BULLETIN OF THE ICA
Volume **104** (2025), 51–79

Flattened Catalan words

Jean-Luc Baril, Pamela E. Harris, and José L. Ramírez

Abstract. In this work, we define flattened Catalan words as Catalan words whose runs of weak ascents have leading terms that appear in weakly increasing order. We provide generating functions, formulas, and asymptotic expressions for the number of flattened Catalan words based on the number of runs of ascents (descents), runs of weak ascents (descents), ℓ-valleys, valleys, symmetric valleys, ℓ-peaks, peaks, and symmetric peaks.

1 Introduction

A word $w = w_1 w_2 \cdots w_n$ over the set of nonnegative integers is called a *Catalan word* if $w_1 = \mathtt{0}$ and $\mathtt{0} \leq w_i \leq w_{i-1}+\mathtt{1}$ for $i = 2, \ldots, n$. Throughout this paper, $|w|$ denotes the length of w, and ϵ denotes the *empty word*, which is the unique word of length zero. For $n \geq 0$, let $\mathcal{C}_n$ denote the set of Catalan words of length n. We set $\mathcal{C} \coloneqq \bigcup_{n\geq 0} \mathcal{C}_n$ and let $\mathcal{C}^+ \coloneqq \bigcup_{n\geq 1} \mathcal{C}_n$ be the set of nonempty Catalan words. For example,

$$\mathcal{C}_4 = \left\{ \begin{matrix} \mathtt{0000},\, \mathtt{0001},\, \mathtt{0010},\, \mathtt{0011},\, \mathtt{0012},\, \mathtt{0100},\, \mathtt{0101}, \\ \mathtt{0110},\, \mathtt{0111},\, \mathtt{0112},\, \mathtt{0120},\, \mathtt{0121},\, \mathtt{0122},\, \mathtt{0123} \end{matrix} \right\}.$$

Note that $|\mathcal{C}_n| = C_n = \frac{1}{n+1}\binom{2n}{n}$ is the n^{th} Catalan number. The exploration of Catalan words has begun with the comprehensive generation of Gray codes tailored for growth-constricted words [12]. Baril et al. [2, 4, 5] have delved into analyzing the distribution of descents and the last symbol in Catalan words avoiding one or two classical patterns of length at most three. Similar findings [1,7,17] emerge in studies of restricted Catalan words avoiding consecutive patterns of length three or pairs of relations. Callan et al. [10] initiate the enumeration of statistics, including area and perimeter, on the polyominoes associated with Catalan words. Furthermore, assorted combinatorial statistics regarding polyominoes associated with both

Key words and phrases: Catalan word; generating function; combinatorial statistic; Dyck path; flattened words
Mathematics Subject Classifications: 05A15, 05A19
Coresponding author: José L. Ramírez <jlramirezr@unal.edu.co>

Received: 10 May 2024
Accepted: 10 August 2024

Catalan and Motzkin terminologies have been scrutinized [6, 13–15]. Next Shattuck [18] initiated an examination into the frequency of distinct subword occurrences, spanning no more than three characters, nestled within Catalan words, like descents, ascents, and levels. In a recent paper [3], Baril et al. provide generating functions, formulas, and asymptotic expressions for the number of Catalan words based on the number of runs of ascents (descents), runs of weak ascents (descents), ℓ-valleys, valleys, symmetric valleys, ℓ-peaks, peaks, and symmetric peaks.

Given a permutation of $[n] = \{1, \ldots, n\}$ in one-line notation $\pi = \pi_1\pi_2\cdots\pi_n$, the *runs* of π are the maximal contiguous increasing subwords of π. If the sequence of leading terms of the runs of π appears in increasing order, then π is called a *flattened partition* of length n. Nabawanda et al. give a recursive formula for the number of flattened partitions of length n with k runs [16, Theorem 1]. Callan gives the number of flattened partitions of length n avoiding a single 3-letter pattern [9]. Elder et al. extended the work of Nabawanda et al. to establish recursive formulas for the number of flattened parking functions built from permutations of $[n]$, with r additional ones inserted that have k runs [11, Theorems 29, 30 and 35]. A further generalization includes the work of Buck et al. [8] who establish that flattened Stirling permutations are enumerated by the Dowling numbers, which correspond to the OEIS entry [19, A007405].

In this work, we define *flattened Catalan words*, which are Catalan words whose maximal contiguous nondecreasing subwords have leading terms in weakly increasing order. For example, the Catalan word

$$\mathtt{0012301222345523343} \in \mathcal{C}_{19}$$

is a flattened Catalan word with four maximal contiguous nondecreasing subwords $\mathtt{00123}$, $\mathtt{012223455}$, $\mathtt{2334}$, and $\mathtt{3}$, whose leading terms satisfy $\mathtt{0} \leq \mathtt{0} \leq \mathtt{2} \leq \mathtt{3}$. Conversely, $\mathtt{012321} \in \mathcal{C}_6$ is not a flattened Catalan word as it has maximal contiguous nondecreasing subwords $\mathtt{0123}$, $\mathtt{2}$, and $\mathtt{1}$, and the leading terms $\mathtt{0}$, $\mathtt{2}$, and $\mathtt{1}$ are not in weakly increasing order. We denote the sets of nonempty flattened Catalan words and flattened Catalan words of length n as $\mathsf{Flat}(\mathcal{C}^+)$ and $\mathsf{Flat}(\mathcal{C}_n)$, respectively.

Let $w = w_1w_2\cdots w_n \in \mathsf{Flat}(\mathcal{C}_n)$. As usual, we say that w has an *ascent* (*descent*) at position ℓ if $w_\ell < w_{\ell+1}$ ($w_\ell > w_{\ell+1}$), where $\ell \in [n-1]$. Similarly, we define *weak ascent* (resp., *weak descent*) at position ℓ if $w_\ell \leq w_{\ell+1}$ ($w_\ell \geq w_{\ell+1}$), where $\ell \in [n-1]$. A *run* (resp., *weak run*) of ascents (resp., *weak ascents*) in a word w is a maximal subword of consecutive ascents (resp., weak ascents). The number of runs in w is denoted by $\overline{\mathsf{runs}}(w)$, and

the number of weak runs in w is denoted by $\overline{\mathsf{wruns}}(w)$. The runs of descents and weak descents are defined similarly, and the statistics will be denoted $\underline{\mathsf{runs}}(w)$ and $\underline{\mathsf{wruns}}(w)$, respectively. An *$\ell$-valley* in a flattened Catalan word w is a subword of the form $ab^{\ell}(b+1)$, where $a > b$ and ℓ is a positive integer and b^{ℓ} denotes ℓ consecutive copies of the letter b. If $\ell = 1$, we say that it is a *short valley*. The number of ℓ-valleys of w is denoted by ℓ-$\mathsf{val}(w)$ and the number of all ℓ-valleys for $\ell \geq 1$ of w is denoted by $\mathsf{val}(w)$. A *symmetric valley* is a valley of the form $a(a-1)^{\ell}a$ with $\ell \geq 1$. The number of symmetric valleys of w is denoted by $\mathsf{symv}(w)$. Analogously, we define the peak statistic. Namely, an *ℓ-peak* in w is a subword of the form $a(a+1)^{\ell}b$, where $a \geq b$ and ℓ is a positive integer. The number of ℓ-peaks of w is denoted by ℓ-$\mathsf{peak}(w)$ and the sum of all ℓ-peaks for $\ell \geq 1$ of w is denoted by $\mathsf{peak}(w)$. If $\ell = 1$, we say that it is a *short peak*; and if $a = b$, it is called a *symmetric peak*. The number of symmetric peaks of w is denoted by $\mathsf{symp}(w)$.

Our contributions include generating functions and combinatorial expressions for the number of flattened Catalan words based on the number of runs of ascents (descents), runs of weak ascents (descents), ℓ-valleys, valleys, symmetric valleys, ℓ-peaks, peaks, and symmetric peaks. We also establish one-to-one correspondences between the following:

- flattened Catalan words of length n with k runs of ascents and k-part order-consecutive partitions of n, which have been studied in [21], see Theorem 3.5;
- flattened Catalan words of length n and compositions of all even natural numbers into $n-1$ parts of at most two where the part 0 is allowed, see Theorem 3.4;
- flattened Catalan words of length n with k runs of weak ascents and binary words of length $n-1$ where $2k-2$ symbols are replaced with a dot $\bullet$, see Theorem 3.11;
- flattened Catalan words of length n with k short valleys and Dyck paths of semilength n with k occurrences of `DDUU`, where the height sequence of occurrences `DDU` (from left to right) is nondecreasing, see Remark 4.3;
- flattened Catalan words of length n and ordered trees with n edges and with $k+1$ nodes having only leaves as children and satisfying two additional conditions, see Remark 4.6.

We aggregate our results and the notation used throughout in Table 1.1.

Table 1.1: Summary of notation and results for statistics considered.

