RECHERCHES INTERDISCIPLINAIRES

Collection dirigée par Pierre Delattre

Séminaires Interdisciplinaires du Collège de France

réalisés avec la collaboration de

l'Institut Collégial Européen (I.C.E.)
et de l'Institut des Sciences Mathématiques
et Économiques Appliquées (I. S. M. E. A.)

sous la direction de

André LICHNEROWICZ
Membre de l'Institut
Professeur au Collège de France
Président de l'I. C. E.

François PERROUX
Professeur au Collège de France
Directeur de l'I. S. M. E. A.
Vice-Président de l'I. C. E.

Gilbert GADOFFRE
Professeur à l'Université de Manchester
Directeur de l'Institut Collégial Européen

INFORMATION ET COMMUNICATION

Introduction de André Lichnerowicz

maloine s.a. éditeur
27, rue de l'Ecole-de-Médecine - 75006 Paris
1983

I

LA THÉORIE DE L'INFORMATION*

par Marcel-Paul Schützenberger

Université de Paris VII

Information, bruit, signal, code..., autant de mots auxquels se sont attachés depuis les dernières décennies des connotations nouvelles qui ne résultent pas seulement des développements des moyens de communication. Qu'ont en commun le *message* que la marquise reçut à cinq heures et celui que sont censés transmettre les affiches du métro ? Bien peu sans doute eut pensé Balzac qui n'ignorait pourtant rien de la publicité mais, au contraire, l'essentiel enseignent aujourd'hui les experts en Sciences des Signes en faisant diligemment référence au petit livre de Claude Shannon, *A mathematical theory of communications* (1949). (Le conférencier brandit un petit volume de couleur rouge.)

La même référence ou une plus vague à une certaine "Théorie de l'Information", figure à maintes pages des oeuvres des Grands Penseurs Contemporains quand ils évoquent la biologie moléculaire, les joies et les peines de la Télévision et l'avenir glorieux (ou béant) de la microinformatique.

Pour en savoir plus nous compulsons les index analytiques et, faites-en l'expérience, ayant pris un ouvrage typique tel que *Recent advances in Information Science and Subtropical Ecology* (en français : "La Théorie de l'Information dans l'approche psychanalytique aux problèmes du Tiers Monde") vous y verrez cité l'ouvrage de Shannon (ainsi que Lewis Carol et Margaret Mead) comme source et fondement d'une pléiade de disciplines originales et d'applications multiples.

Mais en fait de manuel traitant purement et simplement de la théorie de l'information vous ne trouverez que des ouvrages développant les mathématiques des théorèmes découverts par Shannon, à moins que manifestement l'auteur n'ait choisi de mettre le mot magique dans le titre de l'édition renouvelée de son livre habituel. Bref, j'affirme en dépit de bien d'autres plus savants que moi que la théorie de l'information n'a rien à voir avec l'informatique ni même avec l'audio-visuel, et je présente comme prétexte de ce qui suit la citation ubiquitaire (de l'ethnographie structurelle à l'explication de textes) de *A mathematical theory of communications*.

* Texte établi à partir de notes prises lors de l'exposé oral.

Laissez-moi d'abord rappeler très brièvement le contenu de la théorie de Shannon. Pour plus de détails je fais référence au livre de J. Wolfowitz (*Information Theory*, Springer VIg. 1962) qui contient tout ce qu'un mathématicien peut désirer savoir sur le sujet. Des résultats récents peuvent être trouvés, par exemple, dans le livre de Ash. Mais, à vrai dire, le sujet n'a guère évolué depuis quarante ans en ce qui concerne les résultats essentiels, sinon que chaque recoin est devenu une petite théorie plus ou moins autonome.

Ce noyau essentiel est constitué par deux séries de résultats. La première exprime en termes rigoureux l'intuition suivante : supposons que des indications en nombre limité doivent être transmises au moyen de signes convenus ayant des coûts fixés (en temps, en énergie ou en toute autre quantité douée des propriétés formelles voulues), et que l'on s'attache à rendre aussi minime que possible la dépense totale, alors on associera bien sûr aux indications les plus fréquemment utilisées les signes les moins coûteux. Par exemple, si ces indications sont les lettres de l'alphabet dont les fréquences (en français ou en anglais) se trouvent dans l'ordre e, s, a, r, i, ..., q, et si les signaux sont des trains d'impulsions électriques, le code de Morse qui vise à l'économie de temps attribue à "e", "s", "a" les codes respectifs "point", "point, point, point", "point, trait" qui sont parmi les plus brefs. On a laissé à la lettre rare "q" le signal bien plus long "trait, trait, point, trait".

