Langages formels

et suites « automatiques »

Jean-Paul Allouche
CNRS, IMJ-PRG

Introduction

Peut-on engendrer automatiquement toutes les phrases (et seulement elles) qui sont *syntaxiquement* correctes dans une langue naturelle donnée ?

Exemple : Le chien fume intuitivement une chemise rocailleuse.

[cf. Chomsky et Schützenberger 1963]

Par ailleurs, que peut-on dire des suites suivantes

$*$ 0 1 1 0 1 0 0 1 1 0 0 1 ...

$*$ le pliage régulier de papier

$*$ la suite des mouvements pour résoudre le problème de la tour de Hanoï avec une infinité de disques

etc.

Grammaires formelles

Une grammaire —au sens courant— peut être vue comme un ensemble de règles permettant de définir des phrases *admissibles* à partir d'un dictionnaire donné.

Grammaires formelles

Une grammaire —au sens courant— peut être vue comme un ensemble de règles permettant de définir des phrases *admissibles* à partir d'un dictionnaire donné.

Modifions un peu cette phrase en remplaçant *phrases* par *mots* et *dictionnaire* par *alphabet*. Ce qui donne :

Grammaires formelles

Une grammaire (formelle) est un ensemble de règles permettant de définir des mots *admissibles* à partir d'un alphabet donné.

Grammaires formelles

Une grammaire est un ensemble de règles permettant de définir des mots *admissibles* à partir d'un alphabet donné.

Un *alphabet* est un ensemble fini quelconque, un *mot* est une suite finie de symboles appartenant à l'alphabet.

Un *alphabet* est un ensemble fini quelconque, un *mot* une suite finie de symboles appartenant à l'alphabet. Il est représenté comme une concaténation de ces symboles (aussi appelés lettres). Le mot vide est le mot sans aucun symbole noté $\emptyset$.

Enfin on appelle *langage* tout ensemble de mots.

Un *alphabet* est un ensemble fini quelconque, un *mot* une suite finie de symboles appartenant à l'alphabet. Il est représenté comme une concaténation de ces symboles (aussi appelés lettres). Le *mot vide* est le mot sans aucun symbole noté $\emptyset$.

Enfin on appelle *langage* tout ensemble de mots.

Exemples

Alphabet : $\{a, b, c, ..., z\}$; mots $bonjour$, $aabbbbujv$...

Alphabet : $\{0, 1, ..., 9\}$; mots 32188, 17, 4...

Alphabet : $\{0, 3, a, Z, -\}$; mots $3Z - aaaa0$, $ZZ0a3$...

Définition : Une grammaire (formelle) est la donnée

$*$ d'un alphabet de symboles dits *terminaux*,

$*$ d'un alphabet de symboles dits *non-terminaux*, dont l'un des éléments est appelé *axiome* et noté S

$*$ d'un ensemble de *règles de production*, qui associent à un symbole non-terminal une suite finie de symboles terminaux et non-terminaux.

Définition : Une grammaire (formelle) est la donnée

$*$ d'un alphabet de symboles dits *terminaux*,

$*$ d'un alphabet de symboles dits *non-terminaux*, dont l'un des éléments est appelé *axiome* et noté S

$*$ d'un ensemble de *règles de production*, qui associent à un symbole non-terminal une suite de symboles terminaux et non-terminaux.

Par exemple, si les symboles non-terminaux sont $\{S, A, ...\}$ et les symboles terminaux $\{a, b, ...\}$, une règle de production est $A \to ABa$.

Fonctionnement et exemple

Les mots engendrés par une grammaire formelle sont ceux qui ne contiennent que des symboles terminaux et qui sont obtenus en partant de S (l'axiome), et en appliquant un nombre quelconque de règles de production. Leur ensemble est appelé *langage engendré par cette grammaire*.

Fonctionnement et exemple

Les mots engendrés par une grammaire formelle sont ceux qui ne contiennent que des symboles terminaux et qui sont obtenus en partant de S (l'axiome), et en appliquant un nombre quelconque de règles de production. Leur ensemble est appelé *langage engendré par cette grammaire*.

Prenons comme ensemble de symboles non-terminaux le singleton $\{S\}$ (l'axiome), pour symboles terminaux $\{a, b\}$, et comme règles de production les deux règles de production suivantes : $S \to aSb$, $S \to \emptyset$.

