Claude Shannon et Warren Weaver

Théorie mathématique de la communication

Paris, CEPI, 1975
pages : 31 à 50

1 – NOTE INTRODUCTIVE SUR LA SITUATION GENERALE DES ETUDES ANALYSANT LA COMMUNICATION

1.1 - Communication

Le mot communication sera utilisé ici dans un sens très large incluant tous les procédés par lesquels un esprit peut en influencer un autre. Cela, bien sûr, comprend non seulement le langage écrit ou parlé, mais aussi la musique, les arts plastiques, le théâtre, la danse et, en fait, tout comportement humain. Dans quelques domaines proches, il serait souhaitable d'utiliser une définition encore plus large, de façon à inclure les procédés par lesquels un mécanisme (tel un équipement automatique pour dépister un aéroplane et « prévoir » ses positions futures probables) affecte un autre mécanisme (tel un missile téléguidé poursuivant cet aéroplane). Le langage de ce mémoire paraîtra souvent se référer au champ particulier, encore que très vaste et important, de la communication verbale ; mais, en pratique, tout ce qui sera dit est également valable pour la musique et pour toute image fixe ou mobile, comme la télévision.

1.2 - Les trois niveaux des problèmes de communication

Il semble y avoir trois niveaux relatifs au vaste sujet de la communication. Ainsi, il paraît raisonnable de se demander successivement :
Niveau A: avec queue exactitude les symboles de la communication peuvent-ils être transmis? (Problème technique)
Niveau B: avec quelle précision les symboles transmis véhiculent-ils la signification désirée? (Problème sémantique)
Niveau C: avec queue efficacité la signification reçue influence-t-elle la conduite dans le sens désiré? (Problème de l'efficacité)

Les problèmes techniques concernent l'exactitude du transfert depuis l'émetteur jusqu'au récepteur des séries de symboles (langage écrit) ou d'un signal variant de façon continue (transmission de la voix ou de la musique par téléphone ou radio), ou d'un « pattern » bi-dimensionnel continu et variable (télévision), etc. Mathématiquement, le premier problème comprend la transmission d'une série finie de symboles discrets, le second la transmission d'une fonction continue du temps et le troisième la transmission de plusieurs fonctions continues du temps ou d'une fonction continue du temps et de deux coordonnées spatiales.

Les problèmes sémantiques concernent l'identité, ou une approximation suffisamment proche, entre l'interprétation du récepteur et l'intention de l'émetteur. C'est une situation très profonde et complexe, même si l'on traite seulement les problèmes relativement les plus simples de la communication par la parole. Une complication essentielle est illustrée par la remarque que, si M. Y soupçonne M. X de ne pas le comprendre, il n'a d'autre ressource que de parler davantage à M. X pour éclaircir la situation. Si M. Y dit : « Me comprenez-vous ? » et si M. X répond

« Certainement », cela ne constitue pas une certitude que la compréhension est reçue. Cette réponse peut signifier simplement que M. X n'a pas compris la question. Si cela semble niais, essayons-le sous cette forme « Czy pan mnie rozumie ? », avec la réponse : « Hai wakkate imasu. » Je pense que cette difficulté fondamentale[1], du moins dans le domaine restreint de la communication verbale, est réduite (bien que jamais complètement éliminée) par des « explications » qui (a) ne sont vraisemblablement guère plus que des approximations des idées exprimées, mais qui (b) sont compréhensibles parce que formulées en un langage rationnellement clarifié au préalable par des procédés opératoires. Par exemple il faut peu de temps pour traduire le symbole « oui » en quelque langage opérationnellement compréhensible.
Le problème sémantique a de vastes ramifications si l'on pense à la communication en général. Citons l'exemple de la signification des actualités cinématographiques américaines pour un citoyen russe.

Les problèmes d'efficacité concernent le succès avec lequel la signification convoyée jusqu'au receveur provoque chez lui la conduite désirée. II peut sembler à première vue trop limité de proposer comme but de toute communication d'influencer la conduite du receveur. Mais, avec une définition assez large de la conduite, il paraît clair que ou bien la communication affecte la conduite, ou elle ne produit pas d'effet décelable et probablement pas d'effet du tout.

Le problème de l'efficacité implique des considérations esthétiques dans le cas des beaux-arts. Dans le cas du langage, écrit ou parlé, il implique des considérations qui vont de la simple stylistique, en passant par tous les aspects psychologiques et émotionnels de la théorie de la propagande, jusqu'aux jugements de valeur qui sont nécessaires pour attribuer une signification nouvelle aux mots « succès » et « désirée » dans la phrase initiale de ce paragraphe sur l'efficacité.

Le problème de l'efficacité est étroitement lié au problème sémantique et le recouvre d'une façon plutôt vague mais il y a en fait un recouvrement de toutes les catégories de problèmes proposées.

1.3 – Commentaires

Ainsi présenté, on serait enclin à penser que le niveau A est relativement superficiel, concernant seulement les détails techniques d'une bonne conception d'un système de communication, tandis que B et C semblent correspondre à presque tout, sinon tout, le contenu philosophique du problème de la communication. La théorie mathématique des aspects techniques de la communication, telle qu'elle est développée spécialement par Claude Shannon aux Bell Telephone Laboratories, s'applique de façon évidente en premier lieu au problème A, c'est-à-dire au problème technique de la fidélité de la transmission de types varies de signaux d'un émetteur à un receveur.

---

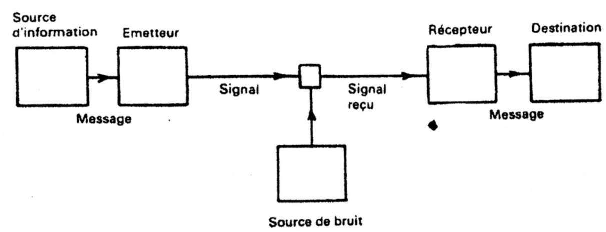
[1] « Lorsque Pfungst (1911) démontra que les chevaux d'Elberfeld, qui faisaient preuve d'extraordinaires capacités linguistiques et mathématiques, réagissaient en fait aux mouvements de tête de l'entraîneur, M. Krall (1911), leur propriétaire, répondit à cette critique de la façon la plus directe qui soit. Il demanda aux chevaux s'ils pouvaient percevoir des mouvements si petits: ils signifièrent leur réponse en émettant un non retentissant. Malheureusement, nous n'avons aucune assurance de nous faire comprendre ou d'obtenir des réponses aussi claires. » Voir K. S. Lashley : « Persistent Problems in the Evolution of Mind », dans Quarterly Review of Biology, vol. 24, mars 1949, p 28.

Mais la théorie a, je pense, une signification profonde, qui prouve que le paragraphe précédent est sérieusement imparfait. Une partie de la signification de la nouvelle théorie vient du fait que les niveaux B et C ne peuvent utiliser que des signaux dont la précision va dépendre des analyses du niveau A. Ainsi, toutes limitations découvertes dans la théorie au niveau A vont nécessairement s'appliquer aux niveaux B et C. Mais une plus large part de la signification vient du fait que l'analyse au niveau A révèle que ce niveau recouvre les autres plus qu'on ne pouvait naïvement le soupçonner. Ainsi, la théorie du niveau A est, au moins jusqu'à un certain point, aussi une théorie des niveaux B et C. J'espère que les parties suivantes de ce mémoire éclaireront et justifieront ces dernières remarques.

2 - Problèmes de communication au niveau A

2.1 - Un système de communication et ses problèmes

Le système de communication considéré peut être symbolisé de la façon suivante :



La source d'information choisit le message désiré parmi une série de messages possibles (cela est une remarque particulièrement importante qui nécessitera plus tard une explication détaillée). Le message choisi peut consister en mots écrits ou parlés, ou en images, musique, etc.

L'***émetteur*** transforme ce message en ***signal*** qui est alors envoyé par le ***canal de communication*** de l'émetteur au récepteur. Dans le cas du téléphone, le canal est un fil métallique, le signal un courant électrique variable parcourant ce fil ; l'émetteur est un ensemble d'éléments (émetteur téléphonique, etc.) qui transforme la pression du son vocal en un courant électrique variable. Dans le cas de la télégraphie, l'émetteur code des mots écrits en séquences de courant interrompues de longueurs variables (points, traits, blancs). Pour le langage parlé, la source d'information est le cerveau, l'émetteur est l'organe vocal qui produit la pression variable sonore (le signal) transmise à travers l'air (le canal). Pour la radio, le canal est simplement l'espace (ou l'éther, si quelqu'un préfère encore ce mot ancien et fallacieux) et le signal est l'onde électromagnétique qui est transmise.

Le ***récepteur*** est en quelque sorte un émetteur inverse, changeant le signal reçu en message et amenant ce message à destination. Quand je vous parle, mon cerveau est la source d'information, et le vôtre la destination; mon système vocal est l'émetteur, et votre oreille avec la huitième paire de nerfs crâniens est le récepteur.

Dans le processus de transmission, il est malheureusement caractéristique que certains éléments qui n'étaient pas voulus par la source s'ajoutent au signal. Ces additions indésirables peuvent être des distorsions du son (au téléphone, par exemple) ou électrostatiques (pour la radio), ou des distorsions de la forme ou du contraste de l'image (télévision), ou des erreurs de transmission (télégraphie, reproduction), etc. Toutes ces altérations du signal transmis sont appelées ***bruit***.

Les questions que l'on se pose au sujet d'un tel système de communication sont :

a) Comment mesurer la quantité d'information?

b) Comment mesurer la capacité du canal de communication?

c) L'action de l'émetteur qui transforme le message en signal implique souvent un processus de codage. Quelles sont les caractéristiques d'un processus de codage efficace? Et, quand le codage est aussi bon que possible, en quelle quantité l'information peut-elle être convoyée par le canal?

d) Quelles sont les caractéristiques générales du bruit? Comment le bruit affecte-t-il l'exactitude du message finalement reçu à destination? Comment peut-on diminuer les effets indésirables du bruit, et jusqu'à quel point peuvent-ils être éliminés?

e) Si le signal transmis est continu (comme pour le langage parlé ou la musique) plutôt que formé de symboles discrets (comme pour le langage écrit, la télégraphie, etc.), comment cela modifie-t-il le problème ?

Nous allons maintenant indiquer, sans preuves et avec un minimum de terminologie mathématique, les principaux résultats obtenus par Shannon.

2.2 - L'information

Le mot information, dans cette théorie, est utilisé dans ***un sens spécial***, à ne pas confondre avec son usage curant. En particulier, ***information*** ne doit pas être confondu avec ***signification***. En fait, deux messages, l'un qui est chargé de sens et l'autre qui ne signifie rien, peuvent être exactement équivalents, du présent point de vue, en ce qui concerne l'information. C'est, sans aucun doute, ce que Shannon entend lorsqu'il dit que « les aspects sémantiques de la communication ne relèvent pas de ses aspects technologiques ». Mais cela ne veut pas dire que les aspects technologiques n'interviennent pas dans les aspects sémantiques. En vérité, le mot information, dans la théorie de la communication, n'est pas tellement lié ce que l'on dit, mais plutôt à ce que l'on pourrait dire.

