

# EXPLORATIONS COMBINATOIRES DES STRUCTURES ARBORESCENTES ET LIBRES

CHRISTOPHE CORDERO

sous la direction de

Samuele Giraudo et Jean-Christophe Novelli

du laboratoire d'informatique Gaspard-Monge  
de l'université Paris-Est Marne-la-Vallée



9 décembre 2019



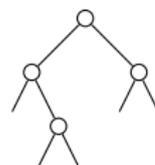
# COMBINATOIRE

## Objectifs

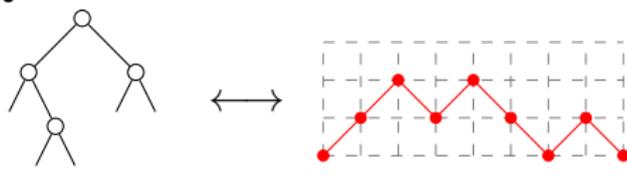
Dénombrer et énumérer des ensembles paramétrés finis.

## Énumérative

Formules :

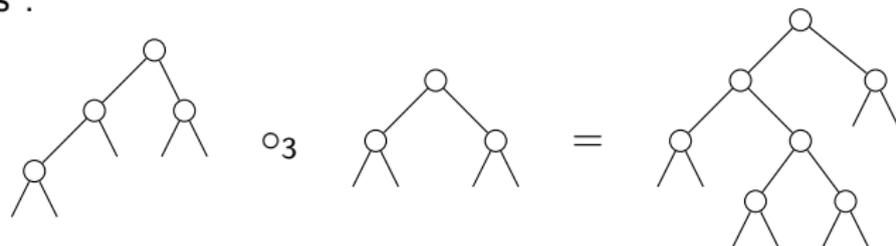

$$: \frac{1}{n+1} \binom{2n}{n}$$

Bijections :



## Algébrique

Opérations :



# OBJETS COMBINATOIRES

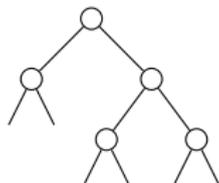
## DÉNOMBREMENT

### Mots

cecinestpasinmot

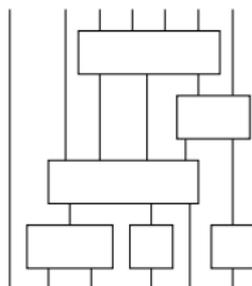
Dénombrement : il y a  $k^\ell$  mots de longueur  $\ell$  sur un alphabet à  $k$  lettres.

### Arbres



Dénombrement : il y a  $\frac{1}{n+1} \binom{2n}{n}$  arbres binaires à  $n$  nœuds.

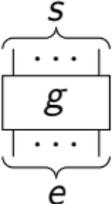
### Circuits



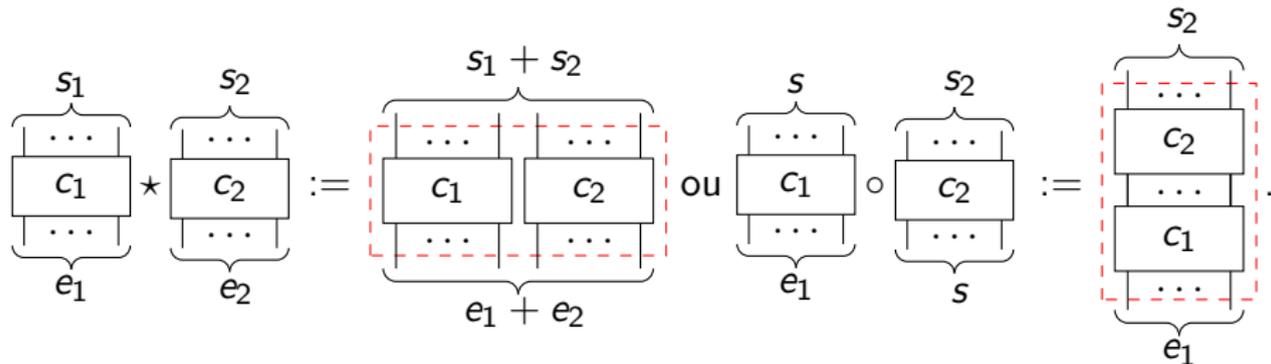
Dénombrement : ?

# I) CIRCUITS

## DÉFINITION

Un **circuit** est soit un *générateur* , soit un *fil*  $|$ ,

soit un des assemblages :



# I) DÉNOMBRER LES CIRCUITS

## PROBLÈME

$$C_{e,n,s}(\mathbb{G}) := \left\{ \begin{array}{c} s \\ \overbrace{\quad \quad \quad} \\ \dots \\ \text{\textit{n} générateurs} \\ \dots \\ \underbrace{\quad \quad \quad} \\ e \end{array} \right\}$$

$$\text{Générateurs : } \mathbb{G} := \left\{ \begin{array}{c} \beta_1 \\ \overbrace{\quad \quad \quad} \\ \text{\textbf{1}} \\ \underbrace{\quad \quad \quad} \\ \alpha_1 \end{array} \right\}, \dots, \left\{ \begin{array}{c} \beta_1 \\ \overbrace{\quad \quad \quad} \\ \text{\textbf{m}_1} \\ \underbrace{\quad \quad \quad} \\ \alpha_1 \end{array} \right\}, \dots, \left\{ \begin{array}{c} \beta_d \\ \overbrace{\quad \quad \quad} \\ \text{\textbf{1}} \\ \underbrace{\quad \quad \quad} \\ \alpha_d \end{array} \right\}, \dots, \left\{ \begin{array}{c} \beta_d \\ \overbrace{\quad \quad \quad} \\ \text{\textbf{m}_d} \\ \underbrace{\quad \quad \quad} \\ \alpha_d \end{array} \right\}.$$

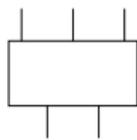
# I) DÉNOMBREMENT DE $C_{2,n,n+2}(\{\square\})$