	Statistics					
	runs of asc.	runs of w. asc.	runs of desc.	runs of w. desc.	ℓ-valleys	short valleys
Statistic on w	$\overline{\mathsf{runs}}(w)$	$\overline{\mathsf{wruns}}(w)$	$\underline{\mathsf{runs}}(w)$	$\underline{\mathsf{wruns}}(w)$	ℓ-$\mathsf{val}(w)$	1-$\mathsf{val}(w)$
Bivariate g. function	$R(x,y)$	$W(x,y)$	$\bar{R}(x,y)$	$\bar{W}(x,y)$	$V_\ell(x,y)$	$V_1(x,y)$
Distribution	$r(n,k)$	$w(n,k)$	$\bar{r}(n,k)$	$\bar{w}(n,k)$	$v_\ell(n,k)$	$v_1(n,k)$
Total occurrences over $\mathsf{Flat}(\mathcal{C}_n)$	$r(n)$	$w(n)$	$\bar{r}(n)$	$\bar{w}(n)$	$v_\ell(n)$	$v_1(n)$
	valleys	sym. valleys	ℓ-peaks	short peaks	peaks	sym. peaks
Statistic on w	$\mathsf{val}(w)$	$\mathsf{symv}(w)$	ℓ-$\mathsf{peak}(w)$	1-$\mathsf{peak}(w)$	$\mathsf{peak}(w)$	$\mathsf{symp}(w)$
Bivariate g. function	$V(x,y)$	$S(x,y)$	$P_\ell(x,y)$	$P_1(x,y)$	$P(x,y)$	$T(x,y)$
Distribution	$v(n,k)$	$s(n,k)$	$p_\ell(n,k)$	$p_1(n,k)$	$p(n,k)$	$t(n,k)$
Total occurrences over $\mathsf{Flat}(\mathcal{C}_n)$	$v(n)$	$s(n)$	$p_\ell(n)$	$p_1(n)$	$p(n)$	$t(n)$

Statistic	Bivariate g. f.	Total occurrences over $\mathsf{Flat}(\mathcal{C}_n)$	OEIS
$\overline{\mathsf{runs}}$	$\frac{xy(1-x-xy)}{1-2x+x^2-2xy+x^2y+x^2y^2}$	$\frac{1}{4}\big(3^{n-1}+1\big)(n+1)$	
$\overline{\mathsf{wruns}}$	$\frac{(1-2x)xy}{1-4x+4x^2-x^2y}$	$\frac{1}{36}\big(27-9n+(5+n)3^n\big)$	
$\underline{\mathsf{runs}}$	$\frac{xy(1-2xy)}{1-4xy-x^2y+4x^2y^2}$	$\frac{1}{36}\big(27n-9+(5n+1)3^n\big)$	
$\underline{\mathsf{wruns}}$	$\frac{yx(1-xy-x)}{x^2y^2+x^2y+x^2-2\,xy-2\,x+1}$	$\frac{n+1}{4}\big(1+3^{n-1}\big)$	
ℓ-val	$\frac{x(1-2x+x^{\ell+1}-x^{\ell+1}y)}{(1-x)(1-3x+x^{\ell+1}-x^{\ell+1}y)}$	$\frac{1}{4}\big(1-3^{n-2-\ell}+2\cdot 3^{n-2\ell}(n-2-\ell)\big)$	
val	$\frac{x-3x^2+x^3(3-y)}{(1-x)(1-4x+4x^2-x^2y)}$	$\frac{1}{36}\big(3^n(n-4)+9n\big)$	A212337
symv	$\frac{x(1-2x)(1-2x+2x^2-x^2y)}{(1-x)(1-5x+8x^2-5x^3-x^2y+2x^3y)}$	$\frac{1}{144}\big(3^n(2n-5)-18n^2+54n-27\big)$	
ℓ-peak	$\frac{x(1-2x)}{(1-x)(1-3x+x^{\ell+1}(1-y))}$	$\frac{1}{4}\big(\big(3^{n-\ell-2}(2n+1-2\ell)\big)-1\big)$	
peak	$\frac{x(1-2x)}{1-4x+4x^2-x^2y}$	$\frac{1}{4}\big(3^{n-2}-1\big)(n-1)$	A261064
symp	$\frac{x(1-x)(1-2x)}{1-5x+8x^2-5x^3-x^2y+2x^3y}$	$\frac{1}{144}\big(63+3^n+2(-45+3^n)n+18n^2\big)\big)$	

2 Basic definitions

Throughout the article, we will use the following decomposition of Catalan words, called *first return decomposition* of a Catalan word w, which is $w = \mathtt{0}(w'+1)w''$, where w' and w'' are Catalan words (w' and w'' could be empty), and where $(w'+1)$ is the word obtained from w' by adding 1 at all these symbols. Note that whenever w' is the empty word, denoted by ϵ, then $(w'+1)$ remains the empty word.

For example, the first return decomposition of

$$w = \mathtt{0122200122322334544} \in \mathsf{Flat}(\mathcal{C}_{19})$$

is given by setting $w' = \mathtt{0111}$ and $w'' = \mathtt{00122322334544}$. For this word w, we have $\overline{\mathsf{runs}}(w)=11$, $\overline{\mathsf{wruns}}(w)=4$, $\underline{\mathsf{runs}}(w)=16$, $\underline{\mathsf{wruns}}(w)=9$, 1-$\mathsf{val}(w)=0$, 2-$\mathsf{val}(w)=2$, ℓ-$\mathsf{val}(w)=0$ $(\ell>2)$, $\mathsf{symv}(w)=1$, 1-$\mathsf{peak}(w)=2$, 2-$\mathsf{peak}(w)=0$, 3-$\mathsf{peak}(w)=1$, ℓ-$\mathsf{peak}(w)=0$ $(\ell>3)$, and $\mathsf{symp}(w)=2$.

Drawing Catalan words as lattice diagrams on the plane proves to be a convenient representation. These diagrams are constructed using unit up steps $(0, 1)$, down steps $(0, -1)$, and horizontal steps $(1, 0)$. Each symbol w_i of a Catalan word is represented by the horizontal segment between the points $(i - 1, w_i)$ and (i, w_i), and the vertical steps are inserted to obtain a connected diagram. For example, in Figure 2.1, we illustrate the lattice diagram associated to the Catalan word w.

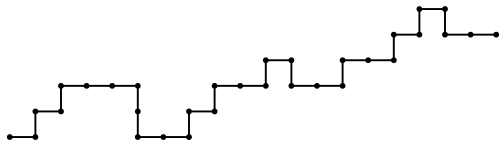

Figure 2.1: Lattice diagram of the word $w = 0122200122322334544$.

Remark 2.1. Let $\mathcal{C}_n^{\uparrow}$ denote the set of weakly increasing Catalan words of length n. Notice that $\left|\mathcal{C}_0^{\uparrow}\right| = 1$ and, for $n \geq 1$, $\left|\mathcal{C}_n^{\uparrow}\right| = 2^{n-1}$; then its generating functions is $1 + x/(1 - 2x)$ if we include the empty word. Note that the set of nonempty weakly increasing Catalan words is precisely the set of flattened Catalan words with a single weak run. Hence, the generating function for the latter set is $x/(1 - 2x)$.

3 The distribution of runs

3.1 Runs of ascents

In order to count nonempty flattened Catalan words according to the length and the number of runs of ascents, we introduce the following bivariate generating function:

$$R(x, y) = \sum_{w \in \mathsf{Flat}(\mathcal{C}^+)} x^{|w|} y^{\overline{\mathsf{runs}}(w)} = \sum_{n \geq 1} x^{|w|} \sum_{w \in \mathsf{Flat}(\mathcal{C}_n)} y^{\overline{\mathsf{runs}}(w)}$$

where the coefficient of $x^n y^k$ is the number of flattened Catalan words of length n with k runs of ascents.

In Theorem 3.2, we give an expression for this generating function, but first we provide an example.

Example 3.1. Consider the flattened Catalan word

$$w = 012230123122 \in \mathsf{Flat}(\mathcal{C}_{12}).$$

Then w has 5 runs of ascents: 012, 23, 0123, 12, and 2.

Theorem 3.2. *The generating function for nonempty flattened Catalan words with respect to the length and the number of runs of ascents is*

$$R(x,y) = \frac{xy(1-x-xy)}{1-2x+x^2-2xy+x^2y+x^2y^2}.$$

Proof. Let w be a nonempty flattened Catalan word and let $w = 0(w'+1)w''$ be the first return decomposition, with $w', w'' \in \mathsf{Flat}(\mathcal{C})$. There are four different types of this word. Figure 3.1 illustrates this case.

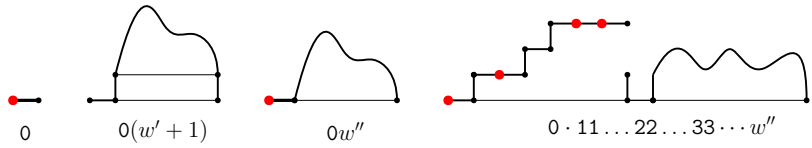

Figure 3.1: Decomposition of a nonempty flattened Catalan word in $\mathsf{Flat}(\mathcal{C})$.

If $w' = w'' = \epsilon$, then $w = 0$. Then its generating function is xy.

If $w'' = \epsilon$ and $w' \neq \epsilon$, then $w = 0(w'+1)$. Then the generating function is $xR(x,y)$.