Cela exprime que l'information est une mesure de la liberté de choix dont on dispose lorsque l'on sélectionne un message. Si l'on a affaire à une situation très élémentaire où le choix se fait entre les deux termes d'une alternative, on dit dans ce cas, par convention, que l'information est égale à l'unité. Notons qu'il est erroné (bien que souvent pratique) d'affirmer que l'un ou l'autre message correspond à une unité d'information. Le concept d'information ne s'applique pas aux messages individuels (comme le ferait le concept de signification), mais plutôt à la situation globale : l'unité d'information indique que l'on dispose, dans cette situation, d'une certaine liberté dans la sélection du message, qu'il convient de considérer comme un degré standard ou unitaire.

Les deux messages entre lesquels on doit choisir, dans une telle sélection, peuvent être de toute nature. L'un peut être le texte de la version biblique de King James, et l'autre peut être « Yes ». L'émetteur a la possibilité de coder ces deux messages de telle sorte que « zero » soit le signal du premier, et « un » le signal du second ; ou encore que la fermeture d'un circuit (passage du courant) soit le signal du premier, et l'ouverture (pas de courant) le signal du second. Ainsi, les deux positions, fermé/ouvert, d'un simple relais peuvent correspondre aux deux messages.

Pour être un peu plus précis, la quantité d'information est définie, dans les cas les plus simples, comme mesurée par le logarithme du nombre des choix possibles. Il est pratique d'utiliser les logarithmes de base 2, plutôt que les logarithmes usuels[2] de base 10; l'information, quand il y a seulement deux choix possibles, est proportionnelle au logarithme de 2 (de base 2). Mais cela correspond à l'unité; ainsi une situation de choix binaire est caractérisée par une unité d'information, comme précédemment établi. Cette unité d'information, est appelée un « bit » ; ce mot, propose par John W. Tukey, est une condensation de « binary digit ». Lorsque les nombres sont exprimés dans le système binaire, il y a seulement deux digits : 0 et 1; tout comme les dix digits 0 à 9 compris, sont utilisés dans le système décimal où 10 est une base. Zéro et un peuvent constituer les symboles de deux choix, ainsi qu'il a été dit plus haut. De sorte que le « digit binaire », ou « bit », est naturellement associé à une situation de double choix constituant une unité d'information.

Soit, par exemple, 16 messages alternatifs pour lesquels un choix équivalent est possible ; étant donné que $16 = 2^4$, $\log_2 16 = 4$, on dit qu'une telle situation se caractérise par une information de 4 bits.
II peut sembler étrange, à première vue, que l'information soit définie par le logarithme du nombre de choix. Mais, dans le développement de la théorie, il devient de plus en plus évident que les mesures logarithmiques conviennent le mieux. Pour l'instant, une seule indication de ce fait sera avancée. Il a été dit précédemment qu'un simple relais « ouvert/fermé » pouvait offrir une situation comportant une unité d'information ; combien d'unités seront apportées par, disons, trois relais? Il paraît logique de pouvoir affirmer que trois relais fourniront trois fois plus d'information. Et, effectivement, il en est ainsi lorsqu'on utilise la définition logarithmique de l'information. Trois relais peuvent, en effet, correspondre à 2^3, c'est-à-dire 8 choix, que l'on peut écrire symboliquement 000, 001, 011, 010, 100, 110, 101, 111, où les trois relais sont ouverts dans le premier cas et fermés dans le dernier. Le logarithme (de base 2) de 2 est 3, de sorte que la mesure logarithmique propose trois unités d'information pour une telle situation, comme il est souhaitable. De même, en doublant le temps nécessaire, on élève au carré le nombre de messages possibles et l'on double le logarithme; l'information logarithmique est ainsi doublée.

Ces remarques nous éloignent des situations artificiellement simplifiées dans lesquelles la source d'information ne peut choisir que certains messages précis - tel un individu qui retirerait, parmi un lot, un télégramme standard de félicitations lots d'un anniversaire. Une situation plus naturelle et importante est celle où la source d'information choisit une séquence de symboles élémentaires qui vont constituer alors le message. Ainsi un homme retient un mot après un autre, et ces mots choisis de façon individuelle s'additionnent pour former le message.

A ce stade, une considération majeure s'impose, qui fut laissée à l'arrière-plan, à savoir le rôle joué par la probabilité dans la genèse du message. Car le choix des symboles successifs, du point de vue du système de communication, est sous la dépendance des probabilités ; et, en fait, des probabilités qui, à chaque étape du processus, dépendent des choix précédents. Ainsi, dans la langue anglaise, si le dernier symbole choisi est « the[3] », la probabilité pour que le mot suivant soit un article ou un verbe est très faible. Cette influence probabiliste s'étend, en fait, à plus de deux

---

[2] Quand $m^x = y$, on dit alors que x est le logarithme de y de base m.
[3] ***le*** en français.

mots. Après ces trois mots : « dans le cas », la probabilité pour que le suivant soit « où » est très grande, alors que pour le mot « éléphant » elle est pratiquement nulle.

En anglais, par exemple, il est évident que les probabilités exercent un certain contrôle sur le langage si l'on pense par exemple que le dictionnaire ne contient aucun mot avec la lettre initiale ***j*** suivie de b, c, d, f, g, k, l, q, r, t, v, w, x ou z ; de sorte que la probabilité est nulle pour que l'initiale ***j*** soit suivie d'une de ces lettres. On conviendra de même qu'il y a une très faible probabilité pour une séquence de mots telle « Constantinople fishing nasty pink[4] ». Cette probabilité est très faible mais cependant pas nulle, car il est possible de concevoir tout un passage avec une phrase se terminant par « Constantinople pêchant » et une suivante débutant par « Horrible rosé ». Cet exemple nous permet de remarquer au passage que cette suite invraisemblable de quatre mots s'est bel et bien produite dans une seule phrase de bon anglais.

