Des mots magiques infinis

La suite de Thue-Morse et ses sœurs sont des structures ubiquitaires : pour les échecs, le dessin et la musique, essayez les mots sans carré et sans cube.

Pour intéresser un auditoire, ou pour écrire un texte en style soutenu, il faut éviter de se répéter. Cette consigne de non-répétition est plus complexe à respecter qu'on ne l'imagine et elle conduit à une série de résultats mathématiques qui décrivent la façon de s'y prendre... même dans le cas de textes infinis ! Lorsqu'il y a trop peu de lettres disponibles, on découvre des impossibilités absolues ; nées de considérations élémentaires, nous allons voir que ces questions constituent un riche domaine de recherche mathématique toujours en progrès.

Mots, sous-mots, carrés, cubes

Les répétitions que nous envisagerons sont uniquement les répétitions côte à côte. Les langages que nous utiliserons seront construits sur des alphabets de symboles distincts dont le nombre d'éléments sera précisé, car ce nombre détermine la difficulté des problèmes.

Le mot *cocorico* – construit avec le langage à quatre symboles *c, o, r, i* – contient la répétition de *co*. Le mot *abracadabra* ne contient pas de répétition : *abra*, présent deux fois, ne compte pas pour nous, car entre les deux occurrences de *abra* il y a une séparation. Le mot *cocorico* contient le carré de *co* et *abracadabra* est sans carré.

En français, les mots *ouistiti, taratata, rififi, barbare, rococo, piripiri, roudoudou, acacia, béribéri, nounours, toutou, nounou, tonton, poupoule, cacatoès, pom-pom-girl* et quelques autres (appréciés des jeunes enfants) sont des mots contenant des carrés. Je ne sais quel est le plus long carré qu'on trouve dans un mot du dictionnaire.

Boutros Boutros Gali, ancien secrétaire général des Nations unies, possède un nom qui contient un carré de 8 symboles (avec le blanc). Cela constitue-t-il le record pour les noms propres ? Le nom du lac péruvien *Titicaca* contient deux carrés, comme le titre de la chanson *Ploum Ploum tralala*.

Les mots avec cubes contiennent trois fois de suite le même sous-mot comme *Hahaha* (ce mot obtient le score de 935 000 chez Google), *chachacha* (505 000 sur Google), *booou* (956 chez Google), *beeeu* (907 chez Google). Le mot *créée* comporte un cube (si on néglige les accents). J'ignore quel est le plus long cube dans un nom, commun ou propre.

Si l'on utilise un alphabet infini, ne pas se répéter est plus que facile : il suffit d'utiliser les uns derrière les autres tous les symboles à sa disposition. À l'opposé, si on n'a que deux symboles (par exemple « 0 » et « 1 »), ne pas se répéter est très difficile puisque les seuls mots sans carré qu'on peut écrire avec 0 et 1 sont les six mots : 0, 1, 01, 10, 101, 010.

Réfléchissez quelques secondes, vous verrez qu'il n'existe aucun mot binaire sans carré de longueur 4 ou plus.

Le problème des mots à deux symboles sans cube est bien moins facile : il va nous plonger dans un tourbillon mathématique. Avant cela, je vous invite à rechercher un instant, un mot sans cube le plus long possible composé uniquement de 0 et de 1.

Le mot infini de Thue-Morse

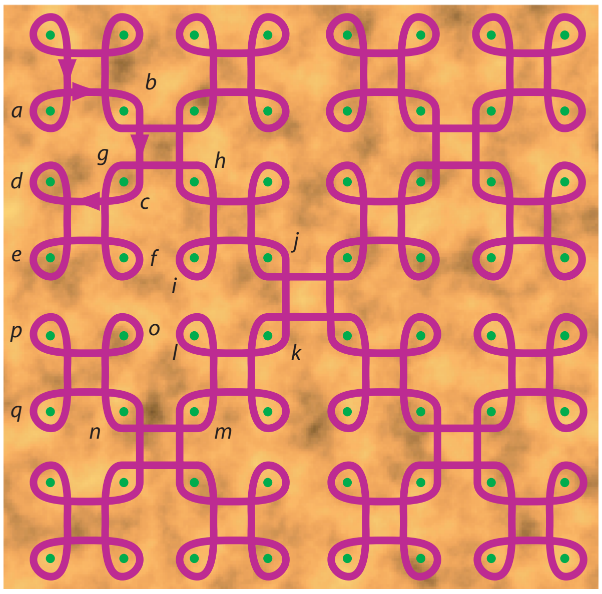
Voici un mot binaire sans cube de longueur 64 : 0110100110010110100101100110100110010110011010010110100110010110

Un regard attentif en livre le secret. Ce mot est construit sur le principe suivant : partir de 0, puis, six fois de suite, faire l'opération de remplacement *r* définie par *r* (0) = 01, *r* (1) = 10. On obtient ainsi : 0 ↦ 01 ↦ 0110 ↦ 01101001 ↦ 0110100110010110 ↦ 01101001100101101001011001101001...