CAS INITIAUX

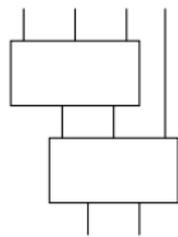
$n = 0$  :



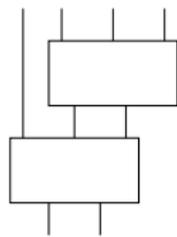
$n = 1$  :



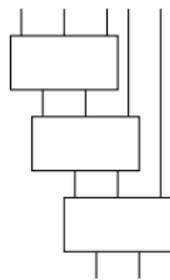
$n = 2$  :



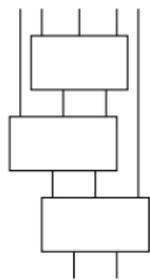
et



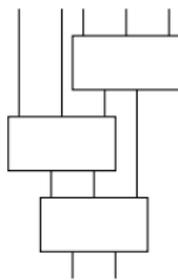
$n = 3$  :



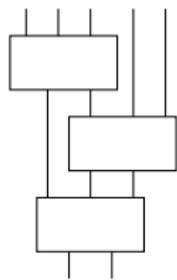
,



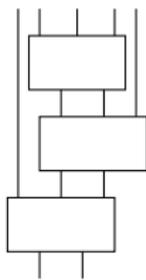
,



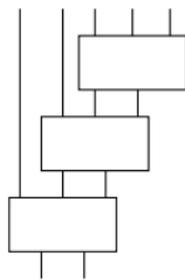
,



,



et

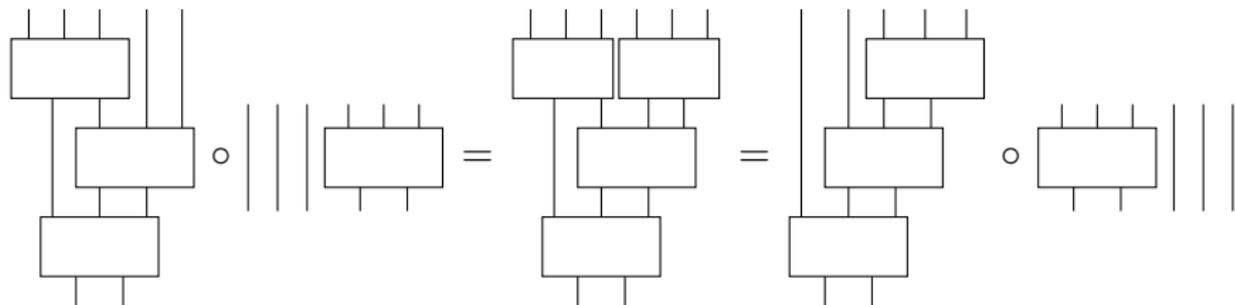


Suite : 1, 1, 2, 6, ?, ...

# I) DÉNOMBREMENT DE $C_{2,n,n+2}(\{\square\})$

DIFFICULTÉ : AMBIGUÏTÉ

$n = 4$  :



# I) DÉNOMBREMENT DE $C_{2,n,n+2}$ ( $\left\{ \begin{array}{c} \square \\ \hline \square \\ \hline \square \end{array} \right\}$ )

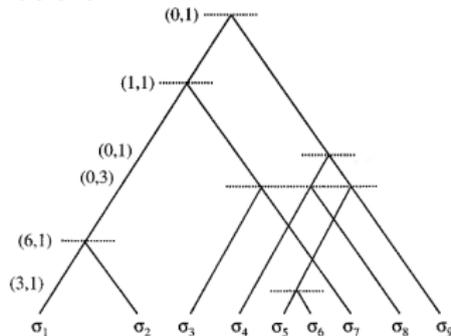
## L'encyclopédie OEIS

[A264868](#) Number of rooted tandem duplication trees on n gene segments.

**1, 1, 2, 6, 22, 92, 420**, 2042, 10404, 54954, 298648, 1660714, 9410772, 54174212, 316038060, 1864781388, 11111804604, 66782160002, 404392312896, 2465100947836, 15116060536540, 93184874448186, 577198134479356, 3590697904513792 ([list](#); [graph](#); [refs](#); [listen](#); [history](#); [text](#); [internal format](#))

REFERENCES Mathematics of Evolution and Phylogeny, O. Gascuel (ed.), Oxford University Press, 2005

## Arbre tandem de duplication

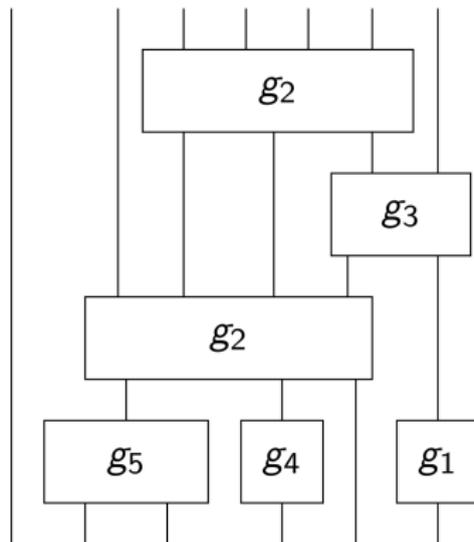


[Image de Gascuel, Hendy, Jean-Marie et Mclachlan]

[C. en 2018]

# I) CONSTRUCTION NON AMBIGUË DES CIRCUITS

« TETRIS »

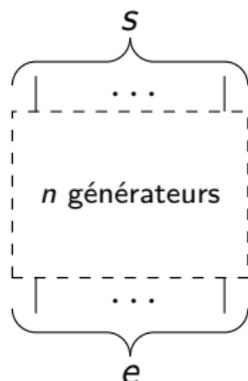


Théorème [C.]