Que peut-on obtenir ?

Prenons comme ensemble de symboles non-terminaux le singleton $\{S\}$ (l'axiome), pour symboles terminaux $\{a, b\}$, et comme règles de production les deux règles de production suivantes : $S \to aSb$, $S \to \emptyset$.

Que peut-on obtenir ?

$S \to aSb \to a\emptyset b = ab$

$S \to aSb \to aaSbb \to aa\emptyset bb = aabb$

...

autrement dit exactement les mots $a^n b^n$, pour $n \geq 0$.

En imposant aux grammaire des propriétés supplémentaires, on définit différentes classes de langages (*hiérarchie de Chomsky*).

Par exemple les *langages rationnels* sont ceux qui peuvent être engendrés par une grammaire formelle *linéaire à droite*, c'est-à-dire dont les règles sont toutes de la forme

$X \to aY$, $X \to a$ ou $X \to \emptyset$

où X, Y sont des symboles non-terminaux et a un symbole terminal.

Exemple de langage rationnel

Considérons la grammaire linéaire à droite définie par

$$
\begin{aligned}
S &\to 0S \\
S &\to 1B \\
B &\to 0B \\
B &\to 1S \\
S &\to \emptyset
\end{aligned}
$$

exemple : $S \to 0S \to 01B \to 010B \to 0100B \to 01001S \to 01001$

Exemple de langage rationnel

Considérons la grammaire linéaire à droite définie par

$$
\begin{aligned}
&S \rightarrow 0S\\
&S \rightarrow 1B\\
&B \rightarrow 0B\\
&B \rightarrow 1S\\
&S \rightarrow \emptyset
\end{aligned}
$$

exemple : $S \rightarrow 0S \rightarrow 01B \rightarrow 010B \rightarrow 0100B \rightarrow 01001S \rightarrow 01001$

Les lecteurs constateront (démontreront) que les mots engendrés par cette grammaire correspondent exactement aux représentations binaires des entiers qui ont un nombre pair de 1 dans cette représentation (on admet des zéros de tête).

Une suite $(a_n)_{n \geq 0}$ à valeurs dans un ensemble fini F est dite k-automatique si pour tout y dans F l'ensemble des représentations en base k des entiers n pour lesquels $a_n = y$ est un langage rationnel sur l'alphabet $\{0, 1, ..., k-1\}$

Une suite $(a_n)_{n\geq 0}$ à valeurs dans un ensemble fini F est dite k-automatique si pour tout y dans F l'ensemble des représentations en base k des entiers n pour lesquels $a_n = y$ est un langage rationnel sur l'alphabet $\{0, 1, ..., k-1\}$

Pourquoi k-automatique ?

Une suite $(a_n)_{n\geq 0}$ à valeurs dans un ensemble fini F est dite k-automatique si pour tout y dans F l'ensemble des représentations en base k des entiers n pour lesquels $a_n = y$ est un langage rationnel sur l'alphabet $\{0, 1, ..., k-1\}$

Pourquoi k-automatique ?

Y a-t-il une caractérisation « plus parlante ».

Prenons $k = 2$. Pour toute suite $(v_n)_{n\geq 0}$ on peut jouer au jeu suivant : on découpe la suite en deux sous-suites, celle correspondant aux indices pairs et celle correspondant aux indices impairs.

$$
\begin{array}{ccc}
 & (v_n)_{n\geq 0} & \\
\swarrow & & \searrow \\
(v_{2n})_{n\geq 0} & & (v_{2n+1})_{n\geq 0}
\end{array}
$$

Prenons $k = 2$. Pour toute suite $(v_n)_{n\geq 0}$ on peut jouer au jeu suivant : on découpe la suite en deux sous-suites, celle correspondant aux indices pairs et celle correspondant aux indices impairs.

$$
\begin{array}{ccccccc}
 & & & (v_n)_{n\geq 0} & & & \\
 & & \swarrow & & \searrow & & \\
 & (v_{2n})_{n\geq 0} & & & & (v_{2n+1})_{n\geq 0} & \\
\swarrow & & \searrow & & \swarrow & & \searrow \\
(v_{4n})_{n\geq 0} & & (v_{4n+2})_{n\geq 0} & & (v_{4n+1})_{n\geq 0} & & (v_{4n+3})_{n\geq 0}
\end{array}
$$