Un système produisant une séquence de symboles (que ce soit des lettres, des notes de musique ou des mots) selon certaines probabilités constitue un processus stochastique et, dans le cas particulier d'un processus stochastique dont les probabilités dépendent d'événements précédents, on a affaire à un *processus de Markoff* ou *chaîne de Markoff* . Parmi les processus markoviens produisant des messages, il en existe un groupe particulièrement important dans la théorie de la communication, appelé processus ergodiques. Les détails d'analyse en sont si complexes et le raisonnement qu'ils exigent si profond qu'ils ont nécessité les efforts des meilleurs mathématiciens pour en tirer une théorie ; cependant, la nature brute d'un processus ergodique est facile à comprendre. Il s'agit d'un processus produisant une séquence de symboles qui pourrait être le rêve d'un candidat électoral, puisque tout échantillon suffisamment grand tend à représenter l'ensemble de la séquence. Supposons que deux personnes choisissent des échantillons de manière différente et étudient le développement des propriétés statistiques au fur et à mesure de la croissance des échantillons. Si la situation est ergodique, quelle que soit la façon dont s'est opéré le choix, les deux personnes tomberont d'accord pour estimer les propriétés de l'ensemble. En d'autres termes, les systèmes ergodiques présentent une espèce de régularité statistique particulièrement sure et satisfaisante.

Revenons au concept d'information. Lorsque nous avons une source d'information produisant un message par choix successifs de symboles discrets (lettres, mots, notes de musique, signes de dimension variable, etc.), la probabilité du choix des différents symboles à un stade donné, dépendant des choix précédents (cas d'un processus markovien), que devient l'information associée ?

La quantité précisément nécessaire à l'établissement de l'« information » correspond exactement à la notion thermodynamique **d'entropie**. Elle s'exprime en divers termes de probabilités : celles nécessaires pour l'accession certains stades du processus de formation des messages et celles, une fois ces stades atteints, prévoyant le choix suivant de certains symboles. La formule, en outre, implique le logarithme de ces probabilités, de sorte qu'elle constitue une généralisation de la mesure logarithmique dont on a parlé plus haut à propos de cas simples.

---

[4] En français « Constantinople péchant horrible rosé ».

Pour ceux qui ont étudié les sciences physiques, il est très significatif qu'une expression du type de l'entropie apparaisse dans la théorie comme mesure d'information. Introduite par Clausius, il y a près de cent ans, étroitement associée au nom de Boltzmann et approfondie par Gibbs dans son travail classique sur la mécanique statistique, l'entropie est devenue un concept de base si étendu qu'Eddington remarque à ce propos « La loi selon laquelle l'entropie augmente sans cesse - seconde loi de la thermodynamique - occupe une place privilégiée parmi les lois de la Nature. »

Dans les sciences physiques, l'entropie associée à une situation donnée constitue une mesure du degré de « hasard » ou, si l'on préfère, de « désordre » de cette situation ; la tendance des systèmes physiques à se désorganiser ou à atteindre un désordre total s'avère si fondamentale qu'Eddington prétend qu'elle donne un sens au temps, nous permettant d'affirmer par exemple qu'un film du monde physique se déroule dans le bon sens ou fait marche arrière.

Ainsi, celui qui rencontre le concept d'entropie dans la théorie de la communication a des bonnes raisons de s'exciter - les bonnes raisons de celui qui a réussi à saisir quelque chose de fondamental et d'important. Que l'information soit mesurée par l'entropie paraît, après tout, chose naturelle, si l'on se souvient que cette information, dans la théorie de la communication, est liée à la liberté des choix lors de l'élaboration des messages. Ainsi, pour une source de communication, on peut dire - tout comme pour un ensemble thermodynamique : « Cette situation est hautement organisée, le hasard ou le choix sont donc limités - c'est-à-dire l'information (ou l'entropie) est faible ». Nous reviendrons sur ce point par la suite, car, si je ne me trompe, il s'agit d'un aspect important de la signification générale de cette théorie.

Ayant calculé l'entropie (ou l'information, ou la liberté de choix) d'une certaine source, on peut la comparer à la valeur maximale qu'elle pourrait avoir, à condition toutefois que la source continue d'employer les mêmes symboles. Le rapport de l'entropie présente à l'entropie maximale est appelé l'***entropie relative*** de la source. Si l'entropie relative d'une source est, disons .8, cela signifie grossièrement que cette source possède, dans le choix de symboles formant un message, environ 80 % de liberté possible pour ces mêmes symboles.

Le complément à Un de l'entropie relative est appelé ***redondance***. Il s'agit de la fraction de structure du message qui est déterminée non par le libre choix de l'émetteur, mais plutôt par les règles statistiques admises gouvernant l'emploi des symboles en question. On l'appelle « redondance », car cette fraction du message est en fait redondante selon un sens proche de l'ordinaire; cette fraction du message n'est pas nécessaire (et par là se révèle répétitive) en ce sens que, si elle manquait, le message resterait complet ou du moins pourrait être complété.

Il est très intéressant de remarquer que la redondance de l'anglais est de l'ordre de 50 pour cent[5], de sorte que la moitié des lettres ou des mots utilisés pour l'écriture ou la parole dépendent d'un libre choix, et l'autre moitié (bien qu'on l'ignore d'ordinaire) s'établit sous le contrôle de la structure statistique du langage.