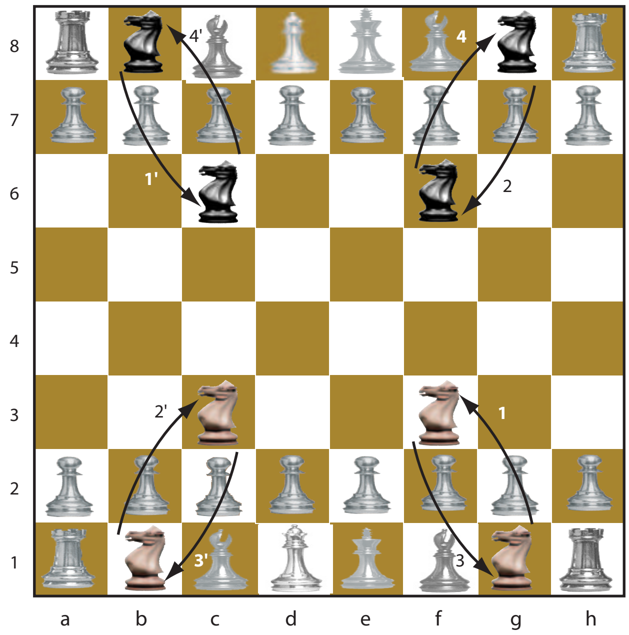
Une autre méthode conduisant au même résultat (mais ce n'est pas la même méthode !) consiste à partir de 0, puis, six fois de suite, à accoler ce que l'on a déjà en changeant les 0 en 1 et les 1 en 0, ce qui donne la suite 0 ↦ 01 ↦ 0110 ↦ 01101001 ↦ 0110100110010110 ↦ 01101001100101101001011001101001

On montre assez facilement (voir par exemple *Logique, informatique et paradoxes*, éditions *Belin/Pour la Science*, 1995, page 35) qu'un mot sans cube auquel on applique l'opération de remplacement *r* donne toujours un mot sans cube. Comme 0 est évidemment sans cube, il en est de même de 01, 0110, 01101001, etc.

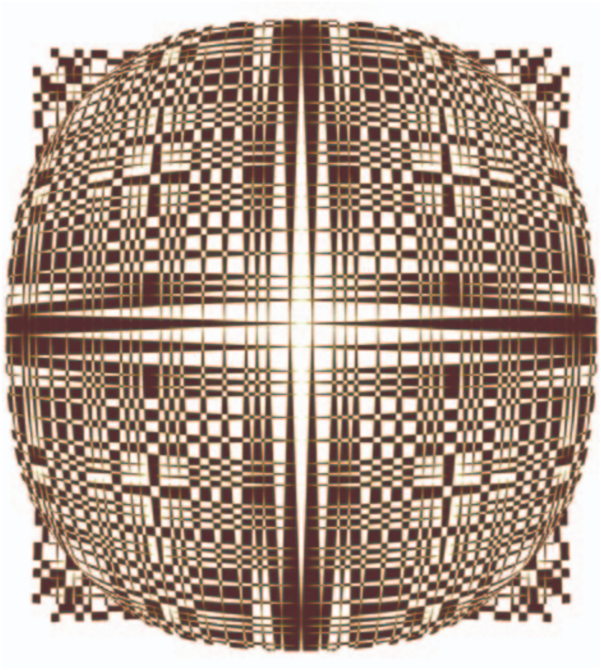
À chaque répétition de l'opération de remplacement *r*, on obtient un nouveau mot qui prolonge le précédent. Cela résulte de la remarque suivante : *r* (0) = 01 prolonge 0, donc *r* (01) prolonge *r* (0), c'est-à-dire 0110 prolonge 01. On en déduit que *r* (0110) prolonge *r* (01), etc. De proche en proche, on obtient que le (*n* +1)-ième élément de la suite de mots considérée



