CROATICA CHEMICA ACTA CCACAA **67** (1) 45–53 (1994)

ISSN 0011-1643
CCA–2164
Conference Paper

A New Coding for Parallelogram Animals*

M. Delest

LaBRI – Université Bordeaux I, 351 Cours de la Libération, 33405 Talence Cedex, France

and

J. P. Dubernard

LIUP – Université de Poitiers, 40 Avenue du Recteur Pinaud, 86022 Poitiers Cedex France

Received April 15, 1994

We describe a new coding for parallelogram animals and Motzkin words. From this new coding, we deduce a relation between the Catalan numbers, Motzkin words and parallelogram animals.

INTRODUCTION

In the cartesian plane IN × IN, a finite connected union of cells (unit squares) without a cut point is called a polyomino. Studying polyominoes presents a long set of problems, see for instance surveys in Refs. 1, 10 and 15. This problem is also well-known in statistical physics. Physicist usually consider animals, instead of polyominoes, an equivalent object obtained by replacing the centre of each elementary cell. Several parameters can be defined on these objects. For an animal A, the most often studied parameters are: the *area* $a(A)$ which is the number of cells, the *perimeter* $p(A)$ which is the length along the border and the site perimeter $s(A)$ which it the number of cells along the border. The general problem is open. Thus, many authors study special cases.

A *column* (resp. *row*) of an animal is the intersection of the animal with an infinite vertical (resp. horizontal) strip of cells. An animal is a *column-convex* (resp. *row-convex*) one if every column (resp. row) is connected. The most recent result on column-convex animals it the one of Ref. 14 which leads to new enumeration results.

*This work was supported by the »PRC de Mathématiques et Informatique« and Sun-Microsystem France.

An animal is said to be *convex* if it is both row- and column-convex. A particular case arises when each column of convex animal has

– no column on its right with a cell lower than its lowest cell,

– no column on its left with a cell higher than its highest cell.

These are named parallelogram animals. An example is displayed in Figure 1. They are classically counted according to their perimeter by the Catalan number, that is the number of parallelogram animals having a perimeter $2n+2$ is

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

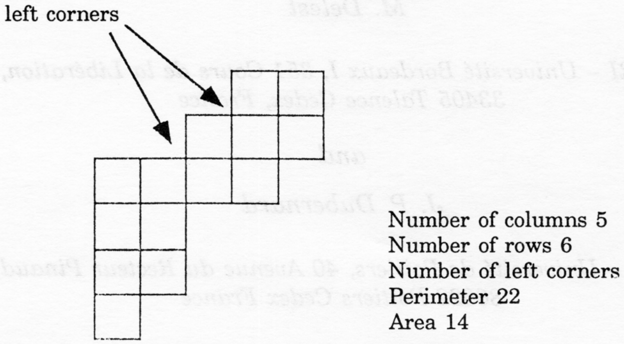

Figure 1. A parallelogram polyomino.

These animals were first studied according to the area by Polya.[12] The Polya formula was recently proved using a bijection by Flajolet.[7] Gessel in Ref. 8. gives a new functional equation for it. Delest, Gouyou-Beauchamps and Vauquelin[3] have given a formula according to both the perimeter and the site perimeter (*i.e.* the number of cells along the border). Recently, Delest and Fedou[2] relied them to q-Bessel functions. These last results are based upon a bijection found by Delest and Viennot and described in Ref. 4. We will recall it in the next section.

Here, a new bijection is presented, from which results can be deduced for the enumeration of parallelogram animals according to the area, the number of columns and the number of left corners (*i.e.* number fo columns where highest cell is higher than the highest cell of the column on its left). This bijection links parallelogram animals whit Motzkin words,[6] that is words written on the alphabet $\{x, \bar{x}, a\}$.

After some definitions and notations, we describe the bijection. Then, in the last paragraph, we construct a bijection between Motzkin words and Dyck words such that the new coding is the usual one for parallelogram animals.

DEFINITIONS AND NOTATIONS

Let X be an alphabet; we denote by X^* the free monoid generated by X, that is the set of words written with finite sequences of letters from X. The *empty word* is denoted by ε. The number of occurrences of the letter x in the word w is denoted by $|w|_x$, the length (number of letters) of w by $|w|$.

The set D of Dyck words w is written over $\{x, \bar{x}\}$ and satisfies the following conditions:

(i) for any left factor f such that $w = fg$, $|f|_x \geq |f|_{\bar{x}}$,
(ii) $|w|_x = |w|_{\bar{x}}$.

The *Motzkin* language M is the »shuffle« product of D and $\{a\}^*$, $M = D \varpi \{a\}^*$, which is the set of words w written over $\{x, \bar{x}, a\}$ satisfying the above conditions (i) and (ii). In the following, we will denote by X the alphabet $\{x, \bar{x}\}$ and by Y the alphabet $\{x, \bar{x}, a\}$. For every word w written on these alphabets, we define $\delta(w) = |w|_x - |w|_{\bar{x}}$. Let w be a Dyck (resp. Motzkin) word; it is prime if and only if for every non empty left factor w' of w not equal to w, $\delta(w') > 0$.

Classically, one associates with the languages their generators which are grammars. An algebraic grammar is a 4-uples $G=\langle T,V,S,P\rangle$ such that

(i) T is a terminal alphabet,
(ii) V is a non terminal alphabet,
(iii) S in V is the axiom,
(iv) P is the set of rules of the form $Y \to w$ with w in $(X \cup V)^*$.

For example, the language M of Motzkin words is generated by the grammar

$$< Y, \{M\}, M, \{M \to \varepsilon, M \to a M, M \to x M \bar{x} M\}>,$$

and the language D of Dyck words by

$$< X, \{D\}, D, \{D \to \varepsilon, D \to x D \bar{x} D\}>.$$

By extension, we denote by the same letter the language L and its enumerative formal power series $\Sigma_{w \text{ in } L}\, w$. Thus, we can write equations in C<<X>>, which are for the Dyck language:

$$D = x D \bar{x} D + \varepsilon$$

and for the Motzkin language:

$$M = x M \bar{x} M + a M + \varepsilon.$$

A path ω in the cartesian plane $\mathbb{N} \times \mathbb{N}$ is a sequence of points $(s_0, s_1, ..., s_n)$. The length of the path ω is n and the pairs (s_i, s_{i+1}) are called elementary steps. They are North-East (resp. South-East, resp. East) if $s_i = (x, y)$ and $s_{i+1} = (x+1, y+1)$ (resp. $(x+1, y-1)$, resp. $(x+1, y)$). The height of the s_i step is y.

The Motzkin paths family are those associated to the Motzkin words in the following way. Let w be a Motzkin word. We associate to it a path ω beginning in $s_0 = (0,0)$ and such that the i^{th} letter y_i of w gives the i^{th} step of ω, that is

– if $y_i = a$, then (s_{i-1}, s_i) is an east step,
– if $y_i = x$, then (s_{i-1}, s_i) is a north-east step,
– if $y_i = \bar{x}$, then (s_{i-1}, s_i) is a south-east step.

Such a path always has steps with a non negative ordinate and its endpoint has a null ordinate. A *Dyck path* is a Motzkin path $w = (s_0, s_1,..., s_{2n})$, such that $s_0 = (0, 0)$, $s_{2n} = (2n, 0)$, having only North-East or South-East steps.

We call a valley (resp. a peak) every factor $\bar{x}\, a^* x$ (resp. $x\, a^* \bar{x}$) of the Motzkin or Dyck words. Classically, Dyck paths (and similarly Dyck words) are associated to parallelogram animals.

The usual bijection in Ref. 4 is described below.

A parallelogram animal A can be defined by the two sequences of integers $(h_1, h_2,..., h_k)$ and $(r_1, r_2, ..., r_{k-1})$, where h_i is the number of cells which belong to the i^{th} column and $(r_i + 1)$ is the number of cells which are adjacent to column i and i+1.

The Dyck word $\beta(A)$ is that associated to the Dyck path having n peaks (resp. n–1 troughs) whose heights are $h_1,...h_k$ (resp. $r_1-1, ..., r_{n-1}-1$). It is very easy to prove that β is a bijection preserving the four parameters.

– if $p(A) = 2n+2$, then $|\beta(A)| = 2n$,
– if $s(A) = b$, then $|\beta(A)| - |\beta(A)|_{\bar{x}\bar{x}x} - |\beta(A)|_{\bar{x}xx} = b$,
– if $a(A) = a$, then the sum of the heights of the peaks in $\beta(A)$ is a,
– if the width of A i k, then $\beta(A)$ has k factors $x\bar{x}$.

For the animal displayed in Figure 1, the associated Dyck words is

$$\beta(A) = xxxx\bar{x}\bar{x}x\bar{x}\bar{x}\bar{x}xx\bar{x}xx\bar{x}\bar{x}x\bar{x}\bar{x}.$$

The Dyck path is displayed in Figure 2. Another classical bijection associates the animal A to a two-coloured Motzkin word. The reader will find a description in Ref. 4.

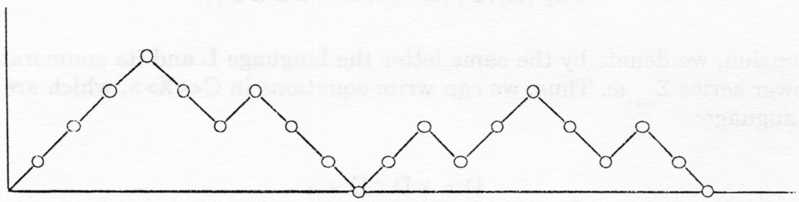

Figure 2. The Dyck word associated to the polyomino displayed in Figure 1.

A NEW CODING FOR PARALLELOGRAM ANIMALS

Let A be a parallelogram animal having k columns. We number them from left to right. Let us define now the two sequences:

$$H(P) = (h_1, h_2, ..., h_k)$$

so that h_i is the height of the i^{th} column and

$$R(P) = (r_0, r_1, r_2, ..., r_k)$$

where $r_0 = r_k = 0$ and for i in $[1, k-1]$, r_i is the number of the common border between the i^{th} and the $(i+1)^{\text{th}}$ columns.

Let us define the word w=Ψ(P) on the alphabet Y by

$$w = u_1 v_1 w_1 \ldots u_k v_k w_k$$

with for every i in [1, k],

- $u_i = x^{h_i - r_i - 1}$
- if $h_i = r_i$ or ($h_i = r_{i-1}$ and there exists $j < i$ such that $h_j > h_i$ and for 1 in $[j, i[$, $r_{i-1} = r_1$), then $v_i = a$ else $v_i = \varepsilon$.
- $w_i = \bar{x}^{h_i - r_i}$

The factor v_i is ε (see Figure 3), except for the two cases displayed in Figure 4, that is if adding the $(i+1)^{\text{th}}$ column adds only one corner. By construction, the word w is a prime Motzkin word so that if A has the perimeter $2p$+2, then w satisfies

$$|w|_x + |w|_a + v = p$$

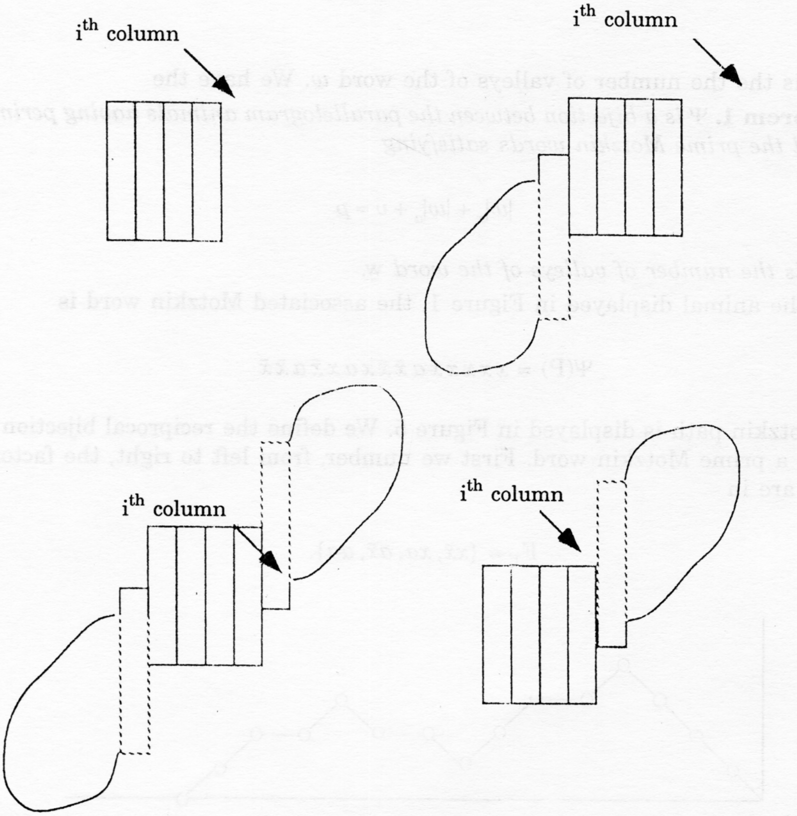

Figure 3. $v_i = \varepsilon$

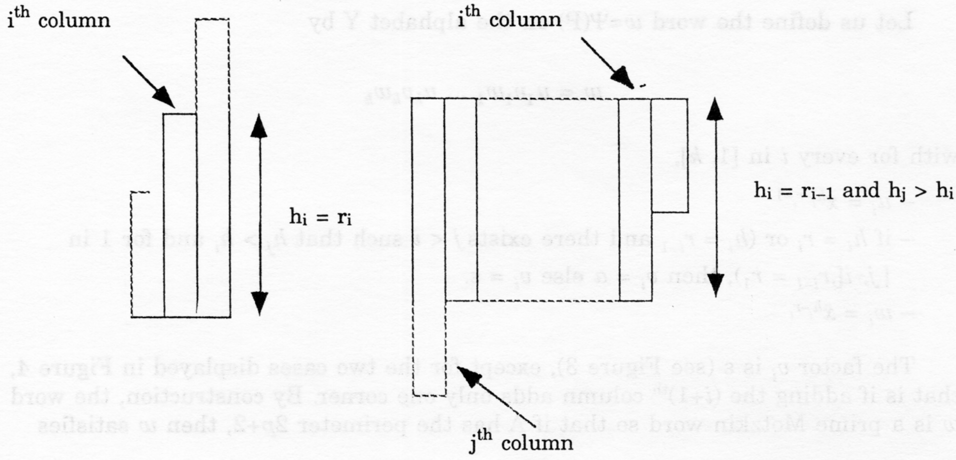

Figure 4. $v_i = a$.

where v is the the number of valleys of the word w. We have the

Theorem 1. *Ψ is a bijection between the parallelogram animals having perimeter* 2p+2 *and the prime Motzkin words satisfying*

$$|w|_x + |w|_a + v = p$$

where v *is the number of valleys of the word* w.

