+1) = \mathcal{A}(CS(i), S(i+1))$.

Example for the triomino: If $S = baababaababaabaababaabab \cdots$ and $CS(1) = b_1$ then as $S(2) = a$ we have $CS(2) = \mathcal{A}(b_1, a) = c_1$ (see Figure 8).

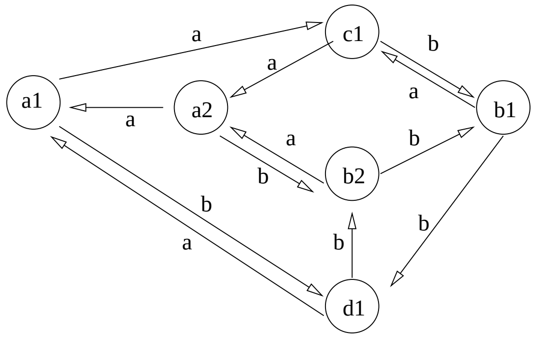

Figure 8: Automaton for the triomino.

Thus we build the infinite word

$$CS = b_1c_1a_2b_2a_2b_2a_2a_1d_1a_1d_1a_1c_1b_1c_1a_2b_2a_2b_2a_2a_1d_1a_2b_2\cdots$$

In order to construct the cutting word $u_{M_0,\alpha}$ associated with Q, M_0 and α, we just remove the letters of E in CS.

Let ρ be the map:
$a_i \to a_i$,
$b_i \to b_i$,
$c_i \to \epsilon$,
$d_i \to \epsilon$.

The cutting word $u_{M_0,\alpha}$ is given by $u_{M_0,\alpha} = \rho(CS)$. The cutting of the example with the triomino is

$$u_{M_0,\alpha} = \rho(CS) = b_1a_2b_2a_2b_2a_2a_1a_1a_1b_1a_2b_2a_2b_2a_2a_1a_2b_2\cdots$$

and the complexity of $u_{M_0,\alpha}$ is $p(n) = 3n + 1$.

We remark that CS has complexity $p(n) = t(n+1)$ where t is the number of squares composing the polyomino Q. Given a word of length n in the original Sturmian word, we build t words of length n in CS, one for each square contained in Q. As there are $(n+1)$ Sturmian words of length n, it follows that there are $t(n+1)$ words of length n on the alphabet $A \cup E$.

To summarize, the preceding construction gives the following proposition.

Proposition 1 *The cutting word $u_{M_0,\alpha}$ is obtained by a deterministic finite automaton and a projection ρ from the associated Sturmian word S of starting point M_0 and slope α.*

6 Conclusion

This article is in fact a first step in the study of cutting words, however many questions remain open. In particular, Crisp, Moran, Pollington and Shiue [8] showed that cutting words on the grid are intimately related to substitutions. Similar work should be done for cutting words on regular tilings. In higher dimensions, the billiard on a cube is also a well-known object [1, 2]. In this context the question is: what is the analogue of our construction for regular tilings of $\mathbb{R}^3$?

References

[1] P. Arnoux, C. Mauduit, I. Shiokawa and J.-I. Tamura, *Complexity of sequences defined by billiards in the cube*, Bull. Soc. Math. France, **122** (1994), 1–12.

[2] Y. Baryshnikov, *Complexity of trajectories in rectangular billiards*, Comm. Math. Phys., **174** (1995), 43–56.

[3] D. Beauquier and M. Nivat, *On translating one polyomino to tile the plane,* Discrete Comput. Geom.,**6** (1991), 575–592.

[4] J. Berstel, *Recent results in Sturmian words,* Developments in Language Theory, (1995), 13-24.

[5] J. Berstel and P. Séébold, *Sturmian words,* in M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, (2002).

[6] V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel eds., *Introduction to finite automata and substitution dynamical systems*, Springer-Verlag, Lecture Notes in Mathematics, 1794, (2002).

[7] J. Cassaigne, P. Hubert and S. Troubetzkoy, *Complexity and growth for polygonal billiards*, Annales de l'institut Fourier, **52** (2002), 1001–1013.

[8] D. Crisp, W. Moran, A. Pollington and P. Shiue, *Substitution invariant cutting sequences*, J. Théor. des Nombres de Bordeaux, **5** (1993), 123–137.

[9] J. Mairesse and L. Vuillon, *Asymptotic behavior in a heap model with two pieces*, Theoret. Comput. Sci., **270** 1–2 (2002), 525–560.

[10] P. Hubert, *Propriétés combinatoires des suites définies par billard dans les triangles pavants*, Theoret. Comput. Sci., **164** (1996) 165–183.

[11] S. Tabachnikov, *Billiards*, Panoramas et Synthèses, Société mathématique de France, n. 1, (1995).