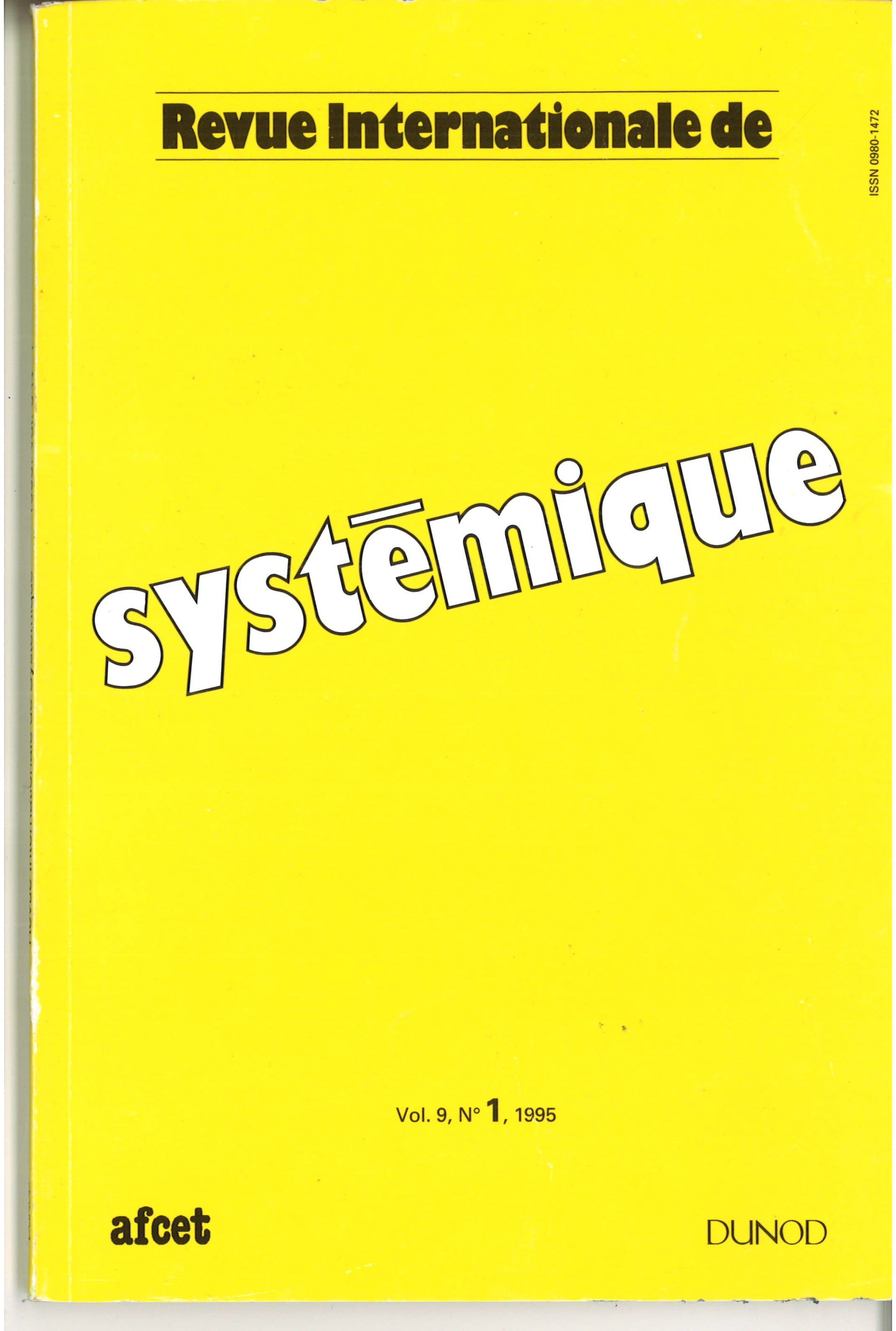

AFSCET

Revue Internationale de systémique

Revue Internationale de Sytémique

volume 09, numéro 1, pages 67 - 98, 1995

Une interprétation objectiviste de la théorie de l'information

François Bonsack

Numérisation Afscet, janvier 2016.

G. SPENCER-BROWN, *Laws of form*, Dutton, New York, 1972.

A. STERN, *Matrix Logic and Mind*, North-Holland/Elsevier, Amsterdam, 1992.

P. TORT, *La raison classificatoire*, Aubier, Paris, 1989.

R. VALLÉE, Sur la modélisation de la perception, in H. B. Lück (dir.): *Biologie Théorique - Solignac 1985*, Ed. CNRS, Paris, 1987, p. 278-288.

R. VALLÉE, Aspects of complementary, in M. E. Carballo (ed.): *Nature, Cognition and System II*, Kluwer Academic Pub, Dordrecht, 1992, p. 77-85.

R. VALLÉE, La « distribution epsilon » comme objet scientifique, *Rev. Int. Systémique*, vol. 8, n° 2, 1994, p. 133-138.

F. VARELA, A Calculus for Self-reference, *Int. J. General Systems*, vol. 2, 1975, p. 5-24.

J. VOGE, Théorie de l'information et économie des systèmes, in *Les renouvellements contemporains de la théorie de l'information*, AFCET, Paris, 1982.

REVUE INTERNATIONALE DE SYSTÉMIQUE
Vol. 9, N° 1, 1995, pp. 67 à 98

UNE INTERPRÉTATION OBJECTIVISTE DE LA THÉORIE DE L'INFORMATION

François BONSACK [1]

Résumé

La quantité d'information n'est ni une entropie (Weaver et Atlan, à la suite de Shannon), ni une néguentropic (Brillouin). Un examen lucide de la manière dont Shannon traite la transmission perturbée par un bruit montre qu'il est amené à définir une nouvelle mesure de la quantité d'information, qui est une *différence de deux entropies*. Cette définition permet d'éviter les pièges dans lesquels tombent ceux (en particulier H. Atlan) qui identifient la quantité d'information (mesure d'un ordre) à l'entropie (mesure d'un désordre). Une telle identification introduit une grande confusion et soulève des difficultés auxquelles on propose des solutions très insatisfaisantes. En outre, l'entropie est envisagée ici non pas comme une incertitude subjective, mais comme une mesure objective de la variété d'un ensemble (pondéré). La quantité d'information apparaît alors comme la spécificité d'un sous-ensemble par rapport à un ensemble, et cette spécificité s'applique très naturellement aux structures biologiques (ADN et protéines), évitant les objections auxquelles se heurtait cette application.

Abstract

The amount of information is neither an entropy (Weaver and Atlan after shannon) nor a neg-entropy (Brillouin). A lucid examination of the way Shannon deals with the noisy discrete channel shows that he is led to define a new measure of information, as a *difference between two entropies*. This definition makes it possible to avoid the pitfalls of those (e.g., H. Atlan) who identify the amount of information (that measures order) with entropy (that measures disorder). Such an

1. Institut de la Méthode, Case postale 1081, CH-2501 Bienne et Université de Neuchâtel.

Rev. intern. systémique. 0980-1472 Vol. 9/95/01/ $ 4.00/© Afcet Gauthier-Villars

identification causes much confusion and creates difficulties to which very unsatisfying answers are given. Another feature of this paper is that it considers entropy not as a subjective uncertainty, but as an objective measure of the variety of a (weighted) set. The amount of information thus appears as the specificity of a subset with respect to a set; this specificity can naturally be applied to biological structures (DNA and proteins), thus aboiding objections previously made to such applications.

INTRODUCTION

Deux raisons me poussent à rappeler la conception de la théorie de l'information que j'avais exposée dans ma thèse *Information, Thermodynamique, Vie et Pensée*, publiée en 1961 chez Gauthier-Villars.
La première, c'est la constatation des nombreuses difficultés que rencontre par exemple H. Atlan avec celle qu'il a adoptée, difficultés pour lesquelles il imagine des solutions souvent fort peu satisfaisantes. J'ai même la nette impression que ces difficultés ont finalement détourné le monde scientifique de la notion d'information selon Shannon, notion dont on ne parle plus guère aujourd'hui et qui est pourtant fort éclairante lorsqu'elle est utilisée à bon escient.
La seconde, c'est la persistance d'une contestation de l'usage de la notion d'information pour caractériser l'organisation des êtres vivants en général et le code génétique en particulier.
Il me semble que la conception que j'ai proposée de la théorie de l'information permet de clarifier le débat sur ces deux problèmes.
Je vais donc
– exposer la théorie de l'information selon Shannon,
– en distinguer différentes interprétations (subjectiviste/objectiviste, dynamique/statique),
– montrer que la manière dont Shannon traite le canal avec bruit suppose en fait une nouvelle définition, plus générale, de la quantité d'information,
– évoquer une autre interprétation, celle de Weaver/Atlan et montrer les nombreuses difficultés qu'elle soulève,
– enfin montrer que l'interprétation objectiviste permet de définir plus généralement des grandeurs purement statistiques (qui ont perdu tout caractère informationnel) : l'*entropie-varié té* et la *spécificité* d'un sous-ensemble par rapport à un ensemble. L'application de ces grandeurs au traitement de l'« information » génétique ne soulève plus aucune difficulté.

MESURE DE LA QUANTITÉ D'INFORMATION

Comment mesurer la quantité d'information apportée par un message ?

Deux approches sont possibles : une approche que je qualifierai d'*objectiviste* parce qu'elle ne fait intervenir que des messages et des ensembles de messages, messages que l'on traite comme des objets et dont on étudie les caractères objectifs. L'autre peut être qualifiée de *subjectiviste* parce qu'elle raisonne sur ce que sait le récepteur et sur la connaissance que le message lui a apportée. J'utiliserai ici presque exclusivement l'approche objectiviste.

Une seconde distinction doit être faite. On peut mettre l'accent sur la *communication*, la *transmission*, c'est-à-dire évoquer une situation où il y a un émetteur, un canal, un récepteur et où ce récepteur est informé de quelque chose. C'est une conception de l'information qu'on peut appeler « dynamique ». Mais on peut aussi considérer la quantité d'information comme un *contenu* statique, indépendant de toute transmission. Dans cette optique, il y a un message au départ, qui contient une certaine information ; lors de la transmission, il perd éventuellement une partie de ce contenu d'information ; à l'arrivée, on a un message qui contient moins d'information, si une partie de l'information a été perdue. (Cette dernière phrase décrit le processus dynamique de la communication en utilisant la quantité d'information en tant que contenu statique.) Il est presque superflu de souligner que cette conception de l'information comme contenu statique est liée à l'approche objectiviste.

La première idée, pour estimer la quantité d'information portée par un message, c'est de compter le nombre de signes. Mais on s'aperçoit immédiatement que cette mesure dépend du code utilisé : un code très riche, tel que celui des idéogrammes chinois, permet de transmettre la même information en moins de signes. La réponse à cette objection n'est pas difficile : on va traduire tous les codes en un code standard, canonique, et on choisit pour cela le plus pauvre possible, celui dont l'alphabet se réduit à deux signes, qu'on peut appeler 0 et 1. Et on dira que la quantité d'information d'un message se mesure par *la longueur minimale de sa traduction en code binaire.*

On voit immédiatement qu'avec un code binaire on peut écrire :

$2^1 = 2$ suites différentes de longueur 1 : 0/1

$2^2 = 4$ suites différentes de longueur 2 : 00/01/10/11

$2^3 = 8$ suites différentes de longueur 3 : 000/001/010/011/100/101/110/111

$2^4 = 16$ suites différentes de longueur 4, etc.

Donc, si l'on veut traduire en code binaire un message écrit dans un alphabet de 32 signes, il faudra, pour traduire chaque signe, au moins 5 signes binaires, puisque $2^5=32$. Et comme le logarithme de base 2 d'un nombre p est la puissance à laquelle il faut élever 2 pour obtenir p, on a immédiatement la règle : *chaque signe d'un alphabet comptant* n *signes transmet une quantité d'information égale à* $\log_2 n$, règle qui ne pose aucun problème pour des alphabets dont l'étendue est une puissance entière de 2 : 32, 64, etc.

PREMIÈRE GÉNÉRALISATION

Mais on s'aperçoit vite que cette règle n'est pas très satisfaisante. Quel code adopter ? – On a déjà 26 lettres, sans compter les majuscules et les lettres accentuées, et 10 chiffres, auxquels il faut ajouter un signe pour la séparation des mots, des signes de ponctuation, des signes mathématiques ou autres signes spéciaux et peut-être quelques signes de service ou de mise en page. Sans majuscules, on peut donc s'en tirer avec 64 signes, mais ils ne suffiront pas pour dactylographier une lettre normale. Faut-il ou non passer à 128 signes, alors que le texte est parfaitement intelligible sans majuscules ? Et, ce qu'il y a de gênant, c'est qu'il y a des signes fréquents et des signes rares. Pour tenir compte de signes rares, il faut allonger les codes de la traduction binaire. Or justement, ces signes sont rares et, dans la plupart des cas courants, on pourrait s'en passer. Va-t-on faire dépendre la quantité d'information de la prise en compte ou non de signes qu'on n'utilisera presque jamais ? Et va-t-on devoir sauter brusquement de 6 signes binaires à 7 lorsqu'on passe d'un alphabet de 64 signes à un alphabet de 65 signes ?

La réponse à cette nouvelle difficulté est un peu plus subtile : on va utiliser des codes binaires de longueur variable, courts pour les signes fréquents, longs pour les signes rares, ceci en prenant des précautions pour que la suite des signes binaires soit univoquement découpable en codes. On montre assez facilement que le meilleur choix pour la longueur de codes binaires, celui qui rend minimale la longueur du message, c'est celui où la longueur l_i du code représentant le signe S_i est égale à $-\log_2 f_i$, f_i étant la fréquence relative de S_i. (Si les f_i sont toutes égales, elles valent $1/n$ et $-\log_2(1/n) = +\log_2 n$: on retombe sur la définition précédente.) La longueur L d'un message long de N signes transcrits en binaire est alors :

$$L = \Sigma_i N f_i l_i = -\Sigma_i N f_i \log_2 f_i = -N \Sigma_i f_i \log_2 f_i$$

et c'est aussi, d'après notre définition, la quantité d'information apportée par le message.

Dernière difficulté : nous avons utilisé ici les f_i, fréquences relatives des signes dans le message qu'on veut transmettre. Ce qui impliquerait que le code devrait changer à chaque message – condition peu satisfaisante, surtout pour des messages courts. Il est donc préférable de se rabattre sur un *type* de messages, par exemple ceux rédigés dans une langue naturelle donnée. Car il est bien évident que le code devra changer selon la langue, puisqu'il dépend des fréquences des lettres et que celles-ci varient considérablement lorsqu'on passe par exemple du français à l'anglais. On fera donc intervenir non pas la fréquence relative des signes dans tel message, mais les fréquences relatives moyennes des signes dans telle langue ou, ce qui revient au même, les probabilités p_i, lorsqu'on choisit au hasard un signe dans un texte rédigé en cette langue, que ce signe soit le signe S_i. On aura alors :

$$L = -N \Sigma_i f_i \log_2 p_i$$

et, si le message est assez long ou si l'on prend un ensemble suffisamment grand de messages, la fréquence relative f_i s'approchera de la probabilité p_i de telle sorte qu'on tendra vers :

$$L = I = -N \Sigma_i p_i \log_2 p_i$$

Comme dans toute théorie faisant intervenir des probabilités, cette formule et la quantité d'information qu'elle définit n'ont de sens que dans le cadre d'un ensemble de messages de même type ; elles n'en ont pas pour un message isolé.

Lorsque les physiciens ont vu cette formule, ils ont immédiatement reconnu une formule utilisée par Boltzmann en mécanique statistique lorsqu'il calculait une grandeur H qui sera interprétée comme représentant l'entropie. On a donc conclu que la quantité d'information n'était qu'une entropie.

Rappelons les caractéristiques de l'entropie :

1. Elle est maximale :
   - lorsque les signes sont équiprobables,
   - lorsqu'ils sont indépendants.
2. Elle augmente avec la variété de l'ensemble des messages à disposition, avec le nombre de complexions différentes possibles.
3. Elle est diminuée par toutes les contraintes, toutes les liaisons, toutes les structures.

4. Elle est donc une mesure :
   – de la variété d'un ensemble de complexions,
   – de la variabilité d'un type de complexions.
5. Elle est en principe additive : l'entropie d'un ensemble de deux complexions, de deux messages ou de deux systèmes est égale à la somme des entropies de chacun d'eux (pour autant qu'ils soient indépendants).

Elle permet donc de caractériser la variété d'un ensemble de messages ou la variabilité d'un type de messages : plus un ensemble de messages est varié, plus grande est l'information qu'*un* de ces messages est capable de porter.

LA TRANSMISSION AVEC BRUIT

Jusqu'ici, cette théorie de l'information semble assez satisfaisante et cohérente ; le seul point étonnant et peut-être un peu inquiétant – la suite montrera qu'on aurait raison de s'inquiéter – c'est qu'on adopte comme mesure de la quantité d'information portée par un message une grandeur qu'on a interprétée en physique comme une mesure du *désordre* moléculaire. Comment l'*ordre* subtil et fragile que constitue un message peut-il être mesuré par une grandeur qui caractérise le désordre ?

Et cette inquiétude est avivée par l'étude de la transmission avec bruit. Le bruit augmente évidemment l'entropie-variété : il crée des mots qui n'existent pas, il détruit les liaisons et les structures. Mais, en même temps, il diminue l'information : le message brouillé a perdu une partie de l'information qu'il contenait, certains passages sont devenus indéchiffrables. Les deux grandeurs *entropie* et *quantité d'information*, dont la première était censée être une mesure de la seconde, varient en sens inverse ! La conclusion est, me semble-t-il, inévitable : l'entropie n'est pas une bonne mesure de la quantité d'information. Elle convient certes dans des cas particuliers, mais il y en a d'autres, par exemple celui de la transmission avec bruit, où elle est prise en défaut. *L'entropie n'est donc pas l'outil conceptuel adéquat pour mesurer la quantité d'information.*

Que faire ?

Avant d'exposer la solution proposée par W. Weaver, le cosignataire avec C. E. Shannon du livre *The Mathematical Theory of Communication*, je crois plus opportun d'aller voir ce qu'a écrit Shannon lui-même. Car j'ai eu la « chance » de lire Shannon dans les articles originaux qu'il a publiés en 1948 dans le *Bell System Technical Journal*, de telle sorte que j'ai dû me forger moi-même une interprétation, sans être influencé par celle de Weaver.

Shannon commence par remarquer que, si un canal est alimenté par une source et s'il est perturbé par un bruit, cela fait deux processus statistiques dont les effets se combinent selon le théorème des probabilités composées :

$$p(x,y) = p(x) \cdot p(y \mid x) = p(y) \cdot p(x \mid y)$$

x étant ici le message au départ, y le message à l'arrivée et $p(x_i, y_j)$ étant la probabilité qu'on ait à la fois le message x_i au départ et le message y_j à l'arrivée. Ce théorème fait intervenir des probabilités conditionnelles : $p(y_j \mid x_i)$, qui est la probabilité d'avoir le message y_j à l'arrivée, étant donné qu'on avait x_i au départ, et $p(x_i \mid y_j)$ qui est la probabilité, si l'on a reçu le message y_j, qu'il provienne du message x_i. (Les x et y sont ici appelés « messages », mais il pourrait tout aussi bien s'agir de signes ou de suites de signes. Même polyvalence pour ce que j'ai appelé un « signe » S_i. Et la quantité d'information doit être la même, qu'on prenne, le message entier ou qu'on le découpe en éléments, ceci pour autant qu'on tienne compte de toutes les liaisons.)

A cause des logarithmes, les produits de probabilités deviennent des sommes pour les entropies, de telle sorte que Shannon peut écrire :

$$H(x,y) = H(x) + H(y \mid x) = H(y) + H(x \mid y) \tag{1}$$

Il faut bien comprendre ce que signifient ces entropies.

$H(x)$ entropie de l'ensemble des messages x au départ ;

$H(y)$ entropie de l'ensemble des messages y à l'arrivée ;

$H(x,y)$ entropie de l'ensemble des couples (x,y) formés d'un message émis x et du message reçu correspondant y ;

$H(y \mid x)$ moyenne pondérée (espérance mathématique) $[\Sigma_i \, p(x_i) \, H(y \mid x_i)]$ des $H(y \mid x_i)$, qui sont les entropies des ensembles de messages y issus d'un message donné x_i (on transmet plusieurs fois le même message x_i, qui peut être perturbé différemment à chaque transmission, et on obtient ainsi un ensemble de messages y dont on calcule l'entropie) ;

$H(x \mid y)$ moyenne pondérée (espérance mathématique) des $H(x \mid y_j)$, qui sont les entropies des ensembles de messages x qui, perturbés, peuvent donner le message y_j (on transmet plusieurs fois tous les messages x possibles et on réunit dans un même ensemble tous ceux qui ont donné le message y_j).