# I) BIJECTION : CIRCUITS $\longleftrightarrow$ CHEMINS

THÉORÈME [C. EN 2018]

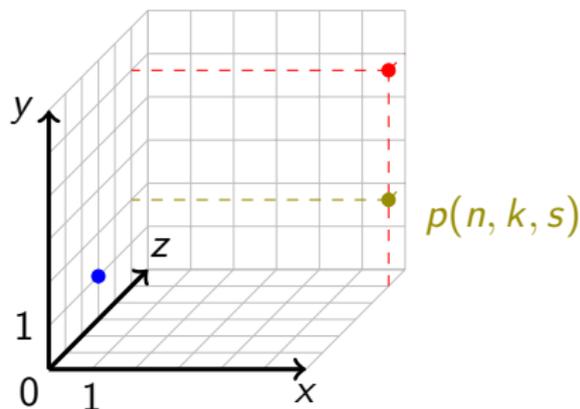
Circuits



Générateurs

$$\left\{ \begin{array}{c} \beta \\ \dots \\ g \\ \dots \\ \alpha \end{array} \right\} \in \mathbb{G}$$

Chemins de  $(0, 1, e)$  à  $(n, s, s)$



Contrainte : «  $1 \leq y \leq z$  »

Pas

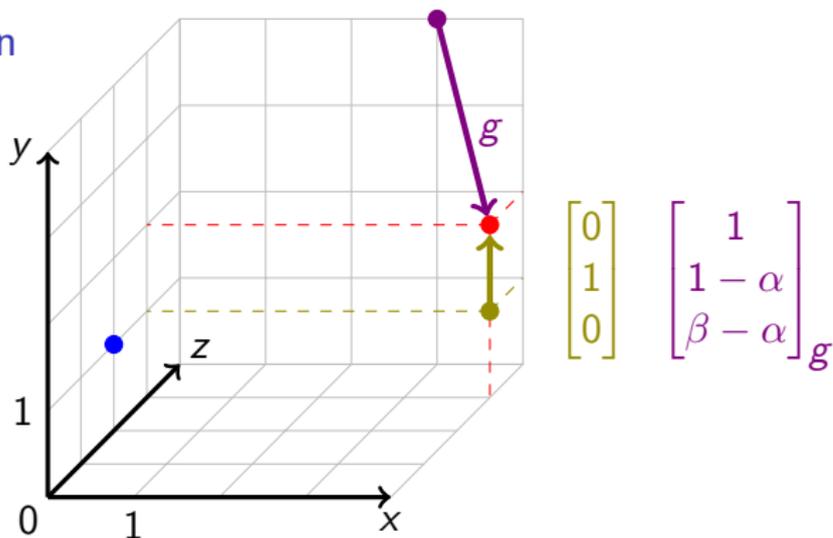
$$\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\} \cup \left\{ \dots, \begin{bmatrix} 1 \\ 1 - \alpha \\ \beta - \alpha \end{bmatrix}_g, \dots \right\}$$

# I) FORMULES DE RÉCURRENCE : 1/2

Théorème [C. en 2018]

$$p(n, k, s) = \begin{cases} 1 & \text{si } n = 0, k = 1 \text{ et } s = e, \\ p(n, k - 1, s) + \sum_{g \in \mathbb{G}(\alpha, \beta)} p(n - 1, k - 1 + \alpha, s - \beta + \alpha) & \text{si } n \geq 0 \text{ et } 1 \leq k \leq s, \\ 0 & \text{sinon.} \end{cases}$$

Démonstration



# I) FORMULES DE RÉCURRENCE : 2/2

## Générateurs (rappel)

$$\mathbb{G} := \left\{ \underbrace{\begin{array}{c} \beta_1 \\ \dots \\ \boxed{1} \\ \dots \\ \alpha_1 \end{array}} , \dots , \underbrace{\begin{array}{c} \beta_1 \\ \dots \\ \boxed{m_1} \\ \dots \\ \alpha_1 \end{array}} , \dots , \underbrace{\begin{array}{c} \beta_d \\ \dots \\ \boxed{1} \\ \dots \\ \alpha_d \end{array}} , \dots , \underbrace{\begin{array}{c} \beta_d \\ \dots \\ \boxed{m_d} \\ \dots \\ \alpha_d \end{array}} \right\}$$

## Théorème [C. en 2018]

Le nombre  $|C_{e,n,s}(\mathbb{G})|$  vérifie la relation de récurrence :

$$\begin{cases} 1 & \text{si } n = 0 \text{ et } s = e, \\ \sum_{\ell=1}^n (-1)^{\ell+1} \sum_{c_1+\dots+c_d=\ell} \binom{\ell}{c_1, \dots, c_d} \binom{s+\ell-\sum_{i=1}^d c_i \beta_i}{\ell} m_1^{c_1} \dots m_d^{c_d} & \left| C_{e, n-\ell, s-\sum_{i=1}^d c_i(\beta_i-\alpha_i)}(\mathbb{G}) \right| \\ 0 & \text{si } n, s \geq 1, \\ 0 & \text{sinon.} \end{cases}$$

# I) CIRCUITS COMPOSÉS D'UN TYPE DE GÉNÉRATEUR

FORMULE CLOSE

## Générateurs

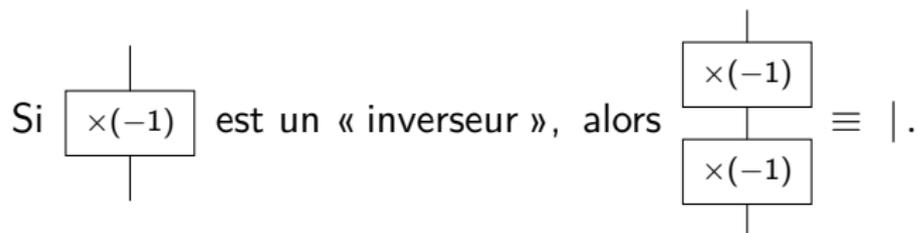
$$\mathbb{G} := \left\{ \begin{array}{c} \beta \\ \cdots \\ \boxed{1} \\ \cdots \\ \alpha \end{array} , \dots , \begin{array}{c} \beta \\ \cdots \\ \boxed{m} \\ \cdots \\ \alpha \end{array} \right\}$$

Remarque :  $s = e + (\beta - \alpha)n$

Théorème [C. en 2018]

$$|C_{e,n,s}(\mathbb{G})| = m^n \det(\mathcal{B}), \text{ où } \mathcal{B}_{i,j} := \begin{pmatrix} e - i(\alpha - 1) + (j - 1)(\beta - 1) \\ i - j + 1 \end{pmatrix}.$$

# DÉNOMBREMENT DES CIRCUITS NON LIBRES ?



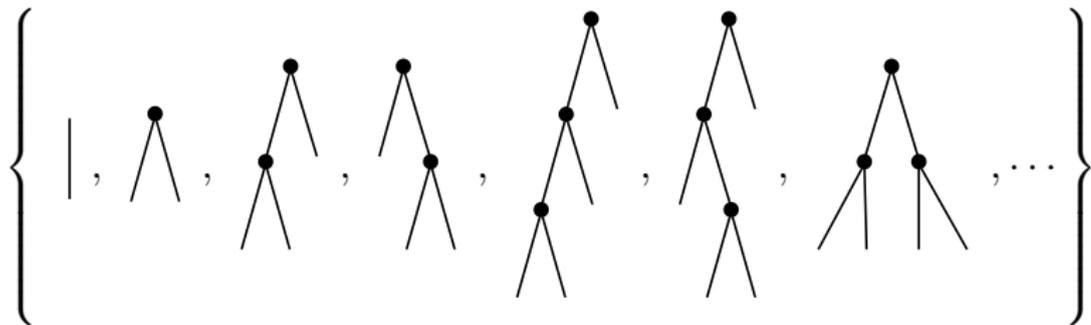
## Problème

Certains circuits équivalents sont comptés plusieurs fois !

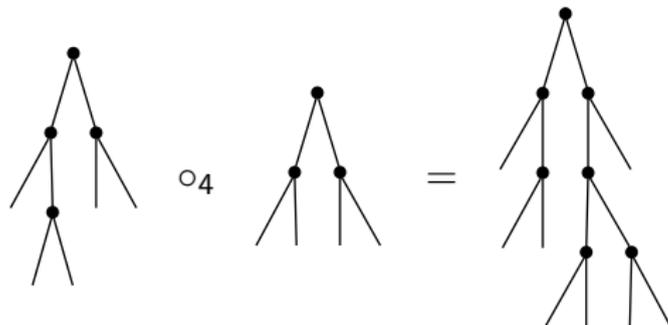
## II) L'OPÉRADE MAGMATIQUE : Mag

### DÉFINITION

L'ensemble des arbres binaires

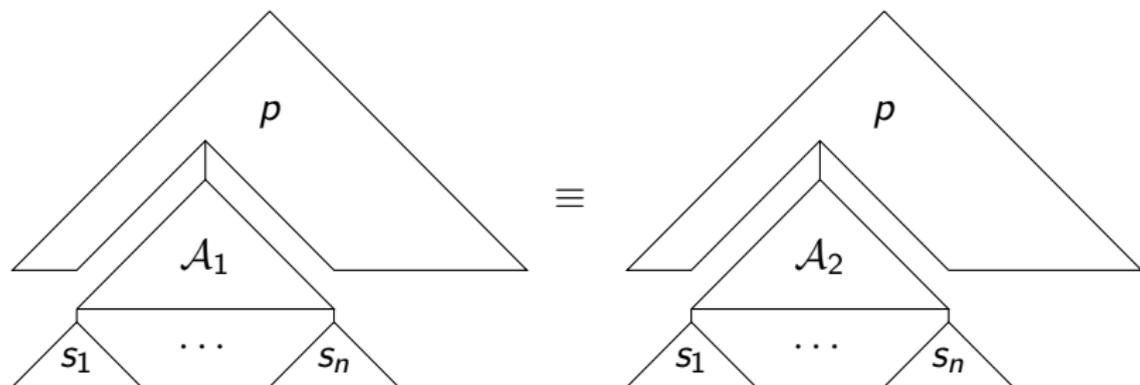


muni des compositions partielles  $\{\circ_1, \circ_2, \circ_3, \dots\}$  :



## II) QUOTIENTS DE Mag : $\text{Mag} / \langle \mathcal{A}_1 \equiv \mathcal{A}_2 \rangle$

DÉFINITION

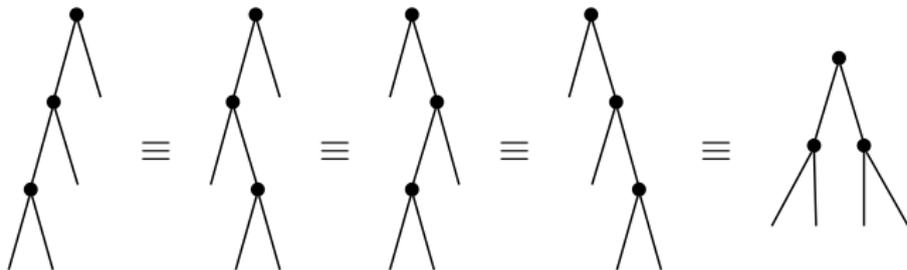


## II) L'OPÉRADE ASSOCIATIVE

Définition

$$\text{As} := \text{Mag} / \left\langle \left\langle \begin{array}{c} \bullet \\ / \quad \backslash \\ \bullet \quad \bullet \\ / \quad \backslash \\ \bullet \quad \bullet \end{array} \equiv \begin{array}{c} \bullet \\ / \quad \backslash \\ \bullet \quad \bullet \\ / \quad \backslash \\ \bullet \quad \bullet \end{array} \right\rangle \right\rangle$$