De telles conventions de bon sens, par exemple les abréviations et les sigles, ont été adoptées dans toutes les cultures. L'apport original de Shannon est d'avoir créé un cadre conceptuel à l'intérieur duquel ces problèmes et d'autres connexes peuvent être soumis à des calculs précis quand les hypothèses de base sont satisfaites. Un autre chapitre traite des mêmes questions dans le cas mathématiquement beaucoup plus difficile où les signaux sont des superpositions d'ondes périodiques. L'ensemble de ces résultats est la théorie de l'information en l'absence de "bruit".

Elle commence par schématiser la communication entre un émetteur (la *"source"*) et un récepteur par les deux données suivantes : une description probabiliste complète des propriétés statistiques avec lesquelles sont choisis (ou "produits") les messages à envoyer et une spécification des coûts des signaux élémentaires qui par composition permettront de transmettre sans ambiguïté des suites de messages.

De fait, on peut ramener la seconde donnée au cas le plus simple où il n'existe que deux signaux élémentaires (disons + et -) ayant même coût. Le dispositif se réduit alors à la seule donnée des probabilités d'émission des messages. A celle-ci est attachée d'une façon naturelle une grandeur numérique que les thermodynamiciens ont introduite au siècle dernier sous le nom d'*entropie* et on prouve qu'avec un choix approprié d'unité cette entropie donne le *nombre minimum moyen* de signaux élémentaires requis pour assurer la transmission.

Par exemple si vous voulez transmettre des messages formés des lettres A, B, C, D, E avec les fréquences respectives 1/2, 1/4, 1/8,

1/16, 1/16 vous pourrez coder A par +, B par -+, C par --+, D par ---+, E par ---- ; le coût moyen sera :

$$(1/2).1 + (1/4).2 + (1/8).3 + (1/16).4 + (1/16).4 = 1,375$$

et la théorie vous garantit que l'on ne peut pas faire mieux.

Bref, je le répète, l'entropie de la source est la mesure du coût minimum moyen de la transmission. D'où la tentation irrésistible de rebaptiser l'entropie et de déclarer qu'elle est l'*information* produite par l'émetteur.

L'obtention de cette borne est un résultat très intéressant en lui-même mais sa valeur est encore accrue par la deuxième série de théorèmes qui traitent de la communication en présence de perturbations, de "bruit" (notez les guillemets).

De nouveau il n'y a au départ qu'une intuition banale : si des parasites (par exemple le bruit ambiant) gênent une communication téléphonique on peut y remédier en répétant le message "56, je dis cinquante sisse") ou en transmettant un signal moins ambigu mais plus long "affirmatif, mon Commandant !")

L'étude mathématique exige une nouvelle donnée : celle des éléments statistiques qui caractérisent le "bruit". On extrait de ceux-ci une autre quantité, que l'on appelle la *capacité* de la ligne de transmission et le théorème fondamental de Shannon consiste en une relation très subtile entre la "capacité" de la ligne et la quantité d'information (au sens précis donné plus haut de ce terme) qui peut être transmise. De fait le résultat est de nature asymptotique c'est-à-dire qu'il décrit la situation limite qui se produit quand la longueur du message à transmettre croît indéfiniment. Très schématiquement, ce théorème d'une admirable profondeur prouve qu'il serait possible de recoder le message de façon de plus en plus complexe de telle sorte que simultanément la probabilité d'erreur tende vers zéro et que l'information transmise atteigne la capacité (une pause pour des applaudissements qui ne viennent pas). On me pardonnera de ne fournir ni énoncé précis, ni exemple. Le théorème affirme la possibilité d'un tel codage (ou mieux d'une telle suite de codages variant en fonction de la longueur du message à transmettre) sans *rien* dire sur la manière de trouver celui-ci. C'est ce que les mathématiciens appellent un théorème d'existence, et depuis quarante ans les recherches se poursuivent sans trop de succès pour en trouver une version constructive, c'est-à-dire pour décrire effectivement ces codages dont ce théorème assure qu'ils existent.