For the animal displayed in Figure 1, the associated Motzkin word is

$$\Psi(\mathrm{P}) = x\,x\,x\,x\,\bar{x}\,a\,\bar{x}\,\bar{x}\,x\,a\,x\,\bar{x}\,a\,\bar{x}\,\bar{x}$$

which Motzkin path is displayed in Figure 5. We define the reciprocal bijection Ψ^{-1}. Let w be a prime Motzkin word. First we number, from left to right, the factors of w which are in

$$\mathrm{F}_1 = \{x\bar{x}, xa, a\bar{x}, aa\}.$$

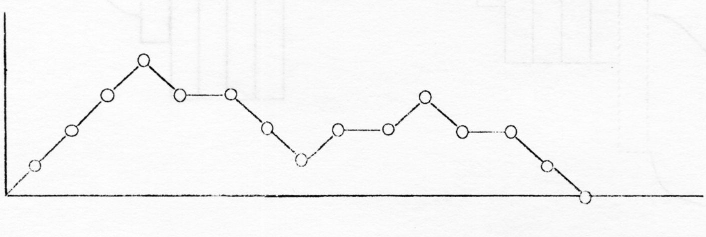

Figure 5. The Motzkin word associated to the polyomino displayed in Figure 1.

Let h_i be the height of the second letter in the i^{th} factor of category F_1 in the word w. Then, in the same manner. we also number the factors of w which are in

$$F_2 = \{\bar{x}x, a\},$$

and let r_i be the height of the last letter in the i^{th} factor of category F_2 in the word w. Then, the sequences

$$h(w) = (h_1, h_2, \dots, h_k) \text{ and } r(w) = (r_1, r_2, \dots, r_{k-1}).$$

define a parallelogram animal. In order to prove this fact, we must prove that for i in $[1, k-1]$,

$$1 \le r_i \le \mathrm{Min}\,(h_i, h_{i+1})$$

Of course, we have for i in $[1, k-1]$ $1 \le r_i$ because the Motzkin words are prime. Suppose that $r_i > h_i$, then the word w must have the form

$$w = w_1 f_1 w_2 f_2 w_3$$

with f_1 (resp. f_2) the i^{th} factor of w in F_1 (resp. F_2). Let us suppose that w_2 is non empty. Then w_2 must contain some letters x and, thus, f_1 cannot be the i^{th} factor since it must be in $w_2 f$. Thus, w_2 is empty and $r_i \le h_i$, The other inequality can be proved in the same manner. We can deduce from this theorem:

Corollary 2. *The number of Motzkin words having* m *letters* x, n *letters* a *and* v *valleys is* $C_{m+n+v+1}$.

LINK WITH THE CLASSICAL BIJECTION

In this section, we link our bijection to the old one in Ref. 4. Let t_1 (resp. t_2) be the morphism which substitutes for every factor $\bar{x}\,a^n x$ (resp. letter a) the sequence $\bar{x}(\bar{x}x)^n\bar{x}x^2$ (resp. $\bar{x}x$). We have

Theorem 3. $t_2 \circ t_1$ *is a bijection between prime Motzkin words having* m *letters* x, n *letters a and* s *valleys and Duck words of length* 2p *so that*

$$p = m + n + s.$$

Proof. First we write out a grammar which is close to the one generating prime Motzkin words. Let M' be the set of words w written on $\{x, \bar{x}, y, a\}$ so that

(i) the word obtained by substituting the letter x for y in w is a Motzkin word,
(ii) a letter y can occur only after a factor $\bar{x}a^n$, $n \ge 0$.

This construction means that we put a mark on the valleys. An unambiguous grammar G for this language is given by the following system of equations:

$$P = x M \bar{x} + \varepsilon,$$
$$M = x M \bar{x} Y + a M + \varepsilon,$$
$$Y = y M \bar{x} Y + a Y + \varepsilon.$$

Applying $t_2 \circ t_1$ to the prime Motzkin words comes up to substituting the factor a (resp. y) by $\bar{x}x$ (resp. $\bar{x}xx$) in G. We obtain the following grammar G'

$$P = x M \bar{x} + \varepsilon,$$
$$M = x M \bar{x} Y + \bar{x} x M + \varepsilon,$$
$$Y = \bar{x} x x M \bar{x} Y + \bar{x} x Y + \varepsilon.$$

We need to prove that this grammar is unambiguous and generates Dyck words. Usually, the first x of the word is paired with the first $\bar{x}$ of the same height. Here, we are pairing it with the last $\bar{x}$ of the same height. Every non empty Dyck word w can be written $xw'\bar{x}$ with w' satisfying

(i) $| w' |_x = | w' |_{\bar{x}}$,
(ii) if $w' = w'_1 w'_2$, then $\delta(w'_1) \geq -1$.

Then M is the set of these words w', Now, let us be in M, then

– if u begins with $\bar{x}$, then this $\bar{x}$ must be followed by letter x because of the property (ii) above and the remaining word is in M,

– if u begins with x, then $u = x\bar{x}$ or $u = xu_1\bar{x}u_2$, where $xu_1\bar{x}$ is a Dyck word and u_2 begins with $\bar{x}$.

This defines the last set of words Y, which are words beginning with $\bar{x}$ and satisfying the above properties (i) and (ii). Then, the third equation derivates straightforwardly.

Let w be a word generated by G such that $|w|$ is $2n$. By construction, we have

$$|w|_x + |w|_a + |w|_y = p$$

and thus the theorem is proved.

Moreover, the following property holds.

Property 6. $\beta = t_2 \circ t_1 \circ \Psi$

Proof. Let w be a Motzkin word and w' a Dyck word so that $w' = t_2 \circ t_1 (w)$. By this bijection, a factor xa (resp. $x\bar{x}$, ax, aa) of height h_i in w becomes factor $x\bar{x}x$ (resp. $x\bar{x}$, $\bar{x}x\bar{x}$, $\bar{x}x\bar{x}x$) of w' which ia a peak of height h_i. On the other hand, factor $\bar{x}x$ (resp. a) of height r_i in w becomes factor $\bar{x}\bar{x}xx$ (resp. $\bar{x}x$) in w' and corresponds to a through of height r_i–1. Thus, the heights of the F_1 and F_2 factors of a Motzkin word are transformed in the heights of peaks and trough of the Dyck word, which are the sequences defining an animal in the bijection β and the property is proved.

REFERENCES

1. M. P. Delest, *J. Math. Chem.* **8** (1991) 3.
2. M. P. Delest and J. M. Fedou, *Discrete Math.* (to appear).
3. M. P. Delest, D. Gouyou-Beauchamps, and B. Vauquelin, *Graphs Combinat.* **3** (1987) 325.
4. M. P. Delest and G. Viennot, *Theor. Comp. Sci.* **34** (1984) 169.
5. R. Donaghey, *J. Comb. Theor. Ser. B* **22** (1977) 114.

6. R. Donaghey and L. W. Shapiro, *J. Comb. Theor. Ser. A* **2** (1977) 291.
7. P. Flajolet, *INRIA*, Research Report, July 1991.
8. I. Gessel, *Trans. Amer. Math. Soc.* **257** (1980) 455.
9. A. J. Guttmann, *J. Phys. A: Math. Gen.* **15** (1982) 1987.
10. A. J. Guttmann, *Computer-aided Statistical Physics*, AIP Conference Proceedings 238, Taipei, Taiwan, 1991.
11. T. Motzkin, *Bull Amer. Math. Soc.* **54** (1948) 222.
12. G. Polya, *J. Comb. Theor.* **6** (1969) 102.
13. M. P. Schützenberger, *Inform. Control* **6** (1963) 246.
14. S. Feretić and D. Svrtan, *SFCA'93 Proceedings*, Florence 1993, 201.
15. X. G. Viennot, *Acts du colloque SFCA'92*, Montreal 1992, 399.

SAŽETAK

Novo kodiranje za paralelogramske životinje

M. Delest i J. P. Dubernard

Opisano je novo kodiranje paralelogramskih životinja i Motzkinovih riječi na osnovi kojega je nađena veza između Catalonovih brojeva, Motzkinovih riječi i paralelogramskih životinja.