$H(x \mid y_j)$ est donc une entropie concernant le problème, qui est celui du récepteur : étant donné le message perturbé reçu y_j, quel est le message x qui a été émis ? Shannon l'appelle *équivoque* et c'est un choix heureux : étant donné le message qui a été reçu, on ne peut plus déterminer avec sécurité quel message le correspondant a émis, il y a plusieurs possibilités, comme lorsqu'on se trouve devant un texte équivoque.

Shannon définit ensuite un « *taux de transmission réelle* » R (*rate of actual transmission*) :

$$R = H(x) - H(x \mid y) \qquad (2a)$$

Mais il ne parle de *taux* que parce qu'il raisonne tout au long de son article sur des *densités* d'information par signe ou par seconde ; c'est un taux qui ne s'exprime pas en pour-cents, il mesure la densité d'information, le débit d'information qui subsiste à l'arrivée. Si l'on raisonne, comme nous, sur la quantité d'information totale apportée par un message, R doit être interprété comme la quantité d'information I_y, à l'arrivée :

$$I_y = H(x) - H(x \mid y) \qquad (2b)$$

On peut parler d'information au départ, à l'arrivée et d'information perdue puisqu'on considère ici la quantité d'information non plus comme quelque chose de dynamique qui est *transmis*, mais comme le *contenu* informatif statique d'un message, contenu dont une partie se perd dans la transmission avec bruit.

D'où tombe cette formule $2b$?

Ici, je quitte le terrain de l'interprétation objectiviste de la théorie de l'information et, pour des raisons pédagogiques, j'adopte passagèrement l'interprétation subjectiviste : l'entropie ne sera plus une mesure de la variété d'un ensemble de messages, mais une mesure de l'*incertitude du récepteur quant au message choisi parmi un ensemble de messages.* Ces deux mesures sont évidemment parallèles : plus l'ensemble des messages est varié, plus grande sera l'incertitude du récepteur quant au message choisi.

Dans cette perspective subjectiviste, la quantité d'information se définit non plus comme la longueur du message transcrit en code binaire, mais, très naturellement, comme *la diminution, due à la réception du message, de l'incertitude du récepteur* (les incertitudes étant mesurées par des entropies).

On voit immédiatement que cette définition correspond tout à fait à ce qu'on attend intuitivement d'une mesure de l'information : avant réception, on avait une grande incertitude ; l'information reçue permet de réduire cette incertitude. Si le message reçu est parfaitement univoque, elle permet de la réduire à zéro, à une certitude quant au message émis. Par contre, si le message a été brouillé, il subsiste après réception une certaine incertitude, l'équivoque. Plus la quantité d'information reçue est grande, plus elle permet de diminuer l'incertitude du récepteur, mais si on annonce à celui-ci une

nouvelle qu'il connaissait déjà avec certitude, le message ne lui apporte plus aucune information : il ne diminue pas son incertitude.

On voit non moins immédiatement que la nouvelle définition ($2b$) de la quantité d'information n'est pas du tout équivalente à l'ancienne. Il n'y a plus une entropie, il y a en a deux : l'incertitude quant au message émis, qui était $H(x)$ avant réception du message, est réduite à l'équivoque $H(x \mid y)$ une fois le message reçu. Et la quantité d'information reçue, c'est la *différence* entre les deux. Enfin, cette nouvelle définition de la quantité d'information peut être aisément appliquée au cas de la transmission sans bruit ou à la quantité d'information à l'émission : si, dans ces cas, on ne rencontre qu'une entropie, qu'on croit pouvoir identifier à la quantité d'information, c'est que la seconde, l'équivoque, est nulle. Tant qu'il y a certitude quant au message émis, $H(x \mid y)$ est nulle et $I = H(x) = H(y)$, puisque $x = y$.

Cette formule $2b$ peut d'ailleurs être également justifiée par un raisonnement objectiviste : Shannon montre en effet, dans son théorème 10, que $H(x \mid y)$ est bien la quantité d'information manquante à la réception, puisqu'il faut un canal ayant au moins une capacité égale à $H(x \mid y)$ pour transmettre en temps réel des messages capables de rétablir à la réception le message émis.

Avant de transposer ce résultat dans l'interprétation objectiviste, je vais montrer, avec Shannon, que cette définition ($2b$) de la quantité d'information est équivalente à une autre. Si l'on tire $H(x) = H(y) + H(x \mid y) - H(y \mid x)$ de (1) et si l'on remplace dans ($2b$), on obtient :

$$\begin{aligned} I_y &= H(x) - H(x \mid y) = H(y) + H(x \mid y) - H(y \mid x) - H(x \mid y) \\ &= H(y) - H(y \mid x) \end{aligned}$$

On peut maintenant revenir à l'interprétation objectiviste, où la transmission peut se représenter par le schéma suivant (les surfaces sont proportionnelles aux entropies, non au nombre d'éléments).

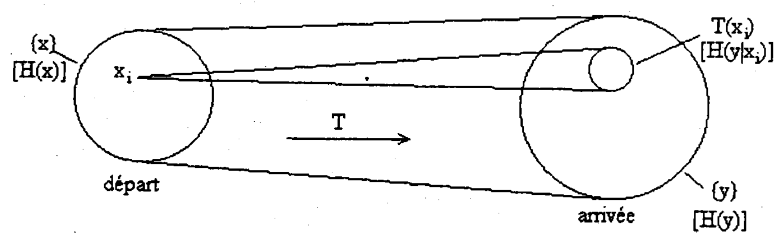

Figure 1.

Si $H(x)$ n'était pas maximale, l'ensemble des messages y à l'arrivée est effectivement plus varié, plus divers, moins lié, plus désordonné que l'ensemble des messages au départ : on a donc bien $H(y) > H(x)$. Mais l'ensemble constitué par le seul message x_i, choisi par l'expéditeur avait une entropie nulle. De telle sorte que la quantité d'information à l'émission était :

$$I_x = H(x) - 0 = H(x)$$

A la réception, $H(y)$ a augmenté. Mais le message x_i est devenu un ensemble $T(x_i)$ de messages perturbés de diverses manières, donc d'une certaine variété $H(y \mid x_i)$. Et l'entropie moyenne de ces sous-ensembles a plus augmenté en passant de 0 à $H(y \mid x)$ que n'a augmenté celle de l'ensemble total en passant de $H(x)$ à $H(y)$, de telle sorte que la quantité d'information à la réception :

$$I_y = H(y) - H(y \mid x)$$

est plus petite que la quantité d'information I_x à l'émission : on a perdu de l'information lors de la transmission avec bruit (démonstration : de (1), on tire $H(y) - H(x) = H(y \mid x) - H(x \mid y) < H(y \mid x) - 0$ si $H(x \mid y) > 0$).

La quantité d'information à la réception vient d'être calculée à partir des messages à la réception. Mais on peut aussi l'estimer à partir des messages à l'émission en considérant l'une des fonctions inverses T^{-1} de la transmission T, qui fait correspondre au message y_j effectivement reçu l'ensemble des messages x qui ont pu lui donner naissance. Le diagramme devient alors celui de la figure 2 ci-après.

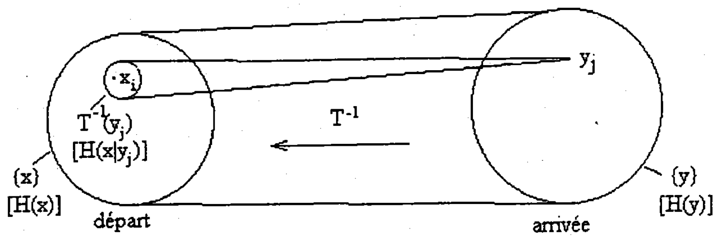

Figure 2.

On voit bien ici que la quantité d'information a diminué puisque l'entropie du sous-ensemble qui se réduisait avant transmission au message x_i était nulle, alors que celle de l'ensemble dont peut être issu y_j a, s'il y a eu perturbation par le bruit, une certaine entropie-variété $H(x \mid y_j)$. En moyenne sur tous les messages :

$$I_y = H(x) - H(x \mid y) < H(x) = I_x$$

UNE PREMIÈRE INTERPRÉTATION DE LA THÉORIE DE SHANNON

Quelle leçon tirer de la manière dont Shannon traite la transmission perturbée par un bruit ? Elle me semble s'imposer :

1. L'entropie qui, dans les transmissions non perturbées, paraissait être une mesure adéquate de la quantité d'information, s'est révélée inadéquate pour la transmission avec bruit : l'entropie augmente alors que la quantité d'information diminue.

2. Shannon, avec un instinct mathématique admirablement sûr, trouve ce qu'il faut faire : il imagine – sans le dire explicitement – une définition plus générale de la quantité d'information qui, elle, décroît effectivement lors de la transmission avec bruit. Il introduit en quelque sorte une nouvelle grandeur, qu'on peut appeler *spécificité* d'un sous-ensemble par rapport à un ensemble. Cette spécificité se mesure par une *différence des entropies de l'ensemble total et du sous-ensemble.* Cette spécificité est d'autant plus grande :

– que l'ensemble de départ, l'ensemble de référence était plus varié (que son entropie était plus grande),

– que le sous-ensemble sélectionné est plus étroit, plus spécifique, plus pointu, moins varié (que son entropie est plus petite).

Si le sous-ensemble sélectionné ne comprend qu'un seul élément, son entropie-variété est nulle et la spécificité se ramène à l'entropie de l'ensemble total.

Cette nouvelle définition de la quantité d'information à partir de la spécificité correspond bien à ce que Shannon dit, au début de son article, vouloir faire. Il écrit par exemple :

« Le problème fondamental de la communication est celui de reproduire exactement ou approximativement en un point un message *sélectionné* en un autre point. [...] L'aspect significatif est que le message en question est *un message choisi parmi un ensemble de messages possibles.* » [c'est moi qui souligne].

La seule nouveauté, c'est qu'on sélectionne un sous-ensemble de messages et non plus un seul.

Le tableau suivant résume les interprétations des différentes grandeurs utilisées :

Tableau

	Interprétation subjectiviste	Interprétation objectiviste
entropie H	incertitude	variété d'un ensemble, variabilité d'un type
entropie $H(x)$ au départ	incertitude quant au message choisi (dans l'ignorance du choix !)	variété de l'ensemble des messages possibles au départ
entropie $H(y)$ à l'arrivée	incertitude quant au message reçu (avant sa réception !)	variété de l'ensemble des messages possibles à l'arrivée
équivoque $H(x \mid y)$	incertitude quant au message émis (message reçu connu !)	variété de l'ens. des messages dont peut être issu le message reçu
$H(y \mid x)$	incertitude quant au message reçu (message émis connu !)	variété de l'ensemble des messages issus d'un message
quantité d'information $H(x) - H(x \mid y)$ $= H(y) - H(y \mid x)$	diminution, due à la réception du message, de l'incertitude du récepteur	spécificité moyenne des ensembles de messages issus d'un message par rapport à l'ensemble des messages à la réception

La nouvelle définition de la quantité d'information ne s'applique bien sûr pas au seul cas de la transmission avec bruit. Il est bien rare qu'il n'existe qu'une manière d'exprimer le message qu'on désire transmettre. En général, il y a donc tout un ensemble de messages qui portent une signification quasiment équivalente. De ce fait, la spécificité du contenu informatif est nettement plus faible qu'elle ne le serait si tous les messages avaient un contenu informatif différent. Ici encore, la quantité d'information est mesurée plus fidèlement par la différence d'entropie entre un sous-ensemble et un ensemble que par la seule entropie de l'ensemble de tous les messages possibles.

Enfin, la notion d'ambiguïté introduite par Shannon dans la transmission avec bruit peut s'appliquer à l'ambiguïté ordinaire des langages naturels. Supposons qu'on ait, avec R. Montague, réussi à construire, à partir du langage naturel, un langage sans ambiguïté. Alors, aux énoncés ambigus de nos langues naturelles correspondra un sous-ensemble d'énoncés non ambigus compatibles avec l'énoncé ambigu, et la quantité d'information apportée par l'énoncé ambigu sera ici encore adéquatement mesurée par la différence d'entropie entre le sous-ensemble des énoncés non ambigus compatibles avec lui, et l'ensemble de tous les énoncés non ambigus.

TROIS DÉFINITIONS, DEUX GÉNÉRALISATIONS

En résumé, on peut dire qu'il y a eu successivement trois définitions de la quantité d'information I, définitions reliées par deux généralisations dans un sens, deux spécialisations dans l'autre sens.

1[re] définition : $I = \log n$ (n = nombre de signes de l'alphabet) (signes équiprobables, codes de même longueur).

La quantité d'information est mesurée par un logarithme.

2[e] définition : $I = H = -N\,\Sigma_i\, p_i \log_2 p_i$ (p_i = probabilité du signe S_i dans le type de messages) (signes non équiprobables, codes de longueurs différentes).

La quantité d'information est mesurée par une entropie H.

Si les signes sont équiprobables, toutes les p_i sont égales à $1/n$, $-\log_2 p_i = \log_2 n$ et on retombe sur la première définition puisque $\Sigma_i p_i = 1$.

3[e] définition : $I = H_O - H_o$ (o = sous-ensemble, O = ensemble de référence) (transmission avec bruit).

La quantité d'information est mesurée par une différence de deux entropies.

Il s'agit de la *spécificité* d'un sous-ensemble (pondéré) – signifiant ce que l'expéditeur a à communiquer – par rapport à un ensemble (pondéré) de référence : l'ensemble de tous les messages possibles de ce type. Cette spécificité se mesure par la différence entre les entropies de l'ensemble et du sous-ensemble. Dans le cas de la transmission avec bruit, à la réception, le sous-ensemble est l'ensemble des messages (avec plus ou moins d'erreurs différentes en différents endroits) pouvant être issus du message envoyé. Tant que ce sous-ensemble est moins varié que l'ensemble des messages issus de n'importe quel message, on peut encore en tirer une certaine information quant au message envoyé. Cette information sera d'autant plus grande que le message est moins perturbé, donc que le sous-ensemble est moins varié.

S'il n'y a pas de perturbation par le bruit, chaque message ne produit qu'un seul message et l'entropie H_o d'un sous-ensemble réduit à un seul élément est nulle, de telle sorte que $I = H_O$.

(Je parle ici de sous-ensembles pour soutenir l'intuition, mais en fait le sous-ensemble peut être l'ensemble entier, avec cependant une distribution de probabilités beaucoup plus pointue, essentiellement concentrée autour d'un point, alors que l'ensemble de référence a une distribution beaucoup plus plate, plus uniforme. On ne peut en général parler de sous-ensembles au sens propre que si l'on sort de ces sous-ensembles des éléments dont la somme des probabilités est quasiment nulle. Et je parle d'ensembles pondérés parce que chaque élément de l'ensemble est muni d'une probabilité, la somme des probabilités de tous les éléments de l'ensemble étant égale à 1.)

L'essentiel du « message » du présent article, c'est que la seconde généralisation et la troisième définition sont passées quasiment inaperçues et que c'est de là que proviennent les difficultés conceptuelles qui embarrassent bien des discussions faisant appel à la quantité d'information de Shannon.

L'INTERPRÉTATION DE WEAVER

Maintenant qu'on a vu ce que fait Shannon et qu'on en a tiré les leçons qui s'imposaient, on peut revenir à l'interprétation proposée par Warren Weaver dans l'article publié généralement comme préface et introduction à celui de Shannon (et qui est, hélas, le seul que lisent ceux qui ont quelque peine à suivre les développements mathématiques de ce dernier).

Weaver voit très bien le problème:

« ... le message reçu présente, du fait du bruit, une incertitude accrue. Mais, si l'incertitude est accrue, l'information est elle aussi accrue et le bruit paraît bénéfique ! »

Mais la solution qu'il propose est tout à fait insatisfaisante:

« C'est une situation qui illustre bien le piège sémantique dans lequel on peut tomber si l'on ne garde pas à l'esprit qu'*information* est ici utilisé dans un sens particulier qui mesure la liberté de choix et donc l'incertitude quant au choix qui a été fait. Il est par conséquent possible que le mot *information* prenne des connotations positives ou négatives. [...] Une incertitude qui provient d'erreurs ou qui est due au bruit est de l'incertitude indésirable.

Il est donc clair où se trouve l'erreur lorsqu'on dit que le signal reçu contient plus d'information. Une partie de cette information est fausse (*spurious*), indésirable et elle s'est introduite à travers le bruit. Pour obtenir l'information utile dans le signal reçu, il faut soustraire cette partie fausse. »

Autrement dit: Weaver ne voit pas que Shannon a changé de définition pour la quantité d'information. Il en reste à la définition initiale, et sous une forme particulièrement vicieuse, puisqu'il identifie *information* et *incertitude* – ce qui est assez exactement le contraire. Et, alors que Shannon avait cherché à donner de la quantité d'information une définition technique qui colle autant que possible au sens intuitif, lorsque Weaver constate que, dans un cas important, l'information dans son sens technique et l'information intuitive varient en sens inverse – ce qui est tout de même assez grave – il ne conclut pas que sa définition technique n'est pas bonne et qu'il faut la changer, mais au contraire qu'elle s'écarte de la signification intuitive, qu'on est tombé

dans un piège sémantique en les identifiant. Ce qui le conduit à désigner comme *information fausse* ou *indésirable* quelque chose qui n'a plus rien d'une information, puisqu'elle la détruit, et revient à transposer le problème sur le plan sémantique en occultant l'élégante solution que Shannon lui avait apportée sur le plan *technique*.

(Il serait cependant injuste de rejeter sur Weaver toute la responsabilité de la mauvaise interprétation). Car Shannon lui-même avait ouvert la voie en écrivant, à propos de l'interprétation intuitive de R:

« La première forme de la définition a déjà été interprétée comme la quantité d'information moins l'incertitude quant à ce qui a été émis. La seconde mesure la quantité reçue moins la part de celle-ci qui est due au bruit ».

Autrement dit: il y a une part de l'information qui est due au bruit. Weaver n'a donc rien inventé et on peut se demander si Shannon lui-même a mesuré toute la portée de ce qu'il avait fait.

On peut penser qu'il y a de ma part quelque outrecuidance à prétendre comprendre ce que Shannon a fait mieux qu'il ne l'a compris lui-même. Mais:

1. Comme dénier à $R = H(x) - H(x \mid y) = H(y) - H(y \mid x)$ le statut de « quantité d'information à la réception » ? Car si, comme l'écrit Shannon, l'équivoque $H(x \mid y)$ est l'information *manquant* (« *missing information* ») à la réception, par rapport à l'information émise $H(x)$, alors R doit être l'information *subsistant* à la réception. Et si, selon Weaver, $H(y \mid x)$ est l'information *fausse* (*spurious*) due au bruit à la réception, alors R doit être l'information *vraie* à la réception (« *the useful information in the received signal* », comme il l'écrit lui-même). Ce n'est que si l'on considère R comme la quantité d'information à la réception qu'on peut parler de perte d'information lors d'une transmission avec bruit.

2. Or si R est une quantité d'information, n'est-il pas impératif de proposer une définition de la quantité d'information suffisamment générale pour l'englober? Ne faut-il pas prendre au sérieux le fait qu'ici, l'information se mesure non plus par une entropie, mais par une *différence* de deux entropies [qui n'est plus une entropie-variété, mais une spécificité !] ?)

ATLAN ET L'« ORDRE PAR LE BRUIT »

Quant à Henri Atlan, il est à la recherche d'un « principe d'ordre à partir du bruit » qui lui permette de rendre compte d'une part de l'évolution biologique selon un mécanisme darwinien, d'autre part de l'« apprentissage non dirigé », c'est-à dire sans un professeur qui transmette des recettes déjà éprouvées.

Pour cela, il s'inspire des idées de Weaver selon lequel le mot « information » peut avoir de bonnes ou de mauvaises connotations et, dans une transmission avec bruit, il y a de l'information indésirable ou fictive qui doit être soustraite. Cette « information », créée par le bruit, qui a une connotation négative dans la transmission avec bruit, ne pourrait-elle pas avoir une connotation positive dans un autre contexte, ce qui expliquerait l'ordre à partir du bruit?

« En fait, Weaver déjà, dans son introduction au travail de Shannon, avait remarqué que les effets du bruit sur des signaux dans une voie augmentent la quantité d'information à la sortie de la voie puisque son incertitude a augmenté. Cela lui apparaissait comme un « effet bénéfique » paradoxal du bruit, inacceptable dans le cadre d'une théorie de la communication où le but est de transmettre l'information avec un minimum d'erreurs. [...] Nous savons donc que cette première intuition de Weaver était correcte et qu'elle peut être le fondement de la solution du problème de la création d'information dans le cadre de la théorie de Shannon. » (H. Atlan: *Entre le cristal et la fumée, essai sur l'organisation du vivant*, abrégé C&F, Seuil, Paris, 1972, pp. 82-83).