Enfin, pour clore, je dois mentionner une théorie jumelle florissante qui par des méthodes toutes différentes se propose de construire des codes ayant a priori des propriétés intéressantes de résistance au bruit : c'est la théorie des codes correcteurs. Le théorème de Shannon donne une limite précise à ce qui peut être obtenu et constitue donc un objectif précis qui oriente la recherche. Il se trouve cependant qu'au point de vue technique cette théorie se développe sans faire le moins du monde recours aux idées de Shannon, ni même mentionner le mot information.

Enfin, les développements récents de la cryptographie doivent tout

aux arcanes les plus élevés de l'arithmétique (et à la célérité des ordinateurs). Les travaux ne partagent techniquement avec la théorie de l'information que l'usage du mot "code" faisant ici référence au secret et non à l'arbitraire.

Il reste à examiner les rapports de la théorie avec ce qui, du monde réel, est hors des mathématiques. Les applications sont fort maigres : je mets à part l'introduction du "bit" comme unité d'information car les "hartleys" étaient connus depuis l'autre guerre et l'usage fait de cette mesure ne nécessite aucune des formules proprement dues à Shannon ou ses continuateurs.

Quelques remarques sur le langage qui figuraient déjà dans les premiers travaux de Shannon et qui donnent comme une mesure de l'écart qui existe entre une phrase appartenant à une langue naturelle et une suite de lettres ou de mots tirés au hasard. Quelques observations sur la capacité du cerveau humain à effectuer la transformation d'un code à un autre, le record étant ici détenu par les sténotypistes, ce qui était d'ailleurs prévisible a priori d'après la manière dont était construite l'expérience. Mais il ne s'est trouvé aucun problème pratique pour requérir le calcul effectif qu'une quantité d'information ou de la capacité d'une ligne de transmission (au sens précis donné à ces termes), et d'ailleurs on eut été bien en peine pour le faire : la mesure de la quantité d'information contenue dans un texte est son écart par rapport au hasard. Mais quel hasard ? Veut-on comparer "le temps a laissé son manteau"... avec les suites aléatoires ayant la fréquence moyenne des lettres du français ? Pourquoi cette base de référence plutôt que celle des phonèmes ? Plutôt que celle des syllabes ? des mots ? ou mieux des mots enchaînés les uns aux autres selon les règles de la grammaire la plus élémentaire (pas d'article devant un verbe, accords de nombre et de genre, etc...) ? Et pourquoi d'ailleurs des mots ? Ne vaudrait-il pas mieux comparer avec les statistiques du discours poétique ou du langage de Cour ? A chaque fois le chiffre obtenu différerait considérablement des autres et ne signifierait en lui-même rien de plus que n'en donne sa définition qui, hors du contexte précis tracé par Shannon, est parfaitement arbitraire. D'ailleurs les statisticiens emploient bien d'autres paramètres (non moins arbitraires d'ailleurs) quand ils veulent vérifier si une suite peut provenir plausiblement ou non d'un tirage au hasard d'après une loi donnée.

Toutes ces questions se posent pour les emplois supposés de la théorie de l'information en biologie, sociologie, ethnologie, écologie, etc, etc... J'espère que quelqu'un plus au fait que moi de ces disciplines m'expliquera pourquoi la formule définissant l'entropie d'une source multinominale apparaît avec son beau " Σ " et ses coquets "$\mathbf{Log_2}$" dans les textes de Haut Niveau (à l'UNESCO très précisément) sous la plume de savants tels que Roman Jokobson ou C. Levi-Strauss. (Une demoiselle au fond de la salle sort, très fermement).

Je soupçonne que le succès de l'informatique doit y être pour quelque chose bien que, comme je l'ai dit, les deux disciplines n'ont en commun que la racine de leur nom (dérivé de l'Italien où en Toscane on parle très judicieusement "d'élucubration de l'information" pour évoquer les techniques électroniques de calcul).

Un cas curieux est celui de la rencontre manquée entre la théorie des codes correcteurs et la biologie moléculaire. Avant que les travaux expérimentaux n'aient livré la structure du code génétique, des hypothèses biologiquement raisonnables et des raisonnements mathématiques avaient suggéré à Gamov une liste précise de codons. Hélas, à la confusion des pythagoriciens (j'en suis un), le code universel utilisé par les êtres vivants n'a rien à voir avec celui prévu et, pire, n'est guère plus inspirant pour un mathématicien que le tarif des P.T.T.

Bien décevant n'est-ce-pas ? Et pourtant j'eusse pu sans me forcer tenir un tout autre propos, celui du mathématicien.