If $w' = \epsilon$ and $w'' \neq \epsilon$, then $w = 0w''$. Then the generating function is $xyR(x,y)$ because we have an extra run.

If $w' \neq \epsilon$ and $w'' \neq \epsilon$, then $w = 0(w'+1)w''$. Note w' is a weakly increasing word because $w \in \mathsf{Flat}(\mathcal{C}^+)$. Then the bivariate generating function for such words w' is

$$\sum_{n\geq 1}\sum_{k=1}^{n}\binom{n-1}{k-1}x^n y^k = \sum_{n\geq 0} y(1+y)^{n-1}x^n = \frac{xy}{1-x(1+y)}.$$

Therefore, the generating function for this case is given by

$$\frac{x^2y}{1-x-xy}R(x,y).$$

Therefore, we have the functional equation

$$R(x,y) = xy + x(1+y)R(x,y) + \frac{x^2y}{1-x-xy}R(x,y).$$

Solving this equation, we obtain the desired result. □

Corollary 3.3. *The generating function for nonempty flattened Catalan words is given by*

$$R(x,1) = \sum_{n\geq 1} f(n)x^n = \frac{x(1-2x)}{(1-3x)(1-x)}.$$

Therefore,

$$f(n) = \frac{1}{2}\big(3^{n-1}+1\big).$$

The first few values of the sequence $f(n)$ $(n \geq 1)$ correspond to the OEIS entry [19, A007051]:

$$1,\quad 2,\quad 5,\quad 14,\quad 41,\quad 122,\quad 365,\quad 1094,\quad 3281,\quad 9842,\ldots.$$

This sequence also counts the compositions of all even natural numbers (from 0 to $2(n-1)$) into $n-1$ parts of at most two (the part 0 is allowed).

Theorem 3.4. *Flattened Catalan words of length n and compositions of all even natural numbers (from 0 to $2(n-1)$) into $n-1$ parts of at most two (the part 0 is allowed) are in bijection.*

Proof. A bijection ψ between flattened Catalan words of length n and this combinatorial class is given by $\psi(\mathtt{0}) = \epsilon$; $\psi\big(\mathtt{0}(w+1)\big) = 2\psi(w)$; $\psi(\mathtt{0}w) = \mathtt{0}\psi(w)$; and $\psi(\mathtt{0}(w+1)w') = 1\psi(w)1\psi(w')$. □

Let $r(n,k)$ denote the number of flattened Catalan words of length n with exactly k runs of ascents, that is $r(n,k) = [x^n y^k]R(x,y)$, which denotes the coefficient of $x^n y^k$ in $R(x,y)$. The first few rows of this array are

$$\mathcal{R} \coloneqq [r(n,k)]_{n,k\geq 1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 6 & \boxed{\mathbf{6}} & 1 & 0 & 0 & 0 & 0 \\ 1 & 10 & 19 & 10 & 1 & 0 & 0 & 0 \\ 1 & 15 & 45 & 45 & 15 & 1 & 0 & 0 \\ 1 & 21 & 90 & 141 & 90 & 21 & 1 & 0 \\ 1 & 28 & 161 & 357 & 357 & 161 & 28 & 1 \end{pmatrix}.$$

For example, $r(4,3) = 6$, the entry boxed in $\mathcal{R}$ above, and the corresponding flattened Catalan words (and lattice diagrams) are shown in Figure 3.2.

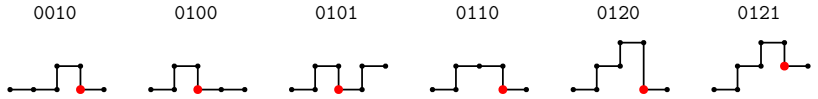

Figure 3.2: Flattened Catalan words of length 4 with 2 runs of ascents. The red marked vertex denotes the start of the second run of ascents.

The array $\mathcal{R}$ corresponds to the OEIS entry [19, A056241]. Notice that this sequence has a different combinatorial interpretation. It counts the number of k-part order-consecutive partitions of n. An *order-consecutive partition* of $\{1, 2, \dots, n\}$ with k parts is a k-uplet $(S_1, S_2, \dots, S_k)$ of subsets such that $S_i \cap S_j = \varnothing$ if $i \neq j$, $\bigcup_{i=1}^{k} S_i = \{1, 2, \dots, n\}$, where every subset S_i are in increasing order relatively to their maximum elements, and satisfying the property: for $j = 1, \dots, k$, $\bigcup_{i=1}^{j} S_i$ is an interval (cf. [21]).

Theorem 3.5. *Flattened Catalan words of length n with exactly k runs of ascents are in bijection with k-part order-consecutive partitions of n.*

Proof. We define recursively a map ψ from the set of words in $\mathsf{Flat}(\mathcal{C}_n)$ and the set $\mathcal{OCP}_n$ of order-consecutive partitions of $\{1, 2, \dots, n\}$. We consider the four cases of Figure 3.1.

- If w belongs to the case (i), then $w = \mathtt{0}$ and we set $\psi(w) = \{1\}$;
- If w belongs to the case (ii), then $w = \mathtt{0}(w'+1)$ and $\psi(w)$ is obtained from $\psi(w')$ by inserting n in the last part; for instance, if $f(w') = \{2, 3\}\{1, 4\}$, then $f(w) = \{2, 3\}\{1, 4, 5\}$;
- If w belongs to the case (iii), then $w = \mathtt{0}w'$ and $\psi(w)$ is obtained from $\psi(w')$ by adding the part $\{n\}$ on the right; for instance, if $f(w') = \{2, 3\}\{1, 4\}$, then $f(w) = \{2, 3\}\{1, 4\}\{5\}$;
- If w belongs to the case (iv), then $w = w'w''$ where w' consists of one weak run starting with $\mathtt{01}$. Using the previous cases, $\psi(w') = S_1 \dots S_k$ where $S_k = \{a_1, \dots a_\ell, |w'|-1, |w'|\}$ ends with a part containing both $|w'| - 1$ and $|w'|$. So, we set $\psi(w) = S_1 \dots S_{k-1}(\psi(w'') + |w'| - 1)\{a_1, \dots, a_\ell, |w'| - 1, |w'| + |w''|\}$. For instance, if $w = \mathtt{0112\ 0120}$, then $w' = \mathtt{0112}$, $w'' = \mathtt{0120}$, $f(w') = \{1, 2\}\{3, 4\}$, and $f(w'') = \{3\}\{1, 2, 4\}$; and hence, $f(w) = \{1, 2\}\{6\}\{4, 5, 7\}\{3, 8\}$. □

Theorem 3.5 and [21, Theorem 6] imply the following combinatorial expression.

Corollary 3.6. *If $n, k \geq 1$, then*

$$r(n,k) = \sum_{j=0}^{k-1} \binom{n-1}{2k-j-2} \binom{2k-j-2}{j}.$$

Let $r(n)$ be the total number of runs of ascents over all flattened Catalan words of length n.

Corollary 3.7. *We have*

$$\sum_{n\geq 0} r(n)x^n = \frac{x - 5x^2 + 8x^3 - 3x^4}{(1-3x)^2(1-x)^2}.$$

Moreover, for $n \geq 1$, we have

$$r(n) = \frac{1}{4}(3^{n-1}+1)(n+1).$$

The first few values of the sequence $r(n)$ $(n \geq 1)$ are

$$1, \quad 3, \quad 10, \quad 35, \quad 123, \quad 427, \quad 1460, \quad 4923, \quad 16405, \quad 54131, \ldots.$$

3.2 Runs of weak ascents

In order to count nonempty flattened Catalan words according to the length and the number of runs of weak ascents, we introduce the following bivariate generating function:

$$W(x,y) = \sum_{w \in \mathsf{Flat}(\mathcal{C}^+)} x^{|w|} y^{\overline{\mathsf{wruns}}(w)} = \sum_{n\geq 1} x^{|w|} \sum_{w \in \mathsf{Flat}(\mathcal{C}_n)} y^{\overline{\mathsf{wruns}}(w)}$$

where the coefficient of $x^n y^k$ is the number of flattened Catalan words of length n with k runs of weak ascents.

Example 3.8. Consider the flattened Catalan word $w = \mathtt{012230123122} \in \mathsf{Flat}(\mathcal{C}_{12})$. Then w has 3 runs of weak ascents: $\mathtt{01223}$, $\mathtt{0123}$, $\mathtt{122}$.

In Theorem 3.9, we give an expression for this generating function.