---

[5] I. L'estimation de 50 pour cent ne tient compte que d'une structure statistique de huit lettres environ, de sorte que la valeur optimale est vraisemblablement un peu supérieure.

Mis à part des implications plus sérieuses - et qui seront reprises lors de la discussion finale - il est intéressant aussi de noter qu'une langue doit disposer au moins de 50 pour cent de liberté réelle (ou entropie relative) dans le choix des lettres pour permettre la construction de mots croisés. S'il y a liberté complète, chaque arrangement de lettres devient mots croisés. S'il y a 20 pour cent de liberté seulement, il est alors impossible de construire des mots croisés suffisamment complexes et variés pour en faire un jeu populaire. Shannon a estimé que, si la langue anglaise n'avait qu'une redondance de 30 pour cent environ, il serait possible de construire des mots croisés tridimensionnels.

Avant de clore ce chapitre sur l'information, il faut remarquer que la véritable raison pour laquelle l'analyse du niveau A utilise un concept d'information caractéristique de la nature statistique globale de la source (et ne se préoccupe pas de messages individuels et encore moins de la signification de tels messages) est que, du point de vue technique, un système de communication doit pouvoir capter tout message que la source est susceptible de produire. S'il est impossible de concevoir un système capable de tout capter parfaitement, alors on concevra un système apte à capter les messages les plus usuels et l'on se contentera de sa moindre efficacité pour les tâches rares. Ce genre de considération conduit immédiatement à la nécessité de caractériser la nature statistique de l'ensemble des messages qu'une source peut et doit produire. C'est justement dans ce sens qu'est utilisé le terme information dans la théorie des communications.

Bien que les détails mathématiques ne constituent pas le but de cet article, il semble néanmoins essentiel d'accéder à une compréhension aussi exacte que possible de l'expression, sous forme d'entropie, qui mesure l'information. Dans un cas simple, soit un ensemble de **n** symboles indépendants ou un ensemble de **n** messages complets indépendants, dont les probabilités de choix sont $p_1, p_2, \ldots, p_n$; l'information présente alors :

$$\mathbf{H = - [p_1 \log p_1 + p_2 \log p_2 + \ldots + p_n \log p_n] \text{ ou } H = - \Sigma\, p_i \log p_i.}$$

Le symbole Σ indique[6], comme usuellement en mathématiques, qu'il s'agit de sommer tous les termes $p_i \log p_i$, semblables au terme utilisé comme échantillon.

Cela parait quelque peu compliqué ; mais voyons ce que devient cette expression dans quelques cas simples. En premier lieu, supposons que nous fassions un choix entre seulement deux messages possibles dont les probabilités sont p_1 pour le premier et $p_2 = 1 - p_1$ pour le second. Si nous calculons dans ce cas la valeur numérique de H, il s'avère que H a sa valeur maximale, à savoir 1, quand les deux messages sont également probables, c'est-à-dire quand $p_1 = p_2 = ½$; en d'autres termes, lorsqu'il y a liberté de choix complète entre les deux messages. Dès qu'un message devient plus probable que l'autre (disons p_1 plus grand que p_2), la valeur de H décroît. Et, quand un message est très probable (disons p_1 voisin de un, p_2 voisin de zéro), la valeur de H est très petite (voisine de zéro).

Dans le cas limite où une probabilité vaut l'unité (cas de la certitude) et toutes les autres sont nulles (cas de l'impossibilité), H devient alors nul (aucune incertitude, pas de liberté de choix, pas d'information).
Ainsi H est le plus grand quand les deux probabilités sont égales (c'est-à-dire quand le choix est complètement libre et indifférent) et il est réduit à zéro lorsque la liberté de choix disparaît.

[6] Il ne faut pas s'inquiéter du signe moins. Toute probabilité est inférieure ou égale à un, et les logarithmes des nombres inférieurs à un sont négatifs. Aussi le signe moins est nécessaire pour que H soit en fait positif.

La situation qui vient d'être décrite est, en fait, typique. Si, au lieu de deux choix, il y en a plusieurs, H est maximal quand les probabilités des différents choix sont aussi équivalentes que possible, c'est-à-dire lorsque la liberté est aussi grande que possible, et l'attirance aussi faible que possible vers des choix particuliers. Supposons, d'autre part, qu'un choix ait une probabilité voisine de un, de sorte que les probabilités des autres choix soient proches de zéro, il est clair que, dans une telle situation, on est fortement influencé par un choix particulier et par conséquent que la liberté de choisir est faible. H, dans un tel cas, a une valeur très petite - l'information (la liberté de choix, l'incertitude) est faible.

Ainsi, lorsque le nombre de cas est fixé, l'information est d'autant plus grande que les probabilités des différents cas s'égalisent. Il y a un autre moyen important d'augmenter H en augmentant le nombre de cas. Plus précisément, si tous les choix sont sensiblement équivalents, plus ils sont nombreux et plus H sera grand. Il y a plus d'« information » si l'on opère un libre choix parmi un ensemble de cinquante messages standards que si le libre choix ne se fait que parmi un ensemble de vingt-cinq.

2.3 Capacité d'un canal de communication

Après la discussion du chapitre précédent, il n'est pas surprenant que la capacité d'un canal ne soit pas évaluée en nombre de symboles susceptibles d'être transmis mais plutôt en information transmise. Ou, mieux, le mot information, dans cette dernière phrase, risquant d'être mal interprété, la capacité d'un canal s'exprime dans sa possibilité de transmettre ce que produit une source d'information donnée.