1. La suite de Thue-Morse et les Kolam indiens. Dans l'État du Tamil Nadu (Sud-Est de l'Inde), les mères enseignent à leurs filles le dessin, avec de la poudre de riz, des figures de Kolam. G. Allouche, J.-P. Allouche et J. Shallit ont récemment étudié une catégorie de dessins Kolam où un tracé *(à droite)* entoure un réseau de points disposés en carré. À chaque point rencontré, le tracé tourne tout autour du point (effectuant un angle de 270°) ou simplement le contourne (angle de 90°). Ce dessin entoure ici un réseau 8×8, mais on peut en faire pour tout réseau carré de côté une puissance de 2. Ces parcours sont secrètement déterminés par le mot de Thue-Morse. En effet, commençons par écrire 0, au point *a*. Ensuite, à chaque fois que nous avançons écrivons 00 ou 11 si nous arrivons à un point où le parcours fait une boucle de 270°, ou écrivons 0 ou 1 si nous arrivons à un point où le parcours fait un angle de 90°, tout cela en prenant soin d'alterner l'utilisation des 0 et des 1. 0. *a* : on rencontre une boucle : 11. *b* : on rencontre un angle de 90° : 0. c : on rencontre un angle de 90° : 1. *d* : on rencontre une boucle : 00. *e* : on rencontre une boucle : 11. *f* : on rencontre une boucle : 00. *g* : on rencontre un angle de 90° : 1. *h* : on rencontre un angle de 90° : 0. *i* : on rencontre une boucle : 11. *j* : on rencontre un angle de 90° : 0. *k* : on rencontre un angle de 90° : 1. *l* : on rencontre une boucle : 00. *m* : on rencontre un angle de 90° : 1. *n* : on rencontre un angle de 90° : 0. *o* : on rencontre une boucle : 11. *p* : on rencontre une boucle : 00. *q* : on rencontre une boucle : 11. Etc. La suite ainsi construite est celle de Thue-Morse (jusqu'à ce qu'on arrive en bas à droite) : 0110100110010110 10010110011.



2. Max Euwe démontra que la règle du jeu d'échecs qui impose de « ne jamais jouer trois fois de suite la même séquence de coups » autorisait des parties infinies (ce qui est « théoriquement » catastrophique). À chaque « 0 » de la suite de Thue-Morse, on associe la séquence de coups Cg1-Cf3, Cg8-Cf6, Cf3-Cg1, Cf6-Cg8 (les Blancs puis les Noirs sortent leur Cavalier gauche, puis le rentrent) et à chaque «1» de la suite, on associe la séquence de coups Cb1-Cc3, Cb8-Cc6, Cc3-b1, Cc6-b8 (les Blancs puis les Noirs sortent leur Cavalier droit, puis le rentrent). En suivant la suite de Thue-Morse 0110100110010110... on ne rejouera jamais trois fois de suite les mêmes coups et donc la partie durera indéfiniment.



3. Dessin géométrique obtenu à partir de la suite de Thue-Morse. On part d'un carré de 64 sur 64 suivant la suite de Thue-Morse (0 = noir, 1 = blanc). On le fait pivoter de deux degrés et on le superpose à son image symétrique en appliquant la règle : noir + noir = blanc. On applique ensuite au résultat obtenu précédemment une transformation qui fait gonfler le centre.

prolonge le n-ième. Il en résulte qu'on peut prolonger à l'infini cette suite de 0 et de 1, et obtenir ainsi une suite illimitée de 0 et 1 qui ne comportera jamais aucun cube.

Thue-Morse, les échecs et Prouhet

Cette suite remarquable est aujourd'hui bien connue des mathématiciens. Elle s'appelle la suite de Thue-Morse (ou mot de Thue-Morse). Elle a été présentée et étudiée en 1906 par Axel Thue (1863-1922) dans un obscur journal norvégien. Elle ne retint l'attention de la communauté scientifique que lorsque le mathématicien américain Marston Morse (1892-1977) la redécouvrit... en 1921. Elle fut retrouvée aussi par le grand maître d'échecs Max Euwe (1901-1981, champion du monde de 1935 à 1937) qui montra que la règle du jeu d'échecs autorisait des parties infinies, et cela dès le départ, sans qu'aucune pièce ne soit prise *(voir la figure 2)*.

La suite de Thue-Morse avait été remarquée avant l'article de Thue. Le mathématicien français E. Prouhet, dans un article de 1851 (*Mémoire sur quelques relations entre puissances d'un nombre*, Comptes Rendus de l'Académie des sciences de Paris, série I, 33, 1851, p. 225), prouvait une jolie propriété arithmétique. Si l'on met à gauche d'une égalité les 8 entiers correspondant aux numéros des 8 premiers emplacements des « 0 » de la suite – ce sont les entiers 0, 3, 5, 6, 9, 10, 12, 15 – et à droite les 8 entiers correspondant aux 8 premiers emplacements des « 1 » – ce sont 1, 2, 4, 7, 8, 11, 13, 14 – alors on obtient trois égalités remarquables simultanées :

$$0+3+5+6+9+10+12+15=1+2+4+7+8+11+13+14$$
$$0^2+3^2+5^2+6^2+9^2+10^2+12^2+15^2=1^2+2^2+4^2+7^2+8^2+11^2+13^2+14^2$$
$$0^3+3^3+5^3+6^3+9^3+10^3+12^3+15^3=1^3+2^3+4^3+7^3+8^3+11^3+13^3+14^3$$

En exploitant la suite de Thue-Morse un peu plus loin, ce résultat se généralise et donne 4 égalités entre deux paquets de 16 nombres, 5 égalités entre deux paquets de 32 nombres, etc. et conduit donc à l'affirmation très surprenante que : pour tout nombre n, on peut trouver deux ensembles de nombres entiers distincts A et B, tels que la somme des nombres de A est la même que la somme des nombres de B, de même pour les carrés, les cubes, ..., les puissances n-ièmes.