Theorem 3.9. *The generating function for the number of nonempty flattened Catalan words with respect to the length and the number of runs of weak ascents is*

$$W(x,y) = \frac{(1-2x)xy}{1-4x+4x^2-x^2y}.$$

Proof. Let w be a nonempty flattened Catalan word and let $w = \mathtt{0}(w'+1)w''$ be the first return decomposition, with $w', w'' \in \mathsf{Flat}(\mathcal{C})$. There are four different types of this word. If $w' = w'' = \epsilon$, then $w = \mathtt{0}$. Then its generating function is xy. If $w'' = \epsilon$ and $w' \neq \epsilon$, then $w = \mathtt{0}(w'+1)$. Then the generating function is $xW(x,y)$. Similarly, if $w' = \epsilon$ and $w'' \neq \epsilon$, then $w = \mathtt{0}w''$. Then the generating function is $xW(x,y)$. If $w' \neq \epsilon$ and $w'' \neq \epsilon$, then $w = \mathtt{0}(w'+1)w''$. Note w' is a weakly increasing word because $w \in \mathsf{Flat}(\mathcal{C}^+)$. Then the generating function is given by

$$x \sum_{k \geq 1} 2^k x^k y W(x,y) = \frac{x^2 y}{1-2x} W(x,y).$$

Therefore, we have the functional equation

$$W(x,y) = xy + 2xW(x,y) + \frac{x^2 y}{1-2x} W(x,y).$$

Solving this equation, we obtain the desired result. □

Let $w(n,k)$ denote the number of flattened Catalan words of length n with exactly k runs of weak ascents, that is $w(n,k) = [x^n y^k]W(x,y)$, which denotes the coefficient of $x^n y^k$ in $W(x,y)$. The first few values of this array are

$$\mathcal{W} := [w(n,k)]_{n,k\geq 1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 & 0 \\ 8 & \boxed{\mathbf{6}} & 0 & 0 & 0 \\ 16 & 24 & 1 & 0 & 0 \\ 32 & 80 & 10 & 0 & 0 \\ 64 & 240 & 60 & 1 & 0 \\ 128 & 672 & 280 & 14 & 0 \\ 256 & 1792 & 1120 & 112 & 1 \end{pmatrix}.$$

For example, $w(4,2) = 6$, the entry boxed in $\mathcal{W}$ above, and the corresponding flattened Catalan words (and lattice diagrams) are shown in Figure 3.3.

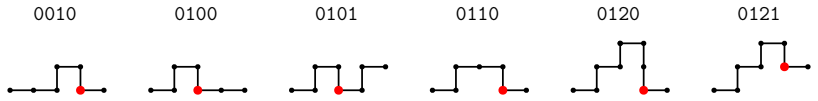

Figure 3.3: Flattened Catalan words of length 4 with 2 runs of weak ascents. The red marked vertex denotes the start of the second run of weak ascents.

Corollary 3.10. *For $n, k \geq 1$, we have*

$$w(n,k) = 2^{n-2k+1}\binom{n-1}{2k-2}.$$

Proof. From Theorem 3.9, we obtain the recurrence relation

$$w(n,k) - 4w(n-1,k) + 4w(n-2,k) - 4w(n-2,k-1) = 0, \quad n \geq 3,\ k \geq 1,$$

with the initial values $w(2,1) = 2$, $w(1,1) = 1$, and $w(n,k)$ for $n < k$. It is not difficult to verify that $2^{n-2k+1}\binom{n-1}{2k-2}$ satisfies the same recurrence relation and the same initial values. Therefore, the sequences are the same. □

We give an alternate proof of Corollary 3.10 through a bijective proof. We state the result formally for ease of reference.

Theorem 3.11. *Flattened Catalan words of length n with k runs of weak ascents and binary words of length $n-1$ where $2k-2$ symbols are replaced with a dot $\bullet$ are in bijection.*

Proof. We now give bijection between flattened Catalan words of length n with k runs of weak ascents and binary words of length $n-1$ where $2k-2$ symbols are replaced with a dot $\bullet$ (Corollary 3.10 and a simple combinatorial argument prove that the two classes of objects have the same cardinality). Let $u = u_1u_2\cdots u_{n-1}$ be such a binary word with $2k-2$ dots ($\bullet$), and let us suppose that the dots ($\bullet$) are on the positions $\{i_1, i_2, \ldots, i_{2k-2}\}$. Then, we define the flattened Catalan words with k runs of weak ascents as follows:

Let $v = v_0v_1\cdots v_{n-1}$ be the word of length n constructed from u by fixing $v_0 \coloneqq \mathtt{0}$; $v_{i_{2a+1}} \coloneqq \mathtt{1}$ and $v_{i_{2a}} \coloneqq \mathtt{0}$ for $a = 0, 1, \ldots, k-1$; and $v_i \coloneqq u_i$ for all other positions i. We fix $i_0 = 0$ and $i_{2k-1} = n$. Now, v consists of the juxtaposition of k nonempty factors of the form $r_a = v_{i_{2a}}\cdots v_{i_{2a+2}-1}$, for $a = 0, 1, \ldots, k-1$, all of them starting with $\mathtt{0}$. We associate to each factor $s = 0s_2\cdots s_p$ the nondecreasing Catalan word $c(s) = \mathtt{0}c_2\cdots c_{|s|}$, where $c_i = c_{i-1}$ if $s_i = 0$ and $c_i = c_{i-1} + \mathtt{1}$, otherwise (for instance, if $s = 011010110$ then $c(s) = \mathtt{012233455}$).

The bijection f is defined as follows:

$$f(u) = c(r_0)\big(a_0 + c(r_1)\big)\big(a_0 + a_1 + c(r_2)\big)\cdots\big(a_0 + a_1 + \cdots + a_{k-2} + c(r_{k-1})\big),$$

where a_j is the number of 1's in the factor $v_{i_{2(j+1)}} \cdots v_{i_{2(j+1)+1}-1}$.

For instance if $n = 29$, $k = 4$, and $u = 10100 \bullet 1010 \bullet 0110 \bullet 01 \bullet 0110 \bullet 0 \bullet 00$, we have

$$v = 01010011010\ \mathbf{0}0110101\ \mathbf{0}011010\ \mathbf{0}000$$

and

$$f(u) = 01122234455\ 22344556\ 4456677\ 666. \qquad \square$$

Let $w(n)$ be the total number of runs of weak ascents over all flattened Catalan words of length n.

Corollary 3.12. *For $n \geq 1$, we have*

$$\sum_{n\geq 1} w(n)x^n = \frac{x(1-2x)^3}{(1-4x+3x^2)^2}.$$

Moreover, for $n \geq 1$, we have

$$w(n) = \frac{1}{36}\big(27 - 9n + (5+n)3^n\big).$$

The first few values of the sequence $w(n)$ $(n \geq 1)$ are

$$1,\quad 2,\quad 6,\quad 20,\quad 67,\quad 222,\quad 728,\quad 2368,\quad 7653,\quad 24602,\ldots$$

3.3 Runs of descents

In order to count nonempty flattened Catalan words according to the length and the number of runs of descents, we introduce the following bivariate generating function:

$$\bar{R}(x,y) = \sum_{w\in \mathsf{Flat}(\mathcal{C}^+)} x^{|w|} y^{\underline{\mathsf{runs}}(w)} = \sum_{n\geq 1} x^{|w|} \sum_{w\in \mathsf{Flat}(\mathcal{C}_n)} y^{\underline{\mathsf{runs}}(w)}$$

where the coefficient of $x^n y^k$ is the number of flattened Catalan words of length n with k runs of descents.

Example 3.13. Consider the flattened Catalan word $w = 012230123122 \in \mathsf{Flat}(\mathcal{C}_{12})$. Then w has 10 runs of descents: 0, 1, 2, 2, 30, 1, 2, 31, 2, and 2.

It is worth noticing that in any flattened Catalan word w of length n, we have $\underline{\mathsf{runs}}(w) = n + 1 - \overline{\mathsf{wruns}}(w)$. Therefore, we can directly deduce Theorem 3.14 and Corollary 3.15.

Theorem 3.14. *The generating function for the number of nonempty flattened Catalan words with respect to the length and the number of runs of descents is*

$$\bar{R}(x,y) = yW\left(xy, \frac{1}{y}\right) = \frac{xy(1-2xy)}{1-4xy-x^2y+4x^2y^2}.$$

Let $\bar{r}(n,k)$ denote the number of flattened Catalan words of length n with exactly k runs of descents, that is $\bar{r}(n,k) = [x^n y^k]\bar{R}(x,y)$, which denotes the coefficient of $x^n y^k$ in $\bar{R}(x,y)$. The first few values of this array are

$$\bar{\mathcal{R}} := [\bar{r}(n,k)]_{n,k\geq 1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{\mathbf{6}} & 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 24 & 16 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10 & 80 & 32 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 60 & 240 & 64 & 0 & 0 \\ 0 & 0 & 0 & 0 & 14 & 280 & 672 & 128 & 0 \end{pmatrix}.$$

For example, $\bar{r}(4,3) = 6$, the entry boxed in $\bar{\mathcal{R}}$ above, and the corresponding flattened Catalan words (and lattice diagrams) are shown in Figure 3.4.

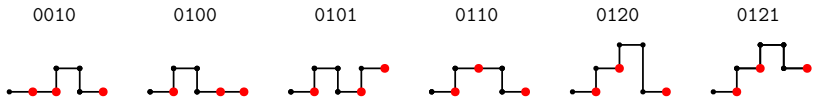

Figure 3.4: Flattened Catalan words of length 4 with 3 runs of descents. The red marked vertices denote the end of a run of descents.