Atlan ignore la seconde généralisation et identifie, tout comme Weaver, information avec incertitude, variété, entropie. D'autre part, il a quelque difficulté à appliquer la théorie de l'information, qui concerne la transmission sur une voie, à des structures statiques, où il n'y a pas à proprement parler de transmission d'information. Pour résoudre cette difficulté, il suppose, après d'autres, que, lorsqu'on observe une structure, celle-ci nous envoie un « message » sensoriel qui « décrit » cette structure, ce qui permet de rétablir un contexte où il y a une voie, entre le système et l'observateur, sur laquelle circule de l'information (« la voie, toujours implicite, qui va du système dans son ensemble vers l'observateur ») (H. Atlan: *L'organisation biologique et la théorie de l'information*, abrégée OBThI, Hermann, Paris, 1972, p. 254). On reviendra plus loin sur cet artifice.

Atlan fait alors le raisonnement suivant (*cf.* figure ci-dessous): supposons qu'un observateur observe un système formé de deux sous-systèmes A et B, le premier ayant déterminé la structure du second en lui envoyant un message (pour fixer les idées, Atlan suppose que le système A est l'ADN et le système B une protéine dont la suite des acides aminés est déterminée par la suite des triplets de bases de l'ADN, *cf.* OBThI, p. 260).

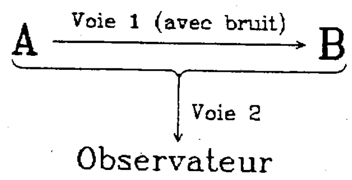

Figure 3.

Si l'on considère maintenant la voie 1, le bruit sur cette voie provoque une perte d'information: on doit soustraire de $H(x)$, l'entropie de A, l'équivoque $H(x\,|\,y)$ pour obtenir la quantité d'information qui parvient à B sur l'organisation de A et déterminera la structure de B. Mais si l'on considère le système total constitué par $A+B$ et si l'on estime la quantité d'information qu'il contient (l'entropie $H(x,y)$), alors la situation change: il faut *ajouter* l'équivoque à $H(x)$ pour obtenir la quantité d'information qui doit être transmise à l'observateur pour qu'il connaisse l'organisation de $A+B$. Comme l'écrit Atlan (C&F, pp. 47-48):

« Alors que la quantité d'information transmise de A à B est égale à celle de B diminuée de l'ambiguïté, la quantité d'information de l'ensemble A et B est égale à celle de B augmentée de cette ambiguïté. En effet, la grandeur qui mesure l'ambiguïté n'est pas autre chose que la quantité d'information de *B en tant que B est indépendant de A*: il est donc normal que cette quantité soit considérée comme perdue du point de vue de la transmission de A à B, et au contraire comme un supplément du point de vue de l'information totale (c'est-à-dire de la variété) de l'ensemble du système. On voit ainsi comment un rôle positif, « organisationnel », du bruit peut être conçu dans le cadre de la théorie de l'information sans pour autant contredire le théorème de la voie avec bruit. [...]

Nous avons donc affaire à deux sortes d'effets de l'ambiguïté produite par un bruit sur l'organisation générale d'un système, que nous avons appelées ambiguïté destructrice et ambiguïté-autonomie, la première devant être comptée négativement, la seconde positivement. »

Et plus loin (p. 67), Atlan parle d'un « *changement de point de vue* sur le rôle de l'ambiguïté dans les communications à l'intérieur du système ».

Critique

A un détail près, et dans le contexte qu'a choisi Atlan, cette solution paraît raisonnable: si la quantité d'information est égale à l'entropie et si la quantité d'information représentée par la structure d'un système se mesure par la quantité d'information qu'il faut transmettre à un observateur pour décrire cette structure, alors la quantité d'information contenue dans $A + B$, c'est $H(x,y)$ qui est égale à $H(x) + H(y\,|\,x) = H(y) + H(x\,|\,y)$, avec un signe plus, alors que la quantité d'information transmise de A à B s'écrit $H(x) - H(x\,|\,y) = H(y) - H(y\,|\,x)$ avec un signe moins. Signalons cependant que la quantité d'information supplémentaire qui apparaît sur la voie 2 n'est pas de l'information significative ou organisée telle que celle qui circule sur la voie 1, elle est de l'information *sur le bruit*, sur la manière particulière dont le message a été perturbé.

D'abord le « détail »: l'examen des formules ci-dessus montre que ce n'est pas la même quantité qui est ajoutée ou soustraite. À $H(x)$ on ajoute $H(y\,|\,x)$, mais on soustrait $H(x\,|\,y)$; à $H(y)$ on ajoute $H(x\,|\,y)$, mais on soustrait $H(y\,|\,x)$.

Or $H(x\,|\,y)$ et $H(y\,|\,x)$ sont deux fonctions différentes, qui n'ont pas la même valeur. $H(x\,|\,y)$, c'est ce que Shannon a pertinemment appelé l'« équivoque »; elle mesure l'incertitude sur le message *émis*, le message reçu étant connu. $H(y\,|\,x)$, c'est ce qu'Atlan appelle moins pertinemment l'« ambiguïté », c'est l'incertitude sur le message *reçu*, le message émis étant connu. Cette dénomination me semble moins pertinente parce que, dans le langage courant et même pour Shannon, équivoque et ambiguïté sont synonymes. En effet, Shannon écrit que l'équivoque « mesure l'ambiguïté moyenne du signal reçu ». Or, alors que l'équivoque mérite son nom, rien ne permet de justifier celui d'« ambiguïté » pour $H(y\,|\,x)$, qui ne représente pas une multiplicité de sens cachés sous une seule forme, mais une multiplicité de formes issues d'un seul sens, si l'on me permet, en accord avec l'usage habituel du mot « équivoque », de désigner par « sens » le message sensé x émis et par « forme » la suite y de signes reçue après perturbation.

$H(x\,|\,y)$ et $H(y\,|\,x)$ ont en général des valeurs différentes: puisque $H(x) - H(x\,|\,y) = H(y) - H(y\,|\,x)$ (que l'on peut transformer en $H(x) - H(y) = H(x\,|\,y) - H(y\,|\,x)$), on voit bien que si $H(y)$ est différente de $H(x)$, $H(y\,|\,x)$ doit être différente de $H(x\,|\,y)$. Il peut même arriver que l'équivoque soit nulle alors que l'« ambiguïté » ne l'est pas: si les messages ou les morceaux de messages sont – comme dans les codes auto-correcteurs – si différents les uns des autres que, même perturbés, ils restent distinguables, l'équivoque est nulle (il n'y a pas d'incertitude quant au message émis) alors que l'« ambiguïté » ne l'est pas.

Par conséquent, la phrase d'Atlan: « Alors que la quantité d'information transmise de A à B est égale à celle de B diminuée de l'ambiguïté, la quantité d'information de l'ensemble A et B est égale à celle de B augmentée de cette ambiguïté » est incorrecte. Ce qui serait correct, c'est que la quantité d'information de l'ensemble $(A + B)$ est égale à celle de A (et non celle de B) augmentée de l'« ambiguïté », ou à celle de B augmentée de l'*équivoque* (et non de l'ambiguïté). Il n'y a pas simple « changement de signe de la fonction ambiguïté », il y a en même temps changement de fonction: l'« ambiguïté » est remplacée par l'équivoque. Il n'y a donc plus changement de signe d'une *même* quantité d'information, changement qui justifiait le changement de *valeur* de cette information et qui joue un rôle central dans les conceptions d'Atlan:

« ... il m'est apparu que cette transformation qui correspond en fait à un changement de signe – l'effet du bruit qui est affecté d'un signe moins à un certain niveau devient affecté d'un signe plus à un autre niveau – n'est qu'un cas particulier de changement de signe qui apparaît chaque fois qu'on passe d'un niveau à un autre, chaque fois qu'on passe d'un niveau élémentaire à un niveau plus englobant. » (H. Atlan *et al.*, *Création et désordre*, L'Originel, Paris, 1987, pp. 23-24).

On me dira que cela détruit certes une partie de la beauté de la symétrie, mais que ce n'est pas tellement grave: l'essentiel est que dans un cas, il y a un signe moins et dans l'autre un signe plus. Et tant pis si ce n'est pas la même quantité qui est ajoutée ou retranchée.

Mais il y a plus grave. Ce qui me paraît non seulement contestable, mais inacceptable, ce n'est pas tellement la manière dont le problème est *résolu*, c'est la manière dont il est *posé*, les hypothèses qui définissent le cadre dans lequel Atlan se situe.

D'abord, comme je l'ai déjà indiqué, il n'a pas vu la seconde généralisation de la notion d'information: pour lui comme pour Weaver, « information » reste équivalente à « entropie », « variété », « incertitude ». Par exemple p. 48 de C&F: « la quantité d'information totale (c'est-à-dire la variété) », p. 50: « Le taux de variation de la quantité d'information $\mathrm{d}H/\mathrm{d}t$ » et « Une diminution de la quantité d'information H du système » (depuis Boltzmann, H désigne traditionnellement une entropie). Cette identification le conduit à des énoncés qui seraient corrects pour des entropies, mais ne le sont pas pour des quantités d'information, par exemple: « Les facteurs de bruit diminuent la redondance du système en général et par là même augmentent sa quantité d'information »; c'est vrai pour l'entropie-variété, mais c'est faux pour la quantité d'information. Dans le cas de la voie avec bruit, $H(y)$, $H(x\,|\,y)$ et $H(y\,|\,x)$ sont des entropies, mais ce ne sont pas des quantités d'information (seules $H(x)$ et $[H(y) - H(y\,|\,x)]$ le sont). Contrairement à ce que pensait

Weaver et à ce qu'écrit Atlan au haut de la p. 47, l'« ambiguïté » n'est *jamais* une quantité d'information, qu'elle soit affectée d'un signe plus ou d'un signe moins.

Ensuite : qui est prêt à admettre que ce qui distingue un bruit désorganisateur d'un bruit organisateur, c'est un changement de point de vue? *Est organisateur ce qui joue un rôle positif dans l'organisation, ce qui améliore les performances fonctionnelles et permet une meilleure survie.* Comme l'écrit von Neumann, cité par Atlan:

« Il existe un concept qui sera très utile ici, dont nous avons une certaine idée intuitive, mais de façon vague, non scientifique et imparfaite. Ce concept appartient de toute évidence au domaine de l'information, et des considérations quasi-thermodynamiques lui sont applicables. Je n'en connais pas de nom adéquat, mais il est le mieux décrit dans le terme de « complication ». Il s'agit de l'efficacité dans la complication, ou de la potentialité de faire des choses. Il ne s'agit pas du degré d'élaboration d'un objet, mais celui de ses opérations intentionnelles. En ce sens, un objet a le plus haut degré de complexité, s'il peut faire des choses très difficiles et très élaborées. » (OBThI, p. 218).

