

II.

"A chaque période de croissance, tous les rameaux essaient de pousser des branches de toutes parts, de dépasser et de tuer les rameaux et les branches environnantes, de la même façon que les espèces et les groupes d'espèces ont, dans tous les temps, vaincu d'autres espèces dans la grande lutte pour l'existence."

Charles Darwin (*L'origine des espèces*, 1859)



Magma et Arborescences.

1. Le Magma binaire monogène.

Définition 1:

Nous appellerons **Magma binaire** monogène, engendré par l'unique générateur x , sur l'alphabet de quatre lettres, ou caractères typographiques

$x (,)$

le langage M des expressions définies par la récurrence suivante:

(1) si les mots m' et m'' appartiennent à M alors, le mot (m', m'') appartient à M

récurrence initialisée par:

(i) le mot x réduit à la seule lettre x appartient à M . \

Ainsi, les mots que voici, séparés par le caractère \,

$x \setminus (x,x) \setminus (x,(x,x)) \setminus ((x,x),x) \setminus ((x,x),(x,x)) \setminus (x,(x,(x,x))) \setminus ((x,(x,x)),x) \dots,$

sont les premiers que l'on obtienne par application raisonnable des définitions ci-dessus.

Représentatifs d'éléments du Magma engendré par le seul générateur x

(le magma est la structure algébrique la plus élémentaire qui soit, simple composition binaire)

ce sont par ailleurs des mots du monoïde engendré par l'alphabet constitué des quatre lettres: $x (,)$

Ce langage est défini par la récurrence générative **(1)** et son initialisation **(i)**.

Ces mots seront pour nous la représentation canonique du Magma engendré par x , et ces expressions seront considérées comme des représentations linguistiques (càd comme suite de caractères) d'arbres binaires (au sens d'arbres dessinés et pointés); nous avons ici la définition axiomatique-algébrique de la notion d'arbre binaire, au sens habituel des programmeurs.

Autrement dit...

x est un arbre binaire,

et tout couple d'arbres binaires est un arbre binaire.

Un allègement du système de notation permet abusivement (car l'abus est permis) de les écrire sous la forme suivante: $x, x^2, xx^2, x^2x, (x^2)^2, xxx^2, (xx^2)x, x^2(xx^2), xx^2x^2, (xx^2)^2 \dots$

Exercice 1:

L'énumération pourrait se poursuivre, en composant par exemple chaque objet, à gauche, puis à droite, avec tous ceux qui le précèdent, puis avec lui-même, et en écrivant ces nouveaux composés à la suite. Programmer ce mode d'énumération exhaustive.
Comparer l'utilisation de divers langages de programmation.

On construit ainsi une bijection des couples d'entiers sur les entiers privés de zéro (et finalement sur les entiers, si l'on veut): l'objet de rang n est composé d'objets de rangs i et j ; x est l'objet de rang 0.

Nous verrons par la suite, dans un long exercice, que ce mode de construction énumère les arbres binaires suivant leur profondeur (dont la définition sera donnée ultérieurement).

Préliminaire: *construire* d'abord les couples d'entiers, soit une bijection f de \mathbf{N}^2 sur l'alphabet \mathbf{N} (càd des mots de degré 2 sur ceux de degré 1). Par exemple,

0 1 2 3 4 5 6 7 8 9 ...
0 00 01 10 11 02 12 20 21 22 ...

Avant tout, établir que l'image du mot ij , c'est l'entier $(i+1)^2$;
pour i distinct de j , si $k=\max(i,j)$, on a $k^2 < f(ij) < (k+1)^2$:

si i diffère de k , alors $f(i,j)=k^2+i+1$,
sinon $f(i,j)=k^2+k+j+1=(k+1)^2-k+j$, valable aussi pour $j=k$.

Tout entier étant repéré par son écart au carré le plus voisin, écrire un programme pour calculer f^{-1} , puis un programme pour calculer un arbre binaire de rang donné.

Il peut être utile de procéder selon un ordre un peu différent:

0 00 01 10 11 02 12 21 20 22 ...

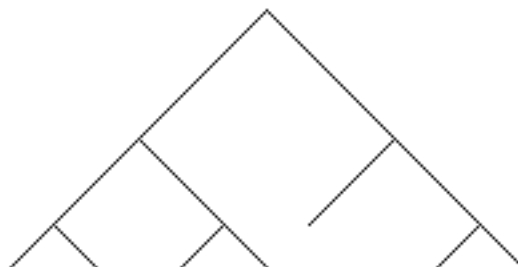
tel que $ij < i'j'$ lorsque
- $\max(i,j) < \max(i',j')$,
-ou $\max(i,j) = \max(i',j')$ et $i < i'$
-ou $\max(i,j) = \max(i',j') = i = i'$ et $j > j'$.

L'ordre suivant

0 00 01 10 11 02 20 12 21 22 03 30 13 31 23 32 ...

possède également, en ce qui concerne les arbres, des propriétés intéressantes.

Voici une représentation dessinée de l'arborescence binaire codée par l'expression magmatique $((x,x),(x,x)),(x,(x,x))$:



Clairement, les expressions magmatiques sont la libre représentation, à partir d'un unique générateur, x , de ce que produit une loi de composition binaire, ni commutative, ni associative.
A ce stade, il n'a pas paru nécessaire d'introduire un élément neutre.

**Un modèle exemplaire de croissance d'arbre,
expansion libre et aléatoire,**
modèle sur lequel nous allons nous arrêter longuement,
objet tant de simulation programmée que d'étude théorique.

Construire (par programme) un arbre binaire défini par le processus aléatoire suivant: soit une probabilité p ("petit p ", nombre réel compris entre 0 et 1), et un mot m du langage M des arbres binaires ci-dessus défini.

Pour *chaque* occurrence de x dans m , soit $1-p$ la probabilité de substitution de (x,x) à x , et donc **p la probabilité de non-substitution**. Nous dirons que p est la probabilité *d'arrêt local*.

Partir de x , et itérer le procédé; programmer.

On s'intéresse à la probabilité P ("grand P ") *d'arrêt global* d'un tel processus de croissance aléatoire d'un arbre binaire. On pourra dire que lorsque sommet x est vivant (assimilons le à un être unicellulaire, une paramécie) il se reproduit par cissiparité avec la probabilité $1-p$, et que la branche de ses descendants se meurt avec la probabilité p .

Ebauche d'étude expérimentale:

Voici un squelette de programme minimum, afin de tester l'éventuel arrêt, expérimentalement (on examine, à chaque unité de temps, le statut d'une quelconque occurrence de x , et l'on compte à chaque instant le nombre d'occurrences de x dont le statut n'a pas été examiné): aléa est une procédure pseudo-aléatoire supposée fournir un nombre entre 0 et 1.

Les machines sont usuellement pourvues de telles procédures. Sans nous interroger ici sur la nature de ce nombre, ni sur ce que peut être la procédure aléa (on peut en imaginer de quantité de manières, et il conviendrait d'étudier le comportement de celle fournie avec votre machine...), en s'intéressant seulement à l'arrêt du processus de croissance, le programme se résume essentiellement à ceci:

Initialisation $n:=1$.

Itération Tant que $n>0$, $n:=n + (-1$ si aléa $<p$, 1 sinon)\

Le symbole $:=$ est l'usuel opérateur de substitution.

Le nombre d'appels à aléa, c'est évidemment le nombre de sommets de l'arbre construit, puisque le statut de tout sommet est examiné une fois et une seule. La capacité de n est limitée, et un éventuel dépassement de capacité provoque l'arrêt du programme; dans un tel cas, la croissance eut continué si la capacité avait été supérieure, mais la probabilité d'arrêt en un temps fini eut été très faible; en pratique, stopper le programme au terme d'un temps t donné que nous ne préciserons pas.

Qu'en est-il de la probabilité d'obtention d'un arbre donné? Que vaut cette fonction, sommée sur tous les arbres? Ou, plutôt: quelle est la probabilité pour que le processus ne s'arrête pas?

Faire d'une part une étude théorique, et d'autre part une étude expérimentale programmée (les résultats expérimentaux sont-ils conformes aux prévisions théoriques?), en utilisant les fonctions pseudo-aléatoires habituellement fournies par les langages implémentés, lesquelles donnent des nombres rationnels particuliers, généralement compris entre 0 et 1 (la procédure aléatoire est-elle fiable ?).

Une question plus compliquée: peut-on parler de délai moyen d'arrêt.

Deux versions de ce problème: 1)-les substitutions sont simultanées (synchrones) 2)-une seule substitution a lieu à chaque unité de temps [mais le statut de tout sommet est examiné en un temps fini lorsque le processus de croissance a une fin].

Envisager que la probabilité p ne soit pas fixe, mais modifiable;

p étant pris aléatoirement dans l'intervalle $[0,1]$, quelle est la probabilité de croissance indéfinie ?

Petite étude théorique:

si p est la probabilité de non-substitution de x^2 à x (c'est notamment la probabilité d'arrêt dès le départ), la probabilité de substitution est $(1-p)$, et la probabilité P d'arrêt à terme est contrainte par

$$(*) P=p+(1-p)P^2,$$

exprimant clairement qu'un arrêt à terme est
 -soit un arrêt dès le départ,
 -soit un arrêt à terme des deux branches engendrées au départ.

P est un polynôme en p dont 1 est racine. L'autre racine est $p/(1-p)$, qui est le produit des racines.
 Pour $p=1/2$, les deux racines sont égales, soit $P(1/2)=1$, et, à fortiori $p>1/2$ entraîne $P(p)=1$,
 tandis que $P=p/(1-p)$ pour $p<1/2$
 En effet, P est nul si p l'est, P vaut 1 si p vaut 1, et l'on postule que P dépend continûment de p.

Pour $p=1/(k+1)$, on a $P=1/k$; la portion de courbe $P(p)$ pour p variant de 0 à 1/2 est une portion d'hyperbole; on vérifiera que la tangente en 0 est de pente 1 et celle en 1/2 de pente 3.
 D'ailleurs, l'équation de l'hyperbole $P=p/(1-p)$ s'écrit aussi bien $(1-p)(P+1)=1$, mettant évidence les deux asymptotes: $p=1$ et $P=-1$. Pour p entre 1/2 et 1, P est constant, de valeur 1.
 Le résultat le plus important est que *pour toute valeur de p comprise entre 1/2 et 1*, la probabilité que la croissance cesse en un temps fini est 1. On peut cependant remarquer intuitivement, après calcul fait, que pour $p=1/2$, le nombre des sommets vivants (et terminaux) de tout arbre a, naturellement, une tendance moyenne à se maintenir.

Afin de ne pas faire une erreur si commune, il est important de remarquer que *la croissance indéfinie, pour p supérieur ou égal à 1/2, n'est pas impossible ; elle est seulement de probabilité nulle.*
 C'est qu'un événement de probabilité nulle peut se réaliser, tandis que la réalisation d'un événement de probabilité 1 n'est pas certaine.

Soit c_n le nombre d'expressions du magma possédant n occurrences de la lettre x.
 Nous allons évaluer le rapport P/p de la probabilité P d'arrêt global à la probabilité p d'arrêt local.

La probabilité d'obtenir un arbre donné de degré n est $p^n(1-p)^{n-1}$, lors d'une telle expérience virtuelle de croissance aléatoire, et la probabilité d'obtenir un quelconque arbre de degré n est $p^n(1-p)^{n-1}c_n$.

On peut donc prévoir la somme (pour $n>0$) de la série numérique de terme général $c_n p^n (1-p)^{n-1}$:
 cette série a pour somme $P/p=1/(1-p)$ pour $0<p<1/2$, et $1/p$ pour $1/2<p<1$. Soit $P/p=1/\max(p, 1-p)$.
 En posant $q[p]=p(1-p)$, ou $q[p]=q[1-p]$, nul pour $p=0$ ou 1, maximum pour $p=1/2$ ($q[1/2]=1/4$).
 Le maximum du rapport P/p est ainsi obtenu pour $p=1/2$, et vaut 2.

Soit la série numérique $s=pP$. On remarquera que $s=q+s^2$.

A partir d'un objet initial x, nous avons initié l'étude de ce phénomène aléatoire itéré: x est remplacé par (x,x) avec la probabilité 1/2. L'étude mériterait d'être complétée avec non pas un objet initial x, mais avec une collection (x_1, \dots, x_n) de n objets initiaux, tout objet x_i étant remplacé, de manière itérative, par le couple (x_i, x_i) avec une probabilité donnée. Ces modèles sommaires sont, entre autre, introductifs aux modèles de la reproduction des gènes chez les êtres vivants.

Epidémiologie :

Nous avons là un modèle, des plus rudimentaires, mais très instructif, simulant une propagation épidémiologique. Il suggère notamment qu'il n'est généralement pas nécessaire de vacciner la totalité d'une population pour empêcher le déclenchement d'une épidémie, ce que l'expérience confirme.

Les épidémies, dont les agents sont souvent spécifiques d'une espèce (les infections dues aux diverses mycobactéries, par exemple la tuberculose d'origine bovine, et la lèpre, spécifiquement humaine, dont les agents sont voisins), les épidémies sont souvent induites par un enchaînement fragile de circonstances favorables: pour les enrayer, il suffit d'abaisser d'une manière ou d'une autre la probabilité de contagion en dessous d'un certain seuil (mais, du fait de l'existence de réservoirs quiescents d'agents pathogènes, ça ne provoque pas leur éradication). Ainsi, par exemple, il est estimé que la diphtérie disparaît rapidement lorsque l'on vaccine seulement les deux tiers de la population.

[A propos de méthodologie en matière d'études statistiques sur des populations, voir l'exposé classique, très critique et exemplaire de Joachim Marcus-Steiff (Pourquoi faire simple quand on peut faire compliqué ? Les Temps Modernes N°482, septembre 1986, pp. 1-50)]

Le cas de la variole a pris une importance particulière, du fait de son éradication par un programme mondial, du fait que l'immunité de la population est désormais très faible et que la réapparition de l'agent causal pourrait provoquer des ravages considérables, du fait qu'il en existe des réservoirs, connus ou occultes, du fait des modes de contamination, du fait des délais d'incubation, du fait de diagnostics nécessairement tardifs (avec des cas rares et importés, les contaminations auraient surtout lieu en milieu hospitalier, comme cela s'est produit en Europe entre 1958 et 1973). On a étudié divers exemples historiques de la dynamique des poussées majeures récurrentes (parfois tous les deux ou trois ans) d'une maladie qui était endémique, maintenant un certain niveau d'immunité (sans compter les vieilles pratiques de vaccins). Nous disposons des judicieux commentaires de Bernoulli sur l'épidémie de 1776 à Paris. On a dépouillé les données disponibles sur les poussées du dix-huitième siècle aux Etats-Unis (Boston 1721) et en Angleterre (Burford 1758, Chester 1774, Warrington 1773), et plus près de nous en Yougoslavie en 1972. On a pu évaluer le nombre de personnes contaminées par chaque personne atteinte, lors de la phase d'expansion de chacune de ces poussées épidémiques; ce nombre était toujours compris entre 3 et 6; on estime qu'il serait de 11 dans une population vaccinée à cinquante pour cent (Nature, vol 414, déc. 2001, pp. 748-751); la question est: quel pourcentage de la population faut-il vacciner, et quelles mesures faut-il prendre pour que ce coefficient descende en dessous de 1 ?

Il y a une écologie de l'agent parasite et de son hôte. La probabilité de contagion est un facteur observable de la sélection naturelle, agissant à la fois sur l'hôte et le parasite. Les modifications d'environnement et de comportement agissent sur le parasite, lequel agit en retour sur l'hôte. Ainsi observe-t-on d'une part la tendance des maladies aiguës à se transformer en maladies chroniques, et d'autre part l'accélération de l'apparition de pathologies nouvelles lorsque les conditions environnementales se modifient. Voir les ouvrages du regretté Mirko Grmek, et les écrits plus anciens du pasteur Charles Nicolle sur le destin des maladies infectieuses.

Désormais, les tentatives de simulations programmées se multiplient lorsque la théorie est muette.

Du simple exercice ci-dessus relatif à la croissance aléatoire d'arbres binaires, on tire deux enseignements fondamentaux: a) il est impossible de saisir un phénomène, fut-il simple, par la seule intuition; b) une variation minime d'un paramètre peut avoir des conséquences considérables, comme le montre ici le voisinage de $p=1/2$; pratiquement, d'un côté l'arbre ne croît pas, de l'autre, dès qu'il commence à se développer tant soit peu, il ne cesse pratiquement pas de croître.

Corrélation n'est pas raison: le virus Ebola infecte principalement des pays francophones; disparaîtrait-il si l'on cessait d'y parler le français?



Comment dessiner sur écran de tels arbres binaires, de croissance aléatoire -en ne définissant que des segments horizontaux ou verticaux de longueur entière ? Ce n'est pas aisé. Bien préciser la méthode; modifier l'affichage de l'arbre lorsque celui-ci croît.

Pour une probabilité p donnée, observer expérimentalement (càd. en itérant les expériences programmées) la taille des arbres obtenus. Calculons la valeur théorique de *la taille moyenne de ces arbres*. La taille d'un arbre est nombre d'occurrences de x et du caractère " $,$ " dans l'expression magmatique, soit $2i-1$ si i est le degré en x . Mesurée en nombre de sommets, la taille T d'un arbre de degré i est $2i-1$; un tel arbre compte i sommets terminaux et $i-1$ sommets internes.

La taille moyenne T est $1/(2p-1)$, pour p entre $1/2$ et 1 :

-en effet, la taille moyenne T doit satisfaire à $T=p+(1-p)(1+2T)$,

-car un arrêt initial donne un arbre de taille 1 avec la probabilité p ,
 -et un arrêt à terme un arbre de taille moyenne $1+2T$ avec la probabilité $(1-p)$.
 Par exemple, si $p=1/2+1/20$, on a $T=10$. Avec $p=1/2+1/10$, $T=5$. Avec $p=1/2+1/4$, $T=2$.
 Comparer à ce que donne l'expérimentation programmée.

Accorder une grande importance à l'examen de la valeur charnière critique: pour $p=1/2$, la probabilité d'arrêt est 1, alors que la taille moyenne de l'arbre est infinie.

Comme pour $p < 1/2$ la probabilité d'arrêt P est inférieure à 1, on convient, pour l'expérimentation, que la probabilité d'arrêt devient négligeable lorsque l'arrêt tarde trop à se produire.
 Observer l'instabilité de l'expérimentation programmée au voisinage de $p=1/2$.

La taille T d'un arbre est le nombre moyen de ses sommets, fonction calculable pour $p > 1/2$.
 Mais qu'en est-il de sa profondeur moyenne ($pr(a,b)=1+\max(pr(a),pr(b))$) sous $pr(x)=0$?

La profondeur est en quelque sorte le nombre de générations avant dépérissement, dans le cas où les sommets vivants, appartenant à une génération donnée, se reproduisent ou se figent simultanément, càd de manière synchrone.

En termes épidémiques, la taille est le volume de la population atteinte, tandis que la profondeur est le temps que dure l'épidémie, càd le nombre de générations affectées.

A chaque taille et à chaque profondeur correspondent des probabilités.

Compter le nombre $f(n)$ d'arbres de profondeur exactement n . Si $F(n)$ est le nombre des arbres de profondeur au plus n , on a $f(n+1)=2F(n-1)f(n)+f(n)^2$, sous $f(0)=F(0)=1$, $f(1)=1$.

Le tout début de la suite f : 1 1 3 21 651 457.653 210.065.930.571....
 et le début de la suite F : 1 2 5 26 677 458.330 210.066.388.901....

La suite F se calcule aisément. F satisfait à la récurrence $F(n+1)=1+F(n)^2$. Et $f(n)=F(n)-F(n-1)$.

Il y a déjà plus de deux cent dix milliards d'arbres binaires de profondeur six....!!!

Il est plus difficile de déterminer la profondeur moyenne des arbres de degré i .

Nous formalisons plus loin les arbres au sens général, dont les arbres binaires font partie.

Qu'en est-il avec des arbres homogènes (d'ordre n : on substitue x^n à x avec la probabilité p), au lieu d'arbres binaires ? Montrer que pour $1 > p > (n-1)/n$, on a $P=1$, et que pour $0 < p < (n-1)/n$, P est soumis à $p^{n-1}+p^{n-2}... +P = p/(1-p)$.

Indications: Pour $p=1$, on a $P=1$, qui est une des racines du polynôme $0=p-P+(1-p)P^n$.

La taille moyenne T de l'arbre aléatoirement engendré, supposée dépendre continuellement de p , vaut 1 lorsque $p=1$, et est contrainte au voisinage de $p=1$ par $T=p+(1-p)(1+nT)$, soit $T=1/(np-(n-1))$, taille moyenne bien définie pour p supérieur à $(n-1)/n$, mais infinie pour p égal à $(n-1)/n$. Il en résulte que lorsque p est inférieur à cette valeur critique, $(n-1)/n$, P prend une valeur inférieure à 1: ainsi, la probabilité d'arrêt en un temps fini est une racine distincte de 1 du polynôme qui contraint P : diviser ce polynôme par $P-1$ pour obtenir le polynôme de degré $n-1$ qui contraint P dans l'intervalle $0 < p < (n-1)/n$.

On vérifie au passage que le polynôme $P(p)$ qui contraint p peut s'écrire $(1-p)(P-P^n)=p(1-P)$.

Une question demeure: pour p entre 0 et $(n-1)/n$, on sait que P varie en croissant de 0 à 1, mais quel est son graphe, quelle racine du polynôme qui contraint P suit-on ? Pour p entre 0 et $(n-1)/n$, ce polynôme possède-t-il une seule racine située entre 0 et 1 ?

Programmation et représentation graphique de l'arbre obtenu: expérimenter tout d'abord le cas $n=3$.

On peut représenter de tels arbres (dits ternaires) de la manière fractalisée suivante:

-partir d'un triangle équilatéral initial, T ;

-si l'on ôte le triangle équilatéral T' , inscrit dans T , déterminé en joignant les milieux des côtés de T , (opération d'évidement de T), il reste alors trois triangles équilatéraux, lesquels sont incidents aux trois sommets de T : chacun de ces trois triangles *peut* subir un évidement comparable, et le processus *peut* être itéré.

Si l'on évide *simultanément* les trois triangles obtenus à partir de T , puis à nouveau simultanément les triangles évidables, et que l'on itère le processus k fois, combien de triangles la figure obtenue comporte-t-elle ? Cette fonction $f(k)$ [somme des puissances de 3] satisfait-elle à la récurrence $f(k+1)=4f(k)-3f(k-1)$, sous $f(-1)=0$, $f(0)=1$?

Le cas $n=4$ est spécialement intéressant: il se représente plus aisément sur un écran: segmenter ou non un carré de côté 2^{i+1} en quatre carrés de côtés 2^i , et itérer ou non sur chaque partie; stopper au niveau du pixel, et afficher.

Etudier ce qui se produit avec des arbres quelconques, non nécessairement homogènes (en dessous de tout sommet, il y a un nombre quelconque de sommets) la probabilité d'avoir k sommets sous un sommet donné étant (par exemple), pour $k>0$, $p/2^k$.

La situation est un peu plus complexe lorsque la probabilité $(1-p)$ de développement d'une branche gauche est distincte de la probabilité $(1-q)$ de développement d'une branche droite: soit donc $(1-p)$ la probabilité de substitution de (x,y) à x , et $(1-q)$ la probabilité de substitution de (x,y) à y .

Partant de x , y , ou (x,y) ,

- 1) écrire un programme de simulation,
- 2) quelle est la probabilité $P(p,q)$ d'arrêt en temps fini,
- 3) quels sont les points (p,q) tels que $P(p,q)=1$?

Dans le même ordre d'idées: sur un alphabet de deux lettres, (x,y) , la probabilité de produire x étant p , celle de produire y étant $1-p$, quelle est la probabilité d'obtenir un mot m de degré n comportant plus d'occurrences de x que d'occurrences de y (problème de Bernouilli -mathématicien suisse, seconde moitié du 17^esiècle).

Enfin, si tout objet disparaît sans aucune descendance avec la probabilité p , $0<p<1$, et produit (statistiquement, soit en moyenne, au sens de la reproduction des populations) $-k$ descendants, avec la probabilité $1-p$, où k est un nombre rationnel ou réel, positif et supérieur à 1, $-et 0$ descendants avec la probabilité p , $-on a pour la probabilité P d'arrêt à terme -càd en un nombre fini de générations- partant d'un unique objet initial,$

$$P=1.p+(1-p).P^k, \text{ et pour la taille moyenne de la population ainsi développée, } T=1.p+(1-p).(1+kT)$$

Nous sommes intéressés à établir que l'on a $p.(1-P)=(1-p).(P-P^k)$ lorsque $p<(k-1)/k$, et $P=1$ sinon.

Il est beaucoup plus difficile d'évaluer par des voies théoriques le temps moyen de l'expérience, avant l'extinction du phénomène de croissance aléatoire. D'où l'utilité des simulations programmées, expérimentations, commodités, dans les cas, si nombreux, où la théorie à quelque difficulté à opérer. Le produit $(1-p)^k$ est le taux de reproduction; le point critique $p=(k-1)/k$, valeur de p pour laquelle la probabilité d'arrêt à terme est 1 -tout en étant compatible avec une population moyenne infinie, càd un nombre de générations moyen infini- ce point critique correspond à un taux de reproduction qui est lui aussi égal à 1; la simulation programmée au voisinage de ce point est instable, sensible aux erreurs de troncature et arrondi.

Dans tous les cas, la probabilité P_{j+1} que la croissance cesse en $j+1$ générations au plus est

$$P_{j+1}=p+(1-p)P_j^k, \text{ sous } P_1=p,$$

et en $j+1$ générations exactement, $P_{j+1}-P_j$.

Naturellement, P s'approche (numériquement) par l'itération $P:=p+(1-p).P^k$, sous l'initialisation $P:=p$. On observe le ralentissement de la convergence au voisinage du point critique $p=(k-1)/k$.

Il n'est pas indigne de se livrer à ce calcul numérique itératif.

Ainsi, pour $k=2$ et $p=1/2$, les dix premières valeurs de la suite P_j sont

0.5 0.625 0.695313 0.74173 0.775082 0.800376 0.820301 0.836447 0.849821 0.861098

tandis que pour $p=1/4$ on obtient

0.25 0.296875 0.316101 0.32494 0.329189 0.331274 0.332307 0.332821 0.333077 0.333205

et pour $p=3/4$

0.75 0.890625 0.948303 0.97482 0.987568 0.993823 0.996921 0.998463 0.999232 0.999616

Par exemple, si l'on avait $P=p+qP+rP^2+sP^3$ (avec $p+q+r+s=1$) on aurait la taille $T=p+qT+2rT+3sT$, soit $T=p/(1-q-2r-3s)$, critique pour $q+2r+3s=1$. Et $P=1$ pour $q+2r+3s<1$, $P<1$ lorsque $q+2r+3s>1$.

La série $P=$ somme, pour i non-négatif, des $p(i)P^i$, $p(i)$ non négatifs, famille sommable de somme 1, conduit à la taille $T=p(0)+$ somme, pour i positif, des $i*p(i)*T$; soit $T=p(0)/(1-\text{somme des } i*p(i), i>0)$, situation critique (càd $P=1$ et taille moyenne T infinie) lorsque $1=\text{somme, pour } i \text{ positif, des } i*p(i)$.

Ensuite, expérimenter la **propagation aléatoire non libre sur un graphe infini**, spécialement

-sur le plan quadrillé, $\mathbb{Z} \times \mathbb{Z}$, sur lequel tout point a quatre voisins (nord, sud, est, ouest)
 -ou huit voisins (ajouter les voisins nord-ouest, sud-est, nord-est, sud-ouest),
 -puis sur le plan régulièrement pavé par des hexagones (tout point a trois voisins)
 -ou sur le plan pavé par des triangles (tout point a six voisins).

-sur $\mathbb{N} \times \mathbb{N}$, le point (i,j) , allumé au temps t , provoquant avec la probabilité q l'allumage de $(i+1,j)$ (resp $(i,j+1)$) au temps $t+1$, $p=(1-q)^2$ est la probabilité du point (i,j) de demeurer sans descendant; initialement, seul $(0,0)$ est allumé.

La probabilité $f(i,j)$ que le point (i,j) soit allumé au temps $i+j$ est contrainte par la récurrence $f(i,j)=q*(f(i-1,j)+f(i,j-1))-q^2*f(i-1,j)*f(i,j-1)$, sous l'initialisation $f(0,0)=1$, $f(0,n)=q^n=f(n,0)$.

Soit donc p la probabilité de n'allumer aucun voisin, P la probabilité d'arrêt au terme d'un temps fini. La valeur critique de p est la plus petite valeur pour laquelle on a $P=1$.

Peut-on la déduire de considérations théoriques, ou seule une simulation est-elle capable d'en fournir une approximation ?

Une fois les bornes dépassées, il n'y a plus de limite. (Sapeur Camenbert) .

Exercice 1:

A.- Soit n un entier positif, et le phénomène aléatoire engendré par la procédure suivante (exercice typique très exemplaire correspondant à une boucle de programme -dans lequel la borne de la boucle est évaluée à chaque cycle, et dans lequel on touche à l'indice de boucle à l'intérieur même de la boucle):

**programme 1: initialement $i:=n$, puis
 "tant que $i>0$,
 substituer à i un entier de l'intervalle $[0,i-1]$ choisi aléatoirement de manière équiprobable"**

Evidemment, n est la racine d'un arbre dont les branches sont les arbres de racines 0 à $n-1$. Soit, en notation magmatique polynômiale, comme nous le verrons (anticipons):

$$A(n)=(x, \text{ suite des } (A(i)) \text{ pour } i \text{ de } 0 \text{ à } n-1), A(0)=x.
 A(1)=(x), A(2)=(x,(x)), A(3)=(x,(x),(x,(x)))...$$

Quel est notamment le temps moyen t d'arrivée en 0 , càd la profondeur moyenne de l'arbre ?

Nous avons $t(0)=0$,
 et $t(n)=1+(\text{somme des } t(k) \text{ pour } k \text{ de } 0 \text{ à } n-1)/n=1+\text{moyenne des valeurs qui précèdent}$;

On montrera que $t(n)=f(n)/n!$, où $f(n)$ =nombre de stirling de première espèce comptant le nombre de

permutations de $[n+1]$ comportant deux cycles, $f(n+1)=(n+1)*f(n)+n!$; pour $n>0$, $f= 1 3 11 50 274...$
 On a $f(n) = n!(\text{Somme des } 1/i \text{ pour } i \text{ de } 1 \text{ à } n)$.

Les événements du programme 1
 sont représentés par les mots strictement décroissants sur l'alphabet [n].

B.- Que cela devient-il avec le programme 2:

**initialement $i:=n$, puis
 "tant que $i>0$,
 substituer à i un entier de l'intervalle $[0,i]$, choisi aléatoirement de manière
 équiprobable" ?**

Les événements engendrés par le programme 2 sont représentés par les mots décroissants au sens large
 (càd non croissants) sur l'alphabet [n].

Quelle est la probabilité de non-arrêt en un temps fini ? Vous pouvez lancer la simulation programmée,
 le programme est tellement simple, mais ici encore, l'étude théorique est possible.