Corollary 3.15. *For $n, k \geq 1$, we have*

$$\bar{r}(n,k) = 2^{2k-n-1}\binom{n-1}{2(n-k)}.$$

A combinatorial interpretation of this last formula can be obtained from the bijection f (see Section 3.2) between flattened Catalan words of length n with $n+1-k$ runs of weak ascents (or equivalently with k descents) and binary words of length $n-1$ with $(2n-2k)$ dots $\bullet$.

Let $\bar{r}(n)$ be the total number of runs of descents over all flattened Catalan words of length n.

Corollary 3.16. *We have*

$$\sum_{n\geq 0} \bar{r}(n)x^n = \frac{x(1-4x+4x^2+2x^3)}{(1-4x+3x^2)^2}.$$

Moreover, for $n \geq 1$, we have

$$\bar{r}(n) = \frac{1}{36}\big(27n - 9 + (5n+1)3^n\big).$$

The first few values of the sequence $\bar{r}(n)$ $(n \geq 1)$ are

1, 4, 14, 50, 179, 632, 2192, 7478, 25157, 83660, ...

3.4 Runs of weak descents

In a flattened Catalan word of length n, the number of runs of ascents plus the number of runs of weak descents equals $n+1$. Hence, the number $\bar{w}(n,k)$ of flattened Catalan words of length n with k runs of weak descents equals the number $r(n,k)$ of flattened Catalan words of length n with k runs of ascents. Moreover, we can defined a simple involution ϕ on $\mathsf{Flat}(\mathcal{C}_n)$ such that $\phi(w) = w'$ with $\underline{\mathsf{wruns}}\big(\phi(w)\big) = \overline{\mathsf{runs}}(w)$, as follows: $\phi(\epsilon) = \epsilon$, $\phi\big(\mathtt{0}(w+1)\big) = \mathtt{0}\phi(w)$, $\phi(\mathtt{0}w) = \mathtt{0}\big(1+\phi(w)\big)$, and $\phi\big(\mathtt{0}(1+w)w'\big) = \mathtt{0}\big(1+\phi(w)\big)\phi(w')$ whenever $w, w' \neq \epsilon$. Then, we the results can be restated as those in Section 3.1.

Theorem 3.17. *The generating function for the number of nonempty flattened Catalan words with respect to the length and the number of runs of weak descents is*

$$\bar{W}(x,y) = R(x,y) = \frac{yx(1-xy-x)}{x^2y^2+x^2y+x^2-2\,xy-2\,x+1}.$$

Therefore,

$$\bar{w}(n,k) = r(n,k) = \sum_{j=0}^{k-1} \binom{n-1}{2k-j-2}\binom{2k-j-2}{j}.$$

Corollary 3.18. *We have*

$$\sum_{n\geq 0} \bar{w}(n)x^n = \sum_{n\geq 0} r(n)x^n = \frac{x(1-3x^3+8x^2-5x)}{(3x^2-4x+1)^2}.$$

Moreover, for $n \geq 1$, we have

$$\bar{w}(n) = r(n) = \frac{n+1}{4}(1+3^{n-1}).$$

4 The distribution of valleys

4.1 ℓ-Valleys

In order to count nonempty flattened Catalan words according to the length and the number of ℓ-valleys, we introduce the following bivariate generating function:

$$V_\ell(x,y) \;=\; \sum_{w\in\mathsf{Flat}(\mathcal{C}^+)} x^{|w|}y^{\ell\text{-}\mathsf{val}(w)} \;=\; \sum_{n\geq 1} x^{|w|} \sum_{w\in\mathsf{Flat}(\mathcal{C}_n)} y^{\ell\text{-}\mathsf{val}(w)}$$

where ℓ-$\mathsf{val}(w)$ denotes the number of occurrences of subwords of the form $ab^\ell(b+1)$, and $a > b$, in w. The coefficient of $x^n y^k$ in $V_\ell(x,y)$ is the number of flattened Catalan words of length n with k distinct ℓ-valleys.

In Theorem 4.1, we give an expression for this generating function.

Theorem 4.1. *The generating function for nonempty flattened Catalan words with respect to the length and the number of ℓ-valleys is*

$$V_\ell(x,y) = \frac{x(1-2x+x^{\ell+1}-x^{\ell+1}y)}{(1-x)(1-3x+x^{\ell+1}-x^{\ell+1}y)}.$$

Proof. Let w be a nonempty flattened Catalan word, and let $w = \mathtt{0}(w'+1)w''$ be the first return decomposition, with $w', w'' \in \mathsf{Flat}(\mathcal{C})$. If $w' = w'' = \epsilon$, then $w = \mathtt{0}$, and its generating function is x. If $w' \neq \epsilon$ and $w'' = \epsilon$, then $w = \mathtt{0}(w'+1)$, and its generating function is $xV_\ell(x,y)$. Similarly, if $w' = \epsilon$ and $w'' \neq \epsilon$, then $w = \mathtt{0}w''$, and its generating function is $xV_\ell(x,y)$. Finally, if $w' \neq \epsilon$ and $w'' \neq \epsilon$, then $w = \mathtt{0}(w'+1)w''$. Because w is a flattened Catalan word, w' must be a weakly increasing word, and we distinguish two cases. If w'' is of the form $\mathtt{0}^{\ell-1}w'''$, where w''' starts with $\mathtt{01}$, then $w = \mathtt{0}(w'+1)\mathtt{0}^{\ell-1}w'''$, and the generating function is

$$\left(\frac{x^{\ell+1}y}{1-2x}\right)\Big(V_\ell(x,y) - \big(x + xV_\ell(x,y)\big)\Big).$$

Notice that $T_\ell(x,y) \coloneqq V_\ell(x,y) - \big(x + xV_\ell(x,y)\big)$ is obtained using the complement of the generating function for the word $\mathtt{0}$ and the words starting with $\mathtt{00}$.

The second case is the negation, so w'' does not start with $\mathtt{0}^\ell\mathtt{1}$. Notice that ℓ is fixed because we are interested in the ℓ-valleys, so the generating function is

$$\frac{x^2}{1-2x}\big(V_\ell(x,y) - x^{\ell-1}T_\ell(x,y)\big).$$

Therefore, we have the functional equation

$$V_\ell(x,y) = x + 2xV_\ell(x,y) + \left(\frac{x^{\ell+1}y}{1-2x}\right)T_\ell(x,y) + \frac{x^2}{1-2x}\big(V_\ell(x,y) - x^{\ell-1}T_\ell(x,y)\big).$$

Solving this equation, we obtain the desired result. □

Let $v_\ell(n,k)$ denote the number of flattened Catalan words of length n with exactly k distinct ℓ-valleys, that is $v_\ell(n,k) = [x^n y^k]V_\ell(x,y)$, which denotes the coefficient of $x^n y^k$ in $V_\ell(x,y)$. For example, the first few values of this array for $\ell = 2$ are

$$\mathcal{V}_2 := [v_2(n,k)]_{n\geq 4, k\geq 0} = \begin{pmatrix} 14 & 0 & 0 & 0 \\ 40 & 1 & 0 & 0 \\ 115 & \boxed{7} & 0 & 0 \\ 331 & 34 & 0 & 0 \\ 953 & 140 & 1 & 0 \\ 2744 & 527 & 10 & 0 \\ 7901 & 1877 & 64 & 0 \end{pmatrix}.$$

For example, $v_2(6,1) = 7$, the entry boxed in $\mathcal{V}_2$ above, and the corresponding flattened Catalan words of length 6 with one 2-valley (and lattice diagrams) are shown in Figure 4.1.

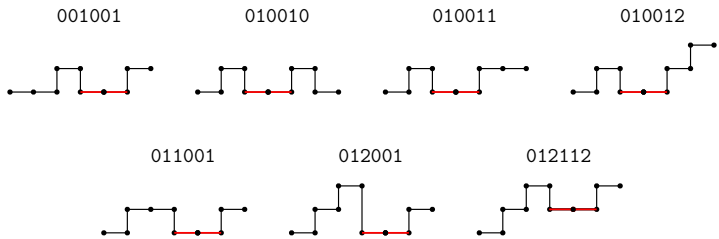

Figure 4.1: Flattened Catalan words of length 6 with one 2-valley. The red edges indicate the location of the 2-valley.

The first column of the array $\mathcal{V}_2$ corresponds to OEIS entry [19, A052963].

Let $v_\ell(n)$ be the sum of all ℓ-valleys in the set of flattened Catalan words of length n.