Prenons $k = 2$. Pour toute suite $(v_n)_{n\geq 0}$ on peut jouer au jeu suivant : on découpe la suite en deux sous-suites, celle correspondant aux indices pairs et celle correspondant aux indices impairs.

$$(v_n)_{n\geq 0}$$

$$\swarrow \qquad\qquad \searrow$$

$$(v_{2n})_{n\geq 0} \qquad\qquad (v_{2n+1})_{n\geq 0}$$

$$\swarrow \quad \searrow \qquad\qquad \swarrow \quad \searrow$$

$$(v_{4n})_{n\geq 0} \quad (v_{4n+2})_{n\geq 0} \quad (v_{4n+1})_{n\geq 0} \quad (v_{4n+3})_{n\geq 0}$$

$$\swarrow\searrow \qquad \swarrow\searrow \qquad \swarrow\searrow \qquad \swarrow\searrow$$

$$(v_{8n})_{n\geq 0} \ (v_{8n+4})_{n\geq 0} \ (v_{8n+2})_{n\geq 0} \ (v_{8n+6})_{n\geq 0} \ (v_{8n+1})_{n\geq 0} \ (v_{8n+5})_{n\geq 0} \ (v_{8n+3})_{n\geq 0} \ (v_{8n+7})_{n\geq 0}$$

L'ensemble de sous-suites ainsi obtenu est appelé le 2-noyau de la suite $(v_n)_{n\geq 0}$.

$$K_2((v_n)_{n\geq 0}) = \{(v_{2^\alpha n+\beta},\ \alpha \geq 0,\ \beta \in [0, 2^\alpha - 1]\}$$

L'ensemble de sous-suites ainsi obtenu est appelé le 2-noyau de la suite $(v_n)_{n\geq 0}$.

$$K_2((v_n)_{n\geq 0}) = \{(v_{2^\alpha n+\beta},\ \alpha \geq 0,\ \beta \in [0, 2^\alpha - 1]\}$$

Théorème : *une suite est 2-automatique si et seulement si son 2-noyau est un ensemble fini.*

L'ensemble de sous-suites ainsi obtenu est appelé le 2-noyau de la suite $(v_n)_{n\geq 0}$.

$$K_2((v_n)_{n\geq 0}) = \{(v_{2^\alpha n+\beta},\ \alpha \geq 0,\ \beta \in [0, 2^\alpha - 1]\}$$

Théorème : *une suite est 2-automatique si et seulement si son 2-noyau est un ensemble fini.*

Idem avec k-automatique et k-noyau.

Exemple : la suite de (Prouhet-)Thue-Morse

0 1 1 0 1 0 0 1 1 0 0 1 0 1 1 0 ...

Exemple : la suite de (Prouhet-)Thue-Morse

$$0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ \ldots$$

Le n-ième terme de cette suite est la parité du nombre de 1 dans la représentation binaire de l'entier n.

Exemple : la suite de (Prouhet-)Thue-Morse

$$0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ \ldots$$

Le n-ième terme de cette suite est la parité du nombre de 1 dans la représentation binaire de l'entier n.

La suite de Prouhet-Thue-Morse est 2-automatique.

Exemple : la suite de (Prouhet-)Thue-Morse

0 1 1 0 1 0 0 1 1 0 0 1 0 1 1 0 ...

Le n-ième terme de cette suite est la parité du nombre de 1 dans la représentation binaire de l'entier n.

La suite de Prouhet-Thue-Morse est 2-automatique.