La suite de Thue-Morse (ou Prouhet-Thue-Morse comme on propose de la dénommer en France...) est un exemple remarquable de la rencontre de propriétés presque contradictoires dans un même objet mathématique : une totale simplicité et une grande complexité.

Nul doute, la suite de Thue-Morse est d'une grande simplicité : les trois égalités $t(0) = 0$, $t(2n) = t(n)$, $t(2n+1) = 1 - t(n)$ la déterminent totalement. Pourtant la suite de Thue-Morse produit un nombre transcendant : le nombre réel qui en base 2 s'écrit 0,1101001100101101001011001101001... est transcendant, c'est-à-dire n'est la solution d'aucune équation polynomiale à coefficients entiers (les nombres π et e dont les écritures sont, elles, totalement désordonnées en apparence sont aussi transcendants). La transcendance du nombre associé à la suite de Thue-Morse est un résultat difficile et remarquable, il a été démontré par Kurt Mahler (1903-1988) en 1929 et a récemment été complété par un autre beau résultat de Martine Queffélec de l'Université de Lille qui indique que la fraction continue :

4. Les propriétés de la suite de Thue-Morse

La suite infinie possède un aspect fractal : certains de ses morceaux la contiennent entièrement. Ainsi, en ne retenant qu'un élément sur deux de la suite, on retombe sur cette même suite (et donc aussi en n'en retenant qu'un élément sur quatre, ou en ne prenant qu'un élément sur huit, etc.). Pour obtenir un schéma régulier comme ci-contre au milieu, on place sur chaque ligne 2^n éléments de la suite ; il y a alors deux types de lignes *A* et *B* qui se suivent selon *ABBA BAAB*... c'est-à-dire selon : 01101001100101101001011001101001 0110011010 0101100110100101.

Le mot de Thue-Morse est un mot sans cube : on ne rencontre jamais trois fois consécutivement la même séquence (jamais on n'y rencontre donc 000, 111, 101010, 101010, etc.). Ce mot est même sans chevauchement, ce qui veut dire que jamais on n'y voit la structure *amama* avec *a* un symbole 0 ou 1, *m* un mot composé de 0 et de 1 (être « sans chevauchement » implique être « sans cube »).

Les apparitions de la suite de Thue

La suite de Thue-Morse apparaît dans des situations mathématiques variées sans rapports apparents. Voici un exemple d'une telle rencontre dans le domaine des partitions d'entiers.

Si on recherche un ensemble E d'entiers ayant la propriété « si *n* est dans E alors 2*n* n'y est pas », et qu'on mette le plus possible d'entiers dans E, on trouve que E est composé des entiers : 0, 1, 3, 4, 5, 7, 9, 11, 12, 13, 15, 17, 19, 20, 21, 23, ... On place 0, on place 1 (qui interdit 2), on place 3 (qui interdit 6), on place 4 (qui interdit 8), etc. Le complémentaire de *E* s'obtient en doublant les éléments de E (sauf 0) : 2, 6, 10, 14, 18, 22, 24, 26, 30, 34, 38, ...

Étrangement, comme l'ont découvert et démontré les mathématiciens J.-P. Allouche, A. Arnold, J. Berstel, S. Brlek, W. Jockush, S. Plouffe et B. Sagan, l'ensemble E des entiers ainsi retenus cache la suite de Thue-Morse. Considérons en effet les écarts entre les éléments successifs de E en notant cet écart alternativement : 0 pour un écart de une unité et 00 pour un écart de deux unités, 1 pour un écart de une unité et 11 pour un écart de deux unités, la suite obtenue est alors exactement celle de Thue-Morse :

0 1 3 4 5 7 9 11 12 13 15 17 19 20 21 23, ...
011 0 1001100 1 0 11 00 11 0 1 00...
= 01101001100101100110100...

On rencontre la suite de Thue-Morse dans de nombreux domaines mathématiques : des figures géométriques décoratives et fractales, des systèmes dynamiques réels, en passant par la géométrie différentielle, les produits infinis, les séries formelles algébriques, l'arithmétique et la théorie des groupes (voir dans la bibliographie l'article de J.-P. Allouche et J. Shallit, *The Ubiquitous Prouhet-Thue-Morse Sequence*).