Par exemple, partant de 1, la probabilité d'arrêt est $1/2+1/4+1/8....=1$, et donc, le non-arrêt en un
 temps fini, bien que possible, est de probabilité nulle, tandis que le temps moyen $t(1)$ est

$$1.1/2+2.1/4+3.1/8... = \text{somme pour } k>0 \text{ de } k/2^k, \text{ ou encore } t(1)=1+t(1)/2, t(1)=2,$$

ce qui constitue d'ailleurs une démonstration du fait que la série de terme général $i/2^i$ est de somme 2.

Quel est le temps moyen de fonctionnement, partant de la valeur initiale n , sinon

$$t(n)=1+(t(n)+t(n-1)...+t(1))/(n+1)$$

soit $nt(n)=(n+1)+ \text{somme pour } i \text{ de } 1 \text{ à } n-1 \text{ des } t(i) \quad ?$

Si $t(n)=g(n)/n!$, nous avons, pour $n>0$

$$g(n)=2 \ 5 \ 17 \ 74 \ 394 \ 2484 \ 18108 \ 149904 \ 1389456 \ 14257440 \ 160460640 \ 1965444480....$$

$$(g(5)=g(4)+4*g(3)+4*3*g(2)+4*3*2*g(1)+(4!)*6=74+4*17+4*3*5+24*2*24*6=394=5!t(5).)$$

et, surtout, g s'exprime en fonction des nombres de strirling évoqués ci-dessus:

Avec $g(0)=1$, pour $n \geq 0$, $g(n)=f(n)+n!=1 \ 2 \ 5 \ 17 \ 74 \ ... = n! (1+ \text{somme de } 1/i \text{ pour } 1 \leq i \leq n)$,
 et la fonction génératrice de t :

$$t = \text{somme des } t(n).x^n = (1 - \log(1-x)).x^{-1} = 1 + 2*x + 5/2*x^2 + 17/6*x^3 + ...$$

Il serait intéressant d'aller voir de plus près comme cette suite qui intervient dans les groupes de
 Coxeter (Coxeter est une référence obligée pour les groupes définis par générateurs, et relations
 d'équivalence sur les mots, calcul et dénombrement des mots réduits, spécialement à propos des
 groupes engendrés par des involutions, comme on aime tant le faire pour les groupes de permutations;
 le problème se présentera au chapitre sur les permutations).

\

Nous appellerons **Magma binaire monogène bicolore**, engendré par l'unique générateur **x**, sur l'alphabet de cinq lettres, ou caractères typographiques

x (, ;)

le langage **Mb** des expressions définies par la récurrence suivante:

**si les mots m' et m'' appartiennent à Mb
alors, les mots (m', m'') et (m'; m'') appartiennent à Mb**

récurrence initialisée par:

le mot x réduit à la seule lettre x appartient à Mb.

Ici, un arbre binaire est

soit x, soit un triplet (g,u,d) où g est le facteur gauche, d le facteur droit, et u la racine.

Nous appellerons **Magma binaire monogène multicolore Mm** de couleurs y dans Y ce que l'on obtient à partir de x par composition (**m'y m''**), où y est dans Y et m et m' dans **Mm**.

C'est dire que x est l'unique générateur, et que Y est une collection d'opérateurs de composition binaire non associative: si Y a pour cardinalité k, on obtient $c(n)k^{n-1}$ objets de degré n. Si maintenant x appartient à un alphabet d'atomes X de cardinalité j, nous voici avec $c(n)j^nk^{n-1}$ objets de degré n, et si les atomes sont tous distincts de 1 à n, $|Y|=2$, $c(n).2^{n-1}.n!$ objets..... etc...

\

Un duel aléatoire très instructif.-

En matière d'événements aléatoires, le premier mouvement, qui est celui de l'intuition irréfléchie, est trompeur. Les bons exemples sont ceux qui éduquent, en mettant en lumière la nécessité d'études rigoureuses. Il faut raisonner, prudemment, mettre en forme avec circonspection, et enfin, calculer. Vous pouvez écrire un petit programme, expérimenter (avec prudence et circonspection), vous ne serez pas toujours en mesure de produire une démonstration étayant vos observations.

Voici un exemple simple, symptomatique, et rigoureux:

- soit une suite de tirages aléatoires, équiprobables, d'événements indépendants issus d'un ensemble alphabétique de deux éléments, disons (0,1) (un tirage à pile ou face équiprobable)
- dans les mots infinis sur (0,1), les facteurs 001 et les facteurs 011 sont de même fréquence
- en outre, le temps d'attente moyen pour voir sortir un facteur de la forme 001 est le même que celui pour voir sortir le facteur 011 (on montrera que c'est l'entier 8).
- cependant, si l'on stoppe la suite des tirages dès la sortie de l'un ou l'autre des facteurs (001, 011), ce sera 001, deux fois sur trois, qui provoquera l'arrêt de l'expérience.

Vous pouvez construire le début de l'arborescence ad hoc, et l'automate fini qui mémorise la situation:

- tant que 1 sort, on ne décolle pas de l'état initial.
- tout commence avec l'apparition d'une occurrence du caractère 0
- alors, quatre situations équiprobables se présentent, en fonction des deux tirages qui suivent cette première occurrence de 0:

01 fait gagner l'un (001),
11 fait gagner l'autre (011), [pour l'instant, il semble qu'il y ait égalité],
 quand à **10**, il ramène à l'état d'attente des deux facteurs qui suivent [c'est un coup pour rien],
mais 00 ruine toute chance de voir apparaître la séquence 011 avant 001: 001 devient certain, à terme.

La probabilité p que le facteur 001 occurre avant 011 satisfait donc à l'équation du premier degré

$$p = 1/4 + 1/4 + p/4, \text{ soit } p = 2/3.$$

Dans un tel duel, l'un des parieurs gagne deux fois plus souvent que l'autre, et l'on en voit bien la raison. Son gain moyen (ce que les probabilistes appellent l'espérance) à chaque coup est d'un tiers de la mise.

Exercice 2 (assez difficile):

Cheminement aléatoire sur l'ensemble \mathbf{Z} des entiers rationnels.

Un point mobile M est en k au temps t , et, de manière équiprobable, en $k-1$ ou $k+1$ au temps $t+1$. On le suppose au point i au temps 0 . Montrer que la probabilité qu'il parvienne au point j en un temps fini est 1 .

Indication: traiter d'abord le cas ($i=1, j=0$)

[on pourra utiliser l'étude sur la croissance aléatoire des arbres binaires, interprétée en termes de cheminement sur l'ensemble N des entiers naturels; bien que la probabilité de retour en 0 soit 1, le délai moyen n'est pas fini].

Si un point mobile M issu d'un point quelconque initial i a la probabilité 1 d'atteindre tout point j , la probabilité qu'il y revienne n fois est toujours 1, quel que soit n : on dira que la probabilité qu'il y revienne indéfiniment est toujours 1.

Evidemment, il ne s'agit pas de certitude: si l'expérience pouvait être réalisée (ce qui n'est pas le cas), rien n'assure que cela se produirait.

Puis, considérer le cas où la probabilité de transition de k vers $k-1$ (resp vers $k+1$) est p (resp $1-p$) [On montrera que le temps moyen de retour de retour en 0 lorsque l'on part de 1, c'est précisément la taille moyenne $T = p/(2p-1)$ de l'arbre binaire aléatoire, pour p entre $1/2$ et 1].

Si l'on se déplace sur \mathbf{Z}^n , au lieu de \mathbf{Z} , l'étude est plus compliquée. Commencer par \mathbf{Z}^2 , et considérer ce qui se passe en projection sur chaque axe, et sur chaque diagonale.

Remarque 1:

En construisant un arbre *généalogique*, à raison de trois générations par siècle, on aboutit à un arbre d'environ un milliard des sommets (2^{30}) en dix siècles. Ce qui excède très largement la population totale de la terre d'il y a mille ans.

Exercice 3 (si facile!... et pourtant!):

Dans l'hypothèse (d'un optimisme dément) où seulement deux donneurs de sang sur mille seraient contaminés par un quelconque agent pathogène, quelle est la probabilité pour qu'un 'pool' (ce que l'on obtient en mélangeant plusieurs lots) de sang ne soit pas contaminé

- lorsque l'on mélange le sang provenant de 5000 donneurs
- lorsque l'on ne mélange que 50 donneurs ?

(*L'affaire du sang, Anne-Marie Casteret, 1992, La découverte, p.148*)

A propos des risques en médecine, des études rationnelles sur les vaccinations, leurs risques éventuels et leurs avantages statistiques, à propos de la politique sanitaire de l'état, recherchez le vieux texte de d'Alembert, datant de 1760, relatif à l'inoculation de ce que l'on appelait "la petite vérole" (ce que l'on appelle maintenant "la variole", qui fut la première maladie probablement éradiquée par l'organisation mondiale de la santé, en attendant la poliomyélite). S'y mêlent raisonnement scientifique probabiliste, inspiré de Bernoulli, et conseils politiques et moraux, mais aussi nécessité de tenir des registres détaillés de surveillance, ce que nous ne faisons à vrai dire pas encore.

Exercice 4 (un magma commutatif):

Considérer les points de la droite (resp. du plan) engendrés à partir de deux (resp.trois) points initiaux a,b (resp.a,b,c) par la récurrence suivante: si les points x et y sont engendrés, leur composition engendre le point noté (x,y), représenté par le milieu du segment d'extrémités x et y. Attention: ce magma n'est pas libre; un même élément peut d'ailleurs s'écrire de plusieurs manières.

Par exemple, en partant de 0 et 1, et en composant chaque rationnel obtenu avec ceux qui précèdent, on a:

0 1 1/2 1/4 3/4 1/8 5/8 3/8 3/8 7/8 5/8 1/2...

et 3/8 est obtenu comme $(1/2+1/4)/2$, et également comme $(0+3/4)/2$,

de telle sorte qu'il faudrait éliminer de la liste tout point déjà obtenu:

0 1 1/2 1/4 3/4 1/8 5/8 3/8 7/8 1/16 9/16 5/16 3/16 7/16 13/16 11/16 15/16...

On obtient ici des nombres rationnels de la forme $k/2^n$ pour k impair. Les obtient-on tous? Observer la suite des numérateurs. On dit que l'ensemble engendré contient 0 et 1, et qu'il est *fermé* pour l'opération demi-somme; c'est la *fermeture* de $[0,1]$ par la demi-somme.

Exercice 5:

Considérer le langage des objets magmatiques binaires engendrés par la suite de suites x, x; x, (x,x), x; x, (x,(x,x)), (x,x), ((x,x),x), x;...

et notamment le nombre d'occurrences de x dans ces objets:

11; 121; 13231; 143525341; 15473857275837451;...

définissant une intéressante suite de fonctions (programmer ça). \

Exercice 6:

Examinez expérimentalement la procédure (nommons la ALEA) de tirage pseudo-aléatoire de votre machine. Supposons qu'elle fournisse un nombre rationnel entre 0 et 1. Par exemple, observer si elle est bien centrée en 1/2 (càd. si elle fournit un nombre comparable de nombres inférieurs et supérieurs à 0.5). Sinon, tentez de la centrer approximativement, par un programme d'apprentissage qui évalue le nombre r tel que l'on obtienne environ autant de nombres inférieurs que de nombres supérieurs à r .

Vous pouvez aussi construire vous-même, d'une infinité de manières, une procédure de tirage pseudo-aléatoire. Suggestion:

x est initialisé par un flottant,
 puis l'on itère l'ensemble des deux instructions:
 $x:=1/x, x:=x-[x]$
 où $[x]$ est la partie entière de x , et "==" l'opérateur d'affectation.

[remarque: observer les valeurs successives que l'on obtient pour $[x]$ en initialisant par exemple par $1+2^{1/2}$; expliquer ce qui se produit] Pour r réel "pris au hasard" entre 0 et 1, établir que la probabilité pour que l'on ait $[1/r]=k$ est $1/(k*(k+1))$.

Si vous le pouvez, avec la même initialisation, comparer les suites obtenues sur diverses machines (avec des mantisses de même longueur). Il peut même se produire que sur une machine donnée, un même programme, initialisé de la même manière, ne donne pas toujours la même suite (autrement dit, le comportement d'une machine à calculer n'apparaît pas obligatoirement déterministe; pourquoi ?).

Pour k donné (demeurer bien en dessous de 2^n , si l'on manipule des mantisses de n caractères binaires !), examiner la suite obtenue par la fonction z de x telle que $y:=k*x, z:=y-[y]$. Est-ce mieux ?

Comment construire (déduire des procédures qui précèdent) une procédure de tirage dans une ensemble de n éléments ? (n pas trop grand, quand même: pas 10 puissance un milliard!).

Voici un petit programme expérimental (CLS vide l'écran, et DRAW trace un trait) utilisant la méthode ci-dessus, *légèrement rectifiée*. L'écran étant de résolution H par L , un tirage est effectué dans les entiers de 0 à L . On affiche la répartition, càd la fréquence obtenue pour chaque entier entre 0 et L .

```
DIM T(L): FOR I=0 TO L:T(I)=0:NEXT
X=RND(1):Im=10000:Cm=100:Ct=0:V=L+1:
WHILE Ct<Cm:Ct=Ct+1:It=0:
  WHILE It<Im:It=It+1:
    X=1/X:X=X-INT(X):
    Y=X*V:Y=Y-INT(Y):Y=INT(Y*V):T(Y)=T(Y)+1:WEND
Tm=T(0):Tn=Tm: FOR I=1 TO L
  IF T(I)<Tn THEN Tn=T(I):ENDIF
  IF T(I)>Tm THEN Tm=T(I):ENDIF : NEXT
CLS : PRINT Ct*Im,(Tm-Tn)/(Tm+Tn):
FOR I=0 TO L: DRAW I,H TO I,H*T(I)/Tm: NEXT:WEND:END
```

On calcule la fréquence maximale (T_m) et la fréquence minimale (T_n), l'ensemble des valeurs se tenant dans une bande de demi-largeur relative $r=(T_m-T_n)/(T_m+T_n)$. Les calculs flottants ont été menés avec des mantisses d'une dizaine d'octets. Avec $L=640$, la valeur que l'on obtient pour r est de l'ordre de 0,5 pour 25.000 tirages, et 0,14 pour 300.000 tirages.

Le diagramme des fréquences est affiché (100 fois) tous les 10.000 tirages (soit au total un million de tirages). Avec un processeur à 8 mégahertz (sans co-processeur) et un programme interprété où tout est global, compter environ une 20.000 tirages par minute (et 30.000 en version compilée).

Exercice 7:

Essayer aussi la méthode suivante, itérant ceci: $z:=(x-y+1)\text{modulo}1$, $y:=x$, $x:=z$.

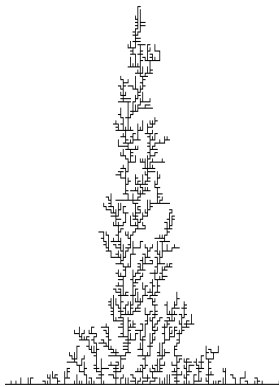
Exercice 8:

Voici une méthode pour construire des structures arborescentes aléatoires.

On considère les $(2n+1)^2$ couples d'entiers (i,j) , où i et j sont compris entre $-n$ et $+n$ (bornes incluses). Chacun de ces $(2n+1)^2$ points admet pour voisins ceux dont une des coordonnées diffère d'une unité. Un point a donc quatre voisins (excepté les points du bord).

Un point peut occuper deux états: allumé ou éteint. Les $8n$ points du bord sont initialement allumés. Le point $(0,0)$ est émetteur: un point mobile M part de $(0,0)$, et se déplace aléatoirement sur ce réseau de point voisin en point voisin, jusqu'à ce que l'un de ses voisins soit allumé: le point M est alors allumé, et il stoppe sa course; un autre point mobile part alors de $(0,0)$.

Le processus est stoppé lorsque le point $(0,0)$ est allumé. Petit programme opérant sur un réseau carré, $n=38$. Les points du bord sont allumés par l'instruction "BOX", et le réseau est dessiné avec un pas $P=5$; l'instruction "POINT" donne l'état d'un point.



```
L=639:H=399:L2=L/2:H2=H/2:
P=5:Ky=38:Kx=38:Khx=Kx*P:Khy=Ky*P:
BOX L2+Khx,H2+Khy TO L2-Khx,H2-Khy: WHILE POINT(L2,H2)=0:Nx=L2:Ny=H2:
WHILE POINT(Nx,Ny)=0:
  V= RND(2):X=Nx:Y=Ny:
  IF RND(2)=0 THEN
    IF V=0 THEN Nx=X+P ELSE Nx=X-P:
  ENDIF:
  ELSE
    IF V=0 THEN Ny=Y+P ELSE Ny=Y-P:
  ENDIF:
  ENDIF :
WEND : DRAW X,Y TO Nx,Ny: WEND :
```

Modifier ce programme pour obtenir une image du genre de celle qui précède.

Remarque 2:

Tout processus permettant
 -de construire une bipartition d'ensemble (bipartition *ordonnée*, c'est-à-dire telle que l'on puisse distinguer la première partie de la seconde),
 -et d'itérer sur les parties jusqu'aux parties élémentaires (atomiques), réduites à un élément unique,
 -est représentable par un arbre binaire.
 Les éléments de cet ensemble correspondent aux sommets terminaux de l'arbre (ou atomes, ou feuilles, ou générateurs, ou objets primitifs, dans d'autres terminologies; on dira aussi "les sommets du fond").

Il en va de même de processus récurrents qui font éclater un ensemble structuré en un couple de parties de structure identique dès que l'on ôte un élément (dit pointé) de cet ensemble; les éléments de cet ensemble sont les sommets de l'arbre (certains sont terminaux; ceux qui ont été ôtés au cours du processus sont internes à l'arbre; le premier ôté est la racine de l'arbre).

Une interprétation géométrique d'un tel processus est fournie par le découpage des polygones en triangles. Un côté du polygone est distingué et orienté. Si l'on ôte le triangle prenant appui sur le côté distingué, la configuration éclate en deux configurations de même type.

\ Exercice 9:

Un magma (géométrique, non libre, fermé par symétries ponctuelles), engendré par (trois) atomes:

Soit trois points, a, b, c, dans le plan euclidien, qui engendrent le réseau suivant:

On note ab l'image du point a par symétrie par rapport au point b (ainsi, b est le milieu du segment déterminé par les points a et ba). On note donc $ab.c$ le point symétrique de ab par rapport à c, et $a.bc$ le symétrique a par rapport à bc. Etc...

Etablir que les six points $a.bc$, $ab.c$, $c.ba$, $cb.a$, et $b.ab=ba.b$, $c.ac=ca.c$ sont alignés.

(évidemment, de même pour les points $a.cb$, $ac.b$, $b.ca$... ainsi que pour $b.ac$, $ba.c$...; clairement, $aa=a$..., et aussi $ba.a=b$...).

Ces trois droites sont parallèles aux côtés du triangle abc et se coupent en trois points dont l'un est $(b.cb).(a.cb)=(b.ab).(c.ab)=(bc.a).(b.cb)=(ba.c).(b.ab)$...

On considèrera la procédure formelle récursive f dont l'usage permet de déterminer si deux expressions sont équivalentes, càd correspondent au même point (quel que soit le nombre d'objets atomiques générateurs); elle envoie toute expression sur un polynôme de degré 1 en a,b,c..., soit $f(g.d)=2f(d)-f(g)$, initialisée par l'identité sur les générateurs, $f(a)=a$, $f(b)=b$...

Le problème de la simplification des expressions (recherche de forme minimale) se pose déjà avec deux générateurs ($a.ab=ab.(b.ab)$).

\ Notation 1: (définition équationnelle compacte, série caractéristique, grammaire générative):

Nous résumerons la définition (1) du début de ce chapitre sous la forme:

(1bis)
$$\mathbf{M} = \mathbf{x} + (\mathbf{M},\mathbf{M})$$

Cette définition peut intuitivement se lire ainsi, de manière ensembliste, de droite à gauche:

tout couple d'éléments de M appartient à M, et x appartient à M.

Ou encore, de gauche à droite: un élément de M est, soit x, soit composé d'un couple d'éléments de M -le premier élément d'un couple en sera la partie gauche, et le second la partie droite.

Ainsi, x est le prototype d'arbre minimal, isécable, atomique, générateur, et, x à part, tout arbre binaire est un couple d'arbres binaire.

[Si le contexte permet d'éviter toute erreur d'interprétation -càd qu'il est sous entendu que les calculs se font dans le magma, non associatif- on peut écrire, de manière encore plus compacte, $\mathbf{M}=\mathbf{x}+\mathbf{M}^2$]

De cette égalité fondamentale se déduit très automatiquement le calcul itératif

$$\mathbf{M} := \mathbf{x} + (\mathbf{M},\mathbf{M}), \text{ sous l'initialisation } \mathbf{M}:=\mathbf{x},$$

ou, pour être plus précis:

la suite des polynômes $M_0=x$, $M_{i+1}=x+(M_i,M_i)$ tend vers la série solution de l'équation **(1bis)**.

Le polynôme M_i est le polynôme caractéristique des arbres binaires

dont la *profondeur* n'excède pas i.

On parle parfois de *hauteur* d'arbre au lieu de *profondeur*.

(on parle plutôt de la profondeur d'une récursivité... et de la hauteur d'une pile...)

Ainsi, la différence $M_i - M_{i-1}$ caractérise les arbres binaires dont la profondeur est exactement i .

La suite des polynômes M_i pourra aimablement être appelée "suite des polynômes (non-associatifs) de Catalan": cette suite de polynômes tend vers la série M ["série de Catalan", non associative] lorsque l'indice i croît indéfiniment. Cette convergence est relativement lente; en effet, alors que le degré de M_i croît exponentiellement (il double rigoureusement à chaque itération), le nombre de coefficients stabilisés de la série de Catalan augmente seulement d'une unité lorsque l'on passe de M_i à M_{i+1} ;

M_i est de degré 2^i , mais ne possède que $i+1$ coefficients stabilisés de la série de Catalan.

\ Exercice 9:

Si l'on considère les choses avec un peu de formalisme (c'est utile), l'égalité **(1bis)** est donc une équation, dont la solution (approchée par une suite de polynômes) est une **série** M sur la structure du magma.

Ensuite, soit s la série formelle sur le *monoïde* engendré par x (cette série ne sert pas, sauf spécification contraire, à définir une fonction numérique:

seule est considérée la suite de ses coefficients), série s telle que $s = s^2 + x$.

Les coefficients de la série s sont les nombres de Catalan. En fait, l'équation $s = s^2 + x$ peut se lire de manière non-associative (sur le magma) ou de manière associative (sur le monoïde).

On approche la série s (série de Catalan associative) par la suite des polynômes (que nous qualifierons encore de polynômes de Catalan associatifs) $p_1 = x, p_{i+1} = x + p_i^2$.

La relation $s^2 = s - x$ permet de calculer les puissances de s . Sur le monoïde, le calcul de tout s^n (dont le terme non nul le plus faible est x^n) se fait par une récurrence telle que le coefficient de x^k dans s^n soit fonction d'un nombre fini de coefficients de s .

Ainsi, $s^3 = s^2 - xs = s(1-x) - x, s^4 = (s-x)^2 = s(1-2x) + x(x-1)$,

et de manière générale $s^n = sp_n - xp_{n-1}$, sous $p_0 = 0, p_1 = p_2 = 1$.

On calcule donc sur les couples de polynômes en x , soumis au produit

$$[p_n, q_n] * [p'_n, q'_n] = [p_n p'_n + p_n q'_n + q_n p'_n, q_n q'_n - p_n p'_n].$$

Calculer les premiers termes des premières puissances de s .

Etablir que le coefficient de $(-x)^i$ dans p_n , c'est le coefficient binomial

$(i, n-2i)$, où (i, j) note le nombre de manière de choisir i objets parmi $i+j$.

Les coefficients des séries s^n comptent les mots sur (x, y) tels que l'on n'ait jamais plus d'occurrences de y que de x en lisant de gauche à droite, classés en fonction du nombre total d'occurrences de x , et de y .

La série s associative peut évidemment être interprétée numériquement comme le développement de Taylor-Mac-Laurin d'une fonction.

On conviendra que cette interprétation est pour nous accessoire, car la série formelle qui nous intéresse, solution d'une équation qui n'est que la forme mathématique d'une grammaire générative, cette série formelle existe combinatoirement comme collection de coefficients, que la série numérique soit convergente ou non.

De quelle fonction numérique la série s est-elle le développement ?

On vérifiera que $1-2s$ est le développement de $(1-4x)^{1/2}$ en série de Taylor au voisinage de zéro.

Exercice 10:

Les coefficients binômiaux (i,j) sont par définition soumis à la récurrence $(i,j)=(i-1,j)+(i,j-1)$, laquelle récurrence est initialisée par $(i,0)=(0,j)=1$ quels que soient les entiers naturels i et j . En calculant modulo 2 (fonction parité), dessiner une partie de l'arbre binaire infini que définissent les binômiaux qui sont congrus à 0 zéro modulo 2 (autrement dit, les coefficients binômiaux pairs). Nous y reviendrons à de multiples occasions.

Exercice 11: (difficile; au centre d'un objet "fractal" célèbre)

Soit x_0 un nombre flottants.

Etudier expérimentalement la suite numérique $x_{i+1}=x_0+x_i^2$,

par le programme $x:=x_0+x^2$ sous l'initialisation $x:=0$.

On étudie ainsi la suite des polynômes de Catalan, pris comme fonctions numériques.

La suite (logarithme base 2 de x_i)/ 2^i converge; l'observer spécialement lorsque la suite (x_i) ne semble pas demeurer dans un domaine borné. Pour certaines valeurs de x_0 (celles comprises entre $1/4$ et $-3/4$),

la suite calculée (x_i) tend vers une racine du polynôme de degré 2, $p(x)=x^2-x+x_0$.

Observer spécialement ce qui se produit au voisinage des valeurs

0.25, 0, -0.75, -1, -1.25, -1/4, -2, de x_0 .

On pourrait associer à tout ensemble de mots E (fini) l'ensemble $f(E)=x+(E,E)$, et noter que le langage M (infini) est point fixe de f (c'est-à-dire invariant par f ; autrement dit, la composition de deux éléments de E est un élément de E : on dit que E est fermé - pour la composition des arbres binaires). La procédure f renvoie en sortie l'union de x et des objets constitués en accouplant de toutes les manières possibles les objets fournis en entrée.

Comme déjà signalé, on approche M par la suite croissante des $M_{i+1}=f(M_i)=f^i(M_0)$, sous l'initialisation $M_0=x$. Programmer cela.

Remarquer que le processus f doit opérer dans un certain ordre

-par exemple, l'ordre lexicographique naturel gauche sur les couples, obtenu en comparant deux couples en priorité par leurs premiers éléments (on dira de préférence 'premier facteur', ou 'facteur gauche'), ou par leur seconds éléments (on dira 'second facteur', ou 'facteur droit') lorsque les facteurs gauches sont identiques.

Notation 2:

Un autre système de notation est traditionnellement utilisé par l'école linguistique. Ce système énumère la liste des substitutions formelles permises.

Dans le cas présent, cela consisterait

à formaliser la récurrence (ou itération) $M:=x+(M,M)$ initialisée par $M:=x$,

convergeant vers la série solution de l'équation $M=x+(M,M)$,

série point fixe de la transformation $f(M)=x+(M,M)$

de la manière suivante:

partir de la lettre M (qualifiée d'axiome), et construire tous les mots possibles par application des deux règles suivantes (appelées productions):

1) on a le droit de remplacer toute occurrence de M par (M,M)

2) on a le droit de remplacer toute occurrence de M par x .

On produit ainsi des mots sur l'alphabet constitué des 5 lettres

$x (,) M$

dont les quatre premières lettres constituent le vocabulaire terminal V^t , tandis que la lettre M appartient au vocabulaire dit auxiliaire V^a [dans le cas présent, le vocabulaire auxiliaire est réduit à une unique lettre; de manière plus générale, l'axiome est une lettre particulière du vocabulaire auxiliaire: l'axiome sert à initialiser le processus de substitution].

Les vocabulaires auxiliaires et terminaux sont supposés disjoints, et les productions [ou règles de substitution, ou règles de grammaire] autorisent le remplacement d'une occurrence d'une lettre auxiliaire par un mot sur l'alphabet V, union des vocabulaires auxiliaires et terminaux [$V = \text{union de } V^a \text{ et } V^t$].

Le langage engendré par cette grammaire est constitué des mots ne comportant pas de lettre du vocabulaire auxiliaire (càd que le langage engendré est constitué des mots dont toutes les lettres appartiennent au vocabulaire terminal).

Ces grammaires, constituées d'un nombre fini de telles règles de substitution, et d'un unique axiome, sont qualifiées de grammaires algébriques, ou parfois de "grammaires de Chomsky".

On voit que les lettres du vocabulaire auxiliaire jouent le rôle de variables, indéterminées, ou inconnues, contraintes par un système d'équations, tandis que les lettres du vocabulaire terminal sont considérées comme des constantes. L'ensemble des règles de substitution constitue en fait un système d'équations polynômiales (il y a autant d'équations que de lettres du vocabulaire auxiliaire).

Nous signalerons certainement, à l'occasion, que ce calcul est un calcul de dérivée, au sens algébrique du terme. Ainsi, lorsque les linguistes parlent de dérivation, il s'agit bien du même mot que lorsque l'on calcule la dérivée d'une fonction. Il semble cependant que le choix de ce vocable (on parle d'une "dérivation" grammaticale, et d'"arbres de dérivation") n'ait pas été fait consciemment.

Facteurs gauches et droits:

Qu'en est-il de l'énumération des expressions du Magma engendré par x? La méthode suggérée plus haut (en exercice) nécessite excessivement de place.

Remarquer que ces expressions peuvent être ordonnées par l'ordre alphabétique sur le Magma. Cet ordre, qui revient à comparer deux objets de M par leur facteurs gauches, ou, si les facteurs gauches sont identiques, par leurs facteurs droits ne peut servir de processus énumératif, car entre deux mots, il peut y avoir (y a-t-il?) une infinité de mots: il en est ainsi entre (x,x) et ((x,x),x), puisqu'il existe une infinité d'expressions admettant x pour partie gauche.

Le générateur x est conventionnellement le plus petit objet de M.

Cet ordre est l'ordre alphabétique *gauche*.

Si l'on compare préférentiellement par les facteurs droits, et par les facteurs gauches seulement lorsque les facteurs droits sont identiques, on parlera d'ordre alphabétique *droit*.

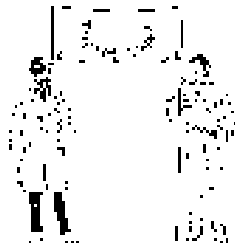
Ainsi, à l'exception de x, tout mot m de M s'écrit identiquement $m=(m',m'')$, où m' et m'' appartiennent à M.

On dira que $Gm=m'$ et $Dm=m''$ sont les facteurs gauche et droit de l'expression m (ou les branches gauche et droite de l'arbre). Il est possible de donner un sens à Gx et Dx , mais ceci sort du cadre de cet exposé.

On a identiquement [pour tout mot m distinct de x] $m=(Gm,Dm)$.

Dans une expression m du magma M, les occurrences de x sont qualifiées d'éléments atomiques, ou atomes (allusion à leur caractère insécable: ils ne peuvent être obtenus par composition binaire) [on dit aussi, oralement, que ce sont les feuilles de l'arbre, ou les sommets du fond, ou encore les sommets terminaux, tous vocables imagés].

Si $m=(m',m'')$, les facteurs gauche et droit de m, m' et m'', sont aussi qualifiés de branches gauche et droite de l'arbre; de manière héréditaire (transitivement) les facteurs de m' et m'' sont également des facteurs de m. Ainsi, tout atome de m est un facteur de m.



Remarque 3:

Il serait possible d'omettre la lettre x -elle serait alors remplacée par le mot vide. Il peut paraître redondant d'utiliser cette lettre x, alors que l'on ne perd aucune information en l'omettant. Cependant, le lecteur remarquera que l'usage veut que le mot vide (que l'on note habituellement e, puisqu'il faut bien le représenter) soit neutre pour la concaténation des mots. Dans la structure du Magma, la lettre x ne se comporte pas ainsi; il est d'autant plus nécessaire de la conserver que l'on pourrait être amené à adjoindre au Magma un objet neutre, évidemment distinct de x (et éventuellement aussi représenté par e); nous éviterons cependant d'introduire cet élément neutre, qui n'est ici pas indispensable, car cela soulève quelques difficultés théoriques.

L'objet x est évidemment l'objet minimal, servant à initialiser la récurrence générative. La virgule symbolise la composition binaire. L'utilisation des parenthèses soulignent le fait que la loi de composition est non-associative. Quand à l'expression représentant un objet du magma, c'est un mot sur quatre lettres.

Deux mots distincts obtenus par la récurrence (1) représentent deux objets distincts (x engendre librement, de manière non-ambigüe, le magma).\

Degré:

A tout mot m du magma binaire M, on associe naturellement son degré (sous-entendu, par rapport à x -on dit aussi : "le degré en x de l'arbre m", ou "le degré de l'arbre codé par le mot m"). C'est un nombre entier, que l'on notera Deg.m, qui dénombre les occurrences de la seule lettre x.

Il est évidemment défini par la récurrence

$$\text{Deg.}(m',m'') = \text{Deg.}m' + \text{Deg.}m'', \text{ sous l'initialisation } \text{Deg.}x = 1.$$

Sommets:

On définit tout aussi naturellement la fonction "Sommets", qui dénombre les sommets par la récurrence $\text{Sommets.}(m',m'') = 1 + \text{Sommets.}m' + \text{Sommets.}m'', \text{ sous } \text{Sommets.}x = 1.$

Note:

Les polynômes et séries formels sur le magma sont des objets très naturels. Une série à coefficients entiers n'est qu'une application du magma dans les entiers. Un polynôme est une série dont seuls un nombre fini de coefficients sont non nuls (on dit "à support fini": le support d'une série est l'ensemble des mots dont les coefficients sont non nuls).

On peut alors définir l'opération linéaire telle que $d[m] = (\text{Deg.}m) m$, affectant pour coefficient, à tout objet du magma, son degré.

Cette opération est une dérivation, au sens algébrique du terme,

$$\text{soit } d[(m',m'')] = (d[m'],m'') + (m',d[m'']).$$

Exercice 12 (les arbres binaires selon leur profondeur):

Sur le magma engendré par x, soit la fonction profondeur définie par $\text{prof}[x] = 0, \text{ prof}[(g,d)] = 1 + \max[\text{prof}[g], \text{prof}[d]].$

Le polynôme $P_{i+1} = x + (P_i, P_i)$, sous $P_0 = x$, caractérise les expressions de profondeur au plus $i+1$. Si r_i est le nombre d'expressions du magma dont la profondeur est inférieure ou égale à i ,

on a $r_{i+1}=1+r_i^2$ [début de la suite r_i : 1, 2, 5, $26=2*43$, 677, $458330=2*5*45833$,
210066388901= $41*1277*4012193...$]

et si h_n est le nombre d'expressions du magma de profondeur strictement n ,

on a $h_{n+1}=r_{n+1}-r_n=r_n^2-r_{n-1}^2=(r_n-r_{n-1})(r_n+r_{n-1})=h_n(r_n+r_{n-1})=(r_n+r_{n-1})(r_{n-1}+r_{n-2})\dots(r_1+r_0)$

Le mode d'énumération lexicographique

$x, (x,x), (x,(x,x)), ((x,x),x), ((x,x),(x,x))\dots$,

[soit sur les couples d'entiers 0, $(0,0)=1$, $(0,1)=2$, $(1,0)=3$, $(1,1)=4$, $(0,2)=5$, $(1,2)=6,\dots$]

consistant à composer tout objet du magma avec ceux qui précèdent, puis lui même, énumère les objets de profondeur n avant ceux de profondeur $n+1$.

On a la récurrence croisée $h_{n+1}=h_n(r_n+r_{n-1})$, $r_{n+1}=h_{n+1}+r_n$ sous l'initialisation $r_0=1$, $r_1=2$, $h_1=1$.

[début de r_n+r_{n-1} : 1, 3, 7, 31, $703=19*37$, 459007, $210066847231=43*313*15607909 \dots$]

[h_n : 1, 1, 3, 21, 651, 457653,...]

En notant $H_n=P_n-P_{n-1}$ le polynôme des objets de profondeur n ,

$H_{n+1}=(H_n,P_{n-1})+(P_{n-1},H_n)+(H_n,H_n)$ sous $H_0=x$, $H_1=(x,x)$,

soit, numériquement, $h_{n+1}=2h_n r_{n-1}+h_n^2$,

(si une expression est de profondeur n , l'un *au moins* des constituants gauche et droit est de profondeur n , et l'autre est de profondeur *au plus* n)

la série magmatique de catalan, solution de $S=x+(S,S)$, est

-la somme des polynômes H_n , -la limite de la suite des polynômes P_i .

Evidemment, le morphisme qui envoie x sur x^2 définit une injection de H_n dans H_{n+1} .

Considérer la suite $s_n=1+h_{n+1}/h_n=1+r_n+r_{n-1}$ [$=3+r_{n-1}^2+r_{n-2}^2\dots$]

[début de s_n : 2^1 , $4=2^2$, $8=2^3$, $32=2^5$, $704=2^6*11$, $459008=2^8*11*163$,

$210066847232=2^9*11*23*163*9949\dots$]

où s_n divise s_{n+1} [s_{n+1}/s_n : 2 , 2^2 , $2*11$, 2^2*163 , $2*23*9949\dots$]

Quand à f : $f(n)=h(n)+h(n-1)$

(nombre d'expressions dont la profondeur est, soit n , soit $n-1$)

[début de $f(n)$: 2, 4, 24, 672, 458304, 210066388224....].

Enfin, on a

[$f(n+1)/f(n)$: 2, $6=2*3$, $28=2^2*7$, $682=2*11*31$, $458356=2^2*19*37*163\dots$].

Question: les règles grammaticales génératives peuvent-elles être appliquées dans le but de produire un mot de degré n (resp. profondeur n) du magma binaire, de manière équiprobable ?

Exercice 13:

Calculer le début du tableau $f(i,j)$ comptant les arbres binaires de degré i et de profondeur j .

On a $f(i,j)=$ somme des produits $f(i_1,j_1)*f(i_2,j_2)$, où $i_1+i_2=i$ et $j=1+\max(j_1,j_2)$; sous $f(1,1)=1$.

1

1
2

Nous avons ici initialisé la profondeur par $\text{prof}(x)=1$.

1 4

sur la ligne $i=2^n$, la profondeur va du minimum $n+1$ au maximum 2^n .

6 8

Sur la diagonale, les puissances de deux.

6 20 16

La somme par ligne donne la suite des nombres de catalan (parfois dits "d'Euler-Catalan)

$c=(1$ 1 2 5 14 42 132 429 1430...), qui satisfont à $c_n=\text{somme des produits } c_i c_j$ pour $i+j=n$.

La somme par colonne donne 1 1 3 21 651 457.653 210.065.930.571...
 suite des différences premières de la suite $s=(1 2 5 26...)$ qui satisfait à $s_{k+1} = 1+s_k^2$.

Exercice 14:

Naturalistes, hydrauliciens, statisticiens, informaticiens, n'ont pas manqué de considérer et étudier l'intéressante fonction suivante, connue sous le nom de fonction de Strahler, essentiellement définie sur les arbres binaires, par la fonction récursive **Str** que voici:

$$\begin{aligned}
 -\mathbf{Str}[x] &= 1 \\
 -\mathbf{Str}[(g,d)] &= \max[\mathbf{Str}[g], \mathbf{Str}[d]] \dots \text{lorsque } \mathbf{Str}[g] \text{ diffère de } \mathbf{Str}[d] \\
 &\text{et } 1 + \mathbf{Str}[g] \dots \text{lorsque } \mathbf{Str}[g] = \mathbf{Str}[d].
 \end{aligned}$$

Autrement dit, au générateur x est affectée la valeur 1, et à tout composé (g,d) le *maximum des valeurs calculées sur g et d si celles-ci sont distinctes, ce maximum plus un lorsque elles sont égales*. Ainsi, cas extrêmes, $\mathbf{Str}[(x,(x,(x,(x,(x,(x,(x,x))))))]=2$, tandis que $\mathbf{Str}[(x,(x,(x,(x,(x,(x,x))))))]=4$

En terme de grammaire générative (ou de série formelle), si $L[k]$ est le langage des arbres binaires de valeur k au sens de la fonction de Strahler, on a la récurrence (ou le système d'équations) $L[k+1] = (L[k], L[k]) +$ somme formelle des $(L[k+1], L[i]) + (L[i], L[k+1])$ pour i inférieur à $k+1$, sous l'initialisation $L[1] = x$.
 D'où, pour le nombre $f(n,k)$ d'arbres de degré n et valeur de Strahler k :
 $f(n,k) =$ somme de $[2f(i,k)f(j,r) + f(i,k-1)f(j,k-1)]$ pour r inférieur à k et $i+j=n$, sous l'initialisation $f(1,1) = 1$.

Evidemment, $\mathbf{Str}[a]$, comme la fonction profondeur, ne dépend que de l'image commutative de l'arbre. Pour tout arbre a , $\mathbf{Str}(a)$ est le **nombre minimum de mémoires (registres) nécessaires** pour évaluer telle "expression arithmétique" représentée selon le mode usuel sur l'arbre a , les sommets terminaux portant les arguments du calcul, et les sommets internes les opérateurs binaires.

C'est aussi bien, comme on le vérifiera sur l'exemple qui suit, le nombre minimum de 'jetons' nécessaires pour 'jouer' au petit jeu combinatoire suivant:
 -on peut placer un jeton sur un sommet terminal,
 -et sur un sommet interne lorsqu'il y a deux jetons sur les sommets placés en dessous
 -les jetons sont récupérables, et déplacés
 jusqu'à ce que l'un d'entre eux puisse être placé sur la racine de l'arbre,
 traitement typiquement récursif, optimisable par le calcul de la fonction de Strahler sur chaque sommet.

Ainsi, sur l'arbre $((x,x),(x,x),(x,(x,x)))$, les **sommets terminaux** correspondent aux occurrences de la lettre 'x', tandis que les **sommets internes** correspondent aux occurrences de la virgule ','.
 En notant par '*' tout jeton placé, on aboutit par exemple à la suite des états, en utilisant trois jetons:

$$\begin{aligned}
 &(((*, *), (x, x)), (x, (x, x))) && (((x * x), (*, *)), (x, (x, x))) && (((x * x), (x * x)), (x, (x, x))) \\
 &(((x, x) * (x, x)), (x, (*, *))) && (((x, x) * (x, x)), (*, (x * x))) && (((x, x) * (x, x)), (x * (x, x))) \\
 &\text{et finalement} && (((x, x), (x, x)) * (x, (x, x))).
 \end{aligned}$$

Les arbres binaires, déjà classables selon degré et profondeur, le sont en outre selon les nombres de Strahler.

Si l'on suit la gènèse standard du magma binaire, en composant tout objet de manière ordonnée et systématique avec ceux qui précèdent, à gauche et à droite, et ensuite avec lui-même:

$x \cdot (x, x) \cdot (x, (x, x)) \cdot ((x, x), x) \cdot ((x, x), (x, x)) \dots$,
 le début de la suite correspondante des nombres de Strahler est
 1 . 2 . 2 2 3 . 2 2 3 3 3 . 2 2 3 3 3 3 3 . 3 3 3 3 3 3 3 3 4....

Exercice 15:

Considérer le langage L ainsi défini sur un alphabet (a,b) de deux lettres:

- ab appartient à L
- si le mot m appartient à L, alors,
le mot déduit de m par substitution de ab à toute occurrence de a dans m appartient à L
et
le mot déduit de m par substitution de ab à toute occurrence de b dans m appartient à L.

Dénombrement:

On note c_n le nombre de mots du magma binaire monogène M dont le degré est n.

Tout mot m de degré $n = \text{Deg.}m > 1$ s'écrit (Gm, Dm) , avec $\text{Deg.}Gm = i > 0$, et $\text{Deg.}Dm = j > 0$.
Il en résulte que $c_n = \sum c_i c_j$ est la somme des produits $c_i c_j$, la sommation étant étendue à tous les couples (i, j) d'entiers strictement positifs tels que $i + j = n$.
Cette récurrence est clairement initialisée par $c_1 = 1$ (d'où $c_2 = 1$).

Les premiers termes en sont

1 1 2 5 14 42 132 429 1430 4862 16796 58768 208012 742900.....

Cette suite, fondamentale, très présente en combinatoire, porte le nom de Catalan.
Nous en faisons par ailleurs des présentations plus formelles.

Si l'on note C_i l'ensemble des mots du magma dont le degré est i, on note naturellement (C_i, C_j) l'ensemble de ceux dont la partie gauche (resp. droite) est de degré i (resp. j), et (C_i, C_j) est ainsi une partie de C_{i+j} . Donc, C_n est la réunion des (C_i, C_j) , étendue aux couples (i, j) d'entiers positifs de somme n. Cette définition est récurrente, et initialisée par $C_1 = x$.

On peut interpréter ça de manière polynômiale: le calcul n'est pas commutatif, ni associatif, mais il est distributif. Il faut alors considérer C_n comme un polynôme dont les monômes sont des expressions du magma. Bien remarquer que l'on peut tout aussi bien interpréter C_n comme une procédure récursive, dépendant d'un paramètre (l'entier n).

Exercice 16: Programmer le calcul la suite des nombres c_n . Nous y reviendrons.

Exercice 17: (binômiaux et nombres de Catalan; formule close)

Autre mode de calcul de ces coefficients (la question se pose de calculer efficacement les éléments de cette suite): comparer la suite de Catalan calculée ci-dessus à celle des coefficients binômiaux (i, i) , diagonale de la récurrence $(i, j) = (i-1, j) + (i, j-1)$, initialisée par $(0, j) = (i, 0) = 1$.

Comparer également à la suite des coefficients binômiaux $(i+1, i)$, et

1) en induire tout d'abord un processus récurrent de calcul des coefficients c_n , (i, i) , et $(i+1, i)$, puis une formule close pour les nombres (i, i) , $(i+1, i)$ et c_n (sans chercher, pour l'instant, à la prouver); les coefficients (i, i) sont les binômiaux diagonaux ou centraux, et les coefficients $(i, i+1)$ les binômiaux quasi-diagonaux, ou quasi-centraux.

2) programmer ensuite quelque procédure de calcul rapide de la suite c_n .

[Note: l'exercice peut se poursuivre -comparer la suite des c_n à la suite des binômiaux de la forme $(i+2, i)$, puis celle des $(i+3, i)$..etc...]

3) observer l'intéressante récurrence $(n+1)c_{n+1}=2(2n-1)c_n$, soumise à l'initialisation $c_1=1$; elle est la source d'une démonstration directe, provenant de l'interprétation des nombres de catalan comme dénombrement des arbres binaires, et permet d'établir que pour tout nombre premier $k=2n-1$, il existe un nombre de catalan multiple de k .

4) notons les coefficients binômiaux centraux: $(i,i)=f(2i)$, $(i,i+1)=f(2i-1)$;
on a $f(2i)=2f(2i-1)$, $f(2i+1)=2f(2i)-c_{i+1}$, $c_{i+1}=f(2i)/(i+1)$

5) signalons enfin que si $b(n)=c_1c_2\dots c_n$, produit des n premiers nombres de catalan, $c_1=1$, $c_2=1$, $c_3=2,\dots$ on a, pour $n>1$, $n(n+1)b(n+1)=4(2n-1)(2n-3)b(n)$.

6) établir que $6.c(n)$ est multiple de $n+1$.

\ Exercice 18: (énumération lexicographique -à degré fixé- des arbres binaires)

Construire l'ensemble C_n des expressions de degré n du magma engendré par x , en l'ordonnant suivant l'ordre naturel qui consiste à comparer deux expressions par leur parties gauches si elles sont distinctes, et par les parties droites lorsque les parties gauches sont identiques.

Une procédure $s[.]$ donne l'expression suivante pour cet ordre.
L'expression minimale de degré i - nous la noterons $\min[i]$ - est de la forme $(x,(x,(...)))$, et l'expression maximale - que l'on notera $\max[i]$ - est de la forme $((...),x,x)$.
Soit $\text{degré}[g]=i$, et $\text{degré}[d]=j$.
Alors, $s[(g,d)]=(g,s[d])$ lorsque $s[d]$ existe.
Lorsque $d=\max[j]$, $s[(g,d)]=(s[g],\min[j])$ lorsque $s[g]$ existe.
Si $d=\max[j]$, $g=\max[i]$, et $j>1$, alors $s[(g,d)]=(\min[i+1],\min[j-1])$.
Si $g=\max[i]$ et $j=1$, alors $(g,d)=\max[i+1]$.

\ Exercice 19: (à propos de la récurrence de Catalan)

Soit le magma (composition binaire non associative) produit de la manière suivante:

- le mot vide, mot de degré 0, n'est pas correct;
- x est l'unique mot correct de degré 1;
- le morphisme v substitue x à y et y à x (ainsi $v(xyyx)=yxyx$):
- si m et m' sont des mots corrects $v(m)v(m')$ est un mot correct.

Certains mots corrects sont produits de plusieurs manières.

S'intéresser à un programme qui donnerait, quel que soit tel mot m , le nombre de manières (éventuellement nul) dont ce mot m est produit.

Soit le monoïde (composition binaire associative) produit de la manière suivante:

- x est l'unique mot correct de degré 1;
- yy est l'unique mot correct de degré 2;
- si m et m' sont des mots corrects, alors mm' est un mot correct;
- on ne tient pas compte du mot vide.

S'intéresser à un programme qui donnerait, quel que soit tel mot m , le nombre de manières (éventuellement nul) dont ce mot m est produit.

\ Exercice 20 :

1) Les nombres de Catalan comptent, en fonction de leur degré, ce que les programmeurs nomment "les arbres binaires", lesquels sont très naturellement représentés par les expressions du magma, dont l'objet x a été l'élément générateur.

Le Magma commutatif M_C (représentation d'arbres commutatifs) s'obtient en considérant que pour tout objet (g,d) du magma, on a $(g,d)=(d,g)$.

Un processus de normalisation de la représentation consiste à conserver une seule de ces deux représentations, par exemple, celle lexicographiquement minimale lorsque l'on dispose d'un ordre total sur M_C : (g,d) si $g < d$. Restent les carrés $[d=g]$.

Supposer ici que cet ordre soit compatible avec l'ordre des degrés.

Dénombrer les expressions du Magma commutatif engendré par x : montrer que le nombre c_n de telles expressions de degré n satisfait à la récurrence

$$c_{2n+1} = (\text{somme des } c_i c_j \text{ pour } i < j \text{ et } i+j=2n+1),$$

$$c_{2n} = (\text{somme des } c_i c_j \text{ pour } i < j \text{ et } i+j=2n) + c_n(c_n+1)/2.$$

Les premiers coefficients sont, pour $n > 0$, 1 1 1 2 3 6 11 23 46 98.

Noter que $c_n(c_n+1)/2 = c_n + c_n(c_n-1)/2$, ce que l'on peut interpréter, par exemple, comme la composition d'objets de même degré, identiques (en nombre c_n , ce que l'on appelle des carrés) ou non.

2) Le Magma commutatif M_C s'injecte donc dans le Magma M , dès que l'on se donne un ordre total sur M , en considérant la condition récurrente sur les objets composés $m=(g,d)$ de M : on exige que la partie gauche g soit inférieure pour cet ordre (ou au plus égale) à la partie droite d . Par exemple, l'énumération se fait naturellement en construisant de manière récurrente une bijection sur les entiers, chaque objet étant composé avec ceux qui le précèdent (et lui-même), soit $x (x,x) (x,(x,x)) ((x,x),(x,x)) (x,(x,(x,x))) ((x,x),(x,(x,x)))...$

3) L'image associative du Magma commutatif (en notant y les occurrences atomiques droites de x) donne la suite $x xy xxy xyxy xxxy xyxy xxyxy xxyxy xyxyxy xyxyxyxy...$

de mots (non tous distincts) et la suite des degrés de ces mots:

1 2 3 4 4 5 6 5 6 7 8 5 6 7 8 8 6 7 8 9 9 10 7 8 9 10 10 11 12.....

4) Un sous-ensemble de l'ensemble des arbres binaires commutatifs.

$$\text{Soit } s \text{ est la série } s \text{ telle que } 2(s-x) = s^2 + s(x^2),$$

$$\text{soit } s(2n) = 2s(2n-1), s(2n+1) = 2s(2n) - s(n) \text{ sous } s(1) = 1 = s(2)$$

La série $s(x^2)$ dénombre des expressions qui sont des carrés.

On établit également que $s(n) =$ somme des $[n/2]$ coefficients qui précèdent.

Soit le langage des partitions de l'entier n représentées par les mots sur l'alphabet des entiers positifs tels que a) la première lettre soit 1 b) toute lettre n'excède pas la somme de celles qui précèdent. La norme d'un mot est la somme de ses lettres. Le nombre de mots de norme n est $s(n)$. Ces mots constituent un monoïde, avec des éléments primitifs. Génèse suivant les degrés:

1 1.1 1.1.1 112 1.1.1.1 1.112 1113 112.1 1122 1123 1124 1.1.1.1.1 1.1.112 1.1113 11114 1.112.1 1.1122 1.1123 1.1124 11125 1113.1 11132 11133 11134 11135 11136...

Par exemple, le mot 1113.1122.1.112, de degré 12, de norme 17, possède trois facteurs primitifs.

Génèse suivant les normes, en récurrant par la droite, par adjonction et contractions de caractères "1":
1; 11; 111; 1111 et 112; 11111 et 1112, 1121; 111111 et 11112 et 1113, 11121, 11211 et 1122;
1111111 et 111112 et 11113, 111121, 11131, 111211 et 11122, 112111 et 11212 et 1123, 11221;...

De chaque mot mx de norme n , où x est la dernière lettre dérivent deux mots de norme $n+1$ lorsque n est impair, les mots $mx1$ et my , où $y=x+1$.

Si n est pair, $n=2*j$, il en va de même, excepté lorsque la dernière lettre est $x=j$, auquel cas m est un mot de norme j .

D'où la récurrence sur le nombre de mots de norme n .

Reste à construire une bijection sur les arbres binaires commutatifs.

On est intéressé au calcul de la fonction $f(n,k)$ donnant le nombre de mots de degré k et norme n .

1
 1
 1 1
 2 1
 2 3 1
 1 5 4 1
 1..... 1

et au calcul du nombre de mots de degré k .

6) Anticipons.

Des constructions, énumérations, et dénombrements comparables, se font sur le Magma généralisé dont il sera question plus loin.

Soit un ordre total sur le Magma généralisé commutatif.

Un objet du Magma généralisé commutatif est une suite croissante (pour cet ordre) d'objet. On essaiera. C'est sensiblement plus compliqué, car le degré utilisé précédemment ne suffit plus à décomposer l'ensemble en sous-ensembles finis.

Les objets du Magma généralisé commutatif sont simplement appelés, en théorie des graphes, arbres pointés, ou enracinés (un sommet de l'arbre est distingué, mais le dessin de la représentation importe peu: les branches de l'arbre commutent).

\

Exercice 20 (une partie du Magma binaire commutatif):

Si l'on compose, dans le magma binaire commutatif, chaque élément avec ceux qui précèdent, à partir de x et $y=(x,x)$, mais sans engendrer les carrés, excepté initialement (x,x) , on obtient

$x (x,x) (x,(x,x)) ((x,x),(x,(x,x))) (x,(x,(x,(x,x)))) ((x,x),(x,(x,x))) ((x,(x,x),(x,(x,(x,x))))$
 ...

un langage d'arbres naturellement ordonnés lexicographiquement suivant les profondeurs.

Suivant cet ordre, on a $x < (x,x)$, et pour $b < d$, $(a,b) < (c,d)$, $(b,h) < (d,h)$.

On calculera le nombre d'arbres à profondeur n :

1 1 1 2 7 56 2212 2595782...

et l'on établira que cette fonction satisfait à la récurrence

$$f(n+1) = f(n) * (f(1) + \dots + f(n-1)) + f(n) * (f(n) - 1) / 2 \text{ sous } f(1) = f(2) = 1.$$

Il est intéressant de dénombrer les arbres de ce langage dont la profondeur est au plus n :

1 2 3 5 12 68 2280 2598062.....,

cette fonction satisfaisant à la récurrence $g(n+1) = g(n)(g(n)-1)/2 + 2$ sous $g(1)=1$.

Le programme itératif suivant n'engendre qu'une partie de ces arbres:

$$[p;q] := [p+q; (p,q)] \text{ sous l'initialisation } [p;q] := [x;(x,x)].$$

Sur le monoïde, on obtient les mots $x y xy x^2y yxy x^3y yx^2y xyx^2y \dots$,

en composant chaque mot avec ceux qui précèdent, les mots ainsi ordonnés sont en bijection avec les entiers naturels; soit $m < m'$; $m.m'$ appartient au langage; tout mot du langage se factorise en deux mots du langage; deux mots se comparent par leurs facteurs droits, et sinon par leurs facteurs gauches.

Peut-on obtenir deux mots de même image commutative ($x y xy x^2y xy^2 x^3y x^2y^2 x^3y^2 \dots$) ?

Oui, car $yxyx$ et $xyxy$ sont produits tous deux.

Nombre de mots de degré donné (ici, pour n de 1 à 27):

2 1 2 4 10 25 68 187 532 1532 4492 13308 39870 120433 366656 1123504 3463050 10729234
 33396202 104381385 327477814 1030903514 3255394686 10309154738 32732315724
 104177891232 332308237666...

fonction satisfaisant à $h(n) = (\text{somme des } h(i).h(j) \text{ pour } i+j=n, i < j) + (\text{si } n=2k, (f(k)-1).f(k)/2)$
 sous l'initialisation $h(1)=2, h(2)=1$.

Combien de mots de degrés partiels (i,j) en x et y ?

Variation: en démarrant la même récurrence avec les mots x et xy (soit $h(1)=h(2)=1$), le nombre $h(n)$
 de mots de degré n est moindre (pour n de 1 à 40):

1 1 1 1 2 3 6 11 22 43 88 179 372 774 1631 3448 7347 15713 33791 72923 158021 343495 749102
 1638103 3591724 7893802 17387931 38379200 84875596 188036830 417284181
 9274698452064465341 4601670625 10270463565 22950838755 51346678941 115001687020
 257839541992 578662327485.....

\

Exercice 21:

Nous savons calculer le nombre des arbres binaires de degré i (les nombres de catalan, $c(i)$), et le
 nombre des arbres binaires de profondeur j (différence première de la suite des $f(n+1)=1+f(n)^2$ sous
 $f(1)=1$). Convenons que le générateur x est de profondeur 1 et degré 1: ainsi, un éventuel élément
 neutre e serait de profondeur 0, et de degré 0.

Construire le tableau $f(i,j)$, $i \leq j$, du dénombrement des arbres binaires de degré i et profondeur j .

[Si $A(i,j)$ est le polynôme des arbres de degré i profondeur j ,

le polynôme $(A(i,j), A(i',j'))$ est une part de $A(i+i', 1+\max(j,j'))$]

On a $f(k,v+1) = \text{sommation pour } i+i'=k, j < v, j' < v, \text{ des } f(i,j)f(i',v) + f(i,v)f(i',j') + f(i,v)f(i',v)$.

On a $f(1,1)=1, f(i,i)=2^{i-2}$ pour $i \geq 2, f(i,j)=0$ pour $i < j, f(i,j)=0$ pour $i > 2j-1$.

\

Dérivations.

Dans un exercice ci-dessus, nous avons suggéré de construire des arbres binaires en substituant aléatoirement (x,x) à une occurrence de x dans les expressions magmatiques. Une énumération systématique des substitutions constitue une méthode générative, non pas des arbres binaires, mais de l'histoire de leur gènèse par ce procédé.

En effet, n substitutions sont possibles sur une expression de degré n : on a ainsi une bijection de l'ensemble de ces histoires (ou gènèses) sur les suites finies d'entiers non-négatifs telles que le k -ième entier soit inférieur à k (et, comme nous le verrons dans les chapitres ultérieurs, une bijection de ces histoires sur les entiers et sur les permutations).

Par exemple, la suite 0 0 1 3, lue caractère par caractère de gauche à droite, conduit à construire les expressions successives: $x (x,x) ((x,x),x) ((x,(x,x)),x) ((x,(x,x)),(x,x))$.

Ce procédé génératif constitue ce que l'on appelle une dérivation, au sens algébrique du terme (mais également, comme signalé plus haut, au sens des grammaires génératives).

Un opérateur de dérivation D est un opérateur linéaire défini
 -sur les objets du magma par $D.(m,m')=(D.m,m')+(m,D.m')$,
 -dans le cas présent, initialisé par $D.x=(x,x)$, mais nous aurons l'occasion de voir bien d'autres situations. Ainsi, par exemple, en utilisant la distributivité du produit par rapport à la somme:

$$D.((x,x),x)=(((x,x),x),x)+((x,(x,x)),x)+((x,x),(x,x)).$$

Si l'on calcule $D.x$, \dots $D^n.x$, etc.. (écrire un programme), des multiplicités apparaissent.

Ainsi, dans $D^3.x$, le monôme $((x,x),(x,x))$ a pour coefficient 2, correspondant à deux histoires distinctes que l'on pourrait mémoriser ainsi, en numérotant les substitutions:

$$D.x=(x1x); D^2.x=((x2x)1x)+(x1(x2x));$$

$$D^3.x=((x2x)1(x3x))+((x3x)1(x2x))+(((x3x)2x)1x)+(x1((x3x)2x))+((x2(x3x))1x)+(x1(x2(x3x))).$$

Nous y reviendrons, notamment en traitant des permutations.

Exercice 21:

A. Les mots m que nous qualifierons (durant cet exercice) de "corrects", seront définis comme satisfaisant à l'une des quatre conditions suivantes [ici exclusives; ceci est une analyse grammaticale]:

- 1) $m=x$
- 2) $m=(x,x)$
- 3) $m=((m',m''),m''')$, avec m' et m'' corrects
- 4) $m=(m',(m',m'''))$, avec m' et m'' corrects.

Etant donné un élément m , décider s'il est correct en calculant (d'abord par procédure récursive) sur un mot quelconque [sur l'alphabet des quatre lettres $x(,)$] la fonction f qui caractérise (dite fonction caractéristique du langage) les éléments corrects, fonction telle que $f(m)=1$ si m est correct, et $f(m)=0$ sinon. S'intéresser au nombre d'éléments corrects de degré n .

Un automate réalisant f est dit 'reconnaître ce langage du magma'

Remarquer qu'une mécanique lisant un mot m (caractère par caractère, de gauche à droite) dans le but de décider s'il est correct ou non, une telle mécanique doit disposer d'une mémoire potentiellement infinie.

[Répétons encore: les automates finis ne peuvent reconnaître que des sous-ensembles particuliers du magma. Pire: il existe des parties qu'aucun automate ne reconnaît.

Voir la théorie de la calculabilité initiée par Alan Turing vers le premier tiers de ce 20^e siècle -théorie dont le génie tient à sa simplicité conceptuelle, typiquement informatique avant l'heure. Voir les diverses formes équivalentes de cette théorie. L'idée fondamentale d'Alan Turing est de considérer une *machine finie et déterministe* qui se déplace sur une *mémoire externe infinie*, lisant cette mémoire, et la modifiant.

C'est conforme au fait que langages, grammaires, automates, se classent et se répondent: à la complexité d'une classe de *langage* correspond la complexité d'une classe de *grammaires* génératives, et celle d'une classe d'*automates* reconnaissants].

B. Considérer maintenant le très important langage L, partie du précédent,

-le langage L, constitué d'arbres binaires, est défini par la récurrence:

si (a,b) appartient à L, alors, $f_g(a,b)=(a,(a,b))$ et $f_d(a,b)=((a,b),b)$ appartiennent à L

-cette récurrence est initialisée par: (x,x) appartient à L.

On définit la fonction profondeur, pr, sur les arbres de L, par

$pr(f_g(a,b))=pr(f_d(a,b))=1+pr(a,b)$, sous l'initialisation $pr(x,x)=1$.

Il y a clairement, dans L, 2^{n-1} arbres de profondeur n.

On remarquera que si (a,b) est dans L, son image dans le miroir $Mir(a,b)$ est dans L, $Mir(a,b)=(Mir(b),Mir(a))$. Evidemment, initialement $Mir(x)=x$, et l'on a $pr(Mir(a,b))=pr(a,b)$.

Il est utile d'initialiser ainsi la récurrence génératrice:

(x,y) est dans L, où x est distinct de y.

Alors, si (a,b) est dans L,

l'image de L par le morphisme injectif $\mu_{(a,b)}$ tel que $\mu_{(a,b)}(x)=a$, $\mu_{(a,b)}(y)=b$, est une partie de L:

on notera $L(a,b)$ ce langage, issu de (a,b).

Tandis que $L=L(x,y)$ est le langage général,

on a évidemment $\mu_{(a,b)}(x,y)=(a,b)$, $L(a,b)=\mu_{(a,b)}L(x,y)$.

Pour tout couple ((a,b),(a',b')) dans $L \times L$,

on a $\mu_{(a,b)} \circ \mu_{(a',b')} = \mu_{(a'',b')}$, où $(a'',b'') = \mu_{(a,b)} \cdot (a',b')$.

A tout arbre (a,b) correspond très naturellement le couple d'entiers (degré(a),degré(b)),

que l'on comparera, en notant d_x et d_y les degrés partiels, au couple d'entiers ($d_x(a,b),d_y(a,b)$).

Quelle est l'image de L (dans $\mathbf{N} \times \mathbf{N}$) par ces applications (degré(.),degré(.)) et ($d_x(\dots),d_y(\dots)$) ?

Toute fonction degré, d, est un morphisme dans les entiers additifs: on a $d(a,b)=d(a)+d(b)$;

initialement, $degré(x)=degré(y)=1$, et $d_x(x)=1=d_y(y)$, $d_x(y)=0=d_y(x)$.

Pour tout couple d'arbres, (a',b') et (a'',b''), il existe un arbre (a,b), de profondeur maximale, tel que les arbres (a',b') et (a'',b'') appartiennent à $L(a,b)$; on notera cet arbre $\inf((a',b'),(a'',b''))$.

Enfin, les arbres de L sont totalement ordonnés par la contrainte $(a,(a,b)) < (a,b) < ((a,b),b)$, et cela entraîne que si $(a,b) < (a',b')$, on ait globalement $L(a,b) < L(a',b')$.

Pour tout (a,b) de L, le morphisme injectif $\mu_{(a,b)}$ est monotone croissant.

Ainsi, L se réalise en un monoïde de morphismes, monotones, croissants, injectifs, de L dans L.

Sauriez vous, pour deux arbres (a,b) et (a',b') de L, décider si l'on a $(a,b) < (a',b')$, ou $(a,b) > (a',b')$?

Quel est le lien avec la suite de mots sur l'alphabet des entiers, où chaque entier n'apparaît qu'un nombre fini de fois, et où tout rationnel apparaît une unique fois comme rapport de deux entiers consécutifs,

1 1, 1 2 1, 1 3 2 3 1, 1 4 3 5 2 5 3 4 1, 1 5 4 7 3 8 5 7 2 7 5 8 3 7 4 5 1,
 1 6 5 9 4 11 7 10 3 11 8 13 5 12 7 9 2 9 7 12 5 13 8 11 3 10 7 11 4 9 5 6 1
 ?

Ici, chaque mot est sous-mot régulier du mot suivant, car chaque mot se déduit du précédent en insérant entre deux caractères consécutifs a et b la somme a+b de ceux-ci. C'est en ce sens que les occurrences de chaque entier n se stabilisent, n'apparaissant qu'un nombre fini de fois. La suite de ces nombres d'occurrences coïncide avec les valeurs de la fonction Phi qui porte le nom d'Euler (définie, pour chaque entier n, comme le nombre des entiers inférieurs à n et premiers à n).

Remarque 4 (et sujet de réflexion classique):

Tout arbre binaire de degré n peut être utilisé afin de coder un alphabet X de n lettres par un alphabet (g,d) de 2 lettres.

Ainsi, ((u,x),(y,(v,z))), arbre binaire sur (u,v,x,y,z), est interprétable comme la représentation du code f tel que:

$$f(u)=gg, f(x)=gd, f(y)=dg, f(v)=ddg, f(z)=ddd.$$

Par exemple, le mot m=yuuvxux sera représenté par le mot m'=f(m):

$$m'=f(yuuvxux)=f(y)f(u)f(v)f(x)f(u)f(x)=dgggggddggdgggd.$$

L'application f est un morphisme [du monoïde engendré par l'alphabet X=(u,x,y,v,x) dans celui engendré par l'alphabet (g,d)].

Un tel codage est dit préfixe gauche, car un unique facteur gauche de m' appartient à l'image de f, de telle manière que le décodage (le calcul de f⁻¹) puisse être produit sans retard lorsque l'on lit le mot m caractère par caractère, de gauche à droite.

Suivant la fréquence d'utilisation des lettres de X, un arbre peut être préféré à un autre, de telle sorte que le choix du code f peut être fait en fonction du mot m (que l'on suppose beaucoup plus long que dans l'exemple ci-dessus), de manière à produire un mot m'=f(m) particulièrement court (le plus court possible?).

En ce sens, le choix de f, ci-dessus, est-il optimal ?

On peut être amené à optimiser f, en le faisant éventuellement dépendre du mot m.

Voici des règles classiques conduisant, par récurrence, d'un alphabet X à un arbre binaire A dont les éléments atomiques sont les lettres de X:

- 1) on associe un poids p(a) à tout arbre binaire a;
- 2) étant donné une collection C d'arbres binaires, on remplace les deux arbres de plus faible poids (soit b et c) par leur composé binaire (b,c), le poids duquel est défini par p(b,c)=p(b)+p(c);
- 3) toute lettre x de X est un arbre de poids connu p(x);
- 4) en prenant X pour collection initiale, et en itérant la règle 2), on aboutit finalement à un arbre binaire A dont X est l'ensemble des éléments atomiques.

Cet arbre A définit un morphisme injectif f de X* dans (g,d)*; il est démontrable que si les poids p(a) [pour a dans X] sont les degrés des lettres de X dans un mot m, alors f est tel que le degré de f(m) soit minimal.

Exercice 22 (et sujet de réflexion):

Une machine est constituée de deux mémoires. Au temps t, l'une contient le mot m, l'autre le mot m'. Au temps t+1, deux possibilités: soit m:=mm', soit m':=m'm; on considère le langage constitué des mots m et m' ainsi engendrés, à partir du temps t=0 auquel m=a et m'=b.

On construit ainsi des mots qui sont les débuts de suites infinies. Comment décider si un mot donné appartient à ce langage ?

Tout d'abord: que dire du nombre de mots de degré n ?
 Cas particulier: on modifie alternativement m et m' .

Réaliser, en cheminant sur un écran, a par un déplacement horizontal (resp. ou un virage à droite), et b par déplacement vertical (resp. ou un virage à gauche)

Modifier ainsi: au temps $t+1$, le couple (m, m') est modifié par $m := mm'$, ou $m' := mm'$.

\

Exercice 23:

A. Considérer le langage des mots (du monoïde engendré par les deux lettres x et y) indexés par les couples d'entiers tels que $f(i, j) = f(i, j-1) \cdot f(i-1, j)$, récurrence initialisée par $f(i, 0) = x$ pour $i > 0$ et $f(0, j) = y$ pour $j > 0$; ainsi, $f(1, 1) = xy$, $f(2, 1) = xxy$, $f(2, 2) = xxyyyx$. Le coefficient binomial (i, j) compte le nombre de lettres (le degré total) du mot $f(i, j)$.

Considérer le nombre d'occurrence $O(x)$ de x (resp $O(y)$ de y) dans le mot $f(i, j)$ de ce langage. Calculer et tabuler ces fonctions.

B. Considérer le langage constitué des arbres binaires $h(i, j)$ tels que $h(i, 0) = x = h(0, j)$, et $h(i, j) = (h(i, j-1), h(i-1, j))$. Le coefficient binomial (i, j) compte le nombre d'occurrences de x (le degré de x) dans l'expression de $h(i, j)$.

Ecrire un programme qui vérifie qu'un arbre donné appartient à ce langage d'arbres; on est amené à examiner de manière récurrente, en partant de l'arbre donné, une suite (a) de n arbres a_1, a_2, \dots, a_n telle que la partie droite $d(a_i)$ de chaque arbre soit identique à la partie gauche de l'arbre suivant, $g(a_{i+1})$; la suite (b) des arbres $b_{i+1} = d(a_i) = g(a_{i+1})$, $b_1 = g(a_1)$, $b_{n+1} = d(a_n)$, est soumise à la même condition; la suite est raccourcie lorsque son premier arbre est x (resp son dernier est y); le processus cesse lorsque la suite d'arbres est vide.

\

Exercice 24:

Soit la suite des polynômes, sur le magma engendré par x [resp sur le monoïde engendré par x], définis par la récurrence $P_0 = x$, $P_1 = x$, $P_{i+1} = x + (P_{i-1}, P_i)$ [resp $P_{i+1} = x + P_{i-1} \cdot P_i$].

Vers quelle série cette suite de polynômes tend-elle ?

\

Exercice 25:

Les langages du style de celui qui suit ont été étudiés par certains mathématiciens durant ce siècle.

Pour tout arbre binaire $a = (b, c)$, distinct du générateur x , on note $g(a)$ son facteur gauche b , et $d(a)$ son facteur droit c . Construire (par programme) les premiers objets d'un langage L , totalement ordonné, défini de proche en proche par une suite d'arbres binaires de la manière suivante:

- 1)- x appartient à L
- 2)- (x, x) appartient à L , et $x < (x, x)$
- 3)- si les arbres a et b appartiennent à L ,
 - avec $a < b$ et $g(b)$ inférieur ou égal à a ,
 - alors (a, b) appartient à L ,
 - et l'on a $a < (a, b)$, ainsi que $b < (a, b)$.

Il est naturel de construire un tel langage de manière compatible avec l'ordre énumératif, c'ad

- le générateur x est inférieur à tout (a, b)
- (a, b) est inférieur à (a', b') dès que b est inférieur à b'
- si $b = b'$ et $a < a'$, alors $(a, b) < (a', b')$.

En fonction des degrés, le nombre d'arbres obtenus est 1 1 2 3 6 9 18 30 99 186 335... \

Exercice 26:

Considérer la suite de fonctions $f(n,t)$ définies sur les entiers naturels \mathbb{N} par
 -l'initialisation $f(n,0) = 0$ pour $n > 0$ et $f(0,0) = 1$
 -la récurrence $f(n,t+1) = f(n-1,t) + f(n+1,t)$ pour $n > 0$ et $f(0,t+1) = f(1,t)$.
 Observer la suite $f(0,t)$.

\

Exercice 27:

Tout arbre binaire, représenté par son expression magmatique binaire, définit un morphisme M_a opérant sur le magma. Par exemple $M_{(x,x)}$ opérant sur $(x,(x,x))$ donne $((x,x),((x,x),(x,x)))$. M_x est l'identité. Tout arbre qui appartient à l'image de M_a pour a distinct de x est non-primitif.

Compter les arbres primitifs en fonction de leur degré.

Nous ferons de même sur le magma général.

Voir Livre III chapitre 7, divers, transformation de Möbius et séries multiplicatives.

\

C'est tout naturellement que l'on peut définir l'axiomatique (par définition récurrente, ou procédure récursive) du produit de mélange, commutatif et associatif, qui sera encore noté "uu", sur le magma binaire monogène engendré par x:

$$x \text{ uu } (g,d) = (x \text{ uu } g,d) + (g, x \text{ uu } d) \quad \text{et} \quad x \text{ uu } x = 2 (x,x)$$

$$(a,b) \text{ uu } (g,d) = ((a,b) \text{ uu } g,d) + (g,(a,b) \text{ uu } d) + (a \text{ uu } g,b \text{ uu } d) + (a \text{ uu } (g,d),b) + (a,b \text{ uu } (g,d))$$

Un élément neutre est éventuellement adjoint, et alors noté e .

C'est dire que

- soit l'un des arbres s'insère en totalité dans le facteur droit ou dans le facteur gauche de l'autre
- soit les parties gauches (resp droites) se mélangent.

Le lecteur, dans un premier temps, s'étonnera peut-être de l'axiome $x \text{ uu } x = 2 (x,x)$.

C'est qu'il faut assurer la compatibilité avec le magma pourvu de plusieurs générateurs, sur lequel nous aurons, naturellement:

$$x \text{ uu } y = (x,y) + (y,x)$$

relation dont $x \text{ uu } x = 2 (x,x)$ est l'image par le morphisme alphabétique qui envoie y sur x .

L'opération de mélange sur les arbres binaires respecte le degré: le mélange d'un arbre de degré i et d'un arbre de degré j produit des arbres de degré $i+1$.

Comme dans le monoïde le mélange de x^i et x^j produit x^{i+j} avec pour multiplicité le coefficient binomial que nous notons (i,j) , le coefficient (en un sens généralisé) du mot m dans le mélange de m' et m'' est bien défini, il dépend des trois mots m, m', m'' , et c'est le coefficient de m dans le polynôme $m' \text{ uu } m''$, que nous noterons $\langle m, m' \text{ uu } m'' \rangle$. De même, la donnée de trois arbres, a, a', a'' , définit un coefficient: le coefficient de a dans le mélange de a' et a'' , $\langle a, a' \text{ uu } a'' \rangle$.

(Plus généralement, pour tout polynôme p et série s , le produit $\langle s,p \rangle$ est bien défini, entraînant la dualité des opérateurs: pour tout opérateur w sur les polynômes, son dual w^t opérant sur les séries est défini, par $\langle w^t s, p \rangle = \langle s, w p \rangle$ pour tout couple (s,p) où p est une série et p un polynôme).

\

Complément important: une voie vers la formule de dénombrement close.

Si c_n est le nombre d'arbres (càd d'expressions du magma binaire) de degré n , on a la récurrence remarquable

$$(n+1)c_{n+1} = 2 * (2n-1)c_n, \text{ sous l'initialisation } c_1 = 1.$$

Il est possible de construire une preuve directe bijective et constructive de cette relation récurrente. A gauche, $(n+1)c_{n+1}$ compte les arbres de degré $n+1$ munis d'un sommet terminal (càd une occurrence de x , dans le mot de codage magmatique) pointé. A droite, $(2n-1)c_n$ compte les arbres de degré n munis d'un sommet (terminal ou non) pointé. En effet, un arbre de degré n compte n sommets terminaux et $n-1$ sommets internes, soit $2n-1$ sommets au total (dans le codage de type magma, les sommets internes correspondent aux occurrences de la virgule).

Le coefficient 2 devant $(2n-1)c_n$ impose de marquer le sommet pointé de deux manières possibles.

On construira, bijectivement, à partir de cet arbre de degré n ainsi pointé et marqué, un arbre de degré $n+1$ muni d'un sommet terminal pointé. Sous le sommet pointé se trouve une branche d'arbre; on fait pousser un chevron sous le sommet pointé, et la branche est raccordée à gauche ou à droite, suivant l'état du marquage.

Ainsi, par exemple, x donne $(x,x)+(x,x)$, puis (x,x) donne $((x,x),x)+((x,x),x)$, (x,x) donne $(x,(x,x))+((x,x),x)$, et (x,x) donne $(x,(x,x))+((x,x),x)$

Dans toute expression de type magma binaire, les occurrences de x représentent les sommets du fond de l'arbre (appelés parfois "feuilles", que l'on pourrait aussi appeler "bourgeons"), les occurrences de la virgule représentent les sommets internes (que l'on pourrait appeler "noeuds" ou "fourches"); tandis que les occurrences du caractère parenthèse ouvrante (resp fermante) représente les "branches gauches (resp droites)".

L'expression magmatique apparaît alors comme la 'projection' de l'arbre binaire.

Pointer un sommet de l'arbre, c'est distinguer soit une occurrence de x , soit une occurrence de virgule, suivant que le sommet pointé est un sommet du fond ou un sommet interne.

Remarquer que la 'racine' de l'arbre est représentée par x lorsque l'arbre est réduit à x , et par une virgule dans tout autre cas.

Il est clair que du dénombrement récurrent $(n+1)c_{n+1} = 2 * (2n-1)c_n$ sous l'initialisation $c_1 = 1$ résulte une expression close pour c_{n+1} , soit $c_{n+1} = 2^{n+1} f(n)/g(n)$,

où $f(n) = (2i-1) =$ produit des $2i-1$ pour i de 1 à $n =$ produit des entiers impairs de 1 à $2n-1$, que l'on peut nommer 'factorielle d'impairs', et $g(n) =$ produit des entiers de 1 à $(n+1)$, càd $(n+1)!$

Nous y reviendrons: en fait, si P_n est le polynôme caractéristique des arbres binaires de degré n (en x), sous leur représentation magmatique, et D la dérivation $Dx = e$ (en introduisant ici l'élément neutre e de la composition binaire qui sert à définir les expressions du magma), on a $DP_{n+1} = 2 * (2n-1)P_n$.

On rappelle que toute opération de dérivation D est une opération linéaire qui satisfait à $D(a,b) = (Da,b) + (a,Db)$. On a, en tout cas, $e.a = a.e$ pour tout arbre a , et $De = 0$ pour toute dérivation D .

Nous n'avons considéré que le magma engendré par un unique générateur, x .

Nous serons amenés à considérer à plusieurs reprises le magma engendré par deux générateurs, x et y , et tout spécialement le langage d'arbres défini ainsi: (x,y) est dans le langage, et si m est une expression du langage, on obtient deux expressions du langage en substituant (x,y) à toute occurrence de x (resp de y).

Exercice 28 (des sous-ensembles particuliers du magma, langages d'arbres munis d'ordre total):

1) Soit, par commodité, deux générateurs atomiques, x et y , et le langage L , totalement ordonné, des arbres engendrés par la récurrence que voici, contrainte par les conditions suivantes:

- l'ordre (noté $<$) est strict, et, notamment, $x < y$

- pour tout arbre a de L distinct de x et y , on a $x < a$ et $a < y$: x est le plus petit élément, et y le plus grand

- si a et b sont deux arbres distincts de L , l'arbre (a,b) appartient à L si et seulement si $a < b$, strictement

-les arbres sont comparés, classiquement, d'abord par leur partie gauche, et, si leurs parties gauches sont identiques, par leurs parties droites: $a < b$ entraîne $(a,c) < (b,d)$ et $(c,a) < (c,b)$. Ainsi, sauf erreur,

$$x < (x, (x, (x, y))) < (x, (x, y)) < (x, ((x, y), y)) < (x, y) \\ < ((x, (x, y)), ((x, y), y)) < ((x, (x, y)), y) < ((x, y), y) < (((x, y), y), y) < y$$

Quel est le nombre $f(n)$ d'arbres ainsi engendrés dont le degré est n ? Sous $f(1)=2, f(2)=1, f(3)=2$, on a $f(n) =$ somme des $f(i)f(j), i+j=n$ pour $0 < i < j$, plus $f(k)(f(k)-1)/2$ lorsque $n=2k$.

Le début est **(2 1 2 4 10 25 68 187...)**.

Dans le monoïde, on obtient en concaténant chaque mot à ceux qui précèdent

$x y xy xxy yxy xxxy yxxx xyxxx xyxy yyxy xyxy xxyyxy xxxxy yxxxx xyxxxx xxyxxx yxyxxx$
 $xyxxx yyxy xyxy xxyyxy yxyxxx xxxxyxxx xxyxxx yxyxxx xyxyxxx xxyyxxx yxyxyxxx$
 $xxxxyxxx yxyxyxxx...$

Et sur trois lettres $x < y < z$,

$x y z xy xz yz xxy yxy zxy xxx yxz zxz xyxz xyz yyz zyz xyyz xzyz xxxy yxy zxxx xyxxx xzxxx$
 $yzxxx xyxy yyxy zyxy xyxy xzyxy yzyxy xxyxy...$

2) En partant d'un unique générateur, x , soit le langage ordonné L' tel que a et b dans L' avec a inférieur ou égal à b entraîne (a,b) dans L' ; l'ordre est la comparaison prioritaire des parties gauches. Ce langage L' est représentatif du magma commutatif engendré par x .

Le nombre $f(n)$ d'arbres de degré n de L' satisfait à

$f(n) =$ somme des $f(i)f(j)$ pour $0 < i < j$, plus $f(k)f(k+1)/2$ lorsque $n=2k$, sous l'initialisation $f(1)=f(2)=1$

(1 1 1 2 3 6 11 23 46 98 207 451 983....).

On en déduit une équation de ce que l'on appelle "la série génératrice", que l'on écrira encore en une variable x (associative), série dont les coefficients comptent le nombre des arbres de degré donné,

$$S^2(x) + S(x^2) = 2(S(x) - x).$$

3) Si l'on se donne initialement les arbres x et $(x,x), x < (x,x)$, et que l'on ne compose deux arbres, a et b , pour produire l'arbre (a,b) , que lorsque le degré de a est strictement inférieur celui de b -càd que l'on exclut notamment de produire des carrés (a,a) - on obtient un certain langage L'' .

En classant suivant leurs degrés les arbres de ce langage L'' , la suite qui les dénombre, $f(1)=1, f(2)=1, f(n) =$ somme des $f(i)f(j)$ pour $i+j=n$ et $i < j$, débute par **(1 1 1 1 2 3 6 11 22 44 89...)**.

4) Si l'on part des mots x et xy , et que l'on construit une suite de mots en concaténant chaque mot avec ceux qui précèdent, on obtient

$x xy xxy xxxy yxy xxxxy yxxxx xyxxxx xxyxxx xyxyxxx xxyyxxx xxxxyxxx xxxxyxxx$
 $xyxxxx xxyxxxx xxxxxxx xyxyxxxx xxyxxx xyxyxxx xxyyxxx xxxxyxxx xxxxyxxx$
 $xyxyxyxxx xxxxyyxxx xxxxyxxx xyxyxxx xxyyxxx xxxxyxxx xyxyxyxxx$
 $xxxxxyxxx xyxyxyxxx....$ langage du monoïde qui comporte cependant des carrés.

Le nombre de mots de degré n est $f(1)=1, f(2)=1,$

$f(n) =$ somme des $f(i)f(j)$ pour $i < j +$ (si $n=2k, (f(k)-1)f(k)/2$),

(1 1 1 1 2 3 6 11 22 43 88 179 372 774 1631 3448 7347 15713 33791 72923...)

Exercice 29 (un découpage naturel des nombres de catalan):

Soit \mathbf{A} le langage d'arbres engendré par x et y de la manière suivante:

-initialisation: (x,y) est un arbre de \mathbf{A} ,

-récurrence: si t et u sont des arbres de \mathbf{A} , $(u,t), (x,t)$ et (t,y) sont des arbres de \mathbf{A} .

Si l'on préfère, tout est résumé par la grammaire générative algébrique, en terme de série formelle,

$$\mathbf{S} = (\mathbf{x}, \mathbf{y}) + (\mathbf{x}, \mathbf{S}) + (\mathbf{S}, \mathbf{y}) + (\mathbf{S}, \mathbf{S})$$

série S que l'on peut approcher par la suite de polynômes

$$S(1)=(x,y), S(k+1)=(x,y)+(x,S(k))+(S(k),y)+(S(k),S(k))$$

mais aussi par la suite énumérative

$$(x,y), (x,(x,y)), ((x,y),y), ((x,y),(x,y)), (x,(x,(x,y))), ((x,(x,y)),y), ((x,(x,y)),(x,(x,y)))...$$

où chaque expression est composée à gauche avec x, à droite avec y, puis avec toute expression qui précède, et enfin avec elle-même.

Les occurrences de x (resp. y) représentent les sommets du fond ou atomes (parfois appelés feuilles de l'arbre, mais aussi pendants) gauches (resp. droits).

En notant $f(i,j)$ le nombre d'arbres $A_{i,j}$ du langage A dont les degrés en x et y sont i et j,

-conventionnellement, nous posons $f(0,1)=f(1,0)=1$ - puis évidemment $f(1,1)=1$;

ensuite, nous avons pour $i > 1, j > 1, f(1,j)=f(i,1)=1$, et, essentiellement

$$f(i,j)=f(i-1,j)+f(i,j-1)+(somme (f(a,b).f(c,d)), a+b=i,c+d=j, a>0,b>0,c>0,d>0).$$

$f(n+1,n+1)$ est le nombre d'arbres de $A_{n+1,n+1}$, càd de degrés en x et y identiques, $n+1$.

$$\text{Ce nombre est } f(n+1,n+1)=(2n+1).c(n+1)^2$$

$$(=1 \ 3 \ 20 \ 175 \ 1764 \ 19404 \ 226512 \ 2760615 \ 34763300 \ 449141836, \dots)$$

où $c(n+1)=(2n)!(n! (n+1)!)$ est le nombre de catalan usuel, càd le nombre des arbres $M_{n+1}(z)$ du magma monogène $M(z)$ engendré par z qui soient de degré $n+1$ en z, tandis que $2n+1$ est le nombre des sommets de ces arbres ($n+1$ sommets terminaux, et n sommets internes).

D'où la recherche d'une construction bijective explicite de $[2n+1].M_{n+1}.M_{n+1}$ sur $A_{n+1,n+1}$.

Ainsi, au produit cartésien $[3]_x \{ (z,z) \}_x \{ (z,z) \}$ correspondra l'ensemble des trois arbres de degré 2 en x comme en y: $\{ ((x,(x,y)),y), ((x,y),(x,y)), (x,((x,y),y)) \}$.

formule close générale: $f(i+1,j+1)=(j+i+1)!(j+i)! / ((j!(j+1)!i!(i+1)!);$

Voici le début, pour i et j de 1 à 10, de la fonction $f(i,j)$, nombres parfois dits "de Narayana":

1	1	1	1	1	1	1	1	1	1
1	3	6	10	15	21	28	36	45	55
1	6	20	50	105	196	336	540	825	1210
1	10	50	175	490	1176	2520	4950	9075	15730
1	15	105	490	1764	5292	13860	32670	70785	143143
1	21	196	1176	5292	19404	60984	169884	429429	1002001
1	28	336	2520	13860	60984	226512	736164	2147145	5725720
1	36	540	4950	32670	169884	736164	2760615	9202050	27810640
1	45	825	9075	70785	429429	2147145	9202050	34763300	118195220
1	55	1210	15730	143143	1002001	5725720	27810640	118195220	449141836

.....

avec, sur la diagonale du tableau,

$$1 \ 3 \ 20 \ 175 \ 1764 \ 19404 \ 226512 \ 2760615 \ 34763300 \ 449141836...$$

sur la ligne 2, $f(2,j)=j(j+1)/(1*2)=p_2(j)/p_2(1)$,

sur la ligne 3, $f(3,j)=j(j+1)^2(j+2)/(1*2^2*3)=p_3(j)/p_3(1)$,

sur la ligne 4, $f(4,j)=j(j+1)^2(j+2)^2(j+3)/(1*2^2*3^2*4)=p_4(j)/p_4(1)$,

$$p_i(j)=j(j+1)^2(j+2)^2 \dots (j+i-2)^2(j+i-1), f(i,j)=p_i(j)/p_i(1)$$

Evidemment, sur les diagonales finies, (somme des $f(i,j)$ pour $i+j=n$)=le nombre de catalan $c(n+1)$.

Anticipons (ceci nécessite la lecture du chapitre sur les permutations):

Sur chaque arbre du langage A, on peut inscrire un certain nombre de permutations, selon la méthode

classique qui associe un arbre binaire à toute permutation de $[n]$: la racine de l'arbre est marquée par la plus petite lettre, 1, et le mot de permutation se factorise en $m1m'$, m marque la granche gauche, m' le branche droite, la plus petite lettre va sur la racine, etc...

Soit $g(i,j)$ le nombre de permutations inscrites sur les $f(i,j)$ arbres de degrés (i,j) en (x,y) ; ces permutations possèdent $i-1$ montées et $j-1$ descentes (une montée en position k dans la permutations p est telle que $p(i)<p(i+1)$), et l'on a $g(i,j)=i.g(i,j-1)+j.g(i-1,j)$, $g(i,1)=g(1,j)$, $i>1,j>1$, $g(1,1)=1$, (aisé à établir par récurrence, par insertion d'une nouvelle lettre, $n+1$, dans les permutations sur $[n]$) Ces nombres portent le nom d'Euler. Début du tableau des $g(i,j)$:

1	1	1	1	1	1	1	1
1	4	11	26	57	120	247	502
1	11	66	302	1191	4293	14608	47840
1	26	302	2416	15619	88234	455192	2203488
1	57	1191	15619	156190	1310354	9738114	66318474
1	120	4293	88234	1310354	15724248	162512286	1505621508
1	247	14608	455192	9738114	162512286	2275172004	27971176092
1	502	47840	2203488	66318474	1505621508	27971176092	447538817472...

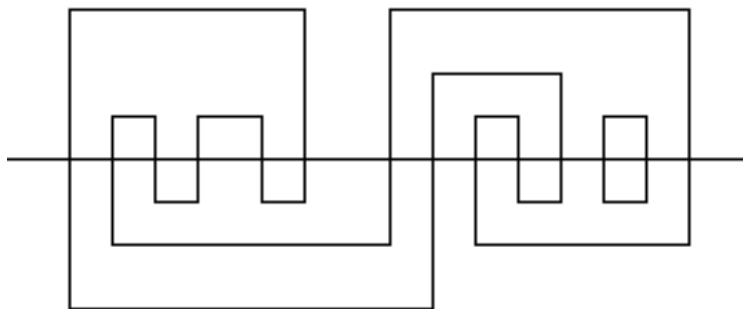
tableau dont la diagonale 1 4 66 2416 156190 15724248 2275172004 447538817472..., compte les permutations de $[2n+1]$ qui présentent autant de montées que de descentes.

La série de terme général $x^i y^j g(i,j)$, pour $i>0, j>0$, est l'image commutative de la série non-commutative s dont le coefficient $\langle s,m \rangle$ d'un mot m se calcule ainsi par récurrence de gauche à droite: $\langle s,xy \rangle=1$, $\langle s,xyz \rangle=\langle s,xym \rangle^*$ (degré de m en y si $z=x$, en x si $z=y$).

Anticipons encore:

$c(n+1)^2$, le carré des nombres de catalan, s'interprètera naturellement comme couple de mots de parenthèses, mots assemblés lettre à lettre de manière à former des configurations parfois appelées méandres qu'il est intéressant de représenter et dénombrer.

Petit dessin interprétatif: au-dessus d'une ligne horizontale, vous dessinez l'emboîtement spécifié par le premier mot de parenthèses, et l'emboîtement du second mot sous la ligne, si bien que l'on obtient la section de courbes planes non sécantes par une droite. Refermez la droite sur elle-même, et vous obtenez un graphe planaire dont tous les sommets sont de degré 4, graphe muni d'un circuit hamiltonien, càd d'un chemin refermé sur lui-même qui passe par tous les sommets. Mais c'est une autre histoire.



(aababbaabbabb, aaababbbaababb)

Note: De même que nous avons défini sur les arbres binaires les sommets du fond gauches et droits, correspondant aux occurrences de x et y dans les expressions magmatiques, on définit naturellement les sommets internes gauches et droits (parfois appelés fils internes gauches et droits) -excluant la racine- : la statistique des arbres suivant le nombre de sommets internes gauches et droits se ramène naturellement à celle des arbres selon les sommets du fond, atomiques, gauches et droits.

Nous savons que les permutations s'inscrivent comme fonction croissante sur les sommets internes; considérer les permutations de $(0,1,\dots,n)$ telles que les entiers pairs s'inscrivent sur les sommets internes gauches (et 0 sur la racine), les entiers impairs sur les sommets internes droits.

Exercice 30:

Etablir de diverses manières que les entiers impairs de la suite fondamentale de catalan
 $c(1)=c(2)=1, c(4)=5, c(8)=429, c(16)=9694845\dots$

sont exclusivement ceux de rang 2^n

(il n'y a un nombre impair d'arbres binaires de degré k que pour k de la forme 2^n).

Autrement dit, la série commutative $s=x+s^2$ à coefficients modulo 2, c'est $s=\text{somme des } x^{2^n}$, comme cela se vérifie dans l'instant, en itérant $s:=x+s^2$, en partant de l'initialisation $s:=x$.

On cherchera une transformation involutive sur les arbres binaires, dans laquelle seuls les arbres homogènes de profondeur n sont invariants.

A un autre moment, nous calculerons la fonction "plus grande puissance de 2 divisant chaque coefficient binomial", et en particulier étudierons sa valeur sur les binômiaux centraux, dont dépendent si étroitement les nombres de catalan.

Exercice 31:

Rappelons que

le magma binaire monogène satisfait à l'équation caractéristique non associative $C=x+C^2$ (que nous qualifierons d'équation fondamentale de catalan), et les nombres de catalan $c(n)$ sont les coefficients de la série solution de l'équation associative $c=x+c^2=\text{somme des } c(n).x^n$ pour $n>0$. Notons F le morphisme $F(x)=C$, et f le morphisme $f(x)=c$.

On cherche à dénombrer (suivant leur degré) les arbres binaires, représentés par les expressions du magma binaire, dont le langage G satisfait à la récurrence:

si a est dans G , alors $(a,F(a))$ est inclus dans G ,

récurrence initialisée par

x est dans G .

$(1 \ 1 \ 1 \ 3 \ 7 \ 20 \ 59 \ 186\dots)=\text{coefficients de la série } g=\text{somme des } g(n).x^n \text{ pour } n>0$.

En effet,

on a pour la série associative g , dénombrant les arbres du langage G suivant leur degré, la récurrence:

$\text{somme pour } n>0 \text{ de } g(n).(xc)^n=\text{somme pour } n>0 \text{ de } g(n).x^n; g(1)=1$.

C'est dire que $f(g)=g-x$.

D'où la disposition du calcul,

1	1	1	2	5	14	42	132...	pour	xc
1		1	2	5	14	42...		pour	le carré de xc
1			1	3	9...			pour	le cube de xc
3					1...			pour	la puissance quatrième de xc
7									puissance cinquième de $xc\dots$
.....et coetera...									

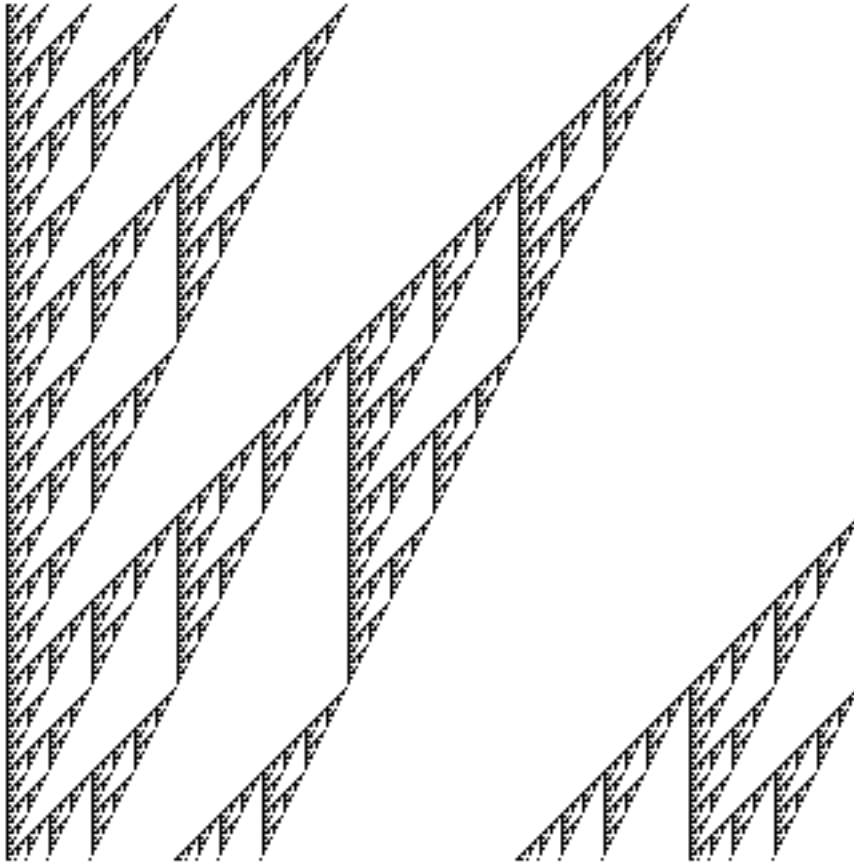
$1 \ 1 \ 3 \ 7 \ 20 \ 59 \ 186\dots$ coefficients $g(n)$ pour $n>1$, obtenus par sommation verticale, chaque ligne de rang i étant multipliée par le coefficient $g(i)$ qui figure à gauche.

Les puissances de C (et donc de même pour l'image associative c) sont exprimables sous la forme $C^{n+1}=p_n C+r_n$, où p_n et r_n sont des polynômes en x de degrés croissants qui satisfont à $(p_{n+1},r_{n+1})=(p_n+r_n, -xp_n)$, soit l'itération $(p,r):=(p+r,-xp)$, initialisée par $(1,0)$.

Exercice 32:

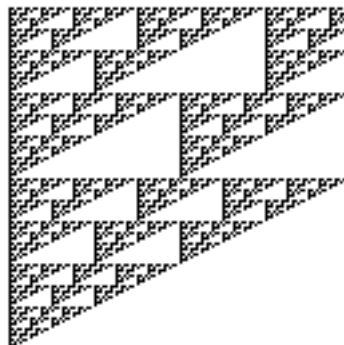
On calcule le tableau dont la première ligne est la suite des coefficients de la série de catalan prise sous la forme $s=e+x+2x^2+5x^3\dots$ et la k-ième ligne la suite des coefficients de s^k .

Ci-dessous, l'image 321*321 de ces coefficients, pris modulo 2:



et ce que l'on obtient avec les puissances de la série des entiers de fibonacci,

$$s=e+x+2x^2+3x^3+5x^4+8x^5\dots$$



tandis que la fractale usuelle de sierpinski est obtenue à partir des puissances $(e-x)^{-1}x^*$, càd le tableau des coefficients binômiaux.

\

Exercice 33:

Soit un alphabet de k lettres (éventuellement identifiées aux entiers de 1 à k) et initialement k mots, de $m(1)$ à $m(k)$; soit le morphisme μ qui opère la substitution, dans chaque mot $m(i)$, du mot $m(j)$ à la lettre j , puis l'itération parallèle synchrone $m(i) := \mu(m(i))$; par exemple, sur deux lettres, 1 et 2, partant de $(m(1)=12, m(2)=21)$, on obtient $(1221, 2112)$, $(12212112211221221, 2112122112212112)$

Dans cet exemple, chaque mot est facteur gauche du suivant auquel il donne naissance, tandis que le second mot du couple se déduit du premier par l'involution qui échange les deux lettres de l'alphabet. Peut-on prévoir ce qu'est la i -ème lettre de la suite infinie obtenue ?

Avec des arbres -binaires- soit initialement $(a(x)=(x,y), a(y)=(y,x))$, puis le couple d'arbres $((x,y),(y,x), ((y,x),(x,y)))$, et à l'étape suivante les deux arbres $((((x,y),(y,x)),((y,x),(x,y))),(((y,x),(x,y)),((x,y),(y,x))))$, $((((y,x),(x,y)),((x,y),(y,x))),(((x,y),(y,x)),((y,x),(x,y))))$

\

Exercice 34 (voir l'exercice 20):

Une bijection -naturelle- des (arbres binaires commutatifs engendrés par x) sur (les entiers positifs),
et, du même coup,
une injection -naturelle- des (arbres binaires commutatifs) dans (les arbres binaires).

La fonction f envoie les entiers positifs dans les arbres binaires représentatifs des arbres commutatifs:

soit $f[1] = x$, et $f[n] = (g[n], d[n])$, $g[n] = f[i[n]]$, $d[n] = f[j[n]]$, $i[2] = j[2] = 1$.

On a la récurrence $f[n+1] = (f[i[n]+1], f[j[n]])$ si $i[n] < j[n]$, et $(f[1], f[j[n]+1])$ si $i[n] = j[n]$.

\

2. Le Magma généralisé.

Définition 1:

On appelle magma général, ou magma généralisé sur l'alphabet de quatre lettres

$x (,)$

le langage M' défini par la récurrence grammaticale générative suivante:

pour tout entier $n > 0$

si les n mots m_1, m_2, \dots, m_n appartiennent à M' ,

le mot (m_1, m_2, \dots, m_n) appartient à M' ;

cette récurrence est initialisée par:

le mot réduit à la lettre x appartient à M' .

Le langage M' est défini par l'équation de sa série génératrice

$$M' = x + (M') + (M', M') + (M', M', M') + \dots$$

soit, en notant $(M')^n$ la composition n -aire d'éléments de M' ,

$$M' = x + \text{somme pour } n > 0 \text{ des } (M')^n.$$

Ce langage peut être considéré comme la définition axiomatisée de la notion générale d'arbre.

Cette définition, ou grammaire générative, est une récurrence initialisée qui est un procédé permettant d'engendrer des arbres, presque un programme ou une procédure de génèse au hasard. Inversement, on peut y voir un procédé d'analyse syntactique.

Nous dirons "arbre", suivant les habitudes des programmeurs, et afin de ne pas alourdir le vocable. Cependant, au sens de la théorie des graphes, il s'agit toutefois d'arbres dessinés sur le plan et munis d'une arête pointée et orientée. Les graphes et les graphes représentés - ou dessinés - sont des objets distincts, qui ne doivent pas être confondus, comme nous serons amenés à le souligner à diverses reprises. En général, le contexte permet d'éviter les confusions, d'autant plus que les informaticiens ont une inévitable prédilection pour les objets dessinés et pointés, seuls à permettre une représentation mathématico-linguistico-informatique canonique.

Autrement dit, suivant la définition des arbres généraux annoncée ci-dessus,

x est un arbre, et toute suite non vide d'arbres est un arbre.

Si l'arbre $m = (m_1, m_2, \dots, m_n)$ est composé de n arbres, cette composition n -aire définit ce que l'on peut appeler un "sommet de degré n ". Ainsi, l'arbre x définit un sommet de degré 0. Par exemple, l'arbre $((x, x), (x))$ possède un sommet de degré 3, un sommet de degré 2, un sommet de degré 1, et trois sommets de degré 0. Le sommet de degré 3 est sa racine, et les sommets de degré 0 sont les sommets terminaux, ou atomes de l'expression, ou feuilles. Tout arbre de la forme (m) est dit "planté".

On remarquera que le Magma binaire constitue tout naturellement une partie du Magma généralisé (plus rigoureusement: il existe une évidente injection -on parle d'injection canonique- du Magma M dans le Magma généralisé M').

Note:

Note:

Dans la théorie des ensembles, il est d'usage d'identifier la suite des entiers dits naturels à la suite des objets,

$$x; (x); (x, (x)); (x, (x), (x, (x))); (x, (x), (x, (x)), (x, (x), (x, (x))))...$$

construits à partir de l'objet atomique x , lequel est alors appelé "ensemble vide". On a là un processus définissant tout nouvel objet comme l'énumération ordonnée de ceux qui précèdent, le procédé étant amorcé par un objet initial.

Les entiers "naturels" n'ont-ils pas été simplement camouflés dans le vocable "suite", ou "passage au suivant", faisant implicitement référence à un temps discrétisé;

sorte de "Deus ex-machina", lequel, pendant qu'il y est, tire également du néant " l'atome x " ?

En termes d'arbres du magma généralisé, chaque arbre de cette suite est le composé de ceux qui précèdent.

Exercice 0:

Faut-il donc écrire un programme qui imprime la suite des entiers, au sens de la théorie des ensembles ? Il serait alors bien utile de disposer d'une machine capable de relire ce qu'elle a écrit (par exemple, une parmi les universelles machines de Turing, machines déterministes mobiles d'états finis, se déplaçant sur un environnement amorphe, capables de lire, relire, et modifier l'état de cet environnement).

Exercice 1:

En partant de x , soit $p/2$ la probabilité de substitution de (x) à x , $p/4$ celle de la substitution de (x, x) à x , $p/8$ celle de la substitution de (x, x, x) à x , etc.. Le processus se voudrait être indéfiniment itéré. Le simuler par programme expérimental.

Les arbres homogènes d'ordre 3 sont définis par les objets du magma général satisfaisant à la récurrence:

- x est un objet correct, et
- si x, y, z sont des objets corrects, alors (x, y, z) est un objet correct.

On définit de manière évidente les arbres homogènes d'ordre k :

- en termes de série, si S_k est la série des arbres homogènes d'ordre k , on a $S_k = x + S_k^k$
- ou encore par le morphisme g tel que $g[x] = x + g[(x, x, \dots, x)]$, où (x, x, \dots, x) , abrégéable en x^k , compte k occurrence de la lettre x .

Les sommets d'un arbre homogène d'ordre k sont donc de degré k ou 0 .

Degré et profondeur:

Le degré d'une arbre m (ou, plus exactement, le degré de l'expression du magma généralisé qui représente cet arbre) est défini comme la somme des degrés de ses constituants, soit

$$\text{deg}[(m_1, m_2, \dots, m_n)] = \text{deg}[m_1] + \text{deg}[m_2] + \dots + \text{deg}[m_n], \text{ sous l'initialisation } \text{deg}[x] = 1.$$

La profondeur de l'arbre $a = (m_1, m_2, \dots, m_n)$ est définie comme

$$1 + \text{le maximum de la profondeur de ses constituants, soit} \\ \text{prof}[(m_1, m_2, \dots, m_n)] = 1 + \max[\text{prof}[m_1] + \text{prof}[m_2] + \dots + \text{prof}[m_n]], \\ \text{sous l'initialisation } \text{prof}[x] = 1.$$

On considérera également la fonction "nombre d'arêtes" définie par

$$\text{ar}[(m_1, m_2, \dots, m_n)] = n + \text{ar}[m_1] + \text{ar}[m_2] + \dots + \text{ar}[m_n], \text{ sous l'initialisation } \text{ar}[x] = 0.$$

\ Exercice 2:

Programmer le calcul du degré et de la profondeur d'une expression du Magma généralisé. Essayer d'énumérer (engendrer), et dénombrer, les arbres de degré n (ou n au plus) et de profondeur p au plus (d'où ceux de profondeur p exactement).

Puis, considérer et dénombrer les arbres à profondeur inférieure à k , en fonction du nombre n d'arêtes. Soit donc $f_k(n)$ le nombre d'arbres a tels que $ar(a)=n$ et $prof(a)<k$. Etudier notamment la suite $f_k(n)$ pour $k=3,4,5,6$.

La grammaire générative (l'équation) des arbres homogènes d'ordre k est $M=x+M^k$. Soit h_n (resp. s_n) le nombre de ces arbres dont la profondeur est exactement (resp. au plus) k .

Par programme, on calculera ces suites, soumise aux relations de récurrence $h_{n+1}=(h_n+s_{n-1})^{k-s_{n-1}k}$, $s_n=h_n+s_{n-1}$.

[note: il a été étudié le domaine de convergence de l'itération $z:=c+z^k$, sous z initialement nul, en fonction de la constante c , c nombre complexe].

\ Remarque 1:

L'arbre (x) est distinct de x . Quel que soit l'arbre m , l'arbre (m) est distinct de m . D'ailleurs, en ce qui concerne la fonction profondeur, on a $prof[(m)]=1+prof[m]$.

Exercice 3:

Définir, sur les arbres généraux, une fonction de Strahler, généralisation naturelle de celle que l'on a définie sur les arbres binaires...

Bijection du Magma sur le Magma généralisé.

A tout arbre $m=(m_1,m_2,\dots,m_n)$, constitué de $n>1$ arbres, correspond d'une part le premier arbre de la suite, $P_m=m_1$, et d'autre part l'arbre composé des arbres qui suivent, $R_m=(m_2,\dots,m_n)$.

Ce qui signifie évidemment que l'arbre m est interprétable comme composé binaire de P_m et R_m .

Ceci suggère une bijection naturelle B des arbres sur les arbres binaires, définie par la récurrence suivante:

si P_m et R_m sont définis, et si l'on note B_m l'image de m par B , la partie gauche de B_m , c'est B_{P_m} ... et la partie droite c'est B_{R_m} , soit $GB_m=B_{P_m}$, et $DB_m=B_{R_m}$ (la partie gauche de l'image, c'est l'image du premier arbre; la partie droite de l'image, c'est l'image du reste).

Cette récurrence est naturellement initialisée par $B_x=x$, ...et pour tout arbre (m) -sur lequel R n'est pas défini-, par $B(m)=(B_m,x)$.

La bijection inverse B' satisfait à $B'_x=x$, $B'(m,x)=(B'_m)$, et pour tout arbre binaire $m=(m',m'')$ tel que m'' ne se réduise pas à x , on a $P_{B'm}=B'_Gm$, $R_{B'm}=B'_Dm$.

On remarquera que la transformation B ne respecte pas les degrés.

Exercice 1: Inévitable: programmer la bijection B de M sur M' , et son inverse B' .

Remarque 2:

Le langage M est inclus dans le langage M' (tout arbre binaire est un arbre au sens général). Pour tout arbre binaire m (du magma général), calculer B_m et -pourquoi pas?- itérer le procédé.

Remarque 3:

Remarque 3:

Le lecteur aura noté que la structure de Magma est la structure fondamentale des langages de programmation de type Lisp. Autrement dit, ces langages de programmation utilisent les codages d'arbres de type magma. On confond parfois - *et c'est là une erreur très commune* - les codages de type magma avec les codages de type parenthésé, que nous considérerons au chapitre suivant.

Exercice 2:

Construire le langage des arbres ternaires, c'ad les expressions du Magma généralisé engendrées à partir de x par la dérivation $Dx=x^3$; autrement dit, le langage L tel que
 - x appartient à L
 - si a,b,c appartiennent à L , alors (a,b,c) appartient à L .

Calculer le nombre d'objet de degré n (degré nécessairement pair).

La série génératrice T les dénombrant satisfait à l'équation $T=x+T^3$. Calculer ces coefficients (par programme). Le début de cette suite est 1 1 3 12 55 273 1428 7752 43263 246675 1430715.

On peut coder ces objets de bien d'autres manières. Par exemple par des mots de parenthèses particuliers, mais aussi par des mots sur un alphabet (x,y) de deux lettres, de degré n en x et $2n$ en y , tels que pour tout facteur gauche de degré i en x et j en y , j n'excède pas $2i$. C'est cette forme que l'on utilise pour établir que le nombre d'objets de degré $2n+1$ est $(3n)!/((n!)(2n+1)!)$.

Calculer par programme le nombre d'arbres ternaires symétriques, dénombrement du sous-langage L_S des arbres ternaires tel que

- x appartient à L_S
- si a et b sont dans L_S , alors (a,b,a) est dans L_S .

Les premiers coefficients: 1 1 2 3 5 8 13 27 51 89 170 300 564.

Observer le rapport de deux coefficients successifs.

Ne pas s'arrêter en si bon chemin: étudier le langage des arbres engendrés par la dérivation $Dx=x^k$. Le nombre des objets de degré $(k-1)n+1$ est $(kn)!/((n!)((k-1)n+1)!)$.

La démonstration directe bijective du dénombrement des mots de parenthèses, que nous verrons à de multiples reprises par diverses méthodes, $[(2n)!/((n!)(n+1)!)]$, s'étend ici sans difficultés.

3. Enumération du Magma.

L'ordre lexicographique.

Comme nous l'avions déjà remarqué, les éléments du Magma M peuvent être ordonnés lexicographiquement par le processus récurrent défini par les relations

- $x < m$ pour tout m distinct de x ,
- $(m,m') < (m,m'')$ lorsque $m' < m''$,
- et $(m,m'') < (m',m''')$ dès que $m < m'$.

Il convient de réfléchir à la définition, dans divers styles de programmation, d'une procédure f décidant si $m < m'$.

Pour cet ordre, il existe généralement une infinité d'arbres entre deux arbres donnés (exception faite de (x,x) , qui succède à x , et de ses conséquences). Or, on voudrait bien disposer d'une procédure d'énumération grammaticale, qui, partant de x , permette d'affecter à tout arbre un successeur [autrement dit, on cherche à construire une bijection F de M sur l'ensemble des nombres entiers].

Exercice 3 (une anticipation sur le codage parenthésé, à passer en première lecture):

L'équation non associative

$$M' = x + \text{somme pour } n > 0 \text{ des } (M')^n,$$

définition de la série M' des expressions générales magmatiques, représentant les arbres au sens général, montre qu'il y a une infinité d'arbres généraux de degré donné (càd d'expressions du magma général de degré donné en x), du fait du terme (M') , qui produit à partir de tout arbre à i sommets (dont k terminaux correspondant aux occurrences de x) un arbre à $i+1$ sommets (dont k terminaux). Les arbres de (M') sont parfois dits plantés, ou encore, pourvus d'un tronc.

Par contre, en excluant la gènèse de (M') , soit M'' :

$$M'' = x + \text{somme pour } n > 1 \text{ des } (M'')^n,$$

on engendre tout naturellement un nombre fini d'arbres de degré k , pour tout k :

$$1 \quad 1 \quad 3 \quad 11 \quad 45 \quad 197 \quad 903 \quad 4279 \quad 20793 \quad 103049 \quad 518859 \quad 2646723 \quad \dots,$$

Cette suite d'entiers impairs, dite de Schröder (fin du 19^e siècle), ou parfois super-catalans, qui fut en fait semble-t-il déjà considérée dans l'antiquité grecque, satisfait à la récurrence de Schröder,

$$[1] \quad (n+1).s(n+1) + (n-2).s(n-1) = 3.(2n-1).s(n), \text{ sous } s(1)=s(2)=1.$$

Une construction bijective en a été donnée, essentiellement fondée sur la bijection canonique des arbres généraux sur les arbres binaires (cette bijection canonique est clairement mise en évidence par les deux interprétations du langage des mots de parenthèses, à venir, au chapitre suivant, comme codage des arbres binaires, ou comme codage des arbres généraux, au choix):

le fait d'exclure les arbres généraux de la forme (M') conduit à une injection de M'' dans les arbres binaires, donnant des arbres binaires particuliers.

Nous anticipons ici sur le codage parenthésé des arbres généraux, ou, si l'on préfère, sur l'interprétation des mots de parenthèses comme représentant le arbres généraux.

Le codage parenthésé S des arbres de M'' satisfait au système de deux équations

$$S = x + a S b a S b T, \quad T = e + a S b T$$

où x représente encore les sommets terminaux des arbres de M'' .

Par morphisme $\mu(a)=e, \mu(b)=e, \mu(x)=x, \mu(S)=s, \mu(T)=t$, on obtient les séries s et t sur la lettre x ,

$$[2] \quad s = x + s^2 t, \quad t = e + s t,$$

les coefficients de s satisfont à la récurrence [1], la série s est approchée par les polynômes s_i :

$$(s_{i+1}, t_{i+1}) = (x + s_i^2 t_i, e + s_i t_i) \text{ sous } (s_0, t_0) = (x, e).$$

ou, programmativement, si l'on préfère ces notations plus compactes

l'initialisation $(s, t) := (x, e)$ suivie de la récurrence $(s, t) := (x + s^2 t, e + s t)$.

D'autres interprétations de cette récurrence sont certainement possibles.

Conjecture: $s(n)$ est multiple de n lorsque n est premier.
(mais pas seulement, car $s(105)$ est multiple de 105)

De [2] résulte $(e-s)t=e, t=s^*, s=x+s^2(e-s)^{-1}$
 soit $s = x - xs + 2s^2,$
 ou encore $s+xs=x+2s^2,$

$s(1)=1,$ et pour $n>1, (s(n)+s(n-1))/2 =$ somme des $s(i).s(j)$ pour $i+j=n, i>0, j>0.$
 (ce qui établit notamment, de proche en proche, que les coefficients sont impairs)

s est approché par (initialisation $s:=x,$ récurrence $s:=x-xs+2s^2,$ donnant le nombre d'objets de profondeur n au plus, $f(1)=1, f(n+1)=1-f(n)+2f(n)^2, 1\ 2\ 7\ 92\ 16837\ 566952302\dots$

La fonction génératrice des nombres de Schröder est donc $(1+x-(1-6*x+x^2)^{1/2})/4,$
 tandis que celle des nombres des nombres de catalan était $(1-(1-4x)^{1/2})/2.$

Clairement, la grammaire

$$[3] \quad A = x - (xuA) + (AuA) + (AvA)$$

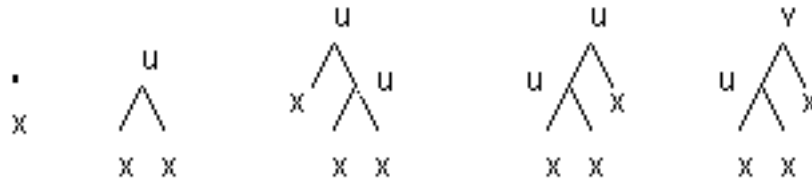
engendre des arbres particuliers du **Magma binaire monogène bicolore:**

des arbres binaires (grd) de facteur gauche $g,$ de facteur droit $d,$ de racine $r,$ sous forme magmatique monogène bicolore, arbres binaires dont les sommets internes sont de deux types, disons u et $v,$ à l'exception des arbres (xua) où a est un arbre du langage $A:$

si la branche gauche se réduit à $x,$ l'arbre (xua) n'est pas produit,
 et le sommet interne correspondant est, obligatoirement, de type $v,$ $(xva).$

Les tout premiers objets de $A :$

(un de degré 1 en $x,$ deux de degré 2, trois de degré 3, construisez les onze de degré 4...)



Les arbres simplement binaires, $C=x+(C,C),$ ou simplement binaires monogènes monocolores, en notant u les sommets internes et x les sommets terminaux, $C=x+(CuC)$ sont susceptibles d'opération de dérivation $Dx=(yux)+(xuy), D(gud)=(yu(gud))+((gud)uy);$

la dérivation D appliquée aux arbres de degré n en x fournit les arbres de degré $n+1, n$ en $x, 1$ en $y,$ ce qui est une démonstration constructive de la récurrence sur les nombres de catalan, $(n+1).c(n+1)=2.(2n-1).c(n).$

Une construction semblable, mais un peu plus compliquée (D. Foata and D. Zeilberger) a été appliquée au langage [3], afin d'établir [1].

Nous verrons que l'on peut préférer exprimer [3] en notations polonaises:

$$[3]' \quad A' = x - xA'u + A'A'u + A'A'v$$

Bijection des arbres binaires sur les entiers.

Etant donné un arbre (m, m') , l'entier $F(m, m')$ ne dépend donc que de F_m et $F_{m'}$: il s'agit donc en fait d'une bijection des couples d'entiers sur les entiers privés de zéro. Problème qui, comme chacun sait, a été quelque peu résolu par Cantor (mathématicien allemand, seconde moitié du 19^{ième} siècle).

On aura naturellement $F_x=0$, (et probablement $F(x, x)=1$).

Ensuite, afin d'alléger les notations, nous conviendrons que si $F_m=i$, et $F_{m'}=j$, alors $F(m, m')=F(F_m, F_{m'})=F(i, j)$ a un sens.

Pour cela, nous identifions abusivement tout arbre m à son image par F , et tout couple d'arbres à un couple d'entiers.

Ainsi F représente ici abusivement tout aussi bien une bijection des arbres sur les entiers qu'une bijection des couples d'entiers sur les entiers.

Ou, si l'on préfère, la procédure F est supposée capable de reconnaître le type des paramètres, et, qu'elle soit appliquée à un arbre ou à un couple d'entiers, elle renvoie un entier.

Première Méthode.

La solution (simple, mais peu esthétique) traditionnellement fournie dans bien des ouvrages, c'est le balayage diagonal des points à coordonnées entières du quart de plan; il consiste à ordonner les couples d'entiers de la manière suivante:

$(i', j') < (i, j)$ dès que $i' + j' < i + j$, ou [lorsque $i' + j' = i + j$] si $i' < i$.

(évidemment, cet ordre équivaut à $F(i, j) < F(i', j')$).

Il en résulte que si $i + j = n$, il y a $1 + 2 + \dots + n = n(n+1)/2$ couples d'entiers (i', j') de somme $i' + j' < n$, de telle sorte que

$F(i, j) = 1 + i + n(n+1)/2$, polynôme en i et j , de degré 2.

Vérifier que l'initialisation est correcte: $F(x, x) = F(F_x, F_x) = F(0, 0) = 1$.

Exercice 4:

Evidemment, programmer F .

Réfléchir à son inverse F' , qui peut se programmer de manière astucieuse (un chapitre ultérieur traitera du calcul des racines de polynômes; il s'agit de calculer la racine positive d'un polynôme de degré 2, avec une précision minimale, car l'on cherche seulement la partie entière de cette racine).

Seconde Méthode:

Il y a tout de même pour l'informaticien un moyen beaucoup plus simple, pratique et triviale pour construire une bijection des couples d'entiers sur les entiers.

Représenter les entiers i et j dans le système de numération de base 2. On obtient ainsi des mots sur un alphabet de deux lettres. Convenons que la première lettre de chaque mot est le chiffre des unités (à l'inverse des conventions usuelles); la dernière est donc le caractère 1 (excepté pour la représentation de l'entier 0). Ces mots, si nécessaire, peuvent impunément être prolongés (à droite) par un nombre indéterminé de caractères 0 (ou plongés dans les suites infinies: on n'hésite pas sur le choix des moyens). Ainsi, le mot vide e représente l'entier zéro.

Alors, construire le mot dont les caractères pairs proviennent de la représentation de i , les impairs de celle de j . Ce mélange alterné est l'évidente représentation d'un entier.

Inversement, il est aisé de séparer un mot en ses caractères de rangs pairs et impairs (petit exercice dans les mots de degré n : concaténer deux tels mots, obtenant ainsi un mot de degré $2n$; puis séparer les caractères de rang pair de ceux de rang impair, ce qui fournit deux mots de degrés n ; itérer le processus).

On pourra tenter de visualiser sur écran coloriable diverses propriétés de cette bijection (par ex., afficher en case (i,j) le composé de i et j , pris modulo p).

Formellement, si x,y sont des lettres, w,w' des mots, et e le mot vide,

$$F(e,e)=e=0, \quad F(xw,yw')=xyF(w,w'), \\ F(xw,e)=x0F(w,e), \quad F(e,yw')=0yF(e,w').$$

et inversement

$$F'(e)=(e,e), \quad F'(xyw)=(x,y)*F'(w), \\ \text{où } (w,w')*(w'',w''')=(ww'',w'w''').$$

Application au Magma:

En supposant encore F défini indifféremment sur les éléments de M et sur les couples d'entiers - lesquels sont d'ailleurs représentés par des mots- , il est aisé de calculer $F(m,m')$, à condition d'initialiser la récurrence par $Fx=0$, et $F(x,x)=1$.

Ainsi, par exemple,

$$F(((x,x),(x,(x,x))))=F[F(x,x),F(x,(x,x))]=F[(1,01)]=1001, \\ \text{soit l'entier 9, en base dix.}$$

Exercice 5:

Programmer F et F' . Faire de même avec un langage de programmation n'autorisant pas les procédures récursives (autrement dit, gérer soi-même la récursivité, en manipulant une pile).

Exercice 6:

Le lecteur apercevra tout de suite les généralisations qui s'imposent (construction de diverses bijections des suites de n entiers sur les entiers).

Il faut faire preuve d'un zest d'astuce supplémentaire pour envoyer sur les entiers les suites (finies) d'entiers (autrement dit, prendre les entiers pour alphabet, et construire une bijection des mots engendrés par cet alphabet sur l'alphabet lui-même). En d'autres termes, construire une bijection des arbres généraux (représentés par les expressions du Magma généralisé) sur les entiers.

[Commentaire: ces exercices résultent immédiatement des grammaires génératives de langages d'arbres (binaires et généraux) telles qu'elles sont énoncées plus haut. Rappelons que, bien que ces grammaires ne soient que des processus récurrents -définissant tout aussi bien la structure générale des mots du langage qu'une méthode de construction aléatoire d'un mot du langage, elles doivent automatiquement se compléter de processus plus précis permettant d'ordonner les mots du langage.... c'est à cela que visent les processus d'énumération systématique, lesquels ne sont que la construction de bijections sur les entiers naturels]

Signalons une **Troisième méthode:**

Un processus célèbre, attribué à Péano (mathématicien italien, fin du siècle passé, début de celui-ci) et Hilbert, (mathématicien allemand, 19^e-20^e siècle, dont la contribution fut décisive), permet de construire, par un processus récursif une bijection non seulement élégante (que l'on interprètera comme la première véritable courbe "fractale", historiquement importante), mais également riche en propriétés, des entiers sur les couples d'entiers.

Il s'agissait d'une parfaite illustration des idées de Cantor, car cette bijection a été construite dans un but précis: établir qu'il existait une bijection continue des nombres réels entre 0 et 1 sur les couples de nombres réels entre 0 et 1.

Ce processus envoyant les couples d'entiers sur les entiers, que nous nommerons encore F , est tel que F^{-1} soit continu, en ce sens que le point d'image $F(i,j)+1$ est l'un des quatre (au plus, à cause des bords) points voisins de (i,j) , c'est-à-dire qu'il diffère de (i,j) par une coordonnée seulement, et d'une unité seulement. Ainsi l'on définit un cheminement, de point en point voisin, sur les points à coordonnées entières du plan.

Cela se construit par récurrence: *supposer* F connue sur le carré $E(n)$ des 4^n points dont les coordonnées sont inférieures à 2^n , le chemin obtenu ayant deux coins du carré pour extrémités; alors, il est possible d'étendre F , par blocs, aux 4^{n+1} points de $E(n+1)$ dont les coordonnées sont inférieures à 2^{n+1} .

L'idée est bien celle-ci: si l'on sait balayer un carré $E(n)$, on doit savoir balayer un carré $E(n+1)$ comportant quatre fois plus de points, en recollant les morceaux (on songe immédiatement à étendre le procédé à une dimension quelconque...). Etant donné son importance, nous y reviendrons.

Exercice 7 (difficile): (d'après Henri Poincaré, "La nature du raisonnement mathématique")

Soit E un ensemble d'objets,
 a un objet particulier de E , et les deux axiomes qui suivent:

- 1) -pour tout objet x de E ,
on suppose défini le composé (x,a) comme étant un objet de E ;
- 2) -on suppose que, si (y,x) est défini pour x et y donnés dans E ,
alors $(y,(x,a))$ est défini comme étant identique à $((y,x),a)$.

Etablir (automatiquement?) que

- si (y,x) est défini,
 (x,y) l'est également, et est identique à (y,x)
- si $((x,y),z)$ est défini,
 $(x,(y,z))$ l'est également, et est identique à $((x,y),z)$.

Que dire du magma (qui est un monoïde commutatif) engendré par l'objet particulier a ?

Un peu plus: on définit une seconde composition $[x,y]$,
contrainte par les deux axiomes

- 3) $[x,a]=x$ pour tout x de E ;
- 4) si $[y,x]$ est défini,
 $[y,(x,a)]$ et $([y,x],y)$ sont définis et identiques.

Etablir que

- si $([y,x],z)$ est défini, il est identique à $([y,z],[x,z])$
 - si $[y,x]$ est défini, il est identique à $[x,y]$.
- A-t-on $[y,x],z]=[y,[x,z]]$, lorsque ces objets sont définis ?

Exercice 8:

Il est un peu plus difficile qu'en termes d'arbres binaires d'établir une démonstration directe de la récurrence sur les nombres de catalan, $[(n+1)c_{n+1}=2*(2n-1)c_n]$, initialisée par $c_1=1$, sous la forme d'une bijection en termes d'arbres généraux. A gauche, $(n+1)c_{n+1}$ compte les arbres généraux à $n+1$ sommets, dont un pointé, et à droite, sur les arbres généraux à n sommets, c'est soit un *sommet*, soit une *arête* qui est *pointé, et marqué* par un objet (gauche ou droit).

Exercice 9:

Exercice 9:

Soit le magma $M[x,y]$, engendré par deux générateurs, x et y .

On appelle basculer d'arbres un ensemble de deux langages d'arbres, G et D , constitué de deux parties disjointes du magma $M[x,y]$, satisfaisant à la contrainte F de fermeture que voici:

(F) tout arbre binaire (g,d) de GxD appartient soit à D , soit à G ,

contrainte qui n'est pas si triviale qu'il y paraît.

En dehors du cas où G (ou D) serait vide, les bascules minimales sont obtenues en fermant, par composition binaire à partir de x et y , plaçant initialement (par exemple) x dans G et y dans D .

Ce sont, de fait, des objets compliqués, que l'on a nommés "bisections", pour des raisons que nous verrons plus loin. Les bascules maximales sont des bipartitions quelconques de la totalité du magma $M[x,y]$, bipartitions nécessairement compatibles avec la règle de fermeture (F).

Exercice 10 (difficile):

Le magma généralisé commutatif est défini de la manière suivante:

l'arbre $m=(m_1,m_2,\dots,m_k)$ ne dépend pas de l'ordre des facteurs m_1,m_2,\dots,m_k (ou, si l'on préfère, il s'agit du même arbre lorsque l'on permute ces facteurs).

Ce qui signifie que l'on peut imposer un ordre sur ces facteurs, de manière à normaliser la représentation de ces arbres commutatifs.

La fonction sommets, $s[m]=1+s[m_1]+s[m_2]...+s[m_k]$, donne le nombre de sommets de l'arbre m .

Il y a un arbre commutatif à un sommet: x , un à deux sommets: (x) , deux à trois sommets: $((x))$ et (x,x) , quatre à quatre sommets: $((((x)))) / ((x,x),x) / ((x,x)) / (x,x,x)$.

Constuire les 9 arbres commutatifs à 5 sommets.

Cayley (mathématicien anglais du 19-ième siècle) a étudié ces arbres sous le nom d'arbres enracinés, et a donné une équation à laquelle satisfait leur dénombrement suivant le nombre de sommets.

Soit g la série en t , de terme général $a(n)t^n$, où $a(n)$ est le nombre d'arbres à $n+1$ sommets.

On note $g_1=t^*$ la série d'inverse $(1-t)$, et pour entier positif k , $g_k=(t^k)^*$ la série d'inverse $(1-t^k)$.

Etablir (c'est la formule de Cayley) que **g est le produit des $g_k^{a(k-1)}$, pour $k>0$.**

Le début de la suite $a(n)$ est 1 1 2 4 9 20 48 115...

Il est à prévoir, du fait que l'on a affaire à des objets commutatifs, que la série $h=tg$ satisfait aussi à une équation de type exponentiel; c'est ce que l'on trouvera dans la littérature:

$$h(t)=t\exp(h(t)+h(t^2)/2+h(t^3)/3+\dots)$$

Exercice 11:

A. Initialement, x est un arbre du langage L .

L représente une partie particulière et ordonnée, ainsi définie, du magma généralisé commutatif:

Soit la récurrence: si a et b sont deux arbres de L avec $a<b$,

alors $(a)<(a,a)<(a,b)<(b)<(b,b)$ sont cinq arbres de L .

Soit la fonction profondeur, p : $p[x]=1$, $p[(a,b)]=p[a]=p[b]=1+p[a]$

La suite f est le nombre d'arbres a de L dont la profondeur est fixée, $p[a]=n$,

$$f(1)=1, f(n+1)=f(n).(f(n)+3)/2.$$

Les premiers arbres de ce langage L : x ; (x) , (x,x) ; $((x))$, $((x),(x))$, $((x),(x,x))$, $((x,x))$, $((x,x),(x,x))$;

Début de la suite f : 1 2 5 20 230 26795...

Cette suite intervient dans le dénombrement de certains arbres de diamètre donné, au sens où on l'entend dans les graphes (arbres non dessinés, disons désordonnés, dont aucun sommet n'est distingué); ces arbres, dont les sommets sont de degré 3 au plus, sont représentés des couples d'arbres de L .

En effet, épluchez un arbre désordonné, en ôtant ses sommets de degré 1 (dits "pendants", càd ceux qui ne sont liés qu'à un autre sommets), et itérez l'épluchage: vous terminez soit sur un sommet, dit central, et le diamètre est pair, soit sur une arête, dite centrale, et le diamètre est impair. Le diamètre est la distance maximale de deux sommets pendants. Dans tout arbres désordonnés ici considéré, la distance entre deux sommets pendants est constante. Appelons les "arbres désordonnés concentriques", les sommets à distance donnée du centre se répartissant sur un même cercle.

Leur nombre $g(n)$, où n est le diamètre, se déduit de $f(n)$:

$$g(2n-1)=f(n)*(1+f(n))/2, g(2n)=g(2n-1)(1+f(n)), g(n)= 1 1 2 3 9 15 90 210 4410 25565..., n \geq 0.$$

B. Dans le magma général, monogène, commutatif, si l'on ne compose que des arbres distincts, les premiers objets sont les suivants:

$$x, (x), ((x)), (x,((x))), (((x))), ((x,(x))), (x,((x))), (((x))),...$$

Ils sont parfois nommés "arbres enracinés -non isomorphes - ou commutatifs- à composantes distinctes"; leur nombre, en fonction du nombre des sommets de l'arbre, est

$$1 1 1 2 3 6 12 25 52 113 247 548 1226 2770 6299 14426...$$

Exercice 12 (magma, nombres rationnels, et approximation discrète de segment):

Sur le magma engendré par x et y , soit l'arbre initial $a=(x,y)$, et les deux morphismes f_x et f_y tels que $f_x x=(x,y)$, $f_x y=y$, $f_y y=(x,y)$, $f_y x=x$.

On fait agir le monoïde de morphismes $(f_x, f_y)^*$ sur l'arbre initial $a=(x,y)$.

Pour f donné dans $(f_x, f_y)^*$, soit (p,q) les degrés de $f(a)$ respectivement en x et y .

On réalise ainsi une importante **injection des nombres rationnels positifs** (p,q) -càd, en fait, des couples d'entiers premiers entre eux- **dans les arbres binaires**.

L'image associative de $f(a)$ donne un mot du monoïde $(x,y)^*$, lequel mot code le chemin d'approximation discrète du segment joignant le point $(0,0)$ au point (p,q) , sur le quart de plan quadrillé $N \times N$. Nous y reviendrons.

Exercice 13:

Sur l'alphabet (a,b) , $a < b$, on considère le langage L ainsi défini:

- a et b appartiennent à L

-si m et m' appartiennent à L , avec m inférieur ou égal à m' , alors mm' appartient à L

Le nombre de mots de L de degré 2^n est $f(n)$: $f(1)=2$, $f(n+1)=f(n)+f(n).(f(n)-1)/2$, soit
2 3 6 21 231 26796 359026206 64449908476890321...

Cette suite compte, bien sûr, des éléments du magma commutatif engendré par un couple de générateurs.

Exercice 14 (profondeur moyenne des objets atomiques des arbres binaires de degré donné):

Dans un arbre binaire, quelle est la distance moyenne à la racine d'un sommet du fond ? (un sommet du fond, ou feuille, comme l'on dit parfois, est une occurrence de x dans l'expression magmatique de l'arbre)

Pour cela, on définit sur chaque arbre binaire $a=(g,d)$, de degré $\deg(a)=n=i+j$, avec le facteur gauche g de degré $\deg(g)=i$ et le facteur droit d de degré $\deg(d)=j$, la fonction $f(a)=f(g,d)=n+f(g)+f(d)$, sous l'initialisation $f(x)=0$,

et l'on définit h comme la somme de f sur la totalité des arbres de degré n , soit

$h(n)=nc(n)+\text{somme des } (c(i)h(j)+h(i)c(j))$ pour $i+j=n$, i et j strictement positifs où les $c(n)$ sont les nombres usuels de catalan,

où les $c(n)$ sont les nombres usuels de catalan,

$c(1)=1, c(2)=1, c(3)=2, \dots, c(n)=\text{somme des } c(i)c(j) \text{ pour } i+j=n, i \text{ et } j \text{ strictement positifs}$
qui dénombrent les expressions du magma en fonction de leurs degrés.

Evidemment, $h(1)=0$.

Les premières valeurs de f sont

0 2 10 44 186 772 3172 12952 52666 213524 863820 3488872 14073060 56708264 228318856...

Etablir que c'est, pour $n>1$, la fonction $(n-1, n-2) \cdot 2^{2n-2}$, où $(n-1, n-2)$ est le coefficient binomial comptant les bipartitions d'un ensemble de $2n-3$ objets en deux parties de $n-1$ et $n-2$ objets.

La moitié de cette suite intervient dans d'autres problèmes, dont le dénombrement de graphes dessinés; elle admet pour série génératrice $c/(1-4x)^{-1}$, où c est la série de catalan, $c=\text{somme des } c(n)x^n, n>0$.

Exercice 15:

L'axiomatique classique de genèse des nombres entiers, tel que les produisent les logiciens (où chaque entier n'est rien d'autre que la liste de ceux qui précèdent)

nous amène à considérer (procédé naturellement exponentiel) le langage des arbres où chaque arbre est tout naturellement le composé de la totalité de ceux qui précèdent (partant de l'objet atomique x):

$x \quad (x) \quad (x, (x)) \quad (x, (x), (x, (x))) \quad (x, (x), (x, (x)), (x, (x), (x, (x))))$
 $(x, (x), (x, (x))), (x, (x), (x, (x))), (x, (x), (x, (x)), (x, (x), (x, (x)))) \dots$

Exercice 16:

Calculer les puissances successives de la série caractéristique du monoïde monogène, $x^*=(e-x)^{-1}$, puis les puissances successives de la série équivalente à la série de catalan $s=(xs)^*=(e-xs)^{-1}$.

Exercice 17 (anticipons; à passer en première lecture):

Si un sommet x terminal peut engendrer un ou deux sommets terminaux, partons de x , notons yx la production d'un unique sommet, zxx la production de deux sommets, l'équation de la série génératrice est $S=x+yS+zS^2$, où y et z codent les sommets internes de l'arbre.

La dérivation $Dx=yx+zxx, Dy=Dz=0$, partant de x , engendre les mêmes mots, mais avec des multiplicités correspondant aux fonctions croissantes sur les sommets internes de l'arbre:

$$D^2x=yDx+z(Dx.x+xDx)=yyx+yzxx+zyxx+xzxxx+zxyx+zxxzxx, \dots$$

En notant $D^*=\text{somme des } D^i$ pour $i \geq 0$, on s'intéressera à la série D^*x .

L'image s de S par le morphisme qui envoie x, y, z sur $t: s=t+ts+ts^2$, est la série caractéristique fonction du nombre total de sommets, internes et terminaux. Les coefficients $a(n)$ de la série s satisfont à la récurrence

$$s(n+1)=s(n)+\text{somme des } s(i)s(j) \text{ pour } i+j=n, i>0, j>0, \text{ sous } s(1)=1,$$

soit 1 1 2 4 9 21 51 127 323 835 2188...

C'est aussi le nombre de manières de relier n points, identifiés aux entiers de 0 à $n-1$, répartis circulairement sur un cercle, chaque point étant éventuellement lié à un seul autre, sans que les liaisons se coupent. Lorsque le nombre des points est pair, et que chaque point est obligatoirement lié à un autre, ces configurations sont évidemment représentées par les mots de parenthèses.

Les configurations considérées ici sont donc représentées par le mélange des mots de parenthèses

$L(x, y)=(xL(x, y)y)^*$ et des mots de z^* , soit le langage $z^* \text{ uu } L(x, y)$.

Ces nombres portent le nom de Motzkin.

Nous rencontrerons par ailleurs le langage $x^* \text{ uu } y^* \text{ uu } L(a, b)$.

Exercice 18 (dénombrons et anticipons encore; à passer en première lecture):

Les arbres ternaires, d'équation magmatique $S=x+(S,S,S)$, sont approchés par la suite des polynômes $S(i+1)=x+(S(i),S(i),S(i))$, sous $S(0)=x$. Il sont $(2n,n)/(2n+1)$ à compter $2n+1$ sommets terminaux, où $(2n,n)$ est le nombre de mots du monoïde $(x,y)^*$ de degrés $2n$ en x et n en y . Et c'est facile à établir.

Les arbres binaires correspondaient à l'équation magmatique $S=x+(S,S)$. La formule close de leur dénombrement s'établit directement en les mettant en bijection avec le langage monoïdal $L(2)$ constitué des mots particulier du mélange de x^{n+1} et y^n :

à partir de l'initialisation: x appartient à $L(2)$

la récurrence: si mxm' appartient à $L(2)$, alors $myxxm'$ appartient à $L(2)$

engendre un langage codant les arbres binaires.

Ce mode de gènèse est une dérivation, la dérivation $Dx=yxx$, $Dy=0$, et l'on ferme par les puissances de D , à partir de l'objet initial x .

Aux occurrences de x correspondent les sommets terminaux de l'arbre binaire, aux occurrences de y les sommets internes. Nous y reviendrons, car les mots de $L(2)$ sont des mots de parenthèses en (x,y) suivis d'une occurrence de x en excès, ils sont primitifs (il n'y a pas répétition d'un motif), et ils sont particuliers dans leur classe de conjugaison.

De même, les arbres ternaires, magmatiquement d'équation $S=x+(S,S,S)$, se codent de manière monoïdale dans le monoïde $(x,y)^*$, par le langage $L(3)$ engendré par

l'initialisation: x appartient à $L(3)$

la récurrence: si mxm' appartient à $L(3)$, alors $myxxxm'$ appartient à $L(3)$

ou, ce qui est mieux:

$L(3)$ est la fermeture de x par la dérivation $Dx=yxxx$, $Dy=0$.

le dénombrement en résulte:

nous avons des mots particulier du mélange de x^{2n+1} et y^n ; les mots du mélange de x^{2n+1} et y^n sont regroupés par conjugaison en classe d'équivalence; il y a dans $L(3)$ un unique mot de chaque classe, terminé par une occurrence de x (il n'est pas difficile de voir lequel: considérer l'évolution de la fonction "double du degré en y - degré en x " sur les facteurs gauches de ces mots).

Les arbres homogènes k -aires seront codables dans $(x,y)^*$ par le langage engendré par la dérivation $Dx=yx^k$, $Dy=0$, à partir de la lettre x . Les occurrences de x correspondant aux sommets terminaux, et les occurrences de y aux sommets internes de ces arbres.

Ainsi, les arbres homogènes k -aires, pour k de 2 à 11, début des séries $S=x+S^k$:

1 2 5 14 42 132 429 1430 4862 16796
 1 3 12 55 273 1428 7752 43263 246675 1430715
 1 4 22 140 969 7084 53820 420732 3362260 27343888
 1 5 35 285 2530 23751 231880 2330445 23950355 250543370
 1 6 51 506 5481 62832 749398 9203634 115607310 1478314266
 1 7 70 819 10472 141778 1997688 28989675 430321633 6503352856
 1 8 92 1240 18278 285384 4638348 77652024 1329890705 23190029720
 1 9 117 1785 29799 527085 9706503 184138713 3573805950 70625252863
 1 10 145 2470 46060 910252 18730855 397089550 8612835715 190223180840
 1 11 176 3311 68211 1489488 33870540 793542167 19022318084 464333035881

Ces coefficients s'obtiennent aisément car leur formule close est connue; nous l'obtiendrons à diverses reprises par la même méthode que celle utilisée dans le cas $S=x+S^2$, en considérant les classes de conjugaison de mots primitifs sur un alphabet de deux lettres. Nous y reviendrons.

Les arbres généraux, muni de l'histoire de leur g n se, en faisant pousser l'arbre par ses sommets terminaux, se pr sentent comme engendr s par la d rivation somme:

$$D = \text{somme des } D_k, \text{ si l'on note } D_k \text{ la d rivation } D_k x = yx^k, D_k y = 0.$$

Quelles que soient les autres interpr tations combinatoires, les s ries alg briques de la forme

$$s = x + s^2 + \dots + s^k$$

sont relatives   l' num ration d'arbres   sommets de degr  born  en fonction du nombre de sommets terminaux (que certains appellent les feuilles), tandis que les s ries de la forme

$$s = x + xs + xs^2 + \dots + xs^k$$

sont relatives   l' num ration en fonction du nombre total des sommets, terminaux et internes.

Les coefficients de $s = x + xs + xs^2 + \dots + xs^k$ co ncident avec les entiers de catalan jusqu'au degr  $k+1$.

Un probl me: le calcul rapide des coefficients de ces s ries. Le cas le plus simple, symptomatique, est celui de la s rie $s = x + (s, s)$ dont les expressions magmatiques sont en bijection avec les mots engendr s   partir de x par la d rivation $dx = yxx$; ces mots sont en bijection avec les classes de conjugaison des mots du m lange y^i ou x^{i+1} ; d'o  la formule close, et une r currence sur les coefficients.

  la limite, lorsque k croit ind finiment, on obtient les s ries

$$s = x + (\text{somme des } s^k \text{ pour } k > 1), \text{ et } s = x + (\text{somme des } xs^k \text{ pour } k > 0).$$

Comparer $s = x + s^2$ et $s = x + (\text{somme des } xs^k \text{ pour } k > 0)$: les coefficients sont les m mes; en effet, nous savons construire des bijections des arbres binaires sur les arbres g n raux.

Exercice 19:

Il est clair que le nombre d'expressions du magma dont les parties gauche et droite sont de m me degr  est compt  par les carr s des nombres de catalan

1 1 4 25 196 1764 17424 184041 2044900 23639044 282105616

mais peu d'interpr tations combinatoires en sont connues, except  ce que l'on a appel  les m andres. Par contre, un certain nombre d'interpr tations ont  t  donn es du produit de deux nombres de catalan successifs en termes de cheminements sur le quart de plan $N \times N$, chemins partant de $(0,0)$ et y revenant, comme les nombres de catalan d nombrent sur N les chemins qui partent de 0, et y reviennent.

1 2 10 70 588 5544 56628 613470 6952660 81662152 987369656

Le nombre d'arbres (au sens g n ral)   n ar tes dont les ar tes sont de deux types est $c(n-1)2^n$:

1 2 8 40 224 1344 8448 54912 366080 2489344 17199104...

Exercice 20 (id e initialement due   Raymond Queneau..., d velopp e par Ulam):

Construire la suite d'entiers positifs d finie par

-l'initialisation $f(1)=1, f(2)=2,$

-et la r currence que voici:

$f(n+1)$ est le plus petit entier qui s' crive d'une unique mani re sous la forme $f(i)+f(j)$, avec $1 \leq i < j \leq n$.
(1 2 3 4 6 8 11 13 16 18 26 28 36 38 47 48 53 57 62 69....).

Construire la suite des diff rences: $f(n+1)-f(n)$: 0 0 0 2 2 3 2 3 2 8....

Tout entier ainsi engendr  est donc composition binaire de deux entiers distincts de rang inf rieur,   partir des g n rateurs 1 et 2. Injecter ces nombres dans les arbres binaires.

Explorer   partir d'autres initialisations. \

Exercice 21:

La série de catalan comme dénombrement d'arbres binaires est solution de l'équation de degré 2:

$$C^2 = C - x$$

Il en résulte que ses puissances C^n s'expriment par $C^{n+1} = p_n C + r_n$ où p_n et r_n sont des polynômes en x de degrés croissant. On se propose de calculer les coefficients de ces polynômes;

$$C^3 = (e-x)C - x, C^4 = (e-2x)C - x(x-e), C^5 = (e-3x+x^2)C - x(e-2x), \dots$$

Le couple (p_{n+1}, r_{n+1}) se déduit de (p_n, r_n) par $(p_{n+1}, r_{n+1}) = (p_n + r_n, -xp_n)$, soit $(p, r) := (p+r, -xp)$, initialisé par $(1, 0)$.

En ne retenant que la suite des coefficients de p et r , on obtient de C à C^7

$$(1, 0) (1, 0 -1) (1 -1, 0 -1) (1 -2, 0 -1 1) (1 -3 1, 0 -1 2) (1 -4 3, 0 -1 3 -1) (1 -5 6 -1, 0 -1 5 -6 1)$$

Donner la règle de formation de ces coefficients:

les exprimer comme coefficients binômiaux, et se contenter de la suite des p_n .

Les signes sont alternés (les coefficients des puissances impaires sont négatifs); en n'en tenant pas compte, la suite des termes constants est 1, la suite des coefficients de x est 1 2 3..., celle de x^2 est 1 3 6..., celle de x^3 est 1 4 10..., soit en présentant décalé le tableau des coefficients du binôme

1	1	1	1	1	1	1	1	1	1	...
	1	2	3	4	5	6	7	8	9	...
		1	3	6	10	15	21	28	...	
			1	4	10	20	35	...		
									

on lit en colonne les coefficients des polynômes p successifs; c'est dire que le coefficient de x^i dans p_n se lit en ligne i colonne n (la première ligne et la première colonne sont d'indice 0), soit le coefficient binomial $\binom{n-i}{i} = \frac{(n-i)!}{i!(n-2i)!}$, affecté de la parité de i .

[Rappelons qu'avec ces notations, le coefficient binomial noté (i, j) est nul dès que $i < 0$ ou $j < 0$, et que la récurrence générative est $(i+1, j+1) = (i, j+1) + (i+1, j)$, sous l'initialisation $(0, 0) = 1$]

Comparer à ce que l'on obtient pour $x^n S^{n+1}$ à partir de $xS^2 = S - e$, (ou $x^2 n S^2 n$ avec $x^2 S^2 = S - e$). Et puis, de $S^{-1} = e - xS$, on tirera les puissances de S^{-1} fonction de S .

Exercice 22 (l'occasion de programmes récursifs):

1) Nous avons vu que les expressions du magma binaire $M = x + (M, M)$ sont tout naturellement en bijection avec les couples d'entiers [c'est une illustration de la dénombrabilité de $N \times N$].

Si le couple d'entiers positifs (i, j) est interprété comme $p(i)j$, où $p(i)$ est le i-ème nombre premier (par exemple, pour l'ordre naturel croissant, 2 3 5 7 11...), l'énumération des expressions magmatiques binaires telle que nous l'avons pratiquée, composant chaque expression avec celles qui précèdent à gauche et à droite, puis avec elle-même, se transforme en une énumération baroque de certaines puissances de premiers:

$$1 \quad (1,1)=2 \quad (1,2)=4 \quad (2,1)=3 \quad (2,2)=9 \quad (1,4)=16 \quad (4,1)=7 \quad (2,4)=81 \quad (4,2)=49 \quad (4,4)=2401$$

$$(1,3)=8 \quad (3,1)=5 \quad (2,3)=27 \quad (3,2)=25 \quad (4,3)=64 \quad (3,4)=((2,1),(1,2))=(((1,1),1),(1,(1,1)))=625$$

$$(3,3) \quad (1,9) \quad (9,1) \quad (2,9) \quad (9,2) \quad (4,9) \quad (9,4) \quad (3,9) \quad (9,3) \quad (9,9), \dots$$

$$1 \quad 2 \quad 4 \quad 3 \quad 9 \quad 16 \quad 7 \quad 81 \quad 49 \quad 2041 \quad 8 \quad 5 \quad 27 \quad 25 \quad 625 \dots$$

2) Nous avons vu que les expressions du magma général sont naturellement en bijection avec les suites finies d'entiers [illustration de la dénombrabilité de \mathbb{N}^* , ensemble des mots, sur \mathbb{N} pris pour alphabet]. Les entiers (factorisés sur les entiers premiers) permettent de construire trivialement une injection f , dans le magma généralisé.

Représenter un entier par la suite strictement croissante des rangs de ses diviseurs premiers (dans la suite des premiers) et associer à chaque diviseur premier sa multiplicité; ainsi, un entier est une suite de couple d'entier ainsi interprétés: (rang d'un premier, exposant de ce premier). Itérer sur tout entier, jusqu'à obtention uniquement d'occurrences de 1 (càd le rang du nombre premier 2).