Exemple



$$\begin{aligned} \text{Série de Hilbert : } \mathcal{H}_{\text{As}}(t) &= \sum_{n \geq 1} |\text{As}(n)| t^n \\ &= t + t^2 + t^3 + t^4 + \dots \\ &= \frac{t}{1-t} \end{aligned}$$

## II) LES OPÉRADES PEIGNES

### GÉNÉRALISATION DE $As$

Définition [Chenavier, C. et Giraud en 2018]

$$OP^{(d)} := \text{Mag} / \left\langle \begin{array}{c} \text{diagram 1} \\ \equiv \\ \text{diagram 2} \end{array} \right\rangle$$

Relation

$$(\cdots ((x_1 * x_2) * x_3) \cdots) * x_{d+1} = x_1 * (\cdots (x_{d-1} * (x_d * x_{d+1})) \cdots)$$

Remarques :  $OP^{(0)} \cong OP^{(1)} \cong \text{Mag}$  et  $OP^{(2)} = As$ .

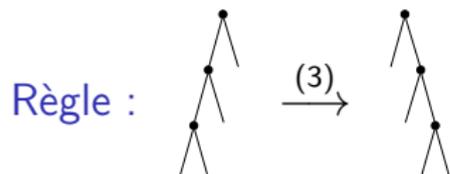
Séries de Hilbert

$$\mathcal{H}_{OP^{(1)} \cong \text{Mag}}(t) = t + t^2 + 2t^3 + 5t^4 + 14t^5 + 42t^6 + 132t^7 + 429t^8 \dots$$

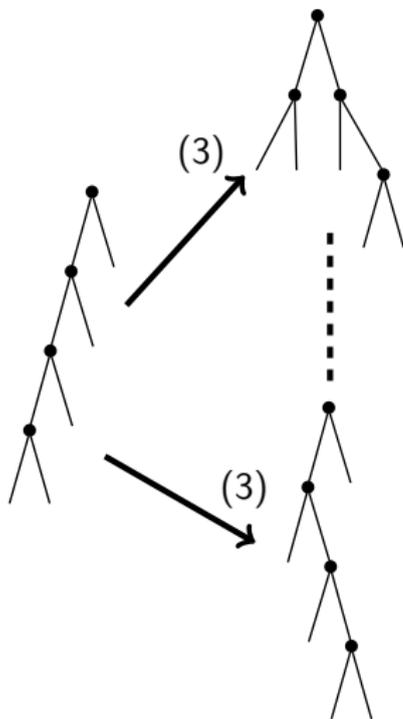
$$\mathcal{H}_{OP^{(2)} \cong As}(t) = t + t^2 + t^3 + t^4 + t^5 + t^6 + t^7 + t^8 \dots$$

$$\mathcal{H}_{OP^{(3)}}(t) = t + t^2 + 2t^3 + 4t^4 + 8t^5 + 14t^6 + 20t^7 + 19t^8 \dots$$