Corollary 4.2. *The generating function of the sequence $v_\ell(n)$ is*

$$\sum_{n\geq 1} v_\ell(n)x^n = \frac{x^{\ell+3}}{(1-x)(1-3x)^2}.$$

Moreover, for $n \geq 1$, we have

$$v_\ell(n) = \frac{1}{4}\big(1 - 3^{n-2-\ell} + 2\cdot 3^{n-2\ell}(n-2-\ell)\big).$$

Taking $\ell = 1$ in Theorem 4.1, we obtain the generating function for nonempty flattened Catalan words with respect to the length and the number of short valleys:

$$V_1(x,y) \;=\; \sum_{w\in \mathsf{Flat}(\mathcal{C}^+)} x^{|w|}y^{1\text{-}\mathsf{val}(w)} \;=\; \frac{x-2x^2+x^3(1-y)}{(1-x)\big(1-3x+x^2(1-y)\big)}.$$

Let $v_1(n,k)$ denote the number of flattened Catalan words of length n with exactly k short valleys, that is $v_1(n,k) = [x^n y^k]V_1(x,y)$, which denotes the coefficient of $x^n y^k$ in $V_1(x,y)$. The first few values of this array are

$$\mathcal{V}_1 = [v_1(n,k)]_{n\geq 1,k\geq 0} = \begin{pmatrix} 1 & 0 & 0 & 0\\ 2 & 0 & 0 & 0\\ 5 & 0 & 0 & 0\\ 13 & 1 & 0 & 0\\ 34 & \boxed{7} & 0 & 0\\ 89 & 32 & 1 & 0\\ 233 & 122 & 10 & 0\\ 610 & 422 & 61 & 1\\ 1597 & 1376 & 295 & 13 \end{pmatrix}.$$

For example, $v_1(5,1) = 7$, the entry boxed in $\mathcal{V}_1$ above, and the corresponding flattened Catalan words of length 5 with exactly one short valley (and lattice diagrams) are shown in Figure 4.2.

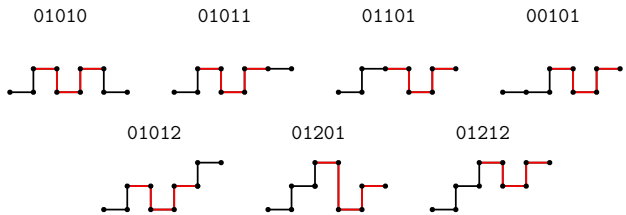

Figure 4.2: Flattened Catalan words of length 5 with one short valley. The red edges indicates the location of the short valley.

Remark 4.3. In [3], we proved that Catalan words of length n with k short valleys are in one-to-one correspondence with Dyck paths of semilength n with k occurrences of `DDUU`. Taking the restriction on flattened Catalan words of this bijection, we obtain a one-to-one correspondence between flattened Catalan words of length n with k short valleys and Dyck paths of semilength n with k occurrences of `DDUU`, where the height sequence of occurrences `DDU` (from left to right) is nondecreasing.

We can also obtain the generating function for the number of flattened Catalan words of length n with respect to the number of valleys (we consider all ℓ-valleys for $\ell \geq 1$).

Theorem 4.4. *The generating function for nonempty flattened Catalan words with respect to the length and the number of valleys is*

$$V(x,y) = \frac{x - 3x^2 + x^3(3-y)}{(1-x)(1-4x+4x^2-x^2y)}.$$

Let $v(n,k)$ denote the number of flattened Catalan words of length n with exactly k valleys, that is $v(n,k) = [x^n y^k]V(x,y)$, which denotes the coefficient of $x^n y^k$ in $V(x,y)$. The first few values of this array are

$$\mathcal{V} = [v(n,k)]_{n\geq 1, k\geq 0} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 \\ 13 & 1 & 0 & 0 \\ 33 & 8 & 0 & 0 \\ 81 & 40 & 1 & 0 \\ 193 & 160 & \boxed{\mathbf{12}} & 0 \\ 449 & 560 & 84 & 1 \\ 1025 & 1792 & 448 & 16 \end{pmatrix}.$$

For example, $v(7,2) = 12$, the entry boxed in $\mathcal{V}$ above, and the corresponding flattened Catalan words of length 7 with exactly two valleys are

`0010101`, `0100101`, `0101001`, `0101010`, `0101011`, `0101012`,
`0101101`, `0101201`, `0101212`, `0110101`, `0120101`, `0121212`.

Corollary 4.5. *For $n \geq 0$ we have*

$$v(n,k) = \begin{cases} (n-1)2^{n-2}+1, & \text{if } k = 0, \\ 2^{n-2k-2}\binom{n-1}{2k+1}, & \text{if } k \geq 1. \end{cases}$$

Note that $v(n,0)$ corresponds to OEIS entry [19, A005183].

Remark 4.6. In [3], we proved that Catalan words of length n with k valleys are in one-to-one correspondence with ordered trees with n edges and having exactly $k+1$ non-leaf nodes all of those children are leaves. Taking the restriction on flattened Catalan words of this bijection, we obtain a one-to-one correspondence between flattened Catalan words of length n and ordered trees with n edges and with $k+1$ nodes having only leaves as children and satisfying the following:

- if $T_1, T_2, \ldots, T_r$ are the subtrees of the root, then T_i, $i \in [1, r-1]$, is nondecreasing (i.e., for any node, its subtrees, except the rightmost, consist of one node only),
- the rightmost subtree of the root again satisfies all these properties.

Let $v(n)$ be the sum of all valleys in the set of flattened Catalan words of length n.

Corollary 4.7. *The generating function of the sequence $v(n)$ is*

$$\sum_{n\geq 0} v(n)x^n = \frac{x^4}{(1-x)^2(1-3x)^2}.$$

Moreover, for $n \geq 4$, we have

$$v(n) = \frac{1}{36}\big(3^n(n-4)+9n\big).$$

For $n \geq 4$, the first few values of the sequence $v(n)$ are

$$1,\quad 8,\quad 42,\quad 184,\quad 731,\quad 2736,\quad 9844,\quad 34448,\quad 118101,\quad 398584,\ldots.$$

This sequence corresponds to OEIS entry [19, A212337].

4.2 Symmetric valleys

A *symmetric valley* is a valley of the form $a(a-1)^\ell a$ with $\ell \geq 1$. Let $\mathsf{symv}(w)$ denote the number of symmetric valleys in the word w. In order to count flattened Catalan words according to the length and the number of symmetric valleys, we introduce the following bivariate generating function generating function:

$$S(x,y) \;=\; \sum_{w\in \mathsf{Flat}(\mathcal{C}^+)} x^{|w|}y^{\mathsf{symv}(w)} \;=\; \sum_{n\geq 1} x^{|w|} \sum_{w\in \mathsf{Flat}(\mathcal{C}_n)} y^{\mathsf{symv}(w)}$$

where the coefficient of $x^n y^k$ in $S(x,y)$ is the number of nonempty flattened Catalan words of length n with k symmetric ℓ-valleys.

In Theorem 4.8, we give an expression for this generating function.

Theorem 4.8. *The generating function of the nonempty flattened Catalan words with respect to the length and the number of symmetric valleys is*

$$S(x,y) = \frac{x(1-2x)(1-2x+2x^2-x^2y)}{(1-x)(1-5x+8x^2-5x^3-x^2y+2x^3y)}.$$

Proof. Let w be a nonempty flattened Catalan word, and let $w = \mathtt{0}(w'+1)w''$ be the first return decomposition, with $w', w'' \in \mathcal{C}$. If $w' = w'' = \epsilon$, then $w = \mathtt{0}$, and its generating function is x. If $w' \neq \epsilon$ and $w'' = \epsilon$, then $w = \mathtt{0}(w'+1)$, and its generating function is $xS(x,y)$. Similarly, if $w' = \epsilon$ and $w'' \neq \epsilon$, then $w = \mathtt{0}w''$, and its generating function is $xS(x,y)$. Finally, if $w' \neq \epsilon$ and $w'' \neq \epsilon$, then $w = \mathtt{0}(w'+1)w''$, we consider three cases.

1. If $w' = \mathtt{0}^k$ and w'' has a nonzero entry, then its generating function is
$$\left(\frac{x^2}{1-x}\right) y \left(S(x,y) - \frac{x}{1-x}\right).$$
2. If w' is a weakly increasing flattened Catalan word different from $\mathtt{0}^k$, and w'' has a nonzero entry, then its generating function is
$$x\left(\frac{x}{1-2x} - \frac{x}{1-x}\right)\left(S(x,y) - \frac{x}{1-x}\right).$$
3. If w' is a weakly increasing flattened Catalan word and $w'' = \mathtt{0}^k$, then its generating function is
$$\frac{x^3}{(1-x)(1-2x)}.$$

Therefore, we have the functional equation

$$\begin{aligned} S(x,y) = x + 2xS(x,y) &+ \left(\frac{x^2}{1-x}\right) y \left(S(x,y) - \frac{x}{1-x}\right) \\ &+ x\left(\frac{x}{1-2x} - \frac{x}{1-x}\right)\left(S(x,y) - \frac{x}{1-x}\right) \\ &+ \frac{x^3}{(1-x)(1-2x)}. \end{aligned}$$

Solving the obtained functional equation yields the desired result. □

Let $s(n,k)$ denote the number of flattened Catalan words of length n with exactly k symmetric valleys, that is $s(n,k) = [x^n y^k]S(x,y)$, which denotes the coefficient of $x^n y^k$ in $S(x,y)$. The first few values of this array are

$$\mathcal{S} = [s(n,k)]_{n\geq 1, k\geq 0} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \\ 13 & 1 & 0 & 0 & 0 \\ 34 & \boxed{7} & 0 & 0 & 0 \\ 90 & 31 & 1 & 0 & 0 \\ 242 & 113 & 10 & 0 & 0 \\ 659 & 375 & 59 & 1 & 0 \end{pmatrix}.$$

For example, $s(5,1) = 7$, the entry boxed in $\mathcal{S}$ above, and the corresponding flattened Catalan words of length 5 with 1 symmetric valley are given in Figure 4.3.