La suite de pliage régulier de papier, la suite de Hanoï sont des suites 2-automatiques

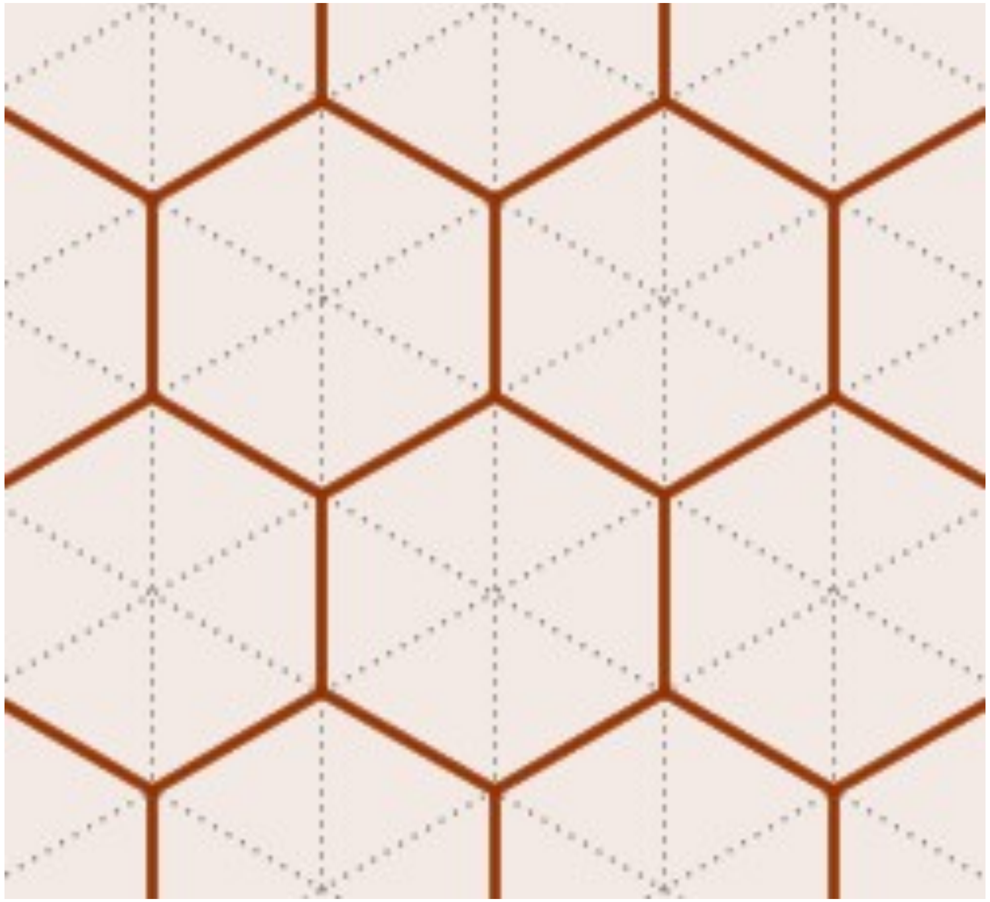
Propriétés ?

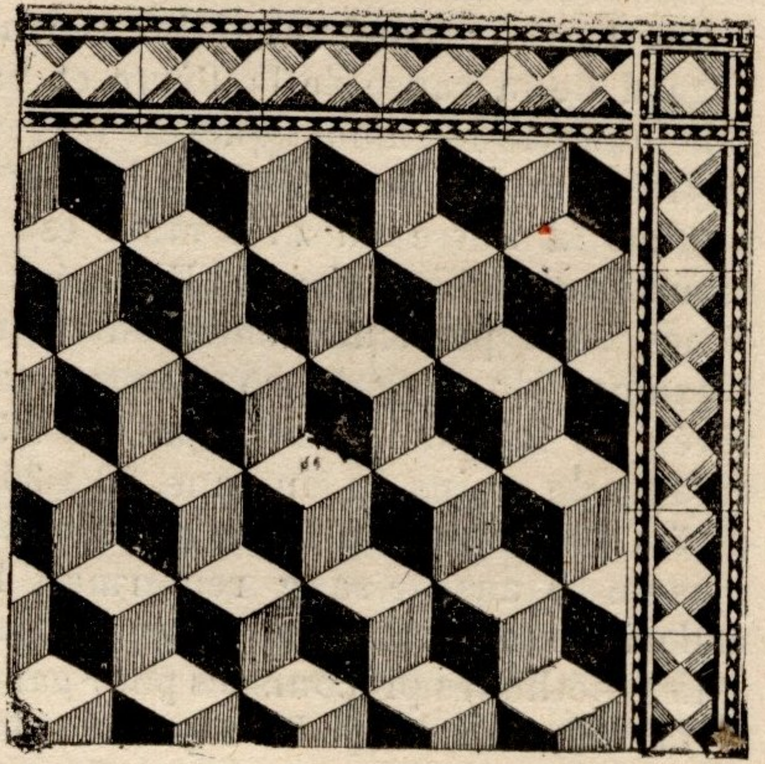
Ubiquité des suites automatiques et tout particulièrement de la suite de Thue-Morse, théorème de Christol, etc...