Von Neumann est très prudent sur certains points, mais très net sur d'autres: il voit clairement qu'on ne peut pas juger de ce qui doit être compté positivement ou négativement du point de vue de l'organisation – c'est-à-dire de ce qui peut être appelé progrès dans l'organisation ou au contraire désorganisation, désordre, perturbation – sans faire intervenir une notion de *valeur* mesurée par une *efficacité fonctionnelle*. Cette efficacité fonctionnelle est, dans les organismes vivants, au service de la survie et de la reproduction. Or Atlan croit pouvoir s'en passer, il veut distinguer le progrès dans l'organisation – ce qu'il appelle « ambiguïté-autonomie » – de la désorganisation et perturbation d'une structure (l'« ambiguïté destructrice ») en ne recourant qu'à des mesures d'entropie-variété laissant de côté la valeur ou l'efficacité fonctionnelle: l'ambiguïté serait négative lorsqu'on considère la transmission d'information d'un sous-système à un autre; elle deviendrait positive lorsqu'on considère le système entier et la transmission d'information de ce système à l'observateur. Qui pourra croire qu'on peut obtenir qu'une mutation défavorable pour le fonctionnement de l'organisme (donc génératrice de désordre) se mette tout à coup à lui devenir favorable (donc à engendrer de l'ordre) simplement parce que l'observateur considère l'ensemble plutôt que l'interaction entre les parties?

APPLICATION DE LA THÉORIE DE L'INFORMATION À DES STRUCTURES

Prenons maintenant un peu de recul et revenons à des considérations plus générales.

a) L'interprétation objectiviste de la théorie de Shannon, le passage de l'entropie-variété à la spécificité et la distinction entre la spécificité du message et l'information éventuelle dont il peut être porteur ouvrent la voie à une application plus large de la notion de spécificité: rien n'empêche de l'utiliser pour d'autres structures spécifiques que des messages.

Par exemple pour des structures fonctionnelles. Supposons qu'on dispose d'un certain nombre d'éléments discrets qu'on puisse assembler d'un grand nombre de manières différentes et que seuls certains de ces assemblages soient capables de certaines performances. On peut alors estimer la spécificité de l'ensemble des structures possédant ces capacités par rapport à l'ensemble des structures constructibles avec ces éléments.

Or cette supposition est précisément satisfaite par les protéines, qui sont toutes formées par l'enchaînement de quelques vingt acides aminés, ainsi que par les ADN et ARN qui portent une suite bien déterminée de triplets de quatre bases puriques ou pyrimidiques.

Certains se sont insurgés contre l'utilisation en génétique des termes d'*information* ou de *message*. Il est vrai qu'il n'y a pas à proprement parler d'expéditeur envoyant un message à un récepteur. Il est également vrai que, si l'on interprète l'information selon le mode subjectiviste, il n'y a personne qui serait incertain et dont l'incertitude serait diminuée par le *message* génétique.

Par contre, dans l'interprétation objectiviste (statique) et surtout si l'on remplace le concept d'*information* par celui de *spécificité*, toute difficulté disparaît: il y a bien une structure spécifique, un ordre qui est copié, transcrit, traduit et qui doit l'être fidèlement si l'on veut éviter que les performances et les capacités fonctionnelles de la nouvelle cellule ne se trouvent altérées. Et cette fidélité pose les mêmes problèmes que celle de la transmission d'un message d'un émetteur à un récepteur; elle peut être décrite dans les mêmes termes. On peut donc dire que l'application du concept d'*information* à la génétique en supposait une interprétation objectiviste, en dehors de laquelle elle perdrait tout sens.

Cette manière de traiter la quantité d'information – plus précisément: la spécificité – de structures complexes envisagées statiquement, c'est-à-dire indépendamment de toute transmission, évite toutes les difficultés rencontrées par Atlan et les auteurs dont il s'inspire pour leur appliquer la théorie de l'information :

« [...] l'application de la théorie de l'information à l'analyse des systèmes implique un glissement de la notion d'information transmise à celle d'information contenue dans un système organisé. » (C&F, p. 46).

Comme nous l'avons vu, pour justifier ce glissement, il « assimile, au moins implicitement, la structure du système à un message transmis dans une voie qui part du système et aboutit à l'observateur ». Mais il n'est pas au bout de ses peines. Car, étant donné qu'il considère « chaque élément ordonné ou organisé (macromolécules, organelles, cellules, organismes) comme la *sortie* d'une voie de communication », il se pose le problème de l'*entrée*, de la source. Il voudrait bien la situer au niveau de l'ADN, au niveau du programme, mais, devant certaines difficultés, il y renonce et déclare qu'on n'a pas besoin de spécifier la source. Pourtant, cette solution ne lui paraît pas entièrement satisfaisante parce que si l'on veut parler d'un bruit, de perturbations qui affectent le système, il faut bien qu'il y ait quelque chose qui soit perturbé :

« En effet, si on peut concevoir que des facteurs de bruit peuvent augmenter l'information transmise, comment envisager qu'ils puissent en même temps détruire cette information *alors même que le message d'entrée n'est pas défini ?* La destruction d'information transmise dans une voie implique que le message de sortie n'est pas conforme à 100% au message d'entrée. Encore faut-il qu'il existe toujours un message d'entrée « idéal » par rapport auquel une modification est « mauvaise ». Or ceci est apparemment contradictoire avec le fait que le message d'entrée n'est pas défini. »

Atlan imagine alors une solution qui, bien qu'elle puisse convenir pour une théorie du vieillissement, me paraît très peu satisfaisante :

« La difficulté peut être levée si on considère ce message idéal comme simplement l'arrangement du système à l'instant $t - dt$, l'observation étant effectuée au temps t. Autrement dit, par rapport à ce qu'il était au temps $t - dt$, les erreurs accumulées pendant l'intervalle dt ont produit une ambiguïté destructrice. » (OBThI, p. 257).

Un peu plus haut, il envisageait une autre solution, à mon avis plus satisfaisante, qui consistait à prendre comme « message idéal » un message de sortie supposé « en tous points conforme au message d'entrée *quel qu'il ait pu être* », c'est-à-dire un message de sortie supposé non perturbé.

On évite toutes ces difficultés et les contorsions nécessaires pour les résoudre en définissant non pas une quantité d'information transmise sur une voie, mais un contenu statique d'information mesuré par la spécificité du message (parmi ses pairs). On n'a plus alors à transformer de façon artificielle une structure en un message transmis, puisqu'on a une théorie *qui s'applique directement et naturellement aux structures*, considérant les messages comme des cas particuliers de structures spécifiques.

LE DIFFUS ET LE POINTU

b) On peut, très généralement, définir deux pôles : un pôle « diffus » et un pôle « pointu » (cf. tableau ci-dessous).

Il y a deux manières d'aborder la théorie de la communication. L'une correspond au point de vue de l'utilisateur qui choisit *un* message bien spécifique et qui veut que ce soit *ce* message qui soit transmis au destinataire qu'il a également désigné, et non à quelqu'un d'autre : la spécificité de l'adressage fait partie intégrante de la spécificité du message. C'est ce qu'écrit Shannon au début de son article, dans le passage que nous avons déjà cité : « Le problème fondamental de la communication est celui de reproduire

Tableau

Pôle diffus	Pôle pointu
Statistique, probabilités sur un ensemble de systèmes, de messages	système ou message individuel
indéterminisme, caractère aléatoire	déterminisme
variété d'un ensemble variété d'un type	choix d'un élément ou d'un sous-ensemble
entropie de l'ensemble de référence	entropie d'un sous-ensemble
entropie-variété	spécificité = différence d'entropies
complexité, richesse	simplicité, redondance, réductibilité à un (à quelques) principe(s) simple(s)
hasard, incertitude	certitude, information
bruit, perturbation	précision, fidélité
désordre	ordre
variation	sélection

exactement ou approximativement en un point un message *sélectionné* en un autre point. [...] L'aspect significatif est que le message en question est *un message choisi parmi un ensemble de messages possibles.* »

De ce point de vue, on souligne donc le pôle pointu : la précision, la spécificité, la sélection, le fait qu'*un* message soit sélectionné.

L'autre manière d'aborder le problème de la communication est celui de l'ingénieur : il a un ensemble de messages d'un certain type à transmettre et il ne s'intéresse pas du tout au message particulier. Mieux encore, il préférerait, pour que l'expérience ne soit pas perturbée par des inhomogénéités accidentelles, avoir affaire à des suites aléatoires de signes produites selon certaines règles et imitant plus ou moins bien des messages écrits dans telle langue naturelle (en tenant compte de la fréquence des lettres et de certaines des liaisons). C'est dans ce contexte qu'apparaissent des probabilités et des entropies-variété d'ensembles calculés à partir de probabilités. C'est également dans ce contexte que Shannon peut écrire : « Si une ligne avec bruit est alimentée par une source, deux processus statistiques interviennent : la source et la ligne ». Ici, les messages sont considérés comme produits par une source de préférence aléatoire (« *a statistical process* ») qui est mise sur le même plan que le bruit, et ces deux processus statistiques conjuguent leurs effets selon le théorème des probabilités composées. Il est presque superflu de souligner que cet autre point de vue relève entièrement du pôle diffus.

Pour des raisons diverses (prédominance des ingénieurs en télécommunication dans le développement et l'utilisation de la théorie, identification de la quantité d'information et de l'entropie – qui, en physique, mesure le désordre moléculaire –, méconnaissance de la grandeur *spécificité*), le pôle diffus a largement dominé la théorie de l'information et son interprétation. Et l'information, qui appartient au pôle pointu (elle restreint l'incertitude), a été assimilée à l'entropie, à la variété et même à l'incertitude qui appartiennent, elles, au pôle diffus. Ce qui a créé une grande confusion conceptuelle : l'information est identifiée avec ce qu'elle est censée réduire et combattre, le bruit semble augmenter l'information et on propose même une théorie de l'ordre par le bruit – autrement dit de l'ordre par le désordre, du pointu produit par du diffus.