La suite de Thue-Morse n'est pas répétitive puisqu'elle ne comporte aucun cube. En même temps, elle est très répétitive, car elle possède la propriété suivante :

– si un mot, par exemple 011, apparaît une fois dans la suite de Thue-Morse alors il apparaît une infinité de fois et les endroits où on le trouve ne sont pas de plus en plus éloignés : la séquence 011 apparaît par exemple au moins une fois dans toute séquence de 14 lettres consécutives de la suite.

```
0110100110010110
1001011001101001
1001011001101001
0110100110010110
1001011001101001
0110100110010110
0110100110010110
1001011001101001
1001011001101001
0110100110010110
0110100110010110
1001011001101001
0110100110010110
1001011001101001
1001011001101001
0110100110010110
1001011001101001
0110100110010110
0110100110010110
1001011001101001
0110100110010110
1001011001101001
1001011001101001
0110100110010110
0110100110010110
1001011001101001
1001011001101001
0110100110010110
1001011001101001
0110100110010110
0110100110010110
1001011001101001
1001011001101001
0110100110010110
0110100110010110
1001011001101001
0110100110010110
1001011001101001
1001011001101001
0110100110010110
0110100110010110
1001011001101001
1001011001101001
0110100110010110
1001011001101001
0110100110010110
0110100110010110
1001011001101001
0110100110010110
1001011001101001
1001011001101001
0110100110010110
1001011001101001
0110100110010110
0110100110010110
1001011001101001
1001011001101001
0110100110010110
0110100110010110
1001011001101001
0110100110010110
1001011001101001
1001011001101001
0110100110010110
1001011001101001
0110100110010110
0110100110010110
100101100110100 ...
```

Ce curieux compromis en ordre et désordre a attiré l'attention de plusieurs musiciens qui ont utilisé la suite de Thue-Morse comme élément structurel pour concevoir des morceaux de musique. Vous pourrez écouter un morceau de musique construit sur la suite de Thue-Morse en : http ://www.geocities.com/SiliconValley/Haven/4386/algorith.html. Ci-dessous une composition de Tom Johnson.

Voir aussi :

• Jean-Paul Allouche et Tom Johnson, *Finite Automata and morphisms in assisted music composition*, in *Journal of New Music Research*, vol. 24, pp. 97-108, 1995.

• Roland Assous, *Deux ou trois remarques sur l'utilisation des mots de Thue-Morse en informatique musicale*, in *Musique et mathématiques*, Grame, Lyon, 1997.

• http://kalvos.org/johness1.html

• http://reglos.de/musinum/

Le compromis subtil entre ordre et désordre que constitue le mot de Thue-Morse peut-être utilisé pour obtenir des dessins géométriques décoratifs *(voir la figure 3)*.

Un mot sans aucun cube et presque sans carré

Pour construire un mot sans cube et sans carré de longueur supérieure à 3, Narad Rampersad, Jeffrey Shallit et Ming-Wei Wang procèdent en deux étapes :

(a) Ils construisent d'abord un mot à quatre symboles qui ne contient aucun des sous-mots 12, 13, 32, 231 et 10302 en appliquant itérativement à partir de 0 la transformation de remplacement *t*(0) = 0310201023, *t* (1) = 0310230102, *t* (2) = 0201031023, t (3) = 0203010201, ce qui leur donne un mot infini.

0 ↦ 0310201023 ↦ 0302010230203010201031 0230102031020102302010310230310201 0230310230102031020102302010310 230203010201...

(b) Dans un second temps, partant du mot infini précédent, ils remplacent 0 par 010011, 1 par 010110, 2 par 011001 et 3 par 011010, ce qui donne :

010011011010010110010011011001010011010110010011011001011010000110101100100110110100100110101100100 110110010100110101 10...

Le mot obtenu ne contient aucun cube et ne contient aucun carré *mm* pour un mot *m* de longueur supérieure à 3. C'est donc un mot sans cube et presque sans carré.

$$\cfrac{1}{u_0+\cfrac{1}{u_1+\cfrac{1}{u_2+\dots}}}$$

où les u_i sont les termes de la suite de Thue-Morse augmentés de 1 (1221211221121221...), est aussi un nombre transcendant.

Le caractère paradoxal de la suite de Thue-Morse se manifeste aussi à propos des répétitions. Cette suite n'est pas répétitive puisqu'elle ne comporte aucun cube (ni de longueur 1, ni de longueur 2, etc.), mais en même temps, elle est très répétitive, car elle possède la propriété suivante : si un mot – par exemple 011 – apparaît une fois dans la suite de Thue-Morse alors il apparaît une infinité de fois, et les endroits où on le trouve ne sont pas de plus en plus éloignés : la séquence 011 apparaît par exemple au moins une fois dans toute séquence de 14 symboles consécutifs de la suite.