Applications?

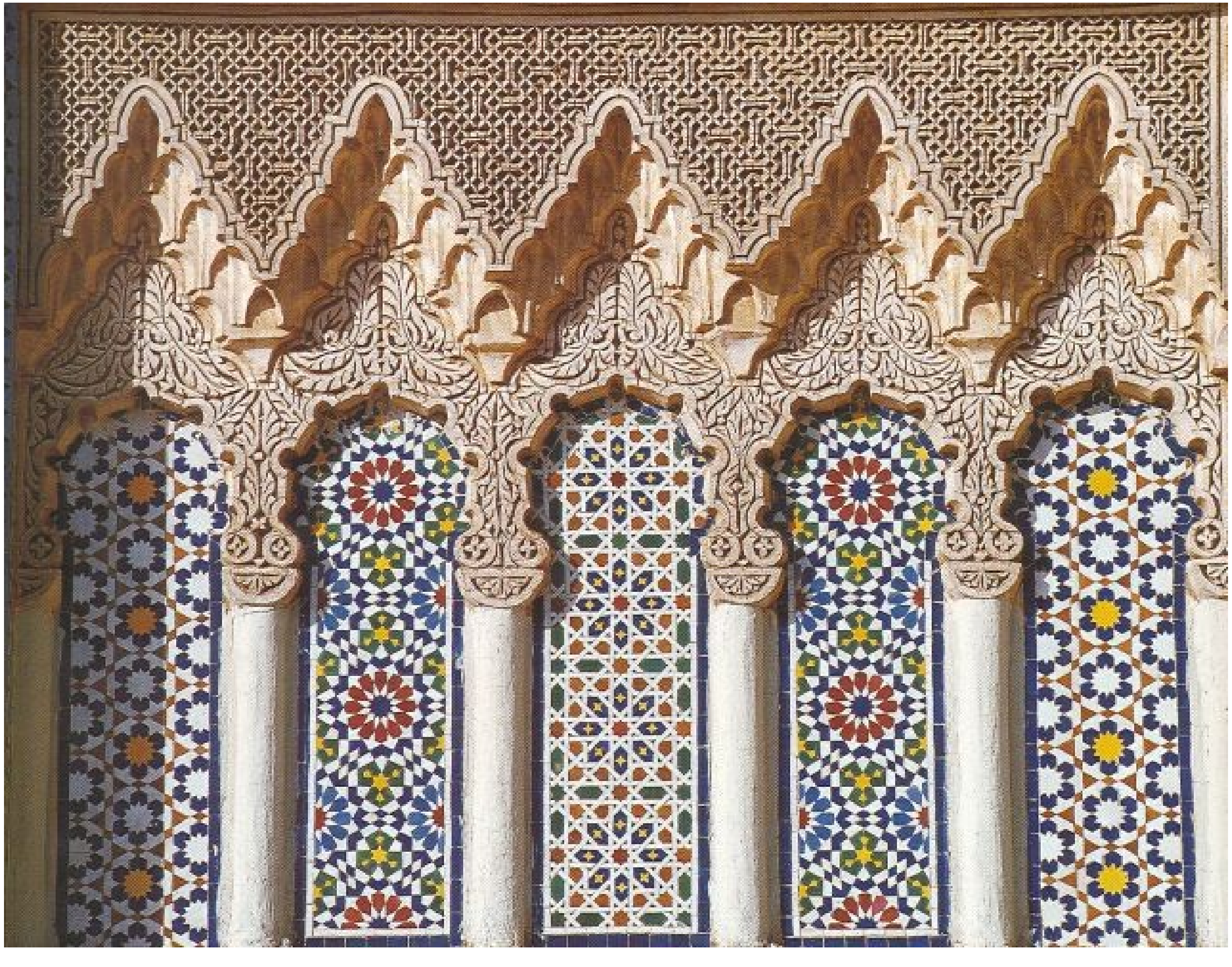
Pour les suites automatiques ou des généralisations : combinatoire des mots, théorie des nombres (en particulier transcendance), analyse harmonique, physique, chimie (cristallographie), biologie (ADN etc.), ethnomathématique, voire musique.

Pavages et quasi-cristaux

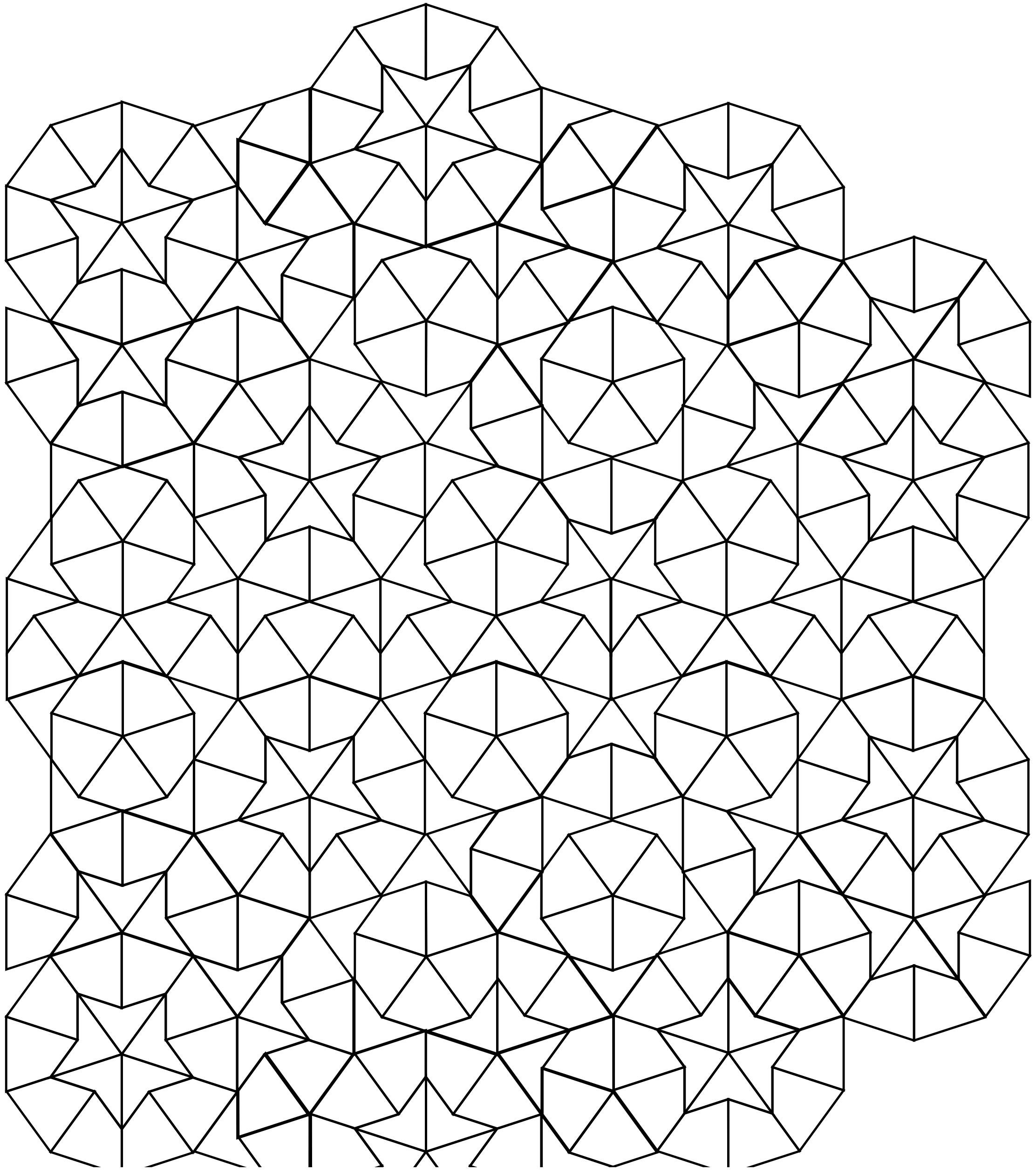




Fragment d'un pavement céramique moderne.



Roger Penrose, The Role of Aesthetics in Pure and Applied Mathematical Research, Bull. Inst. Math. Appl. 10, 266 (1974)



D. Shechtman, I. Blech, D. Gratias, J. W. Cahn, *Metallic Phase with Long-Range Orientational Order and No Translational Symmetry*, Phys. Rev. Lett., vol. 53, no 20, 1984, 1951–1953

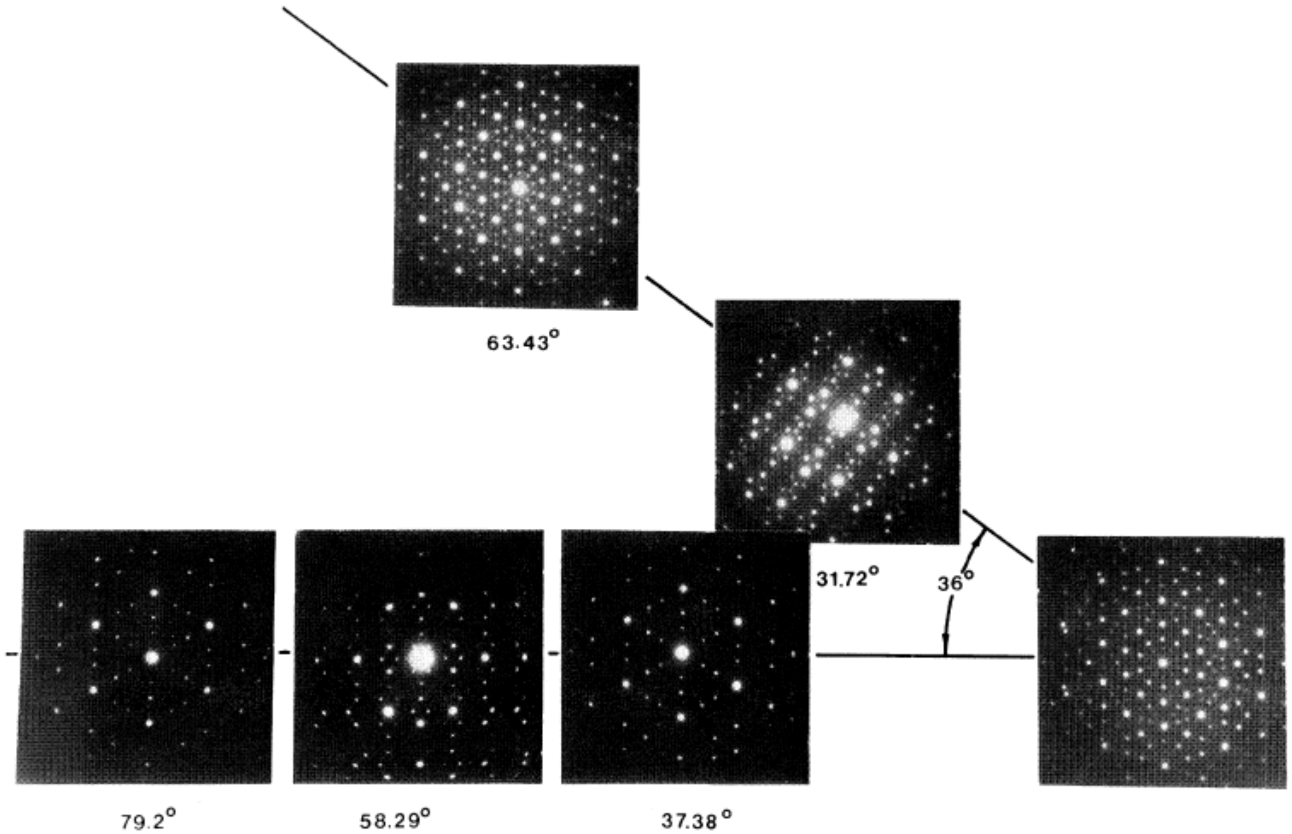


FIG. 2. Selected-area electron diffraction patterns taken from a single grain of the icosahedral phase. Rotations match those in Fig. 1.

En 2011 D. Shechtman obtient le Prix Nobel de chimie

Ethnomathématique

Kolam