## II) RÉÉCRITURE DANS $OP^{(3)}$

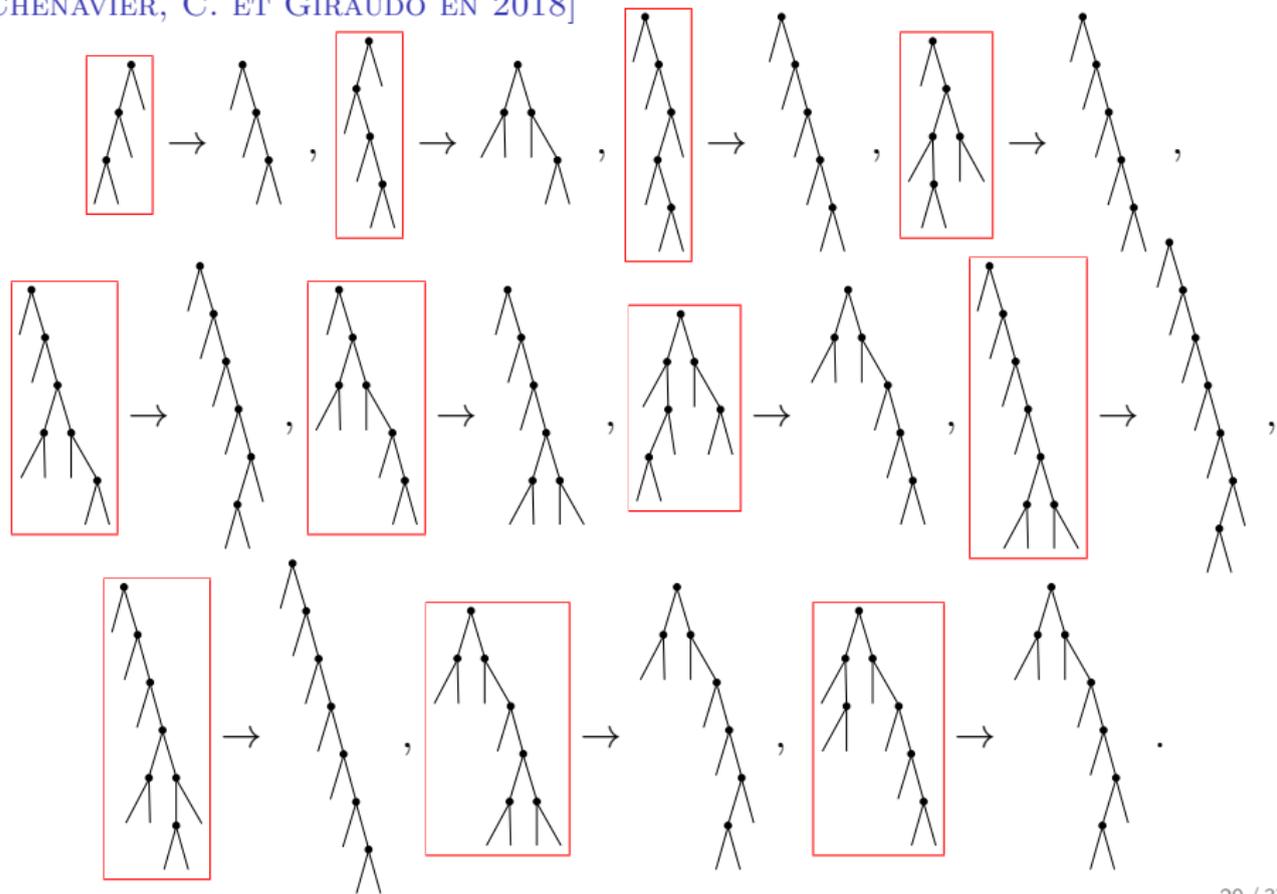


- Pas confluent !
- Calculer une complétion.  
[Dotsenko et Khoroshkin en 2010]



## II) SYSTÈME DE RÉÉCRITURE CONFLUENT POUR $OP^{(3)}$

[CHENAVIER, C. ET GIRAUDO EN 2018]



## II) DÉNOMBREMENT DE $OP^{(3)}$

### SÉRIE DE HILBERT

Proposition [Chenavier, C. et Giraudo en 2018]

$$\mathcal{H}_{OP^{(3)}}(t) = \mathcal{G}_{\text{évite}} \left\{ \begin{array}{cccccccccccc} \begin{array}{c} \wedge \\ \wedge \quad \wedge \end{array}, & \begin{array}{c} \wedge \\ \wedge \quad \wedge \\ \wedge \quad \wedge \end{array}, & \begin{array}{c} \wedge \\ \wedge \quad \wedge \\ \wedge \quad \wedge \\ \wedge \quad \wedge \end{array}, & \begin{array}{c} \wedge \\ \wedge \quad \wedge \\ \wedge \quad \wedge \\ \wedge \quad \wedge \\ \wedge \quad \wedge \end{array}, & \begin{array}{c} \wedge \\ \wedge \quad \wedge \end{array}, & \begin{array}{c} \wedge \\ \wedge \quad \wedge \end{array}, & \begin{array}{c} \wedge \\ \wedge \quad \wedge \end{array}, & \begin{array}{c} \wedge \\ \wedge \quad \wedge \end{array}, & \begin{array}{c} \wedge \\ \wedge \quad \wedge \end{array}, & \begin{array}{c} \wedge \\ \wedge \quad \wedge \end{array}, & \begin{array}{c} \wedge \\ \wedge \quad \wedge \end{array}, & \begin{array}{c} \wedge \\ \wedge \quad \wedge \end{array} \end{array} \right\}$$

En utilisant [Rowland 2010, Khoroshkin & Piontkovski 2015 et Giraudo 2018]

Théorème [Chenavier, C. et Giraudo en 2018]

$$\begin{aligned} \mathcal{H}_{OP^{(3)}}(t) &= \frac{t(1-t+t^2+t^3+2t^4+2t^5-7t^7-2t^8+t^9+2t^{10}+t^{11})}{(1-t)^2} \\ &= t + t^2 + 2t^3 + 4t^4 + 8t^5 + 14t^6 + 20t^7 + 19t^8 + 16t^9 + 14t^{10} + \sum_{n \geq 11} (n+3)t^n \end{aligned}$$

## II) SYSTÈME DE RÉÉCRITURE POUR $OP^{(d \geq 4)}$ ?

Conjecture [Chenavier, C. et Giraud en 2018]

Il n'y a pas de système de réécriture fini et confluent pour  $OP^{(d)}$ , où  $d \geq 4$ .

## II) MORPHISMES D'OPÉRADES ET TREILLIS

### Définition

Un **morphisme d'opérate** est une application  $f : \mathcal{O}_1 \rightarrow \mathcal{O}_2$  telle que :

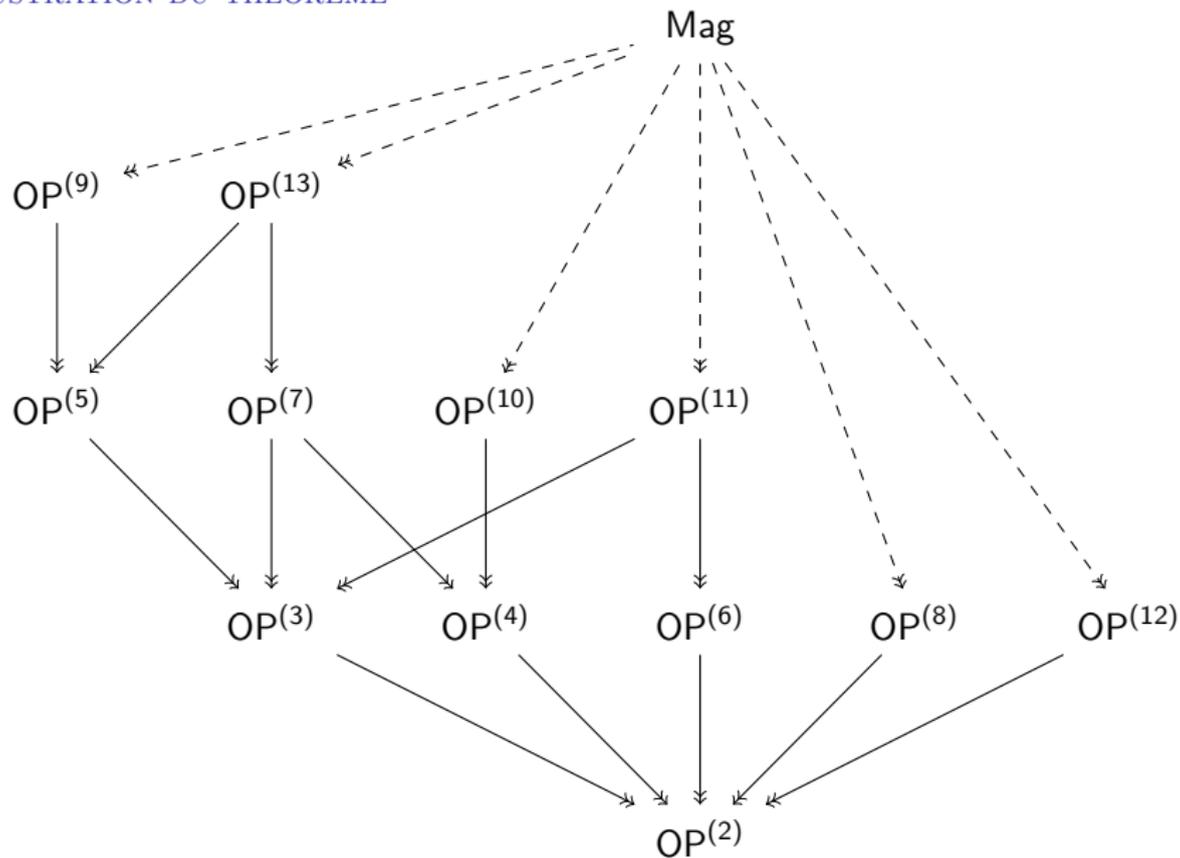
pour tout  $x \in \mathcal{O}_1$ ,  $|x| = |f(x)|$   
et pour tout  $x, y \in \mathcal{O}_1$  et  $1 \leq i \leq |x|$ ,  $f(x \circ_i y) = f(x) \circ_i f(y)$ .