Comme le nombre π ou la suite de Fibonacci 1, 1, 2, 3, 5, 8, 13, ..., la suite de Thue-Morse est un objet ubiquitaire : on le rencontre dans de nombreux contextes mathématiques différents sans que l'on sache précisément pourquoi.

La figure 4 présente des propriétés étonnantes de la suite de Thue-Morse et indique des domaines où le mathématicien l'a rencontrée. G. Allouche, J-P. Allouche et J. Shallit ont récemment montré qu'elle est cachée dans la structure de certains dessins Kolam de l'Inde du Sud qu'on retrouve par ailleurs dans la tradition des figures dessinées sur le sable des peuples des îles Vanuatu *(voir la figure 1)*.

Sans ou peu de carrés ?

Revenons à notre question initiale. Puisque deux symboles ne suffisent pas pour obtenir des mots très longs sans carré, essayons avec trois. Existe-t-il des mots infinis construits avec les trois symboles 0, 1 et 2 et ne comportant aucun carré ?

Assez étrangement, la réponse – positive – est à nouveau donnée par la suite de Thue-Morse. En effet, si on considère la suite des nombres 0, 1 ou 2 indiquant combien il y a de « 1 » entre deux « 0 » de la suite de Thue-Morse, on obtient une suite infinie sans carré :

0 1 1 0 1 0 0 1 1 0 0 1 0 1 1 0 1 0 0 1 0 1 1 0 0 1 1 0 1 0 0 ...
2 1 0 2 0 1 2 1 0 1 2 0 2 1 0 ...

Cette suite ternaire est obtenue par un remplacement analogue à celui utilisé pour la suite binaire de Thue-Morse. On part de 2, puis on effectue itérativement les remplacements $s(2) = 210$, $s(1) = 20$ et $s(0) = 1$, soit : $2 \mapsto 210 \mapsto 210201 \mapsto 210201210120 \mapsto 210201210120210201202101 \mapsto \dots$

La démonstration que ces mots finis et le mot infini ternaire qui en résulte sont sans carré consiste à établir que si on applique les remplacements donnés par s à une suite sans carré, nécessairement la suite déduite n'a pas de carré non plus. Bien que les contraintes semblent fortes, il existe une infinité non dénombrable (comportant autant d'éléments qu'il y a de points sur une droite) de mots de 2 lettres sans cube et de mots de 3 lettres sans carré. Pour en savoir plus, on a étudié le nombre $c(n)$ de mots ternaires sans cube de longueur n.

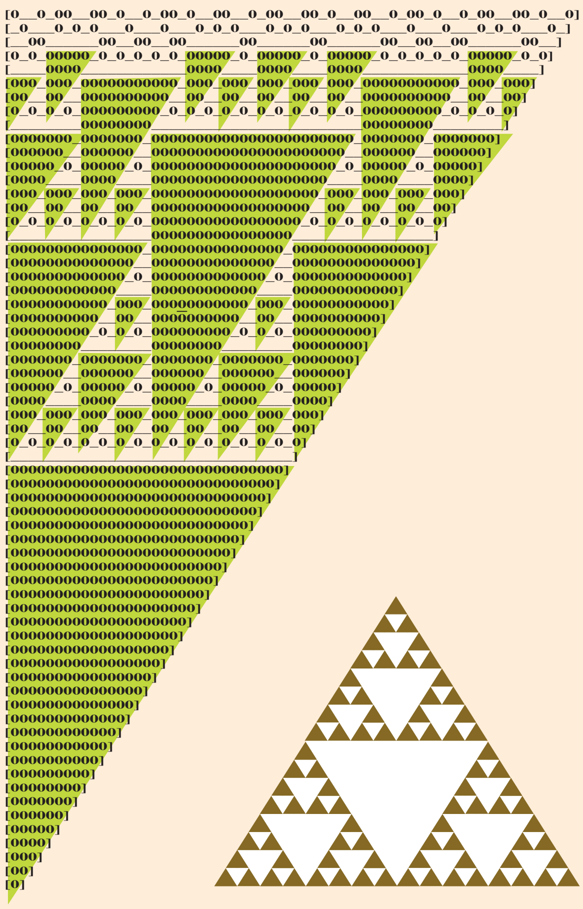
Une longue série de théorèmes de plus en plus précis – et difficiles à démontrer – a conduit au résultat que le nombre $c(n)$ est asymptotiquement supérieur à $110^{n/42}$, soit $1{,}118419132^n$. D'autres travaux se poursuivent comme ceux de Christoph Richard et Uwe Grimm qui ont prouvé qu'il existait des mots infinis ternaires sans carré dont la fréquence limite de chacune des trois lettres est exactement 1/3, et que d'autres fréquences limites (inégales entre les lettres) sont possibles. Une autre direction de recherche a récemment donné des résultats qui ont complété les découvertes d'Axel Thue. L'idée de ces travaux est naturelle : si on ne peut pas éviter tout carré dans un mot binaire, ne peut-on pas en éviter le plus possible et par exemple n'accepter que les carrés 00 et les carrés 11 ou que les carrés des mots de une ou deux lettres ?

Qu'un mot binaire infini n'ait que les carrés 00 et 11 est impossible. Pire encore, n'avoir que les carrés 00, 11 et 1111 est aussi impossible : le plus long mot n'ayant comme carré que 00, 11 et 1111 comporte 24 lettres :

101100111000111100111011.

Au-delà de 24 lettres, une recherche exhaustive par ordinateur montre que tout mot binaire comporte les carrés 00, 11 et un autre carré qui ne sera ni 0000 ni 1111. Le mieux que l'on puisse espérer d'un mot infini binaire est qu'il n'ait que les carrés 00, 11 et 0101 (ou 00, 11 et 1010, mais cela revient au

5. Les différences et Serpinski



Les différences successives des éléments de la suite de Thue-Morse (où les 1 sont remplacés par des tirets-) donnent une structure de triangles en vert « à la Serpinski », un objet fractal dont la structure est obtenue par une itération consistant à évider les triangles au voisinage des sommets.

6. Les suites sans inverse

Cinq résultats ont été démontrés récemment par N. Rampersad et J. Shallit sur les suites de symboles ne contenant jamais – ou presque jamais – l'inverse de leurs sous-mots (l'inverse de *abcd* est *dcba*).

(a) Le mot 012012012... est le seul mot infini ternaire (au nom des symboles près) qui ne contient l'inverse d'aucun de ses sous-mots de longueur 2 ou plus.

Ce mot infini est périodique. Et si nous voulions un mot non périodique ?

(b) Il existe un mot ternaire infini *R* non périodique, qui ne contient l'inverse d'aucun de ses sous-mots de longueur 3 ou plus. Autrement dit, si *R* contient en même temps un mot et son inverse, celui-ci a pour longueur 1 ou 2. On obtient *R* en partant d'un mot binaire non périodique – par exemple 0100100010000100001... – et en remplaçant chaque 0 par 0012 et chaque 1 par 0112, ce qui donne : 001201120012001 2011200120012001201120012001200120012001201120...

(c) Il n'existe pas de mot binaire infini qui ne contienne pas l'inverse d'un de ses sous-mots de longueur 4 ou plus.

Même si l'on autorise les inverses des sous-mots de longueur 1, 2 ou 3, aucun binaire infini ne convient. Il faut être plus tolérant et accepter que certains mots de longueur 4 soient présents avec leur inverse.

(d) Il existe des mots binaires infinis tels que les seuls inverses de leurs sous-mots ont pour longueur 1, 2, 3, ou 4. De tels sous-mots sont tous périodiques à partir d'un certain point.

Si en plus d'éviter (autant que possible) les inverses, on exige que le mot cherché ne soit pas périodique à partir d'un certain point, alors il faut tolérer des inverses de longueur 1, 2, 3, 4, 5.

(e) Il existe un mot binaire infini *J* qui n'est pas périodique à partir d'un certain point et qui ne contient jamais l'inverse d'un de ses sous-mots de longueur 6 ou plus. Autrement dit *J* ne contient que des inverses de mots de longueur 1, 2, 3, 4 ou 5. On construit un tel mot en partant d'un mot non périodique et en remplaçant 0 par 0001011 et 1 par 0010111.

même en échangeant les rôles de 0 et de 1). Un tel mot binaire infini avec pour seuls carrés 00, 11 et 0101 existe-t-il ?

Oui, et voici la description de celui proposé en 2006 par T. Harju et D. Nowotka dans le *Bulletin de la Société Européenne d'Informatique Théorique*. À partir d'un mot à trois lettres *A, B, C* sans carré (nous savons les obtenir) on remplace : chaque *A* par 111000110010110001110010 ; chaque *B* par 111000101100011100101100010 ; chaque *C* par 111000110010110001011100101100. Chacun des trois mots binaires considérés n'a que les carrés 00, 11 et 0101. De plus, on a démontré que leur substitution à *A, B* et *C* dans un mot sans carré n'introduit aucun nouveau carré.