En effet le modèle de communication de Shannon et la définition qu'il implique de l'information n'est pas arbitraire du point de vue du géomètre.

Constituer une théorie n'est pas seulement tracer un cadre axiomatique intuitivement plaisant et définir des quantités par des formules choisies par convenance. En mathématiques ces préliminaires ne comptent pas s'ils ne permettent pas de mettre en place une trame de théorèmes, c'est-à-dire, soyons explicites et pesants, d'énoncés non triviaux requérant des preuves elles aussi non triviales.

Plus riche et serré est ce réseau, plus nous estimons la théorie et plus, l'histoire le montre, sa chance d'efficacité sera grande. Or il se trouve que toutes les tentatives faites pour substituer, en vertu d'a priori extra-mathématiques, au modèle de Shannon d'autres mettant mieux en valeur tel ou tel aspect des communications réputé essentiel, n'ont mené à aucun théorème.

Même, plus précisément, à l'intérieur du modèle de Shannon où comme je l'ai dit la "quantité d'information" contenue dans un message est définie par l'écart de ce dernier au hasard pur, on a proposé d'innombrables formules pour remplacer l'entropie. En vain. Aucune de ces tentatives n'a permis de forger même le premier maillon d'une chaîne de raisonnements intéressants. C'est que l'entropie utilisée par Shannon est un être mathématique jouissant de propriétés privilégiées. A l'opposé des impasses dont je viens de parler, les esprits les meilleurs ont su y voir, bien au-delà de la Thermodynamique, le concept nécessaire pour comprendre et dominer les phénomènes dans lesquels intervient certain type d'interaction entre le tout et ses parties. Oeuvre inachevée qui sera continuée un jour.

Je crois donc qu'effectivement ce modèle qui mène de lui-même à faire intervenir non trivialement l'entropie est le bon, et je pense que la définition qu'il impose de la "quantité d'information" contenue dans un message est la bonne. Le fait que, dans nombre des cas essayés jusqu'ici, le calcul en soit impraticable ou imbécile ne me tourmente pas : la plupart des grandeurs de la physique, (par exemple la température) et les lois qui les relient, étaient dans la même position au siècle d'Archimède faute de techniques de mesure assez précises. Je regrette seulement que cette théorie n'ait pas *encore* pu être largement employée.

Il ne me reste donc plus qu'à présenter les excuses des mathémati-

ciens à ceux qu'ils ont abusés en employant les mots du langage de tout le monde afin de converser entre eux ou avec les ingénieurs d'autres choses, tout aussi ineffables d'ailleurs. C'est une bien dangereuse habitude, et les botanistes plus scrupuleux ne parlent, eux, que d'humifuses ou de pixidules (cf. à ce propos le texte éloquent de notre Maître J. Dieudonné).

J'ai essayé de vous convaincre que "information", "message", "bruit", "etc", "etc" n'étaient en rien ce que feignent de croire certains qui proclament la formule de Shannon. Les liens de ces mots avec les sens qu'ils avaient il y a cinquante ans est plus que ténu. Mais pour moi (moi seul peut-être) il est surtout profond et mystérieux comme celui entre le temps de l'Observatoire de Paris et l'autre de l'attente. Aucune des acceptions ni les deux à la fois ne contient toute la réalité.

A ce même niveau, le modèle de Shannon et la théorie mathématique des jeux de von Neuman me font mieux comprendre (je crois) les liens réciproques qui unissent par exemple, les deux notions évoquées plus haut de secret et d'arbitraire.

Il resterait à commenter, si j'en avais la compétence, le sort de ces expropriations langagières. Parfois, me semble-t-il, c'est le langage courant qui use de métaphore car le courant passe à l'Olympia sans l'intervention des Nymphes. Voire même, disons-le fièrement, c'est le géomètre qui a enrichi le vocabulaire : pendant quelques années la politique intérieure italienne s'est inspirée du mot d'ordre (riémannien) de "parallélisme convergent" (fou-rire de l'enseignant-chercheur en sémiométrie). D'autres cas sont moins clairs : le verbe dans "la vivacité de son expression multipliait les charmes de son visage" renvoie-t-il, (comme on dit) à la Genèse ou à la règle de trois ?

Le plus souvent le larcin s'oublie. Après de longues controverses plus personne ne s'en prend à un Théorème quand (écriture moderne) les $\Sigma m_i v_i^2$ de la Nation tardent à se manifester.

Puisse-t-il en être de même pour la Théorie mathématique des communications.