### Théorème [Chenavier, C. et Giraud en 2018]

Il existe un morphisme de  $OP^{(d')}$  vers  $OP^{(d)}$  ssi  $(d - 1) \mid (d' - 1)$ .  
Un tel morphisme est nécessairement surjectif.

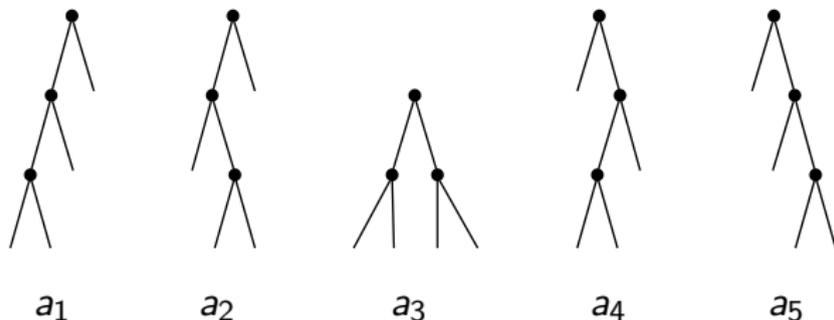
## II) MORPHISMES D'OPÉRADES ET TREILLIS

ILLUSTRATION DU THÉORÈME



## II) QUOTIENTS CUBIQUES

### ÉTUDE EXHAUSTIVE



Définition :  $\text{Mag}^{\{i,j\}} := \text{Mag}/\langle a_i \equiv a_j \rangle$       Remarque :  $\text{Mag}^{\{1,5\}} = \text{OP}^{(3)}$

À équivalence près

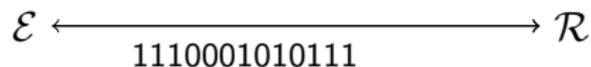
$\text{Mag}^{\{1,2\}}$ ,  $\text{Mag}^{\{1,3\}}$ ,  $\text{Mag}^{\{1,4\}}$ ,  $\text{OP}^{(3)}$ ,  $\text{Mag}^{\{2,3\}}$  et  $\text{Mag}^{\{2,4\}}$

Séries de Hilbert [Chenavier, C. et Giraud en 2019]

$$\mathcal{H}_{\text{Mag}^{\{1,2\}}}(t) = \mathcal{H}_{\text{Mag}^{\{1,3\}}}(t) = \mathcal{H}_{\text{Mag}^{\{1,4\}}}(t) = \mathcal{H}_{\text{Mag}^{\{2,4\}}}(t) = t + \sum_{n \geq 2} 2^{n-2} t^n$$

# III) THÉORIE DES CODES

## UN PROBLÈME DE COMMUNICATION



Lettre	Encodage
$\alpha$	11
$\beta$	10001
$\gamma$	01
$\vdots$	$\vdots$

**Problème** : la trame doit avoir une unique décomposition !

### III) CODE

#### Définition

L'ensemble  $\mathcal{X} \subset \mathcal{A}^*$  est un **code** si et seulement si pour tout

$$x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m \in \mathcal{X}$$

tels que

$$x_1 x_2 \dots x_n = y_1 y_2 \dots y_m,$$

on a

$$n = m \text{ et } x_i = y_i, \text{ pour tout } i \in [1, n].$$

#### Exemple

L'ensemble  $\{aabb, abaaa, b, ba\}$  n'est pas un code car

$$babaaabb = (b)(abaaa)(b)(b) = (ba)(ba)(aabb).$$

### III) CODE PRÉFIXE

#### Définition

L'ensemble  $\mathcal{X} \subset \mathcal{A}^*$  est **préfixe** si aucun élément de  $\mathcal{X}$  n'est le préfixe d'un autre élément de  $\mathcal{X}$ .

#### Exemple

L'ensemble  $\{b, ab, a^2b, a^3b, a^4b, \dots\}$  est un code préfixe.

#### Proposition

Un ensemble préfixe différent de  $\{\varepsilon\}$  est un code.