Toujours dans le cadre de la recherche d'améliorations des résultats de Thue, un théorème remarquable montre que l'on peut dans un mot binaire infini éviter à la fois tous les cubes (ce n'est pas le cas de l'exemple précédent qui comporte les cubes 000 et 111) et presque tous les carrés.

N. Rampersad, J. Shallit et Ming-Wei Wang, de l'Université de Waterloo au Canada, ont proposé tout récemment un mot binaire infini qui ne contient aucun cube et qui ne comporte aucun carré *mm* pour un mot *m* de longueur supérieure à 3. Les seuls carrés présents dans ce mot sont ceux des mots 0, 1, 01, 10, 001, 010, 011, 100, 101, 110.

Une question a été posée il y a 30 ans : en mélangeant parfaitement (le nouveau mot est obtenu en prenant ses lettres alternativement dans les deux mots constituants) deux mots n'ayant que de petits carrés, peut-on obtenir un mot ayant des carrés de grandes tailles ? N. Rampersad, J. Shallit et Ming-Wei Wang ont résolu l'énigme : il existe deux mots binaires infinis *M* et *N* n'ayant aucun carré *mm* avec *m* de longueur supérieure à 3, et tels pourtant qu'en insérant *M* dans *N* on obtient un nouveau mot infini *P* qui possède des carrés aussi grands qu'on veut. La propriété «ne pas posséder de carré » ne se conserve pas dans le mélange.

Mots sans inversion

Indiquons enfin la résolution récente d'une autre jolie question concernant l'existence de suites infinies qui ne contiennent jamais en même temps un mot *m* et son inverse *m'* (sauf éventuellement pour les mots très petits). Les résultats obtenus par N. Rampersad et J. Shallit indiquent : (1) qu'il existe un mot ternaire infini unique (à une permutation près des trois lettres) qui pour tout mot *m* de longueur 2 ou plus ne contient pas *m* ou ne contient pas son inverse *m'* (essayez de deviner ce mot, c'est facile) ; (2) qu'il existe un mot ternaire infini non périodique, qui pour tout mot *m* de longueur 3 ou plus ne contient pas *m* ou ne contient pas son inverse *(voir la figure 6 pour d'autres résultats).*

Le monde mathématique est riche et varié, et à propos de problèmes parfois très simples – qu'on avait oublié de se poser –, les chercheurs y découvrent de nouveaux objets – finis ou infinis – aux propriétés subtiles et déconcertantes. Certains de ces objets, comme la suite binaire de Thue-Morse, interviennent à propos de problèmes en apparence sans lien... ce qui épaissit le mystère qui les entoure. Le comble de l'étrange est atteint quand on s'aperçoit – ce qui vient de se produire pour la suite de Thue-Morse – que de manière indirecte des artistes en avaient fait usage dans les figures géométriques qu'ils dessinaient sur le sable bien des siècles avant que les mathématiciens la redécouvrent et en démontrent les propriétés.

Auteur & Bibliographie

Jean-Paul DELAHAYE est professeur d'informatique à l'Univ. de Lille.

G. ALLOUCHE, J.-P. ALLOUCHE et J. SHALLIT, *Kolam indiens, dessins sur le sable aux îles Vanuatu, courbe de Sierpinski et morphismes de monoïde*, 2006. Ann. Inst. Fourier, à paraître, http://www.lri.fr/~allouche/biblioreente.html ref 45

T. HARJU et D. NOWOTKA, *Binary Words With Few Squares,* in *Bulletin of the European Association of Theoretical Computer Science*, 89, pp. 163-166, 2006.

N. RAMPERSAD et J. SHALLIT, *Words Avoiding Reversed Subwords,* in *J. Combin. Math. Combin. Comput.,* 54, 157-164, 2005.

N. RAMPERSAD, J.SHALLIT et Ming.-Wei Wang, *Avoiding Large Squares in Infinite Binary Words,* in *Theoret. Comput. Sci.,* 339, 19-34, 2005.

J.-P. ALLOUCHE et J. SHALLIT, *Automatic Sequences. Theory, Applications, Generalizations.* Cambridge University Press, Cambridge, 2003.

J.-P. ALLOUCHE et J. SHALLIT, *The Ubiquitous Prouhet-Thue-Morse Sequence*, in C. Ding. T. Helleseth, and H. Niederreiter, eds., *Sequences and Their Applications : Proceedings of SETA '98*, Springer-Verlag, pp. 1-16, 1999.

J.-P. DELAHAYE, *Logique, informatique et paradoxes,* Belin/Pour la Science, 1995,

C. RICHARD et U. GRIMM, *On the Entropy and Letter Frequencies of Ternary Square-Free Words*, in *The Electronic Journal of Combinatorics,* 11, 2004. Voir : http://arxiv.org/abs/math.CO/0302302