

REVUE DE STATISTIQUE APPLIQUÉE

J. OSWALD

Essai sur la théorie de l'information (logodynamique statistique)

Revue de statistique appliquée, tome 20, n° 1 (1972), p. 25-65

http://www.numdam.org/item?id=RSA_1972__20_1_25_0

© Société française de statistique, 1972, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ESSAI SUR LA THÉORIE DE L'INFORMATION (Logodynamique Statistique)

J. OSWALD

Directeur des Laboratoires de MARCOUSSIS (C.G.E.)

SOMMAIRE

	Pages
AVANT-PROPOS	27
CHAPITRE I DEFINITIONS FONDAMENTALES RELATIVES A L'INFORMATION.....	30
Information et états discernables.....	30
Quantité d'information ou énergie diacritique.....	31
Message et travail sémiodynamique.....	33
CHAPITRE II ENERGIE DIACRITIQUE MOYENNE D'UN SYSTEME ET D'UN MESSAGE.....	38
Pression diacritique moyenne d'un système.....	38
Puissance diacritique d'un message - Théorème de Shannon.....	40
CHAPITRE III RELATIONS ENTRE ENERGIE DIACRITIQUE ET TRAVAIL SEMIODYNAMIQUE.....	45
Relation fondamentale.....	45
Théorème.....	50
Problème de Shannon.....	51
Comparaison avec la Thermodynamique.....	53
CHAPITRE IV DEVELOPPEMENT ET EXEMPLES SIMPLES....	55
Groupement ou "multiplexage" des Informations...	55
Premier exemple d'une distribution simple.....	56
Deuxième exemple élémentaire.....	59
Troisième exemple : Alphabet Morse.....	60
CHAPITRE V INFORMATION INCERTAINE : CORRECTION ET DETECTION DES ERREURS.....	
Correction des erreurs.....	
Détection des erreurs.....	

CHAPITRE VI	APPLICATION DE LA THEORIE DES ERREURS AUX MESSAGES BINAIRES.....	
	Système à deux états.....	
	Système à l états différents.....	
	- Définition des erreurs.....	
	- Correction des erreurs.....	
	- Détection des erreurs.....	
ANNEXE I	MAXIMUM DE $H = \sum p_i \log \frac{1}{p_i}$.	
ANNEXE II	DEMONSTRATION DE L'INEGALITE $\mu_0 < 1$	
ANNEXE III	ADAPTATION DU MESSAGE AU SYSTEME LORS- QUE LE NOMBRE DE SYMBOLES N'EST PAS IMPOSE.....	
ANNEXE IV	MAXIMUM DE $H'_i = \sum_{ij} p_i \pi_{ij} \log \frac{\sum_i p_i \pi_i}{p_i \pi_{ij}}$	
ANNEXE V	DETERMINATION DU GRADIENT SEMIODYNAMI- QUE.....	
BIBLIOGRAPHIE.....		

AVANT - PROPOS

C'est courir un grand risque que de créer de nouveaux termes, surtout pour remplacer ceux que l'usage a déjà consacrés. Les savants ne s'en privent pas, mais il faut bien reconnaître que l'introduction d'un mot nouveau peut trouver sa justification dans l'élaboration d'un nouveau concept, ou d'un objet encore inconnu. S'il s'agit au contraire d'un sujet largement traité par des auteurs de premier plan, il convient d'accueillir avec méfiance les néologismes proposés par un quelconque ouvrier de la onzième heure.

La Théorie de l'Information a quarante ans d'âge, si l'on en attribue - apparemment en toute justice - la paternité à Hartley [1]. Elle est majeure depuis plus de vingt ans, grâce à Shannon et à quelques autres. Peut-on encore lui apporter quelque chose d'essentiel ? Il est permis d'en douter. On voudra bien cependant pardonner à l'auteur d'y avoir introduit quelques mots nouveaux, quelquefois pour remplacer ceux qui sont en usage.

Ce n'est ni un bouleversement des concepts, ni une théorie nouvelle, que le lecteur trouvera dans les pages qui suivent. Il s'agit d'un essai, qui prend son départ dans une réflexion sur l'analogie formelle entre l'Energétique et la Théorie de l'Information. L'une comme l'autre décrivent à grands traits des phénomènes très généraux, en n'examinant que le bilan des échanges entre systèmes, d'où résultent diverses transformations de ces grandeurs que sont l'énergie d'une part, l'information d'autre part.

Dès l'origine, la similitude entre les deux théories s'est imposée aux créateurs de la science de l'Information. Ce n'est pas par hasard, ou par caprice, qu'ils ont introduit des termes bien connus des physiciens, telle que l'entropie, pour désigner la quantité d'information produite par une source : c'est bien parce que l'analogie formelle entre les deux grandeurs - j'entends par là la similitude des expressions mathématiques - leur apparaissait avec une évidence totale. Mais sont-ils allés jusqu'au bout de cette "équivalence" ? C'est la question que je me suis posée, et il m'a bien fallu répondre par la négative. Le raisonnement suivi par Shannon est celui d'un mathématicien, l'explication, l'exploitation de ses résultats sont des travaux d'ingénieurs, d'experts en Télécommunications. Il manque à tout cela, me semble-t-il, une comparaison permanente avec une théorie déjà bien ancienne, que les physiciens appellent Thermodynamique Statistique.

Le résultat de cette confrontation est extrêmement frappant : les deux théories se superposent presque complètement. Ce sont, mot à mot, formule après formule, les mêmes raisonnements, les mêmes calculs qui sont développés dans les deux cas. Il est alors inévitable qu'on soit tenté de mettre en regard les grandeurs homologues, de vérifier leur signification, de leur donner des noms semblables.

En premier lieu, il m'a paru utile de donner un nom à cette science des rapports entre les "messages" et les "systèmes" qui jouent respectivement les rôles de véhicules et de sources d'information. Les raisonnements

employés sont exactement ceux de la Thermodynamique Statistique : c'est pourquoi Logodynamique Statistique m'a paru un terme approprié, l'analyse portant sur le "langage" (logos) au sens large du terme, et non plus sur la chaleur (ou plus généralement l'énergie). Il faut bien remarquer au passage que la logodynamique ne prête nullement attention aux propriétés physiques des circuits électriques, aux problèmes de bande passante, de spectