Les débuts de la combinatoire des mots

- Chronologie
- Thue et les mots sans carré
- Mots et arithmétique
- Régularités inévitables
- Dynamique symbolique
- Mots Sturmiens
- Bracelets
- Problèmes de mots
- Codes de Gauss

Chronologie

1736 Théorème d'Euler.

1772 Bernoulli: 'Sur une nouvelle espèce de calcul'.

1823 codes de Gauss.

1851 Suite de Prouhet.

1882 Dyck : 'Gruppentheoretischen Studien'.

1892 Formule de McMahon (nombre de colliers).
Poincaré : 'Les Méthodes Nouvelles de la Mécanique Céleste'.

1894 Flye Sainte-Marie : énumération des mots de de Bruijn binaires.

1898 Hadamard : 'Les surfaces à courbure opposées et leurs géodésiques'.

1902 Le problème de Burnside.

1906 Premier article d'Axel Thue : il y a des mots infinis sans carré.

1912 Deuxième article d'Axel Thue : la suite de Thue-Morse.

1921 Morse : 'recurrent geodesics on a surface of negative curvature'.
Nielsen : sous-groupes du groupe libre.

1927 Birkhoff : 'Dynamical Systems'.
Théorème de van der Waerden.

1930 Théorème de Ramsey.

1936 Conjecture d'Erdös et Turan.

1937 Article de Witt sur les algèbres de Lie libres.
Cours de Morse sur la dynamique symbolique.
Mot infini sans carré d'Aršon.

1938 Morse et Hedlund : 'Symbolic dynamics I'.

1940 Morse et Hedlund : 'Symbolic dynamics II : Sturmian sequences'.

1946 Mots de de Bruijn.

1947 Post : le problème des mots dans les semi-groupes est indécidable.

1954 Lyndon : 'On Burnside problem I'.

1955 Novikov : le problème des mots dans les groupes est indécidable.

1963 Chomsky and Schützenberger : 'Algebraic theory of context-free languages'.

1965 Schützenberger : 'Factorizations of free monoids'.

1968 Adjan et Novikov : solution du problème de Burnside.

Thue et les mots sans carrés

Le premier mot sans carré (Thue 1906)

$$adbcbabcbdabdcbabcdbabdcb\cdots$$

est obtenu comme point fixe de

$$\begin{array}{rcl} a & \to & adbcb \\ b & \to & abdcb \\ c & \to & abcdb \\ d & \to & abcbd \end{array}$$

Le mot de Thue-Morse (Thue 1912)

$$t = abbabaabbaababba\cdots$$

est obtenu en itérant la substitution

$$\begin{array}{rcl} a & \to & ab, \\ b & \to & ba. \end{array}$$

Thue prouve qu'il est sans chevauchement : pas de facteur $uvuvu$ avec u, v non vides. En codant a pour abb, b pour ab et c pour a, on obtient le mot sans carré sur trois lettres

$$abcacba\ldots$$

Mots et arithmétique

On cherche $(a_1, a_2, \ldots, a_k)$ et $(b_1, b_2, \ldots, b_k)$ tels que

$$a_1^i + a_2^i + \ldots + a_k^i = b_1^i + b_2^i + \ldots + b_k^i, \quad (1 \leq i \leq n)$$

Prouhet a obtenu en 1851 une solution avec $k = 2^n$. Pour $n = 2$,

$$\begin{aligned} 0+3+5+6 &= 1+2+4+7 \\ 0^2+3^2+5^2+6^2 &= 1^2+2^2+4^2+7^2 \end{aligned}$$

C'est encore la suite de Thue-Morse :

$$\begin{array}{cccccccc} 0 & & & 3 & & 5 & 6 & \\ a & b & b & a & b & a & a & b \\ & 1 & 2 & & 4 & & & 7 \end{array}$$

Problème de Tarry and Escott : peut-on choisir $k = n + 1$?

$$\begin{aligned} \{0,4,5\} &\equiv_2 \{1,2,6\} \\ \{1,4,5,8\} &\equiv_3 \{2,2,7,7\} \end{aligned}$$

Mots automatiques

Le mot de Thue-Morse

$$t = 01101001\ldots$$

est 2-automatique (Cobham 1972) :

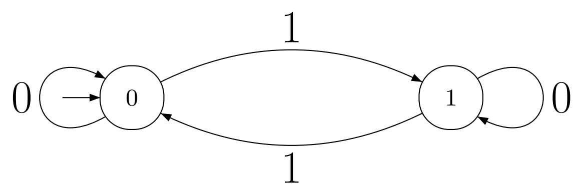


Figure 1: La parité de la somme des chiffres.

Autre caractérisation : la série $F(X) = \sum_{n\geq 0} t_n X^n$ est algébrique sur le corps $\mathbb{F}_2$:

$$(1+X)^3 F^2 + X(1+X)^2 F + X^2 = 0$$

Caractérisation de suites automatiques (Christol, Kamae, Mendès France et Rauzy, 1980).

Régularités inévitables

Motif inévitable p sur k lettres : tout mot infini sur k lettres a un facteur image de p par une substitution non effaçante (Bean, Ehrenfeucht, McNulty 1979, Zimin 1979).

Le motif xx est évitable sur trois lettres.

Le motif xyx est inévitable (comme toutes les sesquipuissances).

Le motif $xyxyx$ est évitable sur deux lettres.

On peut décider si un motif est inévitable (sur tout alphabet) (Zimin 1979)

Théorème de Ramsey

Pour k, r donnés et pour tout coloriage des k-parties d'un ensemble infini X en r couleurs, il existe une partie infinie Y telle que toutes les k-parties de Y sont de la même couleur (Ramsey 1930).

Application aux mots : si on colorie les facteurs d'un mot infini x en r couteurs, il existe une factorisation $x = u_0 u_1 u_2 \cdots$ telle que $u_1, u_2, \ldots$ sont de la même couleur.

Théorème de van Der Waerden

Si on partitionne les entiers en r classes, l'une d'elles contient des progressions arithmétiques de longueur arbitraire (van der Waerden 1927, conjecturé par I. Schur, Baudetsche vermutung).

Pour un mot infini x : il existe pour tout $k \geq 1$ des entiers $nm \geq 1$ tels que $x_n = x_{n+m} = \ldots = x_{n+km}$.

Généralisations :

- le théorème de Hales-Jewett theorem, de type Ramseyen.
- Théorème de Szemeredi (1975) : tout ensemble d'entiers de densité positive contient des progressions arithmétiques arbitrairement longues (conjecture d'Erdös et Turan 1936). Preuve utilisant la théorie ergodique par Furstenberg en 1977.

Systèmes dynamiques

- lois de Newton
- Analyse du XVIIIème siècle (d'Euler à Jacobi)
- de Poincaré (1892) à Hadamard (1898) et Birkhoff (1927)

Notion de stabilité ou de récurrence.

Si X est un espace topologique, et T une transformation continue sur X, un point $x \in X$ est dit récurrent si pour tout voisinage V de x il existe $n \geq 1$ tel que $T^n x \in V$ (on disait aussi : 'stable au sens de Poisson').

Le point x est dit uniformément récurrent si l'ensemble des entiers n comme ci-dessus a des trous de taille borné (Birkhoff ou Morse disent 'récurrence' et on dit aussi 'presque périodicité').

Dynamique symbolique

Marston Morse et Gustav Hedlund : flot symbolique. Les elements, les *trajectoires symboliques* sont des mots infinis . L'idée remonte à Hadamard (1898) : description qualitative des géodésiques sur une surface par un mot infini.

Morse prouve en 1927 l'existence de trajectoires uniformément récurrentes non périodiques sur certaines surfaces (à courbure négative) en utilisant le mot de Thue-Morse (qui est uniformément récurrent).

Mots sturmiens

Un mot infini binaire est sturmien si pour tout $n \geq 0$, le nombre de ses facteurs de longueur n est $n+1$. Par exemple, le *mot de Fibonacci*

$$f = abaababaabaab \cdots$$

qui est l'unique point fixe de $a \to ab, b \to a$ est Sturmien. Facteur spécial (à droite) : peut être

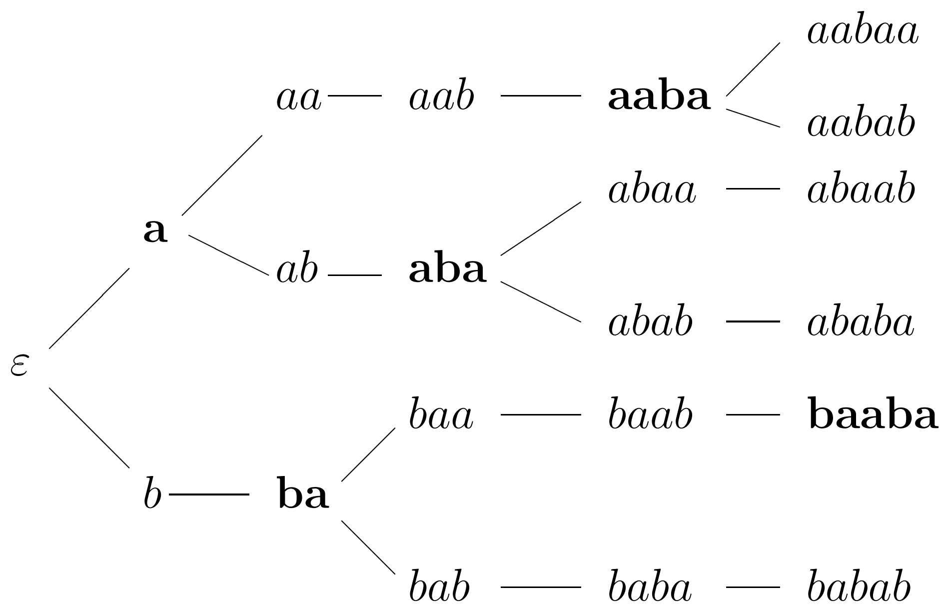


Figure 2: Les facteurs du mot de Fibonacci.

suivi par les deux lettres.

Introduits par Morse et Hedlund en 1940.

Charles François Sturm (1803-55), né à Genève, enseigna à l'Ecole Polytechnique à partir de 1840. Il est célèbre pour sa règle de calcul du nombre de racines réelles d'un polynôme. Son nom est utilisé par référence au thérème de séparation sur les zéros des solutions d'une équation différentielle comme

$$y'' + \phi(x)y = 0$$

Si ϕ est continue de période 1, si k_n est le nombre de zéros d'une solution sur l'intervalle $[n, n+1)$, alors le mot infini $01^{k_0}01^{k_1}\cdots$ est Sturmien (ou ultimement périodique).

Mots caractéristiques

En fait, il existe une autre définition du mot de Fibonacci avec des approximations d'irrationels par des rationels.

Soit $s_\alpha = (s_n)$

$$s_n = \lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor$$

avec $\alpha = 1/\varphi^2$ with $\varphi = (1+\sqrt{5})/2$ Alors $s = 0f$. Tout mot Sturmien peut être défini de cette façon (Morse, Hedlund, 1940). Mots caractéristiques ou

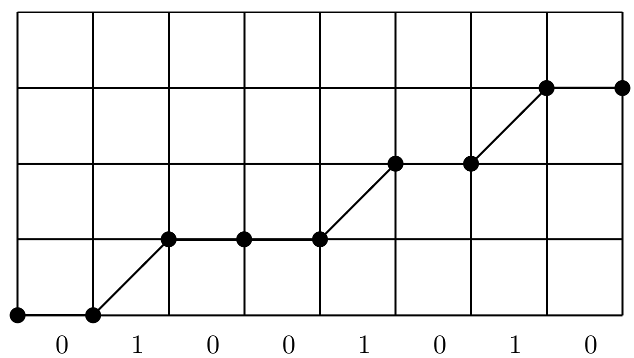


Figure 3: Represéntation du mot de Fibonacci

mécaniques (étudiés par Jean Bernoulli III 1772, en liaison avec les fractions continues).

Fractions continues

$$[n_0, n_1, n_2, \ldots] = n_0 + \cfrac{1}{n_1 + \frac{1}{n_2 + \ldots}}$$

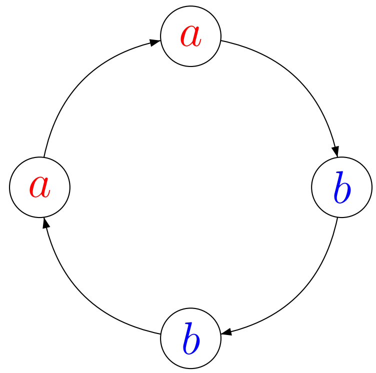
Soit $[0, 1+d_1, d_2, \ldots]$ le développement en fraction continue d'un irrationnel α avec $0 < \alpha < 1$. Soit w_n la suite de mots définie par

$$w_{-1} = 1, \quad w_0 = 0, \quad w_n = w_{n-1}^{d_n} w_{n-2} \quad (n \geq 1).$$

On a alors $s_\alpha = \lim w_n$. Pour le mot de Fibonacci $s_\alpha = 0f$ avec $\alpha = 1/\tau^2$. De fait, $1/\tau^2 = [0, 2, 1, 1, \cdots]$ qui correspond au fait que (w_n) est la suite des mots de Fibonacci.

Colliers

Un mot circulaire, ou collier, est la classe d'équivalence par d'un mot par décalage circulaire. L'énumération



des colliers primitifs de longueur n sur k lettres est donnée par la formule

$$M(n,k) = \frac{1}{n} \sum_{d|n} k^d \mu(n/d)$$

où μ est la fonction de Möbius est appelée formule de Witt 1937 (obtenue en liaison avec le théorème de Poincaré-Birkoff-Witt).

La formule pour le nombre total de colliers

$$N(n,k) = \frac{1}{n} \sum_{d|n} k^d \varphi(n/d)$$

où φ est la fonction d'Euler, est attribuée à MacMahon.

Mots de de Bruijn

Il existe pour tous $n, k \geq 1$ des colliers de longueur n^k sur k lettres tels chaque mot de longueur n sur k apparaît une fois et une seule comme facteur. C'est un mot de de Bruijn d'ordre n. Par exemple

$$x = 00000100011001010011101011011111$$

En fait, ce résultat avait déja été

- obtenu par C. Flye Sainte-Marie en 1894 .
- redécouvert par Martin et par Good .
- une simple conséquence du théorème d'Euler appliqué au graphe de de Bruijn.
- précisé par le BEST théorème (van Aardenne-Ehrenfest et de Bruijn 1951, Smith et Tutte 1941) qui les énumère.
- enigme de Postumus?

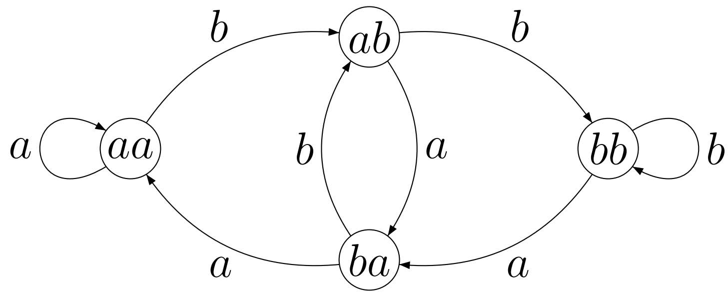


Figure 4: Le graphe de de Bruijn d'ordre $n = 2$.

Mots de Lyndon

Un mot de Lyndon est un mot primitif qui est minimal dans sa classe de conjugués (standard lexicographic sequences Lyndon 1955) .

$$aaaab, aaabb, aabab, aabbb, ababb, abbbbb.$$

Tout mot w se factorise de façon unique comme produit non croissant de mots de Lyndon (Schützenberger 1965).

$$w = x_1 x_2 \cdots x_n \text{ avec } x_1 \geq x_2 \geq \ldots x_n$$

Par exemple

$$abb\ ab\ aabb\ aababb\ a$$

Théorème de Chen, Fox et Lyndon?

Théorie combinatoire des groupes

- permutations, Lagrange 1771.
- 'groupe', Galois 1831.
- générateurs et relations, Dyck 1882. Classe de ϵ pour la réduction = langage de Dyck (Chomsky and Schützenberger 1963).

Problèmes des mots

Question : $u \equiv v$ dans un (semi)groupe finiment présenté $\langle A \mid R \rangle$. Système de Thue (Thue 1910, 1914).

- Décidabilité du problème des mots dans les groupes avec un seul relateur (Magnus,).
- Indécidabilité du problème des mots dans les semigroupes (Post 1947)
- Indécidabilité du problème des mots dans les groupes (Novikov, 1955).

Le problème de correspondance de Post

Soient $f, g : A^* \to B^*$ deux morphismes. Existe-t-il un mot x tel que $f(x) = g(x)$?

- Problème général (pour tous f, g) indécidable (Post 1947)
- Décidable pour $\mathrm{Card}(A) = 2$ (Ehrenfeucht, Karhumaki, Rosenberg 1982)
- Indécidable pour $\mathrm{Card}(A) = 9$

Le problème de Burnside

'Tout groupe finiment engendré satisfaisant l'identité $x^n = 1$ est-il fini'? (Burnside 1902)

- Réponse négative en général (Adjan et Novikov 1968). La solution utilise l'existence de mots infinis sans carrés (mot d'Aršon 1937).
- Un semigroupe finiment engendré satisfaisant l'identité $x^{r+1} = x$ est fini dès que les groupes f.e. satisfaisant $x^r = 1$ sont finis (Green et Rees 1952).
- Les classes du semigroupe f.e. satisfaisant $x^{n+1} = x^n$ sont rationnelles (problème de Brzozowski)? oui pour $n \geq 3$ (Guba).

Codes de Gauss

C'est le collier obtenu en notant les points d'intersection d'une courbe fermée du plan avec elle-même (points doubles transverses).

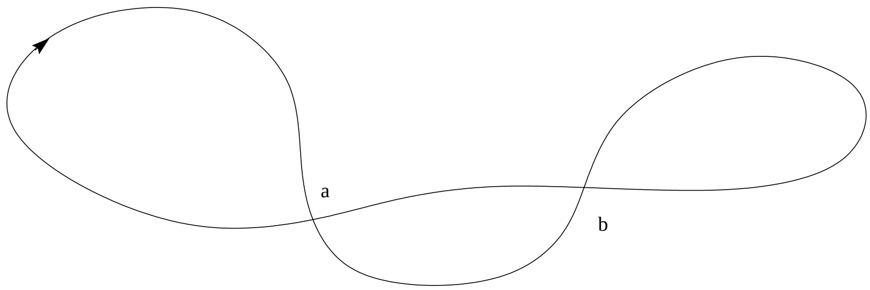
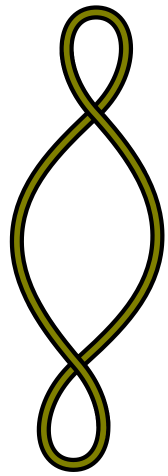


Figure 5: Le code de Gauss est *abba*

Gauss a observé la propriété suivante : entre les deux occurrences d'un symbole, la longueur est paire. Plusieurs caractérisations de codes de Gauss ont été données (Treybig 1968, Marx 1969, Rosenstiehl 1976).

Noeuds

Si on donne la suite $1, 2, -2, -1$ au site knotilus, on obtient



Si on donne la suite $-1, 3, -2, 1, -3, 2$, on obtient

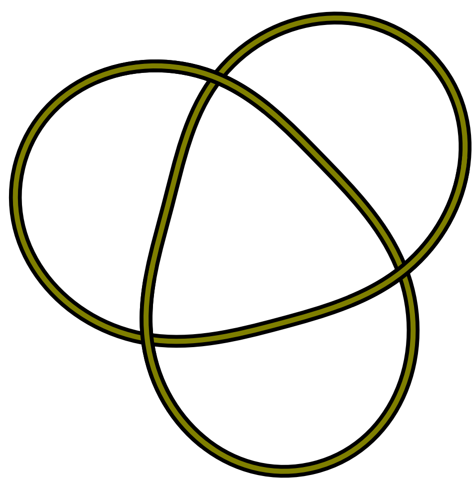


Figure 6: Le noeud de trèfle