Atlan s'est lui aussi aperçu de cette bipolarité contradictoire :

« Pour l'instant, remarquons seulement que ces deux définitions [de l'ordre, par l'entropie et par la redondance] sont contradictoires : pour la première, un degré d'ordre et d'organisation élevé correspond à une quantité d'information élevée, et pour la deuxième à une redondance très grande, donc à une quantité d'information très petite. Le fait que toutes deux aient pu être proposées et justifiées montre le caractère ambigu de la notion d'organisation tirée tantôt vers la régularité et la répétition, tantôt vers la variété et l'hétérogène. »

La mesure de l'ordre par la spécificité, qui est une différence de deux entropies, éclaire cette apparente contradiction : la variété (le pôle diffus) concerne la première des deux entropies, qui doit être aussi grande que possible ; le pôle pointu concerne la seconde entropie, qui doit être aussi petite que possible.

La méconnaissance du pôle pointu a aussi amené à attribuer une valeur à la complexité en elle-même – alors qu'elle a en général une valeur négative : elle introduit une certaine fragilité, sans parler des difficultés accrues de production. La complexité n'a de valeur que lorsqu'elle est le prix à payer pour une performance fonctionnelle irréalisable sans elle (c'est-à-dire lorsque, comme le dit von Neumann, elle est seule capable de « faire des choses très difficiles et très élaborées »).

C'est pourquoi le fait – passé largement inaperçu – que la transmission avec bruit a obligé Shannon à abandonner l'identification de la quantité d'information avec une simple entropie et à définir une nouvelle grandeur qui, elle, replace la notion d'information sur le versant pointu où elle retrouve son lieu naturel, ce fait revêt une certaine importance : il est toujours utile de disposer de concepts adéquats, qui ne soulèvent pas trop de problèmes ou d'interrogations, et un trop grand écart entre le modèle et le domaine qu'il est censé représenter laisse le champ libre à des interprétations peu rigoureuses, tout d'abord nécessaires pour combler l'écart, mais qui risquent bientôt d'échapper à tout contrôle.

L'ORDRE PAR LA SÉLECTION

Reste à expliquer comment de l'ordre peut naître du désordre, de la diversité, par exemple dans l'évolution des êtres vivants.

Ma réponse, c'est que le hasard ne crée que du désordre, en ce sens que, dans leur écrasante majorité, les structures produites par un processus aléatoire à partir d'une structure donnée ont une valeur (par exemple fonctionnelle) inférieure à celle de la structure initiale. Il crée en outre du désordre en cet autre sens que les très rares ordres éventuellement produits sont noyés dans un énorme fatras de structures désordonnées : il faudra donc les identifier, les isoler, les sélectionner et il faudra avoir prélablement défini selon quels critères on les reconnaîtra, lesquelles seront considérées comme étant « en ordre ».

Je dirais donc que la variété et le bruit ne créent que du désordre, mais qu'ils peuvent fournir une matière première à la *sélection* qui, elle, en trouvant la pépite d'ordre parmi tout ce désordre, crée de la spécificité et donc de l'ordre.

Alors que l'interprétation de la quantité d'information ou de l'ordre comme une entropie-variété suggérait une genèse de l'ordre à partir du bruit (qui introduit de l'entropie-variété), l'interprétation de l'ordre comme une spécificité suggère au contraire une origine de l'ordre par la sélection d'un sous-ensemble de structures ordonnées parmi un ensemble plus varié. Le bruit augmente l'entropie de l'ensemble de référence, la sélection diminue celle du sous-ensemble choisi. (Et lorsque je parle de sélection, je n'évoque pas la seule sélection naturelle, mais aussi la sélection critique des idées, technologiques ou artistiques et même celle dont parle Shannon lorsqu'il dit qu'un message est *choisi* parmi un ensemble de messages).

Il n'y a donc pas d'ordre par le bruit, il y a un ordre par la sélection. Atlan semble s'en être lui aussi aperçu : il parle, à propos des petits aimants de von Foerster, de « sélection parmi les composantes du bruit » (OBThI, p. 248), et de « sélection des facteurs de bruit par une structure d'accueil asservissante ». « Seules sont sélectionnées les composantes de bruit efficaces » (p. 249).

Cette réponse – outre qu'elle ne me semble pas solliciter les faits – a l'avantage de ramener sans équivoque l'ordre sur le versant pointu, le bruit et le désordre sur le versant diffus.

DISTINCTION INFORMATION/SPÉCIFICITÉ ET RELATIVITÉ DE LA SPÉCIFICITÉ

c) Il ne faudrait pas cependant conclure hâtivement que, puisqu'on a rapproché la définition technique de la notion intuitive, les deux notions peuvent maintenant être confondues : on peut toujours transmettre fidèlement des « messages » qui n'ont aucune signification ou dont aucun récepteur ne prend connaissance. Il faut donc distinguer la transmission de la spécificité d'un message, de la transmission d'une information. Ce qu'on peut pourtant affirmer, c'est que :

– dans les cas pas trop artificiels, où la spécificité du message est effectivement utilisée pour transmettre une information, il y a un parallélisme satisfaisant,

– l'information éventuelle contenue dans un message est liée à la spécificité du message. Plus cette information est grande, plus spécifique est le message qu'elle exige. Et si le message perd sa spécificité, il perd du même coup l'information qu'il porte.

La spécificité du message est donc une condition nécessaire pour qu'il puisse être porteur d'une information, mais elle n'en est pas une condition suffisante : il ne suffit pas qu'un message soit spécifique pour qu'il soit porteur d'information, et il ne suffit pas que sa spécificité soit conservée par la transmission pour que de l'information soit effectivement transmise.

d) Attention ! L'objectivation de la notion d'information ne représente pas pour autant un dépassement de sa relativité. Bien au contraire, elle met celle-ci en évidence : on ne peut parler de spécificité que lorsqu'on a défini le sous-ensemble auquel on veut l'appliquer. De même qu'on ne peut parler d'ordre et de désordre que lorsqu'on a caractérisé ce qui est considéré comme « ordre ». En général cet ordre est défini à partir d'une certaine valeur : valeur informationnelle, valeur [valeur] fonctionnelle (et pour quelle fonction, dans le cadre de quel tout ?), valeur esthé [tique], etc. Et même spécifiée ainsi, la valeur ne définit pas univoquement le sous-ensemble : un message peut avoir une valeur informationnelle pour son destinataire et n'en avoir aucune pour un autre qui ne la comprend pas, n'en connaît pas le contexte ou, tout simplement, ne s'y intéresse pas.

La spécificité peut de même dépendre de ce qu'on adopte comme référentiel : un message en anglais n'a pas la même spécificité par rapport à l'ensemble des messages en anglais que par rapport à toutes les suites aléatoires des signes qui permettent d'écrire un texte anglais.

Par contre l'entropie physique n'est pas relative ; depuis Nernst et son troisième principe de la thermodynamique, qui fixe le zéro de l'entropie au zéro absolu de la température, depuis Planck, qui a fixé la grandeur des cellules de l'espace de phase, l'entropie d'un système physique est univoquement déterminée. Et l'exemple d'Atlan, où certains atomes d'un gaz sont des isotopes, n'infirme pas ce caractère absolu : on peut toujours se tromper sur la valeur d'une grandeur lorsqu'on ne dispose pas de toutes les données nécessaires à sa détermination.

DEUX SORTES D'ORDRE

Revenons sur le problème posé par les aimants de von Foerster, qui devraient démontrer qu'une agitation désordonnée peut créer de l'ordre. Cet exemple me donnera l'occasion d'utiliser la notion de « spécificité » pour distinguer deux sortes d'ordre.

On se souvient du début de *Le hasard et la nécessité*, où Jacques Monod essaye de caractériser les structures vivantes ou celles qui résultent de l'activité fabricatrice d'êtres intelligents. Il rencontre alors une difficulté, celle de les différencier de structures également « ordonnées » : les cristaux. Qu'est-ce qui distingue l'ordre d'un cristal de l'ordre d'un être vivant ou de celui d'une structure artificielle telle qu'un appareil photographique ?

Voilà ce que je répondrais : c'est que dans le cristal interviennent des forces et que ces forces déterminent un équilibre plus ou moins unique, correspondant à une énergie potentielle minimale, et dans lequel le système se met spontanément (ou par une action qui n'est pas dirigée en fonction de la structure à construire).

Si l'on place de l'eau et de la vapeur d'eau dans un ballon fermé, le système s'« ordonnera » dans un champ de gravitation avec l'eau en bas, limitée par une surface plane horizontale, et la vapeur d'eau au-desus. C'est un ordre, mais ce n'est pas un ordre spécifique, un ordre informationnel : ce type d'ordre n'est pas propre à un sous-ensemble spécifique parmi d'autres systèmes qui n'auraient pas cette structure. En effet, *tous* les systèmes analogues, mis dans une telle situation, présentent spontanément cette même structure. La spécificité de ce type d'ordre est donc nulle : le sous-ensemble coïncide avec le référentiel. Il s'agit d'un ordre obligatoire, sans choix, donc incapable de porter la moindre information.

On peut aussi présenter les choses autrement : Boltzmann répartit les états d'un gaz qui comprend n molécules dans un espace de phase à $6n$ dimensions (3 pour la position et trois pour les composantes de la vitesse de chaque molécule). Il divise cet espace de phase en cellules, identifiant les états où il y a le même nombre d'états dans les mêmes cellules, négligeant donc la position du point-état à l'intérieur de la cellule. Il compte alors le nombre de tels états microscopiques qui correspondent à tel état macroscopique, défini par des grandeurs macroscopiques telles que volume à disposition, énergie interne, nombre de moles. Il y a donc une contrainte sur les états microscopiques : ils doivent tous avoir la même énergie interne (qui se réduit, pour un gaz parfait hors de tout champ, à l'énergie cinétique des molécules).

Mais dès qu'interviennent des forces (de gravitation ou d'attraction-répulsion entre les molécules), l'énergie interne n'est plus uniquement cinétique, elle est en partie potentielle. Par exemple, dans un gaz en équilibre dans un champ de gravitation, la densité et la température (énergie interne cinétique) sont inégalement réparties selon la hauteur : une molécule qui monte perd de l'énergie cinétique et acquiert de l'énergie potentielle ; lorsqu'elle a épuisé son énergie cinétique, elle retombe. Cette part variable d'énergie

potentielle change évidemment la répartition des états et des cellules dans l'espace de phase.

Ce n'est que dans le premier cas (absence de champs) que l'état d'entropie maximale correspond à un désordre homogène. Dès qu'il y a des champs, l'état d'entropie maximale n'est plus homogène, il est structuré par les champs.

