Hartley définit la quantité d'information contenue dans une suite de N symboles élémentaires comme étant égale à $N \log S$, tout en soulignant que ces considérations sont indépendantes de la nature ou de la signification de l'information véhiculée.

C'est dans le contexte de la Seconde Guerre mondiale que les idées de Nyquist, Hartley et d'autres ingénieurs ou chercheurs convergent vers une « théorie de l'information », essentiellement grâce aux travaux de Shannon, également des Laboratoires *Bell*.

Ce mathématicien-ingénieur quelque peu excentrique rédige en 1945 *Une théorie mathématique de la cryptographie*, rapport classé secret jusqu'en 1957. Dans ce document, il a l'intuition de comparer les perturbations du signal engendrées par la clef de cryptage au « bruit » qui parasite les lignes de télécommunication et que les ingénieurs essayent de minimiser. Shannon y définit en outre la quantité d'information, dans le cas discret, comme la moyenne logarithmique des probabilités de sélection des éléments constitutifs du message *(voir l'encadré ci-dessous)*.

Parallèlement à ses recherches cryptographiques, Shannon a travaillé à l'automatisation de la défense antiaérienne. De la Première Guerre mondiale à la Seconde, les avions volaient deux fois plus haut et deux fois plus vite. Les opérateurs des radars n'avaient alors pas le temps de communiquer efficacement avec les équipes d'artillerie. Pour transmettre les données automatiquement, il fallait là encore, pour l'équipe de Shannon, trouver un codage adéquat et définir l'information à transmettre.

Une autre équipe des Laboratoires *Bell*, dirigée par Norbert Wiener, se préoccupait des perturbations dues aux dents des engrenages, qui affectaient la stabilité du bras déterminant l'axe de tir. Assisté de l'ingénieur Julian Bigelow, Wiener eut l'idée de consulter son ami cardiologue, Arturo Rosenblueth, qui avait déjà analysé des mouvements perturbés caractéristiques de certaines formes d'ataxie (difficulté de coordination) dans le tabès, une forme tardive de la syphilis.

De l'information dans les boucles de rétroaction

Ces échanges transdisciplinaires, liés à l'effort de guerre des États-Unis, ont conduit à l'élaboration d'une théorie générale de la régulation, que Wiener nomma cybernétique dès 1947. Cette théorie portant sur des systèmes quelconques, que ce soit des machines ou des organismes vivants, était

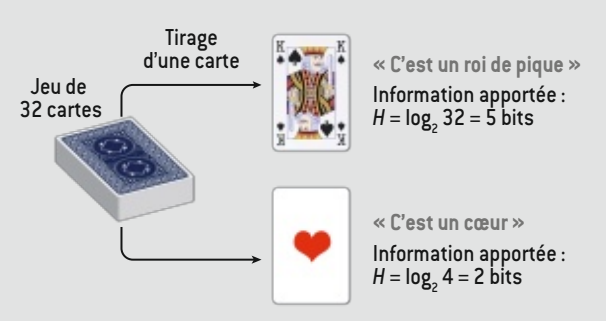
QUANTITÉ D'INFORMATION, INCERTITUDE, ENTROPIE

Pour définir la quantité d'information contenue dans un message d'une certaine longueur et formé à partir d'un certain alphabet de lettres ou symboles, Claude Shannon a cherché à quantifier l'incertitude ou le choix – ou encore le manque d'information – associé à l'ensemble des messages possibles ayant la même longueur et utilisant le même alphabet.

Supposons que l'on ait N messages possibles et que la probabilité de choisir le i-ème message soit p_i, avec $p_1 + p_2 + \ldots + p_N = 1$. Shannon a cherché une fonction $H(p_1, \ldots, p_N)$ qui quantifie l'incertitude associée à ces possibilités. En exigeant qu'elle satisfasse à certaines propriétés naturelles, il a prouvé que la fonction H, nommée entropie informationnelle ou entropie statistique, devait être de la forme : $H = -K(p_1 \log p_1 + p_2 \log p_2 + \ldots + p_N \log p_N)$, où K est une constante positive arbitraire. Quand un seul choix est possible (tous les p_i sont nuls, sauf un), l'incertitude est nulle ; l'information apportée en spécifiant ce choix est donc elle aussi nulle. Dans le cas où les probabilités sont toutes égales (à p), on a $H = K \log (1/p)$; cela correspond à une entropie ou une incertitude maximale, et l'information apportée en spécifiant le choix effectué est maximale.

Un exemple très simple est celui d'un message constitué d'un seul symbole binaire, généralement noté 0 ou 1. Il y a deux choix possibles ; s'ils ont même probabilité ($p_1 = p_2 = 1/2$), l'entropie de Shannon devient $H = K \log 2$. Par commodité, on prend $K = 1$ et des logarithmes de base 2 ; l'entropie de Shannon d'un symbole binaire devient alors $H = 1$, ce qui définit un bit d'information (le terme « bit », contraction de *binary digit*, a été proposé par le statisticien américain John Tukey). De la même façon, l'affirmation « Mon rendez-vous est le matin » apporte une information de un bit, alternative élémentaire, s'il est équiprobable que ce rendez-vous ait lieu le matin ou l'après-midi.

Autre exemple : l'annonce de la carte que l'on a choisie dans un jeu de 32 cartes. Les choix possibles ayant la même probabilité, 1/32, l'entropie de Shannon est égale à $\log_2 32$, c'est-à-dire cinq bits (puisque $32 = 2^5$). La quantité d'information apportée par cette annonce correspond au nombre de questions à réponse binaire (« oui » ou « non ») qu'il faut poser pour trouver la carte : « Est-elle de couleur rouge ? », « Est-elle égale ou supérieure au valet ? », etc.



Shannon a aussi montré qu'on pouvait calculer l'« entropie » d'une langue à partir d'un calcul sur sa redondance. Bien des phrases restent compréhensibles lorsque des lettres sont enlevées ; c'est le principe des techniques de sténographie et à titre d'exemple, sans vouloir flatter les lecteurs, Shannon écrivait : « Is lctrs ntllgnts cmprnnnt fclmnt n txt sns vlls. »

En même temps, Shannon a toujours tenu à préciser que la dimension sémantique de l'information était absente de ses travaux. Que ce soit à travers la définition statistique ou par l'utilisation de la redondance, l'ingénieur en télécommunications aura la même application à transmettre une déclaration de guerre, des cours de la Bourse ou un message dépourvu de sens pour lui.

La formule de l'entropie au sens de Shannon, $H = -K \sum p_i \log p_i$, ressemble beaucoup à la formule qui définit l'entropie en mécanique statistique, telle qu'elle a été établie par les physiciens Ludwig Boltzmann et Josiah Gibbs au XIX^e siècle. Il suffit d'adopter les logarithmes népériens et de prendre pour K la constante de Boltzmann ($k_B = 1{,}38 \times 10^{-23}$ joule/kelvin) pour obtenir une expression formellement identique, p_i désignant alors la probabilité du i-ème état microscopique accessible au système physique considéré. Toutefois, l'identification entre l'entropie/information de Shannon et l'entropie de Boltzmann-Gibbs a été contestée par de nombreux scientifiques ; le physicien et bio-informaticien américain Hubert Yockey, notamment, a souligné que les espaces de probabilités dans les deux situations ne sont pas isomorphes (leur correspondance n'est pas rigoureuse).