Ou encore:

On construit une bijection f des entiers sur les expressions du magma général commutatif (dire que le magma est commutatif, c'est imposer un ordre sur les éléments composants tout objet, de manière à injecter le magma commutatif dans le magma général)

Si m est un mot du magma, le mot (m) du magma général sera dit primitif,

et l'on notera m^k le mot (m, m, \dots, m) , puissance k -aire de m dans le magma général.

Voici la définition de cette bijection f particulière des entiers positifs sur le magma général commutatif

- $f(1)=x$, $f(2)=(x)$,

- $f(p)$ est primitif ssi p est premier: si p est le k -ième premier, alors $f(p)=(f(k))$,

- et $f(p^j)=(f(k), f(k), \dots, f(k))=m^j$, si p est le k -ième premier, $f(p)=m=(f(k))$

finalemt

si $n =$ produit de puissances de premiers = produit des $p_{r(i)}^{j(i)}$, $j(i) > 0$, i de 1 à v ,

où la suite $r(i)$ est une suite strictement croissante d'entiers, $p_{r(i)}$ premier de rang $r(i)$, on a

$$f(n) = (f(p_{r(1)})^{j(1)} \text{ fois}, \dots, f(p_{r(2)})^{j(2)} \text{ fois}, \dots, f(p_{r(v)})^{j(v)} \text{ fois}).$$

Ainsi,

$f(3)=((x))$ $f(4)=((x),(x))$ $f(5)=(((x))))$ $f(6)=((x),((x)))$ $f(7)=(((x),(x))))$ $f(8)=((x),(x),(x))$

$f(9)=(((x)),((x))))$ $f(10)=((x),(((x))))$ $f(11)=((((x))))$ $f(12)=(((x)),((x),(x)))$ $f(13)=((((x)),((x))))$...

où les expressions commutatives ainsi sont convenablement ordonnées pour être représentées par des expressions du magma général.

Chacune de ces expressions représente un arbre commutatif, muni d'un sommets pointé que l'on appelle racine, ce qui permet de représenter des entiers assez grands par de petits graphismes arborescents. Quel est le nombre de sommets de $f(n)$, $n > 0$?

1 2 3 3 4 4 4 4 5 5 5 5 5 6...

Combien d'arbres commutatifs enracinés à n sommets ?

\

Exercice 23:

Le nombre d'arbres binaires de profondeur n au plus se calcule par $n:=1+n*n$ sous l'initial $n:=1$, suite qui débute par 1 2 5 26 677 458330 210066388901....

Pour savoir si un entier de cette suite est multiple du nombre premier k , lancer le programme

[initialement $n:=1$; tant que n diffère de 0, $n:=(1+n*n)$ modulo k].

Les premières valeurs de k : 2 (1) 5 (2) 13 (3) 41 (6) 137 (16) 149 (13)...

sui vies entre parenthèses du rang de l'itération à laquelle ceci se produit.

Evidemment, lorsque k ne divise aucun de ces entiers, la suite obtenue est toujours périodique.

On constatera que les nombres premiers ainsi obtenus sont particulièrement rares.

Tabuler les périodes.

Curieusement, l'itération $n:=1+n*n*n$, traitée de même manière, produit des entiers

(début: 1 2 9 730 389017001..)

très divisibles, par 2 3 5 11 17 19 23 29 37 41 43 47 53 59 71 73 83 89 101...

\

Exercice 24:

Exercice 24:

Les arbres du magma général obtenus par les seules compositions unaire et binaire, de série génératrice $S=x+(S)+(S,S)$, sont approchée par les polynômes $S(n+1)=x+(S(n))+(S(n),S(n))$,

soit à partir de $S(1)=x, S(2)=x+(x)+(x,x)$,
 $S(3)=x+(x)+(x,x)+(x,(x))+((x),x)+((x),(x))+((x),(x,x))+((x,x),x)+((x),(x,x))+((x,x),(x))+((x,x),(x,x)),...$

$S(n)$ est le polynôme des arbres de profondeur (ou hauteur) n au plus (x est de profondeur 1).
 Le nombre $s(n)$ de ces arbres de profondeurs n au plus ($s(n+1)=1+s(n)+s(n)^2$) est obtenu par l'itération $k:=1+k+k*k$, sous l'initialisation $k:=1$, soit la suite
 1 3 13 183 33673 1133904603...,

suite comportant peu de diviseurs premiers, comme on pourra le vérifier.
 Par différence première, le nombre $h(n+1)=s(n+1)-s(n)=1+s(n)^2$ de ces arbres de profondeur n :
 1 2 10 170 33490 1133870930...

Les polynômes $H(n)$ caractéristiques des arbres de profondeur n satisfont à $H(n)=S(n)-S(n-1)$, et
 $H(n+1)=(H(n))+(H(n),H(n))+(H(n),S(n-1))+(S(n-1),H(n))$

ce qui fait que $h(n+1)$ est multiple de $h(n)$: ainsi, $1133870930=1*2*5*17*197*33877$.
 En fait, de $h(n+1)=h(n)+h(n)*h(n)+2h(n)*s(n-1)$ sous $h(1)=1$, résulte $h(n+1)/h(n)=1+s(n)+s(n-1)$.
 La différence première de la suite $1+s(n)+s(n-1)$ semble très factorisable à son tour,
 1 3 3*4 3*4*12 3*4*12*187....

On calculera le nombre $f(i,j)$ de ces arbres dont le degré est i et la profondeur j .

Si l'on considère des arbres signés, suivant la parité des compositions unaires, on a $S=x-(S)+(S,S)$, le nombre d'arbres positifs incluant x , dépasse d'une unité celui des arbres négatifs, qui est $(s(n)-1)/2$, soit 1 6 91 16836 566952301....

On évoquera le cas comparable des arbres binaires stricts, à hauteur n au plus, dont le nombre est $s(n+1)=1+s(n)^2=1 2 5 26 677...$, sous $s(1)=1$; soit $h(n)$ à hauteur n : $h(n+1)=s(n+1)-s(n)$,
 1,1,3,21=3*7,651=21*31,457653=651*703,210065930571=457653*459007
 où $1 3 7 31 703 459007...$ n'est autre que $h(n+1)/h(n)=2*s(n)-h(n)=h(n)+2s(n-1)=s(n-1)+s(n)$,
 suite $s(n)+s(n-1)$ de différence première $1 2 4 24 672 458304...$, factorisable, $2*2*6*28*682...$
 De manière plus générale, $s(n+1+p)-s(n+1)=(s(n+p)-s(n)).(s(n+p)+s(n))$.

[Note: la suite $(k:=1+k+k*k, \text{ sous } k:=1)=(1 3 13 183 33673 1133904603...)$ a fait l'objet d'études particulières, depuis que Lehmer (1935) a défini ce que certains ont considéré comme un analogue du développement en fraction continue (représentation des réels positifs par des suites d'entiers positifs). Utilisant la fonction cotangente, Lehmer produit un objet curieux et compliqué, lequel demanderait quelque nouvel éclairage; tout nombre réel positif r est décrit par une série alternée,

$$\text{somme des } (-1)^n \arccot(a(i)), a(i+1) = a(i)^2 + a(i) + 1, i \geq 0.$$

Sur le bord, lorsque $a(i+1)=a(i)^2+a(i)+1, a(0)=0$, on retrouve la suite $0 1 3 13...$, la convergence est très lente, le nombre r correspondant est appelé constante de Lehmer.]

Nous nous sommes intéressés principalement aux arbres binaire (càd au magma binaire) et aux arbres généraux (càd au magma général) principalement du fait de l'existence de bijections standardisées sur les mots de parenthèses, comme nous le verrons.

Il est naturel de s'intéresser au magma monogène bridé à la composition k -aire, de série génératrice $S=x+(S)+(S,S)+(S,S,S).....+(S,S.....,S)$,

où la composition de degré maximal figurant à droite de l'équation est k -aire, le nombre d'objets de profondeur donnée se calculant par $n:=1+n+n^2.....+n^k$ sous l'initialisation $n:=1$.

Les arbres homogènes k-aires sont obtenus par $S=x+S^k$, en composition magmatique de notation abrégée (en effet, on peut écrire $S=x+S^2$ au lieu et place de $S=x+(S,S)$, s'il n'y a pas d'ambiguïté), dénombrés à profondeur maximale, donnée par $n:=1+n^k$ sous l'initialisation $n:=1$.

Ainsi, les arbres ternaires de profondeur n au plus: 1 2 9 730 389017001... ($n:=1+n^3$ sous $n:=1$, ou $s(1)=1, s(n+1)=1+s(n)^3$), d'équation $S=x+(S,S,S)$, sont en bijection évidente avec les arbres binaires d'équation $S=x+((S,S),S)$ par la transformation inversible $f[x]=x, f[(a,b,c)]=(f[a],f[b]),c$. (On généralisera les transformations de type f à des arbres généraux quelconque).

Les arbres ternaires de profondeur n+1 sont en nombre $h(n+1)=s(n+1)-s(n)=s(n)^3-s(n-1)^3=1\ 7\ 721\ 389016271\dots$ d'où $h(n+1)/h(n)=s(n)^2+s(n).s(n-1)+s(n-1)^2=1\ 7\ 103\ 539551\dots$

Il est également naturel de s'intéresser au magma binaire engendré par un alphabet quelconque X; par exemple, avec deux générateurs, x et y, on a $S=x+y+(S,S)$, arbres binaires à deux générateurs dénombrés à profondeur donnée par $n:=2+n^2$ sous l'initialisation $n:=2$.

Et ainsi de suite. De manière générale, les arbres bridés à la composition k-aire engendrés par un alphabet X de q lettres, $S=X+(S)+(S,S)+(S,S,S)\dots+(S,S,\dots,S)$. Comme nous le verrons, tout ceci peut se représenter agréablement en notations polonaises.

Exercice 25 (à propos de langages d'arbres et de systèmes d'équations différentielles):

1) Sur le magma à deux générateurs, x et y, soit le système de deux équations à deux inconnues
 (1) $S=x+(S,T), T=y+(T,S)$,

[quelque peu différent de $S=x+(S,T), T=y+(S,T)$].

Dans S (resp T), combien d'arbres $f(S,i,j)=f(T,j,i)$ de degrés (i,j) en x et y ?

(la somme des $f(S,i,j)$ pour $i+j=n$, c'est le nombre de catalan que nous avons noté $c(n)$)

Comparer à $S=(x,y)+(x,S)+(S,y)+(S,S)$, conduisant au même nombre d'arbres de degrés gauches en x et droits en y.

(on regardera, dans tous les cas, ce que l'on obtient par le morphisme qui envoie x et y sur x).

1	1	1	1	1	1	1	1	1
1	3	6	10	15	21	28	36	
1	6	20	50	105	196	336		
1	10	50	175	490	1176			
1	15	105	490	1764				
1	21	196	1176					
1	28	336						
1	36							
1								

2) Un bel exemple (par anticipation des chapitres sur les permutations).-

Considérer le système magmatique suivant, avec deux générateurs, x et y, et deux équations

$$(2) \quad S=x+(S,T), T=y+(S,S),$$

et comptons le nombre d'arbres de degrés donnés en x et y dans S et T, par image associative et commutative soit $g(S)=\text{somme des } s(i,j)x^i y^j, g(T)=\text{somme des } t(i,j)x^i y^j$;

par exemple, pour résumer, voici les coefficients de $g(S+T)$, que l'on peut examiner attentivement, où l'on lira sur la ligne i les coefficients des mots de degré i en x:

0	1	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1
1	2	3	4	5	6	7	8	9	
1	4	10	20	35	56	84	120		binômiaux (i,3)
2	10	30	70	140	252	420			
3	21	84	252	630	1386				
7	56	252	840	2310					

12 120 660 2640
 30 330 1980
 55 715
 143

Observer particulièrement la première colonne, coefficients de la série en x^* (obtenue en envoyant y sur 0) où l'on reconnaîtra l'interclassement des parts provenant de S et T , $(2i,i)/(2i+1)$ et $2(2i+2,i)/(3i+2)$ (la seconde part évidemment obtenue par carré de convolution de la première)

Le nombre d'arbres de profondeur donnée, dans S comme dans T est le même que celui des arbres binaires monogènes (dans $S=x+(S,S)$), soit 1 2 5 26 677 458330 210066388901....
 et, dans S comme dans T , le nombre d'arbres de degré n est le nombre de catalan $c(n)$ (on constatera effectivement dans le tableau ci-dessus que les sommes diagonales sont $2.c(n)$).

Ces séries S et T sont liées à la résolution formelle du système d'équations différentielles qui lient les fonctions numériques $tg.t$ et $1/\cos.t$.

Nous y reviendrons à propos de l'algorithme de tri dit "tri en arbres", parfois appelé "tri rapide" (qui est de fait un algorithme d'inversion de permutation), fondé sur la structure d'arbre binaire sous-jacente à chaque permutation.

Les coefficients entiers du développement exponentiel (càd en $t^n/n!$) de $tg.t+1/\cos.t$, dits "nombres d'Euler", (1 1 1 2 5 16 61 272...) comptent les permutations dites alternantes (alternance de montées et descentes) inscrites sur ceux des arbres de $S+T$ qui s'expriment exclusivement en x (l'élimination de toute occurrence de y provient des conditions initiales du système différentiel: $tg(0)=0$ élimine y , $\cos(0)=1$ conserve x). Ces nombres sont reliés aux nombres de Bernoulli et Genocchi (coefficients du développement exponentiel de $tg(t/2)$).

Il faut faire feu de tout bois, calculs associatifs ou non, commutatifs ou non, grammaires génératives de langages formels, séries formelles, structures arborescentes, dérivations et morphismes, et expérimentations programmées.

Exercice 26:

La suite lexicographique, $f(1)=1$, $f(n)=f(i)+f(j)+1$, $0 < i < j$, soit

1 3 5 7 7 9 11 9 11 13 15 9 11 13 15 15 11 13 15 17 17 19 13 15 17 19 19 21 23 11 13 15.....

compte, lors de l'énumération lexicographique des arbres binaires commutatifs

$x (x,x) (x,(x,x)) ((x,x),(x,x)) (x,(x,(x,x))) ((x,x),(x,(x,x))) ((x,(x,x)),(x,(x,x)))...$

le nombre de sommets, terminaux ou internes, de chaque arbre.

S'il y a $g(n)$ arbres commutatifs de degré n , c'est le nombre d'occurrences de $2n-1$ dans f .
 ($g= 1 1 1 2 3 6 11 23...$), nombres de Narayana and co.

On a $g(n)=\text{somme de } g(i)g(j), 0 < i < j, +\text{si } n=2k, g(k)(g(k)+1)/2; g(n)=\text{somme des } [n/2] \text{ qui précèdent}; g(1)=1=g(2), \text{ puis } g(2n)=2g(n), g(2n+1)=2g(2n)-g(n).$ \

Exercice 27 (à propos de l'élagage des arbres):

Nous noterons El l'opérateur (opérant sur les arbres, et sur les polynômes par linéarité) d'élagage des arbres binaires, défini par

$$El[x]=x, El[(g,d)]=x+(El[g],El[d]).$$

Par exemple, $El[((x,(x,x)),(x,x))]=x+(El[(x,(x,x))],El[(x,x)])=x+(x+(x,x+(x,x)), x+(x,x))$
 $=x+(x,x)+(x,(x,x))+((x,x),x)+((x,x),(x,x))+((x,(x,x),x))+((x,(x,x)),(x,x))$

ce qui est le polynôme caractéristique des arbres obtenus par élagage de $((x,(x,x)),(x,x))$.

[Les arbres obtenus par élagage d'un arbre a le sont donc par ablation itérée d'une partie de sommets du fond -càd d'occurrences de x , comme transformées en un élément neutre de la composition binaire- jusqu'à et y compris obtention de l'objet atomique x]

Le nombre $nb[(g,d)]$ des arbres élagués de (g,d) satisfait à $nb[(g,d)]=1+nb[g]*nb[d]$, $nb[x]=1$.

L'ensemble H des arbres homogènes est constitué des carrés d'arbres homogènes (sauf x), soit : x appartient à H ; et si a , appartient à H , alors (a,a) appartient à H .

Si a est de profondeur n , (a,a) est de profondeur $n+1$.

[De manière générale, $prof.(g,d)=1+\max(\text{prof.}g,\text{prof.}d)$]

Si a est homogène de profondeur n , le polynôme $El[a]$ le polynôme caractéristique d'un arbre homogène est le polynôme caractéristique de l'ensemble des arbres de profondeur n au plus, lesquels sont d'ailleurs produits par l'itération polynômiale $p:=x+(p,p)$, sous l'initialisation $p:=x$.

On rappelle que le degré d'un arbre, $deg[(g,d)]$, est défini par le morphisme dans les entiers additifs $deg[(g,d)]=deg[g]+deg[d]$, $deg[x]=1$.

On peut définir un ordre partiel sur les arbres binaires par "l'arbre b' appartient à $El[b]$ ".

Considérer les arbres binaires (g,d) tels que g appartienne à $El[d]$.

Autrement dit, le facteur gauche de l'arbre (g,d) doit appartenir à l'élagage du facteur droit.

Nous cherchons quel est le nombre de tels arbres (g,d) dont le degré est n .

La série C caractéristique des arbres binaires satisfait à $C=x+(C,C)$, et son image associative c définit les nombres de catalan $c(n)$,

$$c=\text{somme formelle des } c(n).x^n \text{ pour } n>0, \text{ avec } (n+1).c(n+1)=2.(2n-1).c(n), c(1)=1.$$

On établit que la série génératrice dénombrant les arbres (g,d) tels que g appartienne à l'élagage de d est

$$\text{somme formelle des } c(n).(x.c)^n \text{ pour } n>0.$$

série associative dont les tout premiers coefficients sont 1 1 3 7 21 62 197 637....

En effet, si (g,d) est un arbre tel g soit dans l'élagage de d , c'est que d est dans l'image de g par le morphisme pouss tel que $\text{pouss}[x]=C$.

[Par exemple, $\text{pouss}[(x,(x,x))]=(C,(C,C))=$

$(x+(x,x)+(x,(x,x))+((x,x),x)....,(x+(x,x)+(x,(x,x))+((x,x),x)....,x+(x,x)+(x,(x,x))+((x,x),x)....)$ ce que l'on développe par distributivité de la composition binaire sur la somme.]

\

Exercice 28 (mots de Meeussen, suites croissantes à croissance bornée):

Sur l'alphabet des entiers positifs, on considère le langage M constitué des mots dont la première lettre $l(1)$ est 1, et dont la lettre $l(i)$ de rang i ne peut être suivie que d'une lettre $l(i+1)$ telle que

$$l(i)<l(i+1) \quad 2^*l(i).$$

Les premiers mots produits sont donc 1, 12, 123 124, 1234 1235 1236 1245 1246 1247 1248,... Ces mots sont appelés "mots de Meeussen".

On peut les classer en fonction de leur degré (1 1 2 7 41 397 6377 171886....), en fonction de leur norme (la somme de leurs lettres), en fonction de la dernière lettre.

Un mot de degré n se terminant par la lettre i engendre naturellement i mots de degré $n+1$, se terminant par les lettres de $i+1$ à $2i$: on range ainsi ces mots sur les sommets d'un arbre infini.

D'où un moyen (assez lourd, il est vrai) pour dénombrer les mots du langage M : à chaque profondeur de l'arbre infini; on note chaque mot par sa dernière lettre, à l'étage suivant, la lettre i est remplacée par le mot $(i+1).(i+2)...(2i)$.

Execution: partant de 1, on obtient par cette substitution 1, 2, 3 4, 4 5 6. 5 6 7 8, 5 6 7 8.6 7 8 9 10. 7 8 9 10 11 12. 6 7 8 9 10. 7 8 9 10 11 12. 8 9 10 11 12 13 14. 9 10 11 12 13 14 15 16,...

la suite des degrés des mots de M étant celle des normes (la somme des lettres) des mots engendrés par cette substitution itérée.
Il convient de dénombrer les mots de M de degré n en fonction de la dernière lettre k , $n \leq k \leq 2^{n-1}$.

Exercice 29 (remarques fondamentales de programmation; faire le tour d'un arbre binaire):

Un exemple symptomatique de programmation: l'énumération des sommets d'un arbre binaire.

Ici, à propos de l'énumération des nombres rationnels (Stern 1858, Brocot 1860):

(0) si (p,q) sont des entiers premiers entre eux, alors $(p+q,q)$ et $(p,p+q)$ le sont.

1) Le programme récursif suivant, avec passage de paramètres, énumère -par exemple- les couples d'entiers (p,q) premiers entre eux tels que $p+q < m$:

définition: $f(p,q)$ =(afficher (p,q) ; si $p+q < m$ alors $f(p+q,q)$, $f(p,p+q)$).
exécution: appel de $f(1,1)$.

C'est dire que, tant que $p+q < m$, l'appel $f(p,q)$ provoque les deux appels $f(p+q,q)$ et $f(p,p+q)$: ce programme est strictement calqué sur la définition (0).

2) Le programme récursif suivant, sans passage de paramètre, tout en global, produit les même sorties:

définition: f =(afficher (p,q) ; si $p+q < m$ alors $p:=p+q$, f , $p:=p-q$, $q:=q+p$, f , $q:=q-p$).
exécution: initialisation $(p,q):=(1,1)$; appel de f .

La procédure f agit sur l'environnement (p,q) , qui est modifié avant l'appel, et restitué après l'appel.

3) Puis la version non récursive (le passage d'un sommet au sommet suivant):

exécution: initialisation $(p,q):=(1,1)$;
tant que $p > 0$ (afficher (p,q) ; si $p+q < m$ alors $p:=p+q$ sinon (tant que $q > p$, $q:=q-p$), $p:=p-q$, $q:=p+q$).

4) Toutes ces versions provoquent la projection des sommets d'un arbre binaire, internes et terminaux, selon un ordre et un algorithme qui seront qualifiés de polonais au chapitre suivant: (dans l'ordre: projeter la racine, projeter le facteur gauche, projeter le facteur droit); ôter la récursivité exige de savoir calculer le sommet qui suit un sommet donné dans cette projection, ce qui est réalisé dans la version non récursive 3).

Dans le représentation (g,d) de type magma des arbres binaires, cette fonction projection, $proj$:

récurrence: $proj[(g,d)] = "(" . proj[g].proj[d]$, sous l'initialisation: $proj[x]=x$.
projette en x les sommets terminaux (atomes, ou feuilles) et en $"$, les sommets internes.

Ceci montre la redondance du codage magmatique.

Nous verrons qu'il s'agit en fait de la traduction du langage magma vers le langage polonais.

La traduction vers le langage parenthésé serait

$trad[(g,d)] = "(" . trad[g].trad[d].")"$ sous l'initialisation: $trad[x]=e$
projetant exclusivement les branches gauches et droites, et non les sommets.

L'on comparera ultérieurement ces programmes à ceux, très voisins, exécutant également le tour d'un arbre binaire, fonctions projetées des sommets d'arbre binaire, mais produisant des sorties différentes, issus de récurrences développant également des arbres binaires, que nous examinons essentiellement dans la suite (Livre II):

- les bipartitions d'une ensemble, les mots sur un alphabet de deux lettres, les codes de gray
- le problème standard du déplacement d'une pile de la tour de Hanoi
- les chemins hamiltoniens sur les hypercubes
- le problème du baguenaudier (voir Lucas, récréations)
-

puis, de la même manière, mots sur un alphabet de trois lettres et fonctions projetées sur les sommets des arbres ternaires, tripartitions d'un ensemble, liées à l'énumération, récursive ou non, des points de la fractale de Sierpinski, càd de la fractale des coefficients binômiaux modulo 2 (injection des mots sur (0,1,2) dans les couples de mots sur (0,1)), ou encore, des états des trois piles de Hanoi.

.....
 \

Exercice 30 (suite de l'exercice précédent; permutations sur les sorties):

a) définition: $f(n)=(\text{afficher } n; \text{ si } n < m \text{ alors } f(2n), f(2n+1)).$
 exécution: appel de $f(1)$.

Par exemple, pour $m=5$, celà provoque la sortie $s(m=5)$:

$$f(1)=1.f(2).f(3)=1.2.f(4).f(5).3.f(6).f(7)=1.2.4.f(8).f(9).5.3.6.7.=1.2.4.8.9.5.3.6.7.$$

dont l'évaluation se fait usuellement de manière latéralisée, par utilisation d'une pile:

$f(i)$ joue le rôle d'une variable remplacée par sa définition, i celui d'une constante évacuée en sortie:
 $f(1)=1.f(2).f(3)=1.f(2).3.f(6).f(7)=1.f(2)*3.6.7=1.2.f(4).*5.3.6.7=1.2.4.f(8).f(9)*5.3.6.7$
 $=*1.2.4.8.9.5.3.6.7$

b) définition: $f=(\text{afficher } n; \text{ si } n < m \text{ alors } n:=2n, f, n:=n/2, n:=2n+1, f, n:=(n-1)/2).$
 exécution: appel de $f(1)$.

où $n/2$ est ici la division entière.

c) Qu'est-ce qui ne va pas dans la version non récursive qui suit ? :

exécution: initialisation $n:=1;$
 tant que $n > 0$
 (afficher n ; si $n < m$ alors $n:=2n$ sinon (tant que $n \bmod 2 = 1, n:=(n-1)/2$), $n:=n/2, n:=2n+1$).

d) Que dire des variantes :

définition: $f(n)=(\text{si } n < m \text{ alors } f(2n), \text{ afficher } n, f(2n+1)).$
 exécution: appel de $f(1)$.

Pour $m=5$, celà provoque la sortie $s(m=5)$:

$$f(1)=f(2).1.f(3)=f(4).2.f(5).1.f(6).3.f(7)=f(8).4.f(9).2.5.1.6.3.7=8.4.9.2.5.1.6.3.7$$

définition: $f(n)=(\text{ si } n < m \text{ alors } f(2n), f(2n+1)), \text{ afficher } n) .$
 exécution: l'appel de $f(1)$ avec $m=5$ produit la sortie 8.9.4.5.2.6.7.3.1.

Et, par exemple,

définition: $f(n)=(\text{ si } n < m \text{ alors } f(2n+1), \text{ afficher } n \bmod k, f(2n)).$
 exécution: appel de $f(1)$.

lorsque m croît, pour $k=2$, la sortie converge vers
 $1\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ \dots$ (stable par morphisme $\mu(1)=110$, $\mu(0)=100$).
 mot infini dont le k -ième caractère $g(k)$ est fonction de l'écriture de k en numération base 3:
 $g(3n)=1$, $g(3n+2)=0$, $g(3n+1)=g(n)$.

Exercice 30 (l'évidente bijection "de Matula"; monoïdes commutatifs libres et générateurs):

Soit par exemple $p[i]$ le i -ème nombre premier, $p[1]=2$, $p[2]=3$, $p[3]=5$, $p[4]=7$, $p[5]=11\dots$
 Si a est une expression du magma général, (a) est une expression dite primitive.
 Notons $f[a]$ l'entier positif associé à l'expression a du magma général commutatif. Initialement, pour le générateur x , $f[x]=2$. Si a est **un mot du magma général commutatif**, composé de $n>0$ facteurs $a[i]$, pour i de 1 à n , soit $a =$ produit magmatique des $a[i]$, alors $f[a]$ est défini par

$$f[a] = \text{produit des } p[f[a[i]]], \text{ pour } i \text{ de } 1 \text{ à } n, \\ \text{et si } a = (b), a \text{ est facteur primitif, } f[a] = p[f[b]].$$

Nous avons une évidente bijection du monoïde commutatif des arbres généraux commutatifs, (considéré comme engendré par les arbres primitifs) sur le monoïde multiplicatif entiers, les nombres premiers correspondant aux expressions magmatiques primitives.
 Ainsi, les entiers associés aux arbres x , (x,x) , (x,x,x) , (x,x,x,x) , (x,x,x,x,x) sont les puissances de 2.
 Les entiers associés aux arbres x , (x) , $((x))$, $((((x))))$, $(((((x))))))\dots$ sont les nombres premiers engendrés par l'itération $q:=p(q)$, sous l'initialisation $q:=2$.

Certains se sont intéressés à tabuler et examiner les entiers associés aux arbres commutatifs pointés comportant n sommets (le premier à avoir dénombré les arbres commutatifs généraux dont un sommet est pointé, c'est Cayley, en 1857, sous le nom d'arbres enracinés), et diverses conjectures ont été avancées à propos du plus petit entier (resp. plus grand entier) associé à un arbre de n sommets, résultant de l'examen des suites dont le début a été calculé. Par exemple,

-le plus petit entier, $\min(n)$, associé aux arbres de n sommets:
 $1\ 2\ 3\ 5\ 9\ 15\ 25\ 45\ 75\ 125\ 225\ 375\ 625\ 1125\ 1875\ 3125\dots$
 suggère que l'on aurait $\min(n+3)=5.\min(n)$, - à partir de $15=\min(2)*5$ - Pourquoi ?

-le plus grand entier $\max(n)$ associé aux arbres de n sommets:
 $1\ 2\ 4\ 8\ 19\ 67\ 331\ 2221\ 19577\ 219613\ 3042161\dots$
 on a $\max(n+1)=$ le nombre premier de rang $\max(n)$, - à partir de $n=3$ - Immédiat.

On s'est évidemment intéressé à relier les propriétés d'un arbre à celles de l'entier associé.

Exercice 31 :

L'équation des arbres binaires, $S=x+S*S$, conduit à l'itération $P:=x+P*P$ sous l'initialisation $P:=x$, laquelle engendre les arbres selon profondeur maximale.

Que représente l'itération numérique $s:=a+s^2$ sous l'initialisation $s:=a$, convergeant pour $0<a \leq 1/4$.
 s converge vers $1/2$ pour $a=1/4$, diverge pour $a>1/4$.
 (pour l'initialisation $a=1$, la suite des valeurs de s décompte les arbres de profondeur inférieure à n).

note: Les nombres $P(i)$ des arbres de profondeur i satisfont à $P(0)=1$, et $P(i+1)=S(i)P(i)+P(i)S(i)+P(i)P(i)$, où $S(i)$ est la somme des $P(n)$ pour $n<i$.
 Il a été établi que l'on a, remarquablement

$$P(n+1)/P(n) = 1 + P(n) + (4*P(n)-3)^{1/2}, \text{ sous } P(0)=1. \\ \text{soit} \\ P(n)^2(4*P(n)-3) = (P(n+1)-P(n))^2 - P(n)^2.$$

\

Eponymie: Eugène Charles Catalan (1814-1894, de Bruges à Liège) a donné son nom à ces nombres si présents en combinatoire. Probablement avait-il compté les triangulations de polygones avec plus d'élégance que ses prédécesseurs ou ses contemporains.

Une assez célèbre conjecture arithmétique, que l'on doit à Catalan, paraît avoir été hissée au statut de théorème: le couple (8,9) d'entiers consécutifs est le seul qui soit constitué de puissances consécutives (cela avait été établi dès le quatorzième siècle en ce qui concerne les seules puissances de 2 et 3).

Compléments combinatoires formels: arbres, séries formelles, et opérateurs.

L'inversion des séries par substitution, ou "inversion de Lagrange".