### III) CONJECTURE COMMUTATIVEMENT PRÉFIXE

#### Définition

L'ensemble  $\mathcal{X} \subset \mathcal{A}^*$  est **commutativement préfixe** s'il existe un code préfixe  $\mathcal{P}$  tel que

$$\{ \{ (|x|_a, |x|_b) : x \in \mathcal{X} \} \} = \{ \{ (|p|_a, |p|_b) : p \in \mathcal{P} \} \}.$$

#### Exemple

L'ensemble

$$\{ a, ba, aabb, baabb, ababb \}$$

est commutativement préfixe, car il est équivalent au code

$$\{ a, ba, bbaa, bbaba, bbbaa \}.$$

#### Conjecture [Schützenberger avant 1965]

Les codes maximaux finis sont commutativement préfixes.

### III) CODE NON-COMMUTATIVEMENT PRÉFIXE

#### Théorème [C.]

L'ensemble  $\mathcal{X} := \{x_1, \dots, x_n\} \subset \mathcal{A}^*$  avec  $|x_1| \leq |x_2| \leq \dots \leq |x_n|$  est commutativement préfixe si et seulement si

$$\sum_{i=1}^j \binom{|x_j| - |x_i|}{|x_j|_a - |x_i|_a} \leq \binom{|x_j|}{|x_j|_a},$$

pour tout  $j$  tel que  $1 \leq j \leq n$ .

#### Exploration informatique [C., Rao]

Il n'existe pas de code non commutativement préfixe dont les mots sont de longueurs au plus 6.

### III) CODE NON-COMMUTATIVEMENT PRÉFIXE

#### BAÏONNETTE

#### Exemple [Shor en 1984]

Le code  $\left\{ \begin{array}{ccccc} b, & ba, & ba^7, & ba^{13}, & ba^{14}, \\ a^3b, & a^3ba^2, & a^3ba^4, & a^3ba^6, & \\ a^8b, & a^8ba^2, & a^8ba^4, & a^8ba^6, & \\ a^{11}b, & a^{11}ba, & a^{11}ba^2 & & \end{array} \right\}$

est non-commutativement préfixe.

#### Exploration informatique [C. en 2019]

69 nouveaux codes baïonnettes non-commutativement préfixes dont

$\{ b, ba^2, ba^8, ba^{10}, aba^8, aba^{10}, a^4b, a^4ba^2, a^5b, a^5ba^3, a^5ba^6, a^9b, a^9ba^2 \}$ .

### III) LE RATIO DE SHOR

Problème [Shor en 1984]

Quelle est la valeur maximale de  $\frac{|\mathcal{X}|}{n}$ , où  $\mathcal{X} \subset a^*ba^* \cap \mathcal{A}^{\leq n}$  est un code ?

Réponse partielle de Shor, C. et Hansel

Cette valeur est comprise entre  $\frac{13}{12}$  et  $1 + \frac{1}{\sqrt{2}}$ .

### III) NON INCLUSION DANS UN CODE MAXIMAL FINI

#### Théorème [Restivo en 1977]

Le code

$$\{aaaaa, ab, b, baa\}$$

n'est pas inclus dans un code maximal fini.

#### Conjecture [C.]

Le code

$$\{aa, abab, baaa, b\}$$

est le plus petit code non inclus dans un code maximal fini.

### III) INCLUSION DANS UN CODE MAXIMAL FINI ?

#### CONDITION NÉCESSAIRE

#### Propriété

Si  $\mathcal{X}$  est un code maximal fini alors il existe  $n$  tel que  $a^n \in \mathcal{X}$ .

#### Corollaire d'un théorème de Perrin et Schützenberger [1977]

Si  $\mathcal{Y} \subseteq a^*ba^*$  est inclus dans un code maximal fini, alors il est inclus dans un code  $\mathcal{X} \subseteq a^{<n}ba^{<n}$  tel que

$$\{a^n\} \cup \mathcal{X} \text{ soit un code et } |\mathcal{X}| = n.$$

#### Exploration informatique [C. en 2019]

Aucun des 70 codes non-commutativement préfixes ne vérifient cette condition pour  $n \leq 32$ .

### III) INCLUSION DANS UN CODE MAXIMAL FINI ?

#### BORNES INFÉRIEURES

#### Théorème [C. en 2019]

Pour chacun des 70 codes,  $n$  est nécessairement supérieur à 33 et pour 20 d'entre eux :

Nombre de codes	$n$
2	$30k$ , où $k \geq 3$
6	$66k$ , où $k \geq 3$
2	$330k$ , où $k \geq 4$
4	$390k$ , où $k \geq 4$
4	$390k$ , où $k \geq 3$
2	$130k$ , où $k \geq 3$

### III) NON INCLUSION DANS UN CODE MAXIMAL FINI

Théorème [C. en 2019]

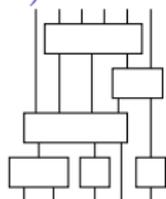
Le code

$$\left\{ ba, ba^3, ba^9, ba^{11}, aba^8, aba^{10}, a^4b, a^4ba^2, a^5b, a^5ba^3, \right. \\ \left. a^5ba^6, a^9b, a^9ba^2 \right\} \cup \left\{ a^{16} \right\}$$

est non-commutativement préfixe et non inclus dans un code maximal fini.

# CONCLUSION ET PERSPECTIVES

## I) CIRCUITS

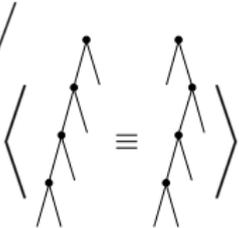


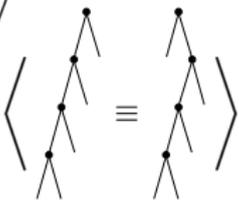
Nature (rationnelle, algébrique, holonome, etc) des séries génératrices des circuits ?

Pistes : GFUN, équations fonctionnelles et singularités.

## II) OPÉRADES

Étudier les quotients d'arbres à 4 nœuds.

Exemple :  $\text{Mag}/$   donne les *animaux dirigés* [A005773].



## III) CODES

Rechercher des codes non-commutativement préfixes de la forme

$$\left\{ \omega \in \{a, b\}^*, |\omega|_b \leq 2 \right\}.$$