Si la source est d'un type simple délivrant des symboles de même durée (cas d'un télétype, par exemple), si dans cette source chaque symbole choisi représente **s** bits d'information (pris librement parmi 2^s symboles) et si enfin le canal peut transmettre, disons, **n** symboles par seconde, la capacité **C** du canal est alors définie comme étant de **ns** bits par seconde.

Dans un cas plus général, on doit tenir compte des longueurs variables des différents symboles. L'expression générale de la capacité d'un canal comprendra alors le logarithme des nombres de symboles d'une certaine durée (Ce qui introduit naturellement l'idée d'information et correspond au facteur **s** dans le cas simple du paragraphe précédent); cette expression comprendra également le nombre de symboles manipulés (qui correspondent au facteur **n** dans le paragraphe précédent). Ainsi, pour un cas général, la capacité mesure non pus le nombre de symboles mais la quantité d'information transmise par seconde et utilise le hit par seconde comme unité.

2.4 - Le code

Au début, il a été souligné que l'***émetteur*** acceptait le ***message*** et le transformait en un ***signal***; ce signal traverse le ***canal*** pour parvenir au ***récepteur***. L'émetteur, dans le cas du téléphone, transforme simplement le signal, constitué par la voix audible, en autre chose (une variation du courant électrique dans le fil téléphonique), qui est à la fois

nettement différent et nettement équivalent. Mais l'émetteur peut effectuer une opération bien plus complexe sur le message pour produire le signal. Il peut, par exemple, prendre un message écrit et utiliser un certain code pour le chiffrer, disons en une séquence de nombres ; ces nombres envoyés dans le canal constituent alors le signal.
On dit donc, de façon générale, que la fonction de l'émetteur est de coder et celle du récepteur de décoder le message. La théorie s'applique aux émetteurs et récepteurs très complexes comme ceux qui possèdent des « mémoires », de telle sorte que la manière dont ils codent un certain symbole du message dépend non seulement de ce symbole mais aussi des symboles précédents et de la façon dont ils ont été codes.

Nous pouvons maintenant énoncer le théorème fondamental de cette théorie, dans le cas d'un canal sans bruit transmettant des symboles discrets. Ce théorème concerne un canal de communication ayant une capacité de C bits par seconde, recevant des signaux d'une source d'entropie (ou d'information) de H bits par seconde. Le théorème montre que, grâce à des procédés de codage appropriés de l'émetteur, il est possible de transmettre des symboles dans le canal à un débit[7] moyen voisin de C/H, mais ce débit, queue que soit l'ingéniosité du codage, ne peut excéder C/H.

La signification de ce théorème sera discutée de manière plus profitable par la suite, lors du cas le plus général, c'est-à-dire quand le bruit est présent. Pour l'instant, il est important de souligner le rôle critique tenu par le codage.
Souvenons-nous que l'entropie (ou l'information) associée au processus qui produit messages et signaux est déterminée par le caractère statistique de ce processus - autrement dit par les diverses probabilités présentées pour atteindre les situations du message et choisir ensuite les symboles suivants. La nature statistique des messages est entièrement déterminée par le caractère de la source. Mais le caractère statistique du signal réellement transmis dans un canal, donc l'entropie dans le canal, est déterminé à la fois par ce qui est fourni au canal et par l'aptitude de ce dernier à utiliser les différents signaux. Par exemple, dans le cas du télégraphe, des espaces doivent être ménagés entre points, d'autres entre points et traits et d'autres entre traits afin de pouvoir distinguer ceux-ci de ceux-là.

Il apparaît donc maintenant qu'un canal ayant des contraintes de cette sorte, contraintes qui limitent la complète liberté du signal, possède certaines caractéristiques statistiques conduisant à une entropie du signal plus grande que pour tout autre structure statistique ; et, dans ce cas important[8], l'entropie du signal est exactement égale à la capacité du canal.

A l'aide de ces considérations, il est maintenant possible de caractériser avec précision le mode le plus efficace de codage. En fait, le meilleur émetteur est celui qui code le message de façon que le signal possède exactement les caractéristiques statistiques optimales pouvant être utilisées au mieux par le canal - caractéristiques qui, en fait, majorent l'entropie du signal (ou peut-on dire du canal) et la rendent égale à la capacité C du canal.

Cette façon de coder conduit, selon le théorème fondamental énoncé plus haut, au taux maximal de C/H pour la transmission des symboles. Mais, pour ce gain dans le taux de transmission, il faut payer le prix. Car, fort

[7] On se souvient que la capacité C implique l'idée d'information transmise par seconde. L'entropie H mesure ici i'information par symbole, de sorte que le rapport de C à H mesure les symboles par seconde.
[8] Du canal sans bruit (N.D.T)

insidieusement, il se trouve que plus la façon de coder s'approche de l'idéal, plus les délais deviennent longs lors du processus de codage. Ce dilemme est dû en partie au fait qu'en électronique « long » peut correspondre à une très faible fraction de seconde et en partie au fait qu'un compromis se réalise entre le gain dans le taux de transmission et la perte de temps de codage.

2.5 - Le bruit

Comment le bruit affecte-t-il l'information ? Il faut toujours se rappeler que l'information mesure la liberté de choix d'un message. Plus cette liberté de choisir est grande, plus l'information est grande et plus grande est l'incertitude d'avoir réellement choisi tel message particulier. Ainsi, une grande liberté de choix, une grande incertitude et une grande information vont de pair.

Si du bruit est introduit, le message reçu contient certaines distorsions, erreurs ou matériel surajouté qui amènent à penser que le message possède une incertitude accrue, due en fait au bruit. Mais si l'incertitude est augmentée, l'information l'est également, et il semblerait ainsi que le bruit soit bénéfique !