(à passer impérativement en première lecture)

Un morphisme g ,
des séries sur la structure du magma engendré par x ,
satisfait à $g[(m,m')] = (g[m], g[m'])$, et $g[m+m'] = g[m] + g[m']$.

Si l'on considère le morphisme g envoyant
-le générateur x
-sur la série formelle $g(x)$,
série caractéristique des expressions du magma binaire
on a, par définition, $g[x] = x + (g[x], g[x])$,
manière seulement préférable d'écrire $M = x + (M, M)$,
exprimant qu'un arbre binaire, c'est soit x , soit un couple d'arbre binaires.

D'où $x = g^{-1}[x] + (x, x)$, soit encore $g^{-1}[x] = x - (x, x)$.

Les séries $g[x]$ et $g^{-1}[x]$ sont inverses l'une de l'autre pour le produit magmatique, l'inversion étant prise au sens de la substitution.

Détaillons: ceci signifie que si l'on considère l'ensemble des expressions magmatiques (soit, ici, $g[x]$), et que l'on remplace dans $g[x]$ toute occurrence de x par $x - (x, x)$, et que l'on développe ces expressions linéairement... il ne reste que x (toute expression magmatique comptant n occurrences de x - nous dirons "de degré n en x " - a pour image par g^{-1} un ensemble de 2^n expressions, chacune de celles-ci étant affectée d'un coefficient positif ou négatif).

Ainsi, par exemple, l'image par g^{-1} de l'objet (x, x) du magma, c'est le polynôme $(x, x) - ((x, x), x) - (x, (x, x)) + ((x, x), (x, x))$.

Si l'on calcule soudain dans le monoïde, c'est-à-dire si l'on rend brutalement le calcul associatif, et si l'on introduit un élément neutre, e [c'est-à-dire que l'on prend l'image associative de la série], alors $g(x) = x + g(x^2)$ équivaut à $g^{-1}(x) = x(e - x)$, d'où $x * g^{-1}(x) = x$, et finalement $g(x) = xg(x^*)$.

Si c_n est le coefficient de x^n dans $g(x)$, de $g(x) = xg(x^*)$ on déduit ceci: c_n est formellement égal à la somme des mots (sur l'alphabet des coefficients (c_j)) dont la somme des indices est $n-1$

(bien sûr, après ce calcul formel, l'interprétation numérique éventuelle consiste à remplacer chaque lettre par sa valeur numérique);

par exemple, puisque $c_1 = 1$, on a $c_2 = c_1 = 1$, $c_3 = c_2 + c_1^2 = 2$, et $c_4 = c_3 + c_2c_1 + c_1c_2 + c_1^3 = 2 + 1 + 1 + 1 = 5$.

Il est clair que cette récurrence ne s'interprète pas en termes d'arbres binaires; ces coefficients comptent ici des objets structurés d'une autre façon: on compte ici les arbres généraux, en fonction du nombre de leurs sommets, au lieu de dénombrer les arbres binaires en fonction de leurs sommets terminaux.

Le problème général:

Si le morphisme g était tel que $g(x) = xg(s)$, où s serait une série de coefficients (a_k) , la même récurrence permettrait de calculer les coefficients (c_n) de $g(x)$, à condition de multiplier les mots de degré $n-1$ (sur l'alphabet (c_j) des coefficients de $g(x)$) par a_{n-1} .

Ce qui signifie finalement que l'on énumère la totalité des arbres généraux, en inscrivant a_i sur tout sommet composé de i sommets. On regroupe ainsi les arbres généraux comptant n sommets pour calculer le coefficient de x^n dans $g(x)$; ces arbres sont codés par les mots polonais sur l'alphabet (a_i) .

Or, les conjugués de ces mots polonais sont obtenus comme coefficients de x^{n-1} dans s^n . On déduit donc, de $g(x) = xg(s)$, le calcul des coefficients de $g(x)$:
 $n * (\text{coefficient de } x^n \text{ dans } g(x)) = (\text{coefficient de } x^{n-1} \text{ dans } s^n)$.

Lorsque $s=x^*$, le coefficient de x^{n-1} dans s^n , c'est le n-ième coefficient binomial central, que nous noterons plus loin (n,n) =nombre de manières de partager un ensemble de $2n$ éléments en deux sous-ensembles de n éléments. Tel est le problème formel de l'inversion des séries, inversion au sens de la substitution, dite "inversion de Lagrange".

On aura remarqué que la série de catalan sous la forme $g(x)=x+g(x^2)$ réfère à la gènèse des arbres binaires, tandis que la forme $g(x)=xg(x^*)$ réfère à la gènèse des arbres généraux.

Bissections et trisections (*anticipation à passer également en première lecture*) :

Les structures de monoïde et de magma sont souvent étroitement liées. Par exemple, on appelle bissection du monoïde X^* engendré par l'alphabet X la donnée de deux parties, A et B , de X^* , telles que l'on ait $X^*=A^*B^*$.

Autrement dit, tout mot sur l'alphabet X se factorise en une suite de mots de A suivie d'une suite de mots de B .

En termes de séries formelles, $X^*=A^*B^*$ équivaut à

- $(e-X)^{-1}=(e-A)^{-1}(e-B)^{-1}$,
- où $(e-X)^{-1} = \sum X^n$ est la somme formelle des puissances X^n de X
- où X^n est l'ensemble des mots de degré n sur l'alphabet X
- où $(e-A)^{-1} = \sum A^n$ est la somme des puissances A^n de A
- où A^n est l'ensemble des mots constitués de n facteurs dans A ...

Donc, la relation $X^*=A^*B^*$ équivaut à $(e-X)^{-1}=(e-A)^{-1}(e-B)^{-1}$, qui équivaut elle-même à $(e-X)=(e-B)(e-A)$, soit $A+B=X+BA$, ce qui signifie que le produit d'un mot de B par un mot de A est soit dans A , soit dans B , et que les seuls mots de $A+B$ qui ne s'obtiennent pas ainsi sont ceux de X .

Ce qui signifie que, hormis les lettres de X , tout mot de A (resp de B) se factorise en un mot de B suivi d'un mot de A , et ceci de manière récurrente: les mots de A comme ceux de B apparaissent comme des objets du magma engendré par X , ce qui entraîne au passage que les intersections de X avec A et B ne sont pas vides.

Ainsi, les objets de A et B apparaissent comme des arbres binaires dont les sommets terminaux gauches (resp droits) sont marqués par des lettres de X appartenant à B (resp à A).

Un moyen de construire des bissections est le suivant. Limitons nous à un alphabet de deux lettres $X=[a,b]$, où a est dans A et b dans B . Le produit ba est un mot nouveau, appartenant soit à A , soit à B . Peu importe: tout mot nouveau m est ainsi placé soit dans A , soit dans B ; s'il est dans A , il sera composé avec tout mot m' de B défini avant lui, pour constituer un nouveau mot, $m'm$; et s'il est dans B , il sera composé avec tout prédécesseur m' placé avant lui, dans A , afin de constituer le mot nouveau mm' .

De fait, nous construisons là *une* solution de l'équation $BA+X=A+B$, par approximation polynômiale, en forçant l'itération $A_n+B_n=X+B_{n-1}A_{n-1}$ (sous l'initialisation $A_0+B_0=X$, bipartition initiale de X).

[On peut établir ainsi une bijection de l'ensemble des mots de $A+B$, constituant une solution de $BA+X=A+B$, sur l'ensemble \mathbb{N} des entiers naturels]

En notant $f(n)$ le nombre de mots de A_n-A_{n-1} , et $g(n)$ le nombre de mots de B_n-B_{n-1} , on a $f(n+1)+g(n+1)=f(n)g(n)+(somme des (f(n)g(i)+f(i)g(n)) pour i<n)$, sous $f(0)+g(0)=|X|$.

La bissection ne peut dissimuler sa structurefondamentamentarborescente; dans le magma engendré par x et y , deux langages d'arbres, G et D , sont disjoints; initialement x est dans G , y dans D , puis, si g est dans G et d dans D , l'objet (g,d) est placé soit dans G, soit dans D (exclusivement).

Inversement, soit donné un arbre binaire, (a,b) , du magma engendré par $[x,y]$.
Peut-il être engendré par une bissection ?

Plaçons a dans G , b dans D , et itérons l'analyse syntactique sur a et b , jusqu'au placement dans G ou

Plaçons a dans G , b dans D , et itérons l'analyse syntactique sur a et b , jusqu'au placement dans G ou D d'occurrences de x et y . Pour que (a,b) appartienne à une bissection, il faut et il suffit que tout arbre binaire figurant dans G (resp D) ne figure pas dans D (resp G). On engendre donc ce que l'on appelle de manière imagée les branches de l'arbre (a,b) , et l'on veut que l'ensemble des branches gauches de (a,b) soit disjoint de l'ensemble des branches droites de (a,b) .

Associons donc à tout arbre (a,b) le polynôme de ses branches gauches, défini par la récurrence $g((a,b))=a+g(a)+g(b)$, et de même $d((a,b))=b+d(a)+d(b)$, polynôme des branches droites; évidemment, $g((x,y))=x$, $d((x,y))=y$, tandis que l'on conviendra de $g(x)=g(y)=d(x)=d(y)=\emptyset$, vide. L'arbre (a,b) appartient à une bissection si et seulement si les polynômes $g((a,b))$ et $d((a,b))$ sont sans terme commun.

D'où l'existence de ce langage d'arbres binaires, constitué de ceux qui appartiennent à une bissection. Qui sait combien d'arbres binaires de degré n (resp de profondeur k) appartiennent à ce langage ? La contrainte est sévère, et ces arbres sont évidemment relativement rares. Mais encore ? Comment les engendrer, sinon en représentant tout arbre par le couple de polynômes $(g(a),d(a))$?

Enfin, signalons ici, sur un alphabet de deux lettres, $X=a+b$, la remarquable trisection du monoïde libre en trois sous-monoïdes libres $X^*=(Lb)^*(aLb)^*(aL)^*$, qui équivaut à $e-X=(e-aL)(e-aLb)(e-Lb)$, laquelle relation résulte (vérifiez) de la définition du langage L des mots de parenthèses comme somme de langages rationnels, $L=(aLb)^*$, équivalant à $L^{-1}=e-aLb$, soit (produit par L à gauche ou à droite), $L-e=LaLb=aLbL$, qui sont les équations génératrices algébriques du langage des mots de parenthèses.

Cette trisection a de multiples conséquences, certaines assez inattendues, dans divers domaines, probabilités et statistiques, logiciel de documentation automatique. Nous y reviendrons certainement. Autrement dit, tout mot $m=m_g m_c m_d$ sur l'alphabet X se factorise en un facteur gauche m_g dans $(Lb)^*$, un facteur central m_c dans $L=(aLb)^*$, un facteur droit m_d dans $(aL)^*$, soit $m=m_1 b m_2 b \dots m_i b m_{i+1} a m_{i+2} a m_{i+3} \dots m_{2i+1}$, où les facteurs m_1 à m_{2i+1} sont des mots de parenthèses, séparés par, tout d'abord des occurrences de b , ensuite des occurrences de a .

De la trisection $X^*=(Lb)^*(aLb)^*(aL)^*=A^*B^*C^*$ résulte $A+B+C+CBA=CB+CA+BA$.

En en prenant l'image commutative [$\gamma(L)$ =somme des $c_{n+1} a^n b^n$ pour n positif ou nul] on obtient quelques relations entre nombres de catalan c_n ($c_1=c_2=1, c_3=2, \dots$).

C'est l'occasion d'établir que le nombre de mots de degré $2n$ de $(Lb)^*$ (resp $(aL)^*$), c'est (n,n) , ce coefficient binomial qui compte le nombre d'équipartitions (en deux parts de n éléments chacune) d'un ensemble de $2n$ éléments.

De même, le nombre de mots de $(Lb)^*$ (resp $(aL)^*$) de degré $2n+1$, c'est le binomial $(n+1,n)$.

Pour cela, on interprète tout binomial (i,j) comme dénombrant les mots du mélange $a^i \mathbf{u} \mathbf{u} b^j$, ce polynôme caractéristique des mots comportant i occurrences de a (resp j occurrences de b), (axiomatique du mélange: $a^i \mathbf{u} \mathbf{u} b^j = a^{i-1} \mathbf{u} \mathbf{u} b^j + a^i \mathbf{u} \mathbf{u} b^{j-1}$, $a^0 = b^0 = e$, e neutre pour le produit $\mathbf{u} \mathbf{u}$).

Cette trisection fournit également le moyen d'une démonstration (élégante) du fait que le maximum de l'ensemble des binomiaux (p,q) contraints par $p+q=2n$ est atteint lorsque $p=q=n$. On construit pour cela, grâce à cette trisection, une injection des mots du mélange $a^{n+i} \mathbf{u} \mathbf{u} b^{n-i}$ dans ceux du mélange $a^n \mathbf{u} \mathbf{u} b^n$.

Pendant que nous y sommes, établir que la somme des $(i,j)^2$ contraints par $i+j=2n$, c'est (n,n) .

Cette trisection $X^*=(Lb)^*(aLb)^*(aL)^*=A^*B^*C^*$ est-elle également sous-tendue par une structure arborescente? On comparera à ce que donne le calcul dans le magma engendré par $[x,y]$ contraint par le système d'équations sur trois séries magmatiques $C=x+(C,B)$, $A=y+(B,A)$, $B=(C,A)$, soluble par les approximations polynômiales

$C_{n+1}=x+(C_n, B_n)$, $A_{n+1}=y+(B_n, A_n)$, $B_{n+1}=(C_n, A_n)$, $C_0=x, A_0=y$.

Attention: dans le monoïde, un mot de B ne se factorise pas d'une unique manière en un produit de deux mots de C et A. Dans la structure non associative du magma, le système de ces trois équations engendre des arbres binaires sur lesquels s'inscrivent des permutations (voir le chapitre "combinatoire des permutations") dont on étudiera les particularités combinatoires, conséquence de la forme des arbres. Le dénombrement de ces permutations donne les coefficients du développement en série de Taylor-Mac-laurin de fonctions numériques f,g,h, contraintes par le système différentiel $f'=fg, h'=gh, g=fh$, soit $f'=f^2h, h'=fh^2$, sous la donnée initials de $f(0),h(0)$.

Sur $X=a+b+c$, construisez par approximations le début d'une trisection équilibrée $X^*=A^*B^*C^*$, relation qui équivaut à $A+B+C=X+CB+CA+BA-CBA$, en récurrant à partir de $A_0=a, B_0=b, C_0=c$.

Sur $X=a+b$ avec b dans A et a dans C, vous retrouvez la trisection ci dessus, $X^*=(Lb)^*(aLb)^*(aL)^*$.

Signalons ici que le monoïde libre X^* , même pour le cas d'un alphabet fini minimal X, réduit à deux lettres, peut se factoriser (et d'une infinité de manières (!!!)) en un produit ordonné croissant au sens large (pour l'indice i dans un ensemble totalement ordonné) de sous-monoïdes libres A_i^* (ce que l'on appelle des codes) en nombre éventuellement infini (!!!), ces sous-monoïdes, finiment engendrés ou non, étant éventuellement monogènes (!!!), càd éventuellement engendrés par un mot unique,

-où, si l'on préfère, e-X s'écrit comme le produit ordonné (pour l'ordre inverse sur l'ensemble ordonné des indices) des polynômes, ou séries, $(e-A_j)$

-où encore: l'ensemble des produits ordonnés (pour l'ordre inverse sur les indices) d'un nombre impair de facteurs -strictement décroissant- coïncide avec l'ensemble des produits ordonnés d'un nombre pair de facteurs (les éléments de X sont rangés parmi les objets constitués de facteurs en nombre pair).

Par exemple, nous avons $A+B+C+CBA=X+CB+CA+BA$, pour la trisection $X^*=A^*B^*C^*$: tout mot de l'un des codes A,B,C est produit inversé (soit dans CB, CA ou BA) de deux mots de deux de ces codes. C'est encore vrai pour la factorisation en un nombre infini de codes.

Inversement (hors le cas des bisections) ces produits inversés ne sont pas tous des objets de l'union des codes (disjoints) A_i .

Voyons l'exemple de la construction de la factorisation infinie sur deux lettres,

$X^*=A_0^*A_1^*A_2^*...A_i^*... , X=a+b$, où les A_i sont des polynômes homogène -de degré i- (excepté A_0 , de degré 1 comme A_1), donc codes bipréfixes:

$$A_0=x, A_1=y, A_2=A_1A_0=y.x, A_3=A_2(A_1+A_0)=A_1A_0(A_1+A_0)=yx(x+y),$$

$$A_4=A_3(A_1+A_0)-A_2A_1A_0=A_3(A_1+A_0)-A_2^2=A_1A_0(A_1+A_0)^2-A_1A_0A_1A_0=yx(xx+xy+yy),$$

$$A_5=A_4(A_1+A_0)+A_3(A_2-A_1A_0)=A_4(A_1+A_0)=yx(xx+xy+yy)(x+y),$$

$$A_6=A_5(A_1+A_0)+A_4(A_2-A_1A_0)-A_3A_2(A_1+A_0)=(A_5-A_3A_2)(A_1+A_0)$$

$$=(yx(xx+xy+yy)(x+y)-yx(x+y)yx)(x+y)=yx(xxx+xyy+yyy)(x+y),$$

$$A_7=A_6(A_1+A_0)+A_5(A_2-A_1A_0)+A_4(A_3-A_2(A_1+A_0))+A_3A_2A_1A_0$$

$$=A_6(A_1+A_0)+A_3A_2^2=yx((xxx+xyy+yyy)(x+y)^2+(x+y)yxyx),$$

$$A_8=A_7(A_1+A_0)+A_6(A_2-A_1A_0)+A_5(A_3-A_2(A_1+A_0))+A_4(A_2A_1A_0-A_3(A_1+A_0))$$

$$=yx((xxx+xyy+yyy)(x+y)^2+(x+y)yxyx)(x+y) - yx(xx+xy+yy)yx(x^2+y^2+xy) \dots$$

.....

incitant à l'énumération des suites d'indices -soit, en première approximation, les partitions strictement monotones de l'entier n, partitions en parties distinctes, en tenant compte ensuite de la parité du nombre des parties (ainsi, $8=7+1=6+2=5+3=5+2+1=4+3+1$, six partitions)

de fonction caractéristique $(e+x)(e+x^2)(e+x^3)(e+x^4)...$

série dont les premiers coefficients sont 1 1 1 2 2 3 4 5 6 8 10 12 15 18 22 27 32 38 46 54 64 76 89...

Il y a ainsi seulement dix telles partitions en parts distinctes de l'entier 10 (y compris la partie unique réduite à l'entier 10 lui-même), et 64 de l'entier 20.

On s'intéresse naturellement à la suite des écarts (différences premières)

0 0 1 0 1 1 1 1 2 2 2 3 3 4 5 5 6 8 8 10 12 13 15 18 20 23 27 30 34 40 44 50 58 64...

Sur l'alphabet infini constitué de l'ensemble des lettres (A_j) , alphabet ordonné par les entiers naturels figurant en indice, la série $A_0^* A_1^* A_2^* \dots A_j^* \dots$ rassemble les mots croissants, au sens large, incluant le mot vide.

La série inverse $\dots(e-A_j)\dots(e-A_2)(e-A_1)(e-A_0)$ énumère les mots décroissants, au sens stricts, pondérés par la parité de leur degré, incluant le mot vide e .

D'un côté, le langage ND des mots non-décroissants (càd croissants au sens large), de l'autre la différence DP-DI, DP (resp DI) langage des mots strictement décroissants de degré pair (resp impair), et une bijection immédiate de DP.ND sur $e+DI.ND$, en considérant la plus petite lettre du mot $m.m'$ (la dernière de m ou la première de m') où m est dans DP et m' dans Nd'ou $DP.ND=e+DI.ND$.

Cette démonstration vaut pour tout alphabet (A_j) totalement ordonné.

Alors,

peut-on ne pas suivre l'ordre croissant sur les facteurs A_j^* de la décomposition du monoïde X^* ?

Par exemple, en postant les facteurs d'indices pairs successifs à gauche de A_0 , les impairs à droite,

peut-on avoir $X^* = \dots A_{2i}^* \dots A_4^* A_2^* A_0^* A_1^* A_3^* A_5^* \dots A_{2i+1}^* \dots$?

où A_j serait un polynôme homogène de degré j ,

série dont l'inverse est-elle ou non

$\dots(e-A_{2i+1})\dots(e-A_5)(e-A_3)(e-A_1)(e-A_0)(e-A_2)(e-A_4)\dots(e-A_{2j})\dots$?

Aurement dit, ce qui était vrai avec l'ordre induit par \mathbf{N} le serait-il encore pour un ordre induit par \mathbf{Z} ?

Ensuite, utiliser, par exemple, un ordre où A_0 serait le facteur le plus à gauche, A_1 le facteur le plus à droite, $X^* = A_0^* \dots A_1^* \dots A_1^*$, avec un ordre spécial \ll sur les entiers, $0 < i < 1$.

Essayer avec un seul mot dans chaque A_j , $X = (x,y)$, et l'ordre de l'alphabet **initialisé par $x < m < y$** pour tout mot m non vide de degré supérieur à 1.

Voici donc un exemple de factorisation, épatant, où X^* apparaît comme le produit de monoïdes monogènes A_i^* , pour l'indice i dans un ensemble infini dénombrable pourvu d'un plus petit et d'un plus grand élément : soit l'alphabet $X=(x,y)$, partir de $x < y$, utiliser l'ordre alphabétique usuel sur X^* , placer un seul mot dans chaque A_j , et pour tout couple (m,m') de mots engendrés, avec $m < m'$, engendrer le mot mm' , le placer entre m et m' , de préférence à sa place alphabétique usuelle.

En voici le début:

$x < xy < y, x < xxy < xy < xyy < y, x < xxxy < xxy < xxyxy < xxyxyy < xxyy < xy < xyxy < xyy < xyyy < y, \dots$

On produit ainsi de proche en proche la totalité de ces mots que l'on appelle primitifs et standards, totalement et strictement ordonnés par l'ordre croissant alphabétique usuel, et il se trouve que tout mot

de X^* est, d'une unique manière, produit d'une suite décroissante (au sens large) de mots primitifs standards: c'est faire apparaître X^* comme produit infini ordonné de monoïdes monogènes A_i^* , où,

entre deux indices, $i < j$, il existe une infinité d'indices k , $i < k < j$.

En particulier, observer le sous-langage des mots obtenus par composition des seuls mots au contact lors de cette construction, excluant par exemple $xxyxyy$, composé de xxy et xyy , qui ne sont pas au contact, tandis que $xxyxy$ est composé de xxy et xy , qui sont au contact: tout mot de ce sous-langage est composé, d'une manière unique de deux mots du langage: la composition est donc magmatique, càd que tout mot de ce sous-langage est naturellement la représentation d'un arbre binaire.

Par exemple, $xxyxyy = ((x,(x,y)),((x,y),y))$. Nous y reviendrons.

Rappel de quelques définitions:

-un mot m' est primitif ssi il ne peut s'écrire m^k , quels que soient le mot m et l'entier k supérieur à 1,

-pour deux mots m' et m'' quelconques, les mots $m'm''$ et $m''m'$ sont dits conjugués,

-un mot primitif est dit standard ssi il est le plus petit de sa classe de conjugaison

(pour l'ordre habituel, induit par celui de l'alphabet).

L'algorithme ci-dessus de genèse des mots primitifs standards est fondé sur le fait que tout mot m est primitif et standard (exception faite des générateurs x et y) ssi il peut s'écrire $m'm''$ avec $m' < m''$, où m' et m'' sont eux mêmes primitifs et standards (mais cette factorisation n'est évidemment pas unique;

Définitions comparées de l'ordre usuel et des objets standards, en ce qui concerne les mots et les arbres binaires.-

L'ordre alphabétique, un ordre total sur les mots.-

x et y sont des lettres, m, m', m'' des mots.

l'initialisation: $x < y$ entraîne $xm < ym'$.

les récurrences: d'une part m' non vide implique $m < mm'$, et d'autre part $m < m'$ implique $m''m < m''m'$.

L'ordre lexicographique, un ordre total sur les arbres binaires.-

l'initialisation: x distinct de a implique $x < a$.

la récurrence: $a < b$ implique $(a, c) < (b, d)$ et $(c, a) < (c, b)$.

Définition et génèse des mots standards.-

l'initialisation: x et y sont standards.

la récurrence: si m et m' sont standards avec $m < m'$, alors mm' est standard.

Définition et génèse des arbres binaires standards.-

l'initialisation: x et y sont standards.

la récurrence: si a et b sont standards avec $a < b$, alors (a, b) est standard.

le dénombrement:

initialisation: $f(1) = 2$.

récurrence: $f(i)$ (resp $f(j)$) arbres standards de degré i (resp j) produisent $f(i) * f(j)$ arbres de degré $i + j$.

$f(i)$ arbres standards de degré i produisent $f(i) * (f(i) - 1) / 2$ arbres standards de degré $2i$.

Petite introduction au formalisme des Dérivations et Morphismes.

Séries d'opérateurs. (à passer en première lecture)

Soit à nouveau la série magmatique de Catalan, ni commutative, ni associative, caractéristique du magma binaire, définie par l'équation $S=x+(S,S)$ [série approchée par la suite des polynômes du magma construits par la récurrence de Catalan ($P:=x+(P,P)$ initialisée par $P:=x$)].

Soit **une série** u , quelconque, mais dans laquelle, si l'on veut éviter les problèmes de convergence, on pourra tout d'abord supposer nul le coefficient de x .

Soit ensuite **la dérivation** $Dx=u$ [on pourra prendre comme exemples $u=(x,x)$ et $u=S-x$], opérateur linéaire tel que $D(a,b)=(Da,b)+(a,Db)$ sur tout objet (a,b) du magma.

Il est facile d'établir, formellement, que l'exponentielle de D ,

l'opérateur $T=\exp(D)=\sum D^n/n!$ [la sommation formelle d'opérateurs portant sur n entier naturel]

est un morphisme. [$D^0/0!=I$ est l'opérateur identité], comme il en va de l'exponentielle de toute dérivation, dès qu'elle a un sens.

En effet, sur tout objet (a,b) du magma, nous avons,

[en notant, comme nous le faisons toujours, (i,j) le coefficient binomial = nombre de bipartitions d'un ensemble à $i+j$ éléments en deux parties de respectivement i et j éléments = $(i+j)!/(i!*j!)$ = nombre de mots comportant i occurrences de la lettre x et j occurrences de la lettre y]

$D^n(a,b)=$ somme formelle des $(i,j)(D^i a, D^j b)$, pour $i+j=n$, i,j entiers naturels.

Donc $D^n(a,b)/n!=$ somme des $(D^i a/i!, D^j b/j!)$, pour $i+j=n$.

Et finalement, $T(a,b)=(Ta, Tb)$, ce que l'on avait en vue.

Le morphisme inverse, c'est évidemment $T^{-1}=\exp(-D)$, tandis que $D=\text{Log}(T)=(T-I)-(T-I)^2/2+\dots$

De $S=x+(S,S)$ résulte $DS=Dx+(DS,S)+(S,DS)$, et $TS=Tx+(TS,TS)$.

Avec l'image associative s de S (et, comme x est l'unique générateur, cette image est nécessairement commutative), on a des relations de même type.

Ainsi, $Ds=Dx+2sDs$. [on sait que les coefficients de s sont les nombres de Catalan c_n]

En prenant par exemple $Dx=x$, on a, pour $n>1$, $nc_n=2^*($ somme des $j c_j c_i$ sur $i+j=n$).

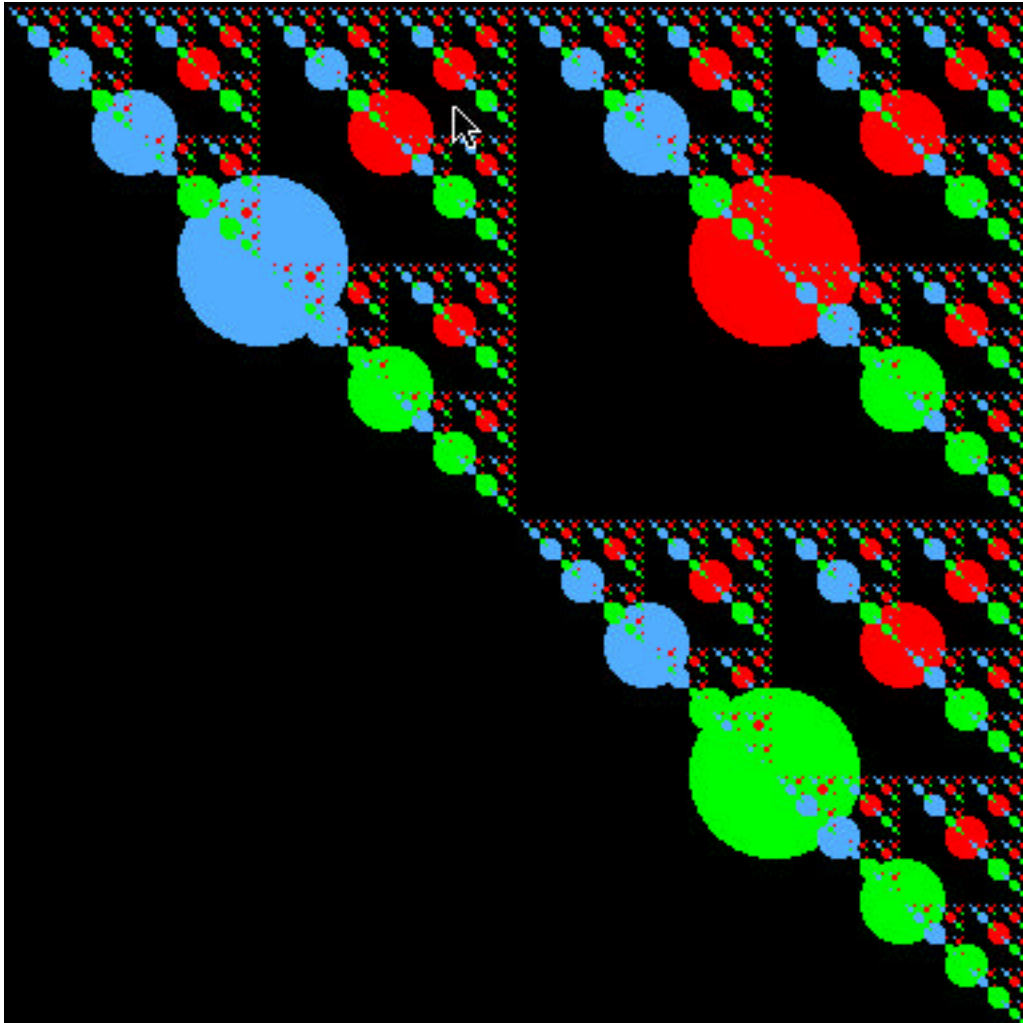
[$j c_j$ est le coefficient binomial central $(j-1, j-1)$]

En considérant $D^2s=x+2(Ds)^2+2sD^2s$, on a, pour $n>1$

$n^2 c_n=2^*($ sommation des $(i+j)j c_j c_i$, sur $i+j=n$), etc....



Fin du chapitre "Magma et arborescences".



Comment faire ça? (...il manque un gros disque, au centre)