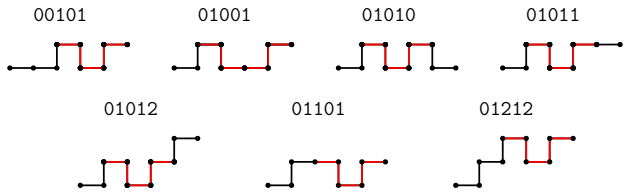

Figure 4.3: Flattened Catalan words of length 5 with one symmetric valley. In red we mark the location of the symmetric valley.

Let $s(n)$ be the sum of all symmetric valleys in the set of flattened Catalan words of length n.

Corollary 4.9. *The generating function of the sequence $s(n)$ is*

$$\sum_{n\geq 0} s(n)x^n = \frac{x^4(1+2x)}{(1-3x)^2(1-x)^3}.$$

Moreover, for $n \geq 4$, we have

$$s(n) = \frac{1}{144}\big(3^n(2n-5) - 18n^2 + 54n - 27\big).$$

The first few values of the sequence $s(n)$ $(n \geq 4)$ are

$$1,\quad 7,\quad 33,\quad 133,\quad 496,\quad 1770,\quad 6142,\quad 20902,\quad 70107,\quad 232489,\ldots.$$

5 The distribution of peaks

5.1 Peaks

In order to count flattened Catalan words according to the length and the number of ℓ-peaks, we introduce the following bivariate generating function:

$$P_\ell(x,y) = \sum_{w\in \mathsf{Flat}(\mathcal{C}^+)} x^{|w|} y^{\ell\text{-}\mathsf{peak}(w)} = \sum_{n\geq 1} x^{|w|} \sum_{w\in \mathsf{Flat}(\mathcal{C}_n)} y^{\ell\text{-}\mathsf{peak}(w)}$$

where ℓ-$\mathsf{peak}(w)$ denotes the number of occurrences of subwords of the form $a(a+1)^\ell b$, and $a \geq b$, in w. The coefficient of $x^n y^k$ in $P_\ell(x,y)$ is the number of flattened Catalan words of length n with k distinct ℓ-peaks.

In Theorem 5.1, we give an expression for this generating function.

Theorem 5.1. *The generating function for nonempty flattened Catalan words with respect to the length and the number of ℓ-peaks is*

$$P_\ell(x,y) = \frac{x(1-2x)}{(1-x)\big(1-3x+x^{\ell+1}(1-y)\big)}.$$

Proof. Let w be a nonempty flattened Catalan word, and let $w = \mathtt{0}(w'+1)w''$ be the first return decomposition, with $w', w'' \in \mathcal{C}$. If $w' = w'' = \epsilon$, then $w = \mathtt{0}$, and its generating function is x. If $w' \neq \epsilon$ and $w'' = \epsilon$, then $w = \mathtt{0}(w'+1)$, and its generating function is $xP_\ell(x,y)$. Similarly, if $w' = \epsilon$ and $w'' \neq \epsilon$, then $w = \mathtt{0}w''$, and its generating function is $xP_\ell(x,y)$. Finally, if $w' \neq \epsilon$ and $w'' \neq \epsilon$, then $w = \mathtt{0}(w'+1)w''$, and its generating function is

$$x\left(\frac{x}{1-2x} - x^\ell - \frac{x^{\ell+1}}{1-2x}\right)P_\ell(x,y) + xy\left(x^\ell + \frac{x^{\ell+1}}{1-2x}\right)P_\ell(x,y).$$

Therefore, we have the functional equation

$$P_\ell(x,y) = x + 2xP_\ell(x,y) + x\left(\frac{x}{1-2x} - x^\ell - \frac{x^{\ell+1}}{1-2x}\right)P_\ell(x,y) + xy\left(x^\ell + \frac{x^{\ell+1}}{1-2x}\right)P_\ell(x,y).$$

Solving the obtained functional equation yields the desired results. □

Let $p_\ell(n)$ be the sum of all ℓ-peaks in the set of flattened Catalan words of length n.

Corollary 5.2. *The generating function of the sequence $p_\ell(n)$ is*

$$\sum_{n\geq 1} p_\ell(n)x^n = \frac{x^{\ell+2}(1-2x)}{(1-3x)^2(1-x)}.$$

Moreover, for $n \geq 1$ we have

$$p_\ell(n) = \frac{1}{4}\big(3^{n-\ell-2}(2n+1-2\ell)-1\big).$$

Taking $\ell = 1$ in Theorem 5.1, establishes that the generating function for flattened Catalan words with respect to the length and the number of short peaks is

$$P_1(x,y) = \frac{x(1-2x)}{(1-x)\big(1-3x+x^2(1-y)\big)}.$$

Let $p_1(n,k)$ denote the number of flattened Catalan words of length n with exactly k short peaks, that is $p_1(n,k) = [x^n y^k]P_1(x,y)$, which denotes the coefficient of $x^n y^k$ in $P_1(x,y)$. The first few values of this array are

$$\mathcal{P}_1 = [p_1(n,k)]_{n\geq 1,k\geq 0} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 & 0 \\ 9 & 5 & 0 & 0 & 0 \\ 22 & 18 & 1 & 0 & 0 \\ 56 & 58 & \boxed{8} & 0 & 0 \\ 145 & 178 & 41 & 1 & 0 \\ 378 & 532 & 173 & 11 & 0 \\ 988 & 1563 & 656 & 73 & 1 \end{pmatrix}.$$

For example, $p_1(6,2) = 8$, the entry boxed in $\mathcal{S}$ above, and the corresponding flattened Catalan words of length 6 with 2 short peaks are

`001010`, `010100`, `010101`, `010010`, `010120`, `010121`, `012010`, `012121`.

For $n \geq 1$ we have $p_1(n,0) = F_{2(n-1)} + 1$, where F_m is the m^{th} Fibonacci number with initial values $F_1 = F_2 = 1$. For $n \geq 1$, the sequence $p_1(n,0)$ corresponds to the OEIS entry [19, A055588].

Using a similar proof as for Theorem 5.1, we generalize the result in order to obtain the following generating function for the number of flattened Catalan words of length n with respect to the number of peaks (we consider all ℓ-peaks for $\ell \geq 1$).

Theorem 5.3. *The generating function for flattened Catalan words with respect to the length and the number of peaks is*

$$P(x,y) = \frac{x(1-2x)}{1-4x+4x^2-x^2y}.$$

Let $p(n,k)$ denote the number of flattened Catalan words of length n with exactly k peaks, that is $p(n,k) = [x^n y^k]P(x,y)$, which denotes the coefficient of $x^n y^k$ in $P(x,y)$. The first few values of this array are

$$\mathcal{P} = [p(n,k)]_{n\geq 1, k\geq 0} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 & 0 \\ 8 & \boxed{\mathbf{6}} & 0 & 0 & 0 \\ 16 & 24 & 1 & 0 & 0 \\ 32 & 80 & 10 & 0 & 0 \\ 64 & 240 & 60 & 1 & 0 \\ 128 & 672 & 280 & 14 & 0 \\ 256 & 1792 & 1120 & 112 & 1 \end{pmatrix}.$$

For example, $p(4,1) = 6$, the entry boxed in $\mathcal{P}$ above, and the corresponding flattened Catalan words of length 4 with 1 peaks are

$$\texttt{0010}, \quad \texttt{0100}, \quad \texttt{0110}, \quad \texttt{0101}, \quad \texttt{0120}, \quad \texttt{0121}.$$

Let $p(n)$ be the sum of all peaks in the set of flattened Catalan words of length n.

Corollary 5.4. *The generating function of the sequence $p(n)$ is*

$$\sum_{n\geq 0} p(n)x^n = \frac{(1-2x)x^3}{(1-4x+3x^2)^2}.$$

Moreover, for $n \geq 3$, we have

$$p(n) = \frac{1}{4}(3^{n-2}-1)(n-1).$$

The first few values of the sequence $p(n)$ $(n \geq 3)$ are

$$1, \quad 6, \quad 26, \quad 100, \quad 363, \quad 1274, \quad 4372, \quad 14760, \quad 14760, \quad 49205, \ldots.$$

This sequence corresponds to the OEIS entry [19, A261064]. Our combinatorial interpretation is new.

5.2 Symmetric peaks

A *symmetric peak* is a peak of the form $a(a+1)^\ell a$ with $\ell \geq 1$. Let $\mathsf{symp}(w)$ denote the number of the symmetric peaks of the word w. In order to count flattened Catalan words according to the length and the number of symmetric peaks, we introduce the following bivariate generating function:

$$T(x,y) = \sum_{w \in \mathsf{Flat}(\mathcal{C}^+)} x^{|w|} y^{\mathsf{symp}(w)} = \sum_{n \geq 1} x^{|w|} \sum_{w \in \mathsf{Flat}(\mathcal{C}_n)} y^{\mathsf{symp}(w)}$$

where the coefficient of $x^n y^k$ in $T(x,y)$ is the number of flattened Catalan words of length n with k symmetric peaks.