De façon générale, il est exact que lorsqu'il y a bruit le signal reçu contient plus d'information - ou, mieux, que le signal reçu est extrait d'un ensemble plus varié que le signal émis. Une telle situation illustre merveilleusement le piège sémantique dans lequel on tombe lorsqu'on oublie que l'« information » a, ici, un sens spécial : elle mesure la liberté de choix et par là l'incertitude quant au choix effectué. Il est par conséquent possible au mot information d'avoir soit de bonnes, soit de mauvaises connotations. L'incertitude provenant de la liberté de l'émetteur est souhaitable. Celle provenant d'erreurs ou de l'influence du bruit est une incertitude indésirable.

On peut alors dissiper l'équivoque en disant que le signal reçu contient plus d'information mais qu'une partie de cette information introduite par le bruit est falsifiée et indésirable. Pour obtenir l'information utile dans le signal reçu, il faut soustraire cette part falsifiée.

Avant d'éclaircir ce point, nous devons effectuer un petit détour. Supposons deux groupes de symboles : ceux du message émis par la source d'information et ceux réellement reçus. Les probabilités de ces deux groupes de symboles sont liées, car il est clair que la probabilité de réception d'un certain symbole dépend de l'émission de ce symbole. Sans les erreurs dues au bruit ou à des causes diverses, les signaux reçus correspondraient exactement aux symboles du message envoyé ; mais quand les erreurs sont possibles, les probabilités des symboles reçus sont de toute évidence modifiées par rapport à celles qui correspondaient aux symboles du message envoyé.

Dans une telle situation, on peut calculer l'entropie d'un groupe de symboles, relative à celle d'un autre. Considérons, par exemple, l'entropie du message par rapport à celle du signal. Il est regrettable de ne pouvoir sortir d'une telle situation sans entrer dans certains détails. Supposons pour l'instant que l'on sait qu'un symbole du signal est reçu. Chaque symbole du message reçoit alors une certaine probabilité - relativement élevée pour les symboles identiques ou les symboles semblables à celui reçu et relativement faible pour les autres. Utilisant cet ensemble de probabilités, on calcule une valeur approchée de l'entropie. C'est l'entropie du message en supposant connu le symbole reçu ou signal.

Dans des conditions favorables, cette valeur est faible puisque les probabilités en jeu ne se répartissent pas également dans les différents cas mais se concentrent fortement sur un ou quelques cas. Cette valeur serait nulle dans le cas où le bruit serait totalement absent, car, alors, le symbole signal étant connu, toutes les probabilités du message seraient nulles, à l'exception de celles d'un seul symbole (celui reçu) qui aurait une probabilité égale à l'unité.

Pour chaque hypothèse relative au symbole signal reçu, on peut calculer une entropie approchée du message. Une entropie moyenne peut alors être évaluée, en pondérant chaque entropie en accord avec la probabilité du symbole signal suppose. Les entropies ainsi calculées, lorsqu'on considère deux ensembles de symboles, sont appelées entropies relatives. L'entropie particulière qui vient d'être décrite constitue l'entropie relative du message par rapport au signal ; Shannon l'a aussi appelée équivoque.

Selon le mode de calcul de cette équivoque, nous pouvons connaître sa signification. Elle mesure l'incertitude moyenne dans le message lorsque le signal est connu. En absence de bruit, il n'y a aucune incertitude concernant le message si le signal est connu. Si la source d'information a quelque incertitude résiduelle après la connaissance du signal, il doit alors y avoir une incertitude indésirable due au bruit.

La discussion des quelques paragraphes précédents a trait à la quantité d'« incertitude moyenne dans la source du message quand le signal reçu est connu ». On peut également exprimer une « quantité d'incertitude moyenne » similaire « relative au signal reçu lorsque le message envoyé est connu ». Cette incertitude serait évidemment nulle en absence de bruit.

La relation existante entre ces quantités montrerait que :

$\mathbf{H(x) - H_y(x) = H\,(y) - H_x(y)}$, dans laquelle H(x) est l'entropie ou information de la source des messages; H (y) l'entropie ou l'information des signaux reçus; $H_y(x)$ l'équivoque ou l'incertitude dans la source du message quand le signal est connu: $H_x(y)$ l'incertitude dans les signaux reçus quand les messages émis sont connus - ou la partie falsifiée par le bruit de l'information du signal reçu. La partie droite de cette équation constitue l'information utile transmise en dépit de l'effet nocif dû au bruit.

II est maintenant possible d'expliquer ce que l'on entend par capacité C d'un canal où il y a du bruit. Cette capacité est en fait définie comme étant égale au débit maximal (en bits par seconde) d'information utile pouvant être transmise dans le canal.

Pourquoi parle-t-on ici d'un débit « maximal » ? Que peut-on faire pour augmenter ou diminuer ce débit ? Il peut être modifié en choisissant une source dont les caractéristiques statistiques s'accordent aux contraintes imposées par la nature du canal, à savoir qu'il est possible d'augmenter au maximum le débit de transmission de l'information utile en se servant d'un codage approprié (voir pages 49-50).

Considérons, pour terminer, le théorème fondamental pour un canal avec bruit. Supposons que ce canal ait une capacité C, telle qu'elle vient d'être décrite, et qu'il soit alimenté par une source d'information caractérisée par une entropie de H(x) bits par seconde, l'entropie des signaux reçus étant de H(y) bits par seconde. Si la capacité C du canal est égale ou supérieure à H(x) grâce à un codage approprié, l'émission de la source pourra être transmise avec un minimum d'erreur. Aussi petite que soit la fréquence d'erreur demandée, il existe un code satisfaisant. Mais si la capacité C du canal est inférieure à H(x), entropie de la source dont il reçoit les messages, il devient alors impossible d'imaginer des codes pouvant réduire autant qu'on le désire la fréquence de l'erreur.