Il ne faut donc pas juger la structure due aux champs comme étant improbable en la référant à la répartition des cellules en l'absence de champs. Cette structure due aux champs n'est comparable qu'aux structures de même genre imposées par les mêmes conditions. Dans ces circonstances, l'état d'entropie ou de désordre maximaux est partiellement « en ordre » en ce sens que la répartition des molécules n'est pas spatialement homogène, qu'il y a des endroits privilégiés où les molécules se placeront plus volontiers.

Il y a donc deux sortes d'ordres, qui ont des caractères différents :

– des ordres spontanés, obligatoires, dus aux champs (ordre par gravitation, ordre cristallin, etc.) ;

– des ordres spécifiques qui, parmi toutes les configurations physiquement possibles (donc éventuellement aussi parmi tous les ordres spontanés) en privilégient certains qui ne sont en rien favorisés par l'évolution physique, qui présenteraient la même stabilité s'ils étaient différents. Ces ordres ne sont pas imposés par des contraintes internes, ils sont arrangés ou sélectionnés en fonction de critères externes (efficacité fonctionnelle, valeur informationnelle ou esthétique, etc.). Par exemple, la forme d'une bouteille ou d'une clé ne résulte pas de contraintes internes de la matière dont elles sont faites, mais de la volonté de leur fabricant de leur donner cette forme particulière, parmi d'autres formes également possibles. De même, la structure d'un être vivant, déterminée par son programme génétique, pourrait physiquement et chimiquement être toute différente, mais le programme a été sélectionné, par suite de l'efficacité fonctionnelle de l'ensemble, parmi d'autres tout aussi possibles.

Les cristaux appartiennent évidemment à la première catégorie : dans des circonstances favorables, les atomes se placent spontanément selon les mailles d'un réseau cristallin ; cet état correspond à un état d'équilibre, avec transformation d'énergie potentielle en chaleur (énergie de liaison, chaleur dégagée lors de la solidification, « travail interne » de Clausius). A cette chaleur cédée par le corps lors de la solidification correspond certes une diminution d'entropie. Mais cette diminution traduit la perte de l'énergie de liaison (il suffit de diviser celle-ci par la température à laquelle se fait la solidification) plutôt que l'arrangement régulier des atomes ; elle se produit également lorsqu'on passe de l'état gazeux à l'état liquide, dans lequel les atomes ne sont pourtant pas ordonnés.

Selon l'interprétation statistique de l'entropie, sa diminution correspond à une réduction du nombre des états microscopiques compatibles avec l'état macroscopique, mais cette diminution ne signifie pas nécessairement que les états microscopiques sont plus ordonnés au sens d'un ordre régulier, elle traduit simplement les contraintes nouvelles exercées par les forces internes. Et on peut se demander si le terme de *disgrégation* proposé par Clausius pour interpréter la partie de l'entropie qui est liée à l'arrangement des molécules, ne prêterait pas moins à confusion que celui de *désordre* utilisé habituellement.

Avec les cristaux, on n'a pas affaire à un ordre spécifique : tous les systèmes se mettent sous cette forme et l'ensemble des cristaux coïncide avec le référentiel. Il n'y a donc pas de spécificité de l'ordre cristallin, celui-ci ne fait que résulter des liaisons entre les atomes (contrairement à l'ADN qui est, comme le dit Schrödinger dans *What is life?*, un cristal *apériodique*, donc non répétitif, non redondant, où l'ordre des bases n'est pas spontané).

Il en est bien sûr de même avec les aimants de von Foerster : l'agitation ne fait que permettre à liaisons spontanées de se manifester et l'« ordre » qui se constitue est dû non pas à l'agitation, mais *aux liaisons entre les aimants.* Il n'y a donc pas d'ordre par le bruit, il y a un ordre par les liaisons magnétiques, et cet ordre n'est pas de nature informationnelle, c'est un ordre spontané, latent, qui n'est que révélé par l'agitation (s'il ne se manifestait pas avant, c'était à cause des frottements et de la gravitation). Atlan ne dit d'ailleurs pas autre chose :

« En fait, [ce modèle] montre qu'il n'y a pas de véritable auto-organisation d'un point de vue physique, en ce sens que les états de structure successifs sont potentiellement contenus dans la magnétisation initiale des cubes et dans l'état d'énergie libre minimum qui en résulte. » (OBThI, p. 248).

En fait, les aimants de von Foerster sont presque l'exact équivalent des « spectres » magnétiques. Lorsqu'on saupoudre de limaille de fer une plaque et qu'on approche celle-ci des pôles d'un aimant, il ne se passe rien : la limaille reste en place à cause du frottement. Ce n'est qu'en tapotant sur la plaque qu'on voit se former le spectre qui dessine les lignes de forces du champ. Dira-t-on que cet ordre a été créé par les secousses imprimées à la plaque ? – Ce serait ridicule, l'ordre est évidemment dû au champ. Il n'y a pas plus d'ordre par le bruit dans les aimants de von Foerster que dans l'expérience du spectre magnétique ; la seule différence, c'est que dans le premier cas, le champ est créé par les aimants eux-mêmes alors que dans le second, il est créé par un aimant extérieur à la limaille.

Cette différence fondamentale entre les ordres spontanés et les ordres arbitraires condamne toute tentative d'explication de la vie, de l'organisation biologique, qui est un ordre arbitraire, par des ordres spontanés tels que les cellules de Bénard, ordres qui, même s'ils présentent une certaine variabilité, sont dus à des effets internes à la matière (différences de densité, viscosité) qui s'organise spontanément de cette manière. Dans les ordres biologiques (chaînes d'ADN et protéines), la matière accepte n'importe quel ordre, elle est indifférente à l'ordre, comme le montrent bien les manipulations génétiques : on peut mettre n'importe quelle suite de bases ou n'importe quel acide aminé n'importe où. Les forces internes à la matière ne privilégient d'aucune manière les ordres fonctionnels, au service de l'organisme et favorisant la survie, ceci même si l'on place cette matière dans un état situé loin de l'équilibre.

RÉSUMÉ SOUS FORME DE THÈSES

1. La quantité d'information ne se mesure ni par une entropie, ni par une néguentropie, mais par une différence de deux entropies.

2. Elle peut être interprétée comme la spécificité d'un sous-ensemble par rapport à un ensemble de référence. Cette notion est applicable en dehors du domaine de l'information.

3. Alors que l'interprétation courante associe l'information à l'entropie, donc à la variété, à la complexité, à l'incertitude, au désordre et même au bruit, créant ainsi une grande confusion conceptuelle, l'interprétation proposée la replace sur son lieu naturel, qui est celui de la spécificité, de l'ordre et de la détermination.

4. Cette interprétation évite, sans devoir recourir à aucun artifice, la plupart des difficultés auxquelles se heurte la conception habituelle et elle permet d'écarter les objections faites à l'application de la théorie de l'information à des structures organisées (par exemple en biologie).

5. Le bruit n'est jamais organisateur : il n'y a pas d'ordre par le bruit, il y a un ordre par la sélection. Quant au « changement de signe de la fonction ambiguïté », qui serait dû à un changement de point de vue, il repose essentiellement sur une confusion de l'équivoque avec ce qu'Atlan appelle l'« ambiguïté ». Or il s'agit de deux entropies distinctes, qui n'ont généralement pas la même valeur.

Bibliographie

H. ATLAN, *L'organisation biologique et la théorie de l'information*, Hermann, Paris, 1972.

H. ATLAN, *Entre le cristal et la fumée*, Seuil, Paris, 1972.

H. ATLAN *et al.*, *Création et désordre*, ouvrage collectif préfacé par M. Cazenave, L'Originel/Radio France, Paris, 1987.

F. BONSACK, *Information, Thermodynamique, Vie et Pensée*, Gauthier-Villars, Paris, 1961. Sera republié en 1996 aux Éditions de l'Aire, Vevey (Suise).

J. MONOD, *Le hasard et la nécessité*, Seuil, Paris, 1970.

E. SCHRÖDINGER, *What is Life ? : the physical aspect of the living cell*, Cambridge Univ. Press, Cambridge, 1945.

C. E. SHANNON, A Mathematical Theory of Communication, *Bell Syst. Techn. J.*, vol. 27, 1948, pp. 379-423 et 623-656.

C. E. SHANNON and W. WEAVER, *The Mathematical Theory of Communication*, Univ. of Illinois Press, Urbana, 1949.

MISSION POSSIBLE

Pierre CALAME

Paris, Éditions Lieu Commun, 1993

Préface d'Edgard PISANI

Curieux ouvrage que ce livre au champ gigantesque puisqu'il traite de la plupart des grands sujets de l'heure : de la défense de l'environnement à la réforme de l'État, en passant par la paix, le rôle de la science... L'auteur lui-même présente un cursus pour le moins particulier : appartenant à un grand corps technique de l'État, il a surtout œuvré dans les études et le social (au sens large) ; actuellement, il préside une fondation d'origine française mais de droit suisse qui appuie des actions de types très divers, sur quatre continents.

Une approche pragmatiquement systémique

Ce qui attirera en premier l'attention d'un lecteur de la *Revue Internationale de Systémique* dans cet ouvrage, c'est son approche résolument systémique. Pierre Calame considère en effet les objets auxquels il s'intéresse comme les systèmes éco-socio-techniques et applique avec pertinence à leur examen toutes les ressources qu'offre cette approche qu'il veut globale, transnationale et transdisciplinaire. Quelques exemples : il insiste sur la variété des représentations possibles du même objet en fonction de la situation et de la personnalité de l'observateur ; il affirme nécessaire d'admettre le flou et l'incertain ; il recourt largement aux « autos » : auto-analyse, auto-régulation, auto-évaluation et évidemment autonomie.

Plutôt qu'un résumé, donnons quelques citations : « assumer la complexité, la saisir à bras le corps » ; « il faut en finir avec la dictature du simple » ; « une société concrète, complexe, diverse » ; l'intelligibilité de l'ensemble doit prendre le pas sur la mesure des parties » ; c'est moins l'abondance de l'information qui compte que sa sélectivité, sa crédibilité, sa spécificité et sa durée ».

Une éthique exigeante

Cette solide méthodologie est mise au service d'un système de valeurs qui doit permettre de s'attaquer à sept défis majeurs : l'avenir de la planète, les rapports entre science et progrès, la croissance de l'exclusion sociale, la difficile construction de la paix, le poids des lo-

Rev. intern. systémique. 0980-1472 Vol. 9/95/01/ $ 4.00/© Afcet Gauthier-Villars