In Theorem 5.5, we give an expression for this generating function.

Theorem 5.5. *The generating function of the nonempty flattened Catalan words with respect to the length and the number of symmetric peaks is*

$$T(x,y) = \frac{x(1-x)(1-2x)}{1-5x+8x^2-5x^3-x^2y+2x^3y}.$$

Proof. Let w be a nonempty flattened Catalan word, and let $w = \mathtt{0}(w'+1)w''$ be the first return decomposition, with $w', w'' \in \mathsf{Flat}(\mathcal{C})$. If $w' = w'' = \epsilon$, then $w = \mathtt{0}$, and its generating function is x. If $w' \neq \epsilon$ and $w'' = \epsilon$, then $w = \mathtt{0}(w'+1)$, and its generating function is $xT(x,y)$. Similarly, if $w' = \epsilon$ and $w'' \neq \epsilon$, then $w = \mathtt{0}w''$, and its generating function is $xT(x,y)$. Finally, if $w' \neq \epsilon$ and $w'' \neq \epsilon$, then $w = \mathtt{0}(w'+1)w''$, and we have two cases to consider.

1. If w' is all $\mathtt{0}$'s, its generating function is

$$\frac{x^2y}{1-x}T(x,y).$$

2. Otherwise, the generating function is

$$x\left(\frac{x}{1-2x} - \frac{x}{1-x}\right)T(x,y).$$

Therefore, we have the functional equation is

$$T(x,y) = x + 2xT(x,y) + \frac{x^2y}{1-x}T(x,y) + x\left(\frac{x}{1-2x} - \frac{x}{1-x}\right)T(x,y).$$

Solving this equation yields the desired result. □

Let $t(n,k)$ denote the number of flattened Catalan words of length n with exactly k symmetric peaks, that is $t(n,k) = [x^n y^k]T(x,y)$, which denotes the coefficient of $x^n y^k$ in $T(x,y)$. The first few values of this array are

$$\mathcal{T} = [t(n,k)]_{n\geq 1, k\geq 0} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 & 0 \\ 9 & \boxed{\mathbf{5}} & 0 & 0 & 0 \\ 23 & 17 & 1 & 0 & 0 \\ 63 & 51 & 8 & 0 & 0 \\ 176 & 149 & 39 & 1 & 0 \\ 491 & 439 & 153 & 11 & 0 \end{pmatrix}.$$

For example, $t(4,1) = 5$, the entry boxed in $\mathcal{T}$ above, and the corresponding flattened Catalan words of length 4 with 1 symmetric peak (and lattice diagrams) are shown in Figure 5.1.

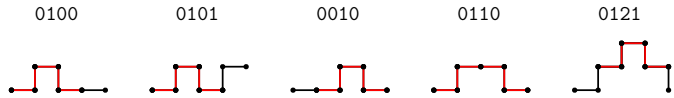

Figure 5.1: Flattened Catalan words of length 4 with 1 symmetric peak. In red we mark the location of the symmetric peak.

The first and second column of the array $\mathcal{T}$ coincides with OEIS entries [19, A369328, A290900].

Let $t(n)$ be the sum of all symmetric peaks in the set of flattened Catalan words of length n.

Corollary 5.6. *The generating function of the sequence $t(n)$ is*

$$\sum_{n\geq 0} t(n)x^n = \frac{(1-2x)^2x^3}{(1-3x)^2(1-x)^3}.$$

Moreover, for $n \geq 3$, we have

$$t(n) = \frac{1}{144}\big(63 + 3^n + 2(-45+3^n)n + 18n^2\big).$$

For $n \geq 3$, the first few values of the sequence $t(n)$ are

$$1,\quad 5,\quad 19,\quad 67,\quad 230,\quad 778,\quad 2602,\quad 8618,\quad 28303,\quad 92275,\ldots.$$

Acknowledgments

The authors thank an anonymous referee for some advice and comments that helped to improve the exposition. Jean-Luc Baril was supported by the University of Burgundy. Pamela E. Harris was supported in part by a Karen Uhlenbeck EDGE Fellowship. José L. Ramírez was partially supported by the Universidad Nacional de Colombia. The authors thank Kimberly J. Harry and Matt McClinton for their helpful discussions during the completion of this manuscript.

References

[1] J.-L. Baril, D. Colmenares, J. L. Ramírez, D. Silva, L. M. Simbaqueba, and D. Toquica, Consecutive pattern-avoidance in Catalan words according to the last symbol, *RAIRO Theor. Inform. Appl.* **58** (2024), Paper No. 1. https://doi.org/10.1051/ita/2024001

[2] J.-L. Baril, J. F. González, and J. L. Ramírez, Last symbol distribution in pattern avoiding Catalan words, *Math. Comput. Sci.* **18** (1) (2024). https://doi.org/10.1007/s11786-023-00576-5

[3] J.-L. Baril, P. E. Harris, K. J. Harry, M. McClinton, and J. L. Ramírez, Enumerating runs, valleys, and peaks in Catalan words, arXiv:2404.05672 (2024).

[4] J.-L. Baril, C. Khalil, and V. Vajnovszki, Catalan words avoiding pairs of length three patterns, *Discret. Math. Theor. Comput. Sci.* **22** (2) (2021), #5. https://doi.org/10.46298/dmtcs.6002

[5] J.-L. Baril, S. Kirgizov, and V. Vajnovszki, Descent distribution on Catalan words avoiding a pattern of length at most three, *Discrete Math.* **341** (2018), 2608–2615. https://doi.org/10.1016/j.disc.2018.06.001

[6] J.-L. Baril, S. Kirgizov, J. L. Ramírez, and D. Villamizar, The combinatorics of Motzkin polyominoes, arXiv:2401.06228 (2024).

[7] J.-L. Baril and J. L. Ramírez, Descent distribution on Catalan words avoiding ordered pairs of relations, *Adv. in Appl. Math.* **149** (2023), 102551. https://doi.org/10.1016/j.aam.2023.102551

[8] A. Buck, J. Elder, A. A. Figueroa, P. E. Harris, K. J. Harry, and A. Simpson, Flattened Stirling permutations, arXiv:2306.13034 (2023).

[9] D. Callan, Pattern avoidance in "flattened" partitions, *Discrete Math.* **309** (12) (2009), 4187–4191. https://doi.org/10.1016/j.disc.2008.11.019

[10] D. Callan, T. Mansour, and J. L. Ramírez, Statistics on bargraphs of Catalan words, *J. Autom. Lang. Comb.* **26** (2021), 177–196. https://doi.org/10.25596/jalc-2021-177

[11] J. Elder, P. E. Harris, Z. Markman, I. Tahir, and A. Verga, On flattened parking functions, *J. Integer Seq.* **26** (2023), Article 23.5.8. https://cs.uwaterloo.ca/journals/JIS/VOL26/Harris/harris3.pdf

[12] T. Mansour and V. Vajnovszki, Efficient generation of restricted growth words, *Inform. Process. Lett.* **113** (2013), 613–616. https://doi.org/10.1016/j.ipl.2013.05.008

[13] T. Mansour and J. L. Ramírez, Enumerations on polyominoes determined by Fuss-Catalan words, *Australas. J. Combin.* **81** (3) (2021), 447–457.

[14] T. Mansour and J. L. Ramírez, Exterior corners on bargraphs of Motzkin words. In: F. Hoffman et al. (eds) Combinatorics, Graph Theory and Computing. SEICCGTC 2021. *Springer Proceedings in Mathematics & Statistics*, **448** (2024), 355–367.

[15] T. Mansour, J. L. Ramírez, and D. A. Toquica, Counting lattice points on bargraphs of Catalan words, *Math. Comput. Sci.* **15** (2021), 701–713. https://doi.org/10.1007/s11786-021-00501-8.

[16] O. Nabawanda, F. Rakotondrajao, and A. Bamunoba, Run distribution over flattened partitions, *J. Integer Seq.* **23** (2020), Article 20.9.6.

[17] J. L. Ramírez and A. Rojas-Osorio, Consecutive patterns in Catalan words and the descent distribution, *Bol. Soc. Mat. Mex.* **29** (2023), Article #60. https://doi.org/10.1007/s40590-023-00532-0.

[18] M. Shattuck, Counting subword patterns in Catalan words, *Art Discrete Appl. Math.* Accepted, (2024). https://doi.org/10.26493/2590-9770.1695.4da.

[19] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, http://oeis.org/.

[20] R. Stanley, *Catalan Numbers*, Cambridge University Press, Cambridge, 2015.

[21] F. K. Hwang and C. L. Mallows, Enumerating nested and consecutive partitions, *J. Combin. Theory Ser. A* **70** (2) (1995), 323–333.

Jean-Luc Baril
LIB, Université de Bourgogne,
B.P. 47 870, 21078,
Dijon Cedex, France
barjl@u-bourgogne.fr

Pamela E. Harris
Department of Mathematical Sciences,
University of Wisconsin-Milwaukee,
Milwaukee, WI 53211 United States
peharris@uwm.edu

José L. Ramírez
Departamento de Matemáticas,
Universidad Nacional de Colombia,
Bogotá, Colombia
jlramirezr@unal.edu.co