Quelle que soit l'adresse déployée dans le processus de codage, il n'en restera pas moins vrai qu'après la réception du signal il demeurera quelque incertitude indésirable (bruit) concernant ce que fut le message ; et cette incertitude indésirable - cette équivoque - sera toujours égale ou supérieure à H(x) - C. Par ailleurs, il y a toujours un code, au moins, susceptible de réduire cette incertitude indésirable du message, à une valeur qui n'excède H(x) - C que d'une quantité arbitrairement faible. L'aspect le plus important est naturellement que ce minimum d'indésirable ou fallacieuse incertitude ne puisse être réduit en deçà de cette valeur, et cela queue que soit la complexité ou l'excellence du processus de codage adopté. Ce théorème important permet une description précise et étonnamment simple du degré de fidélité que l'on peut accorder à un canal de communication opérant en présence de bruit.

Une conséquence pratique, relevée par Shannon, doit être remarquée. Etant donné que l'anglais a une redondance de l'ordre de 50 pour cent, il serait possible d'économiser la moitié du temps dans la télégraphie ordinaire grâce à un processus de codage approprié, à condition que la transmission s'effectue dans un canal sans bruit. Lorsqu'il y a bruit, cependant, il y a un avantage substantiel à ne pas utiliser un processus de codage qui élimine toute redondance, car la redondance restante permet de lutter contre le bruit. Il est très aisé de le constater; l'importance de cette redondance en anglais fait, par exemple, qu'on éprouve peu ou pas d'hésitation dans la correction d'erreurs commises en épelant lors de la transmission.

2.6 - Messages continus

Jusque-là, nous nous sommes préoccupés de messages formés de symboles discrets, comme les mots formés de lettres, les phrases de mots, une mélodie de notes ou une peinture en demi-teintes d'un nombre déterminé de touches. Qu'advient-il de la théorie si l'on considère des messages continus, telle la voix parlée avec sa variation continue de hauteur et d'intensité?

De façon globale, on peut dire que l'extension de la théorie devient plus difficile et mathématiquement plus complexe, mais ne diffère pas fondamentalement. Un grand nombre de considérations établies précédemment pour un cas discret demeurent sans modifications et d'autres n'exigent qu'un faible changement. La considération suivante permet de comprendre ce phénomène : dans la pratique, l'intérêt est toujours porte à un signal continu compose de constituants harmoniques simples de certaines fréquences seulement, fréquences contenues, globalement, dans une bande allant de 0 à W cycles -seconde. Ainsi, bien que la voix humaine comporte des fréquences plus élevées, une communication très satisfaisante peut être obtenue dans un canal téléphonique utilisant seulement des fréquences de l'ordre, disons, de quatre mille cycles. Avec des fréquences de dix ou douze mille cycles, une transmission radiophonique haute fidélité de musique symphonique devient possible, etc.

Il existe un théorème mathématique fort commode qui montre qu'un signal continu de durée T secondes, dans une bande de fréquences de 0 à W, peut être entièrement déterminé en établissant 2 TW nombres. Le théorème est très remarquable. Couramment, une courbe continue ne peut être caractérisée que de façon approximative en établissant un nombre fin de points qu'elle traverse, mais un nombre infini serait nécessaire pour obtenir une information complète à son sujet. Cependant, si la courbe est construite à partir de simples constituants harmoniques avec un nombre limité de fréquences, tout comme un son complexe est compose d'un nombre limité de sons purs, alors un nombre fini de paramètres devient seulement nécessaire. II en résulte un avantage considérable permettant de réduire le caractère du problème de la communication des signaux continus, comportant un nombre infini de variables, à un nombre fini (quoique grand) de variables.

Dans la théorie du cas continu, sont développées des formules précisant la capacité maximale C du canal de bande passante de fréquence W, quand la puissance moyenne de transmission est P, le canal étant soumis à un bruit de puissance N; ce bruit ou « bruit blanc thermique » particulier est défini par Shannon. Ce bruit blanc thermique est lui-même dans une bande de fréquences limitée, et les amplitudes des diverses fréquences constituées relèvent d'une distribution de probabilité normale (gaussienne). Compte tenu de ces considérations, Shannon obtient un théorème remarquablement simple, donnant le débit de transmission de digits binaires, avec un codage approprié : $\mathbf{W \log_2 P + N / N}$ bits par seconde, avec une fréquence d'erreur arbitrairement faible. Mais ce débit ne peut être dépassé, quelle que soit l'habileté du codage, sans donner lieu à une fréquence d'erreurs déterminée. Dans le cas de bruit arbitraire, au lieu de bruit « blanc thermique » particulier dont il a été question, Shannon n'a pas réussi à élaborer une formule explicite relative à la capacité du canal ; il a cependant obtenu les limites inférieure et supérieure de cette capacité. II a également trouvé les limites de la capacité du canal lorsque la puissance moyenne de l'émetteur n'est pas déterminée mais que la puissance de pointe instantanée est spécifiée.

Pour terminer, on doit constater que les résultats obtenus par Shannon, tout en étant nécessairement moins spécifiques, revêtent une signification éminemment profonde et large; dans le cas général d'un message ou signal continu, ils permettent de caractériser la fidélité du message reçu, le débit d'information émis par la source, celui de la transmission et de la capacité du canal, le tout en fonction du degré de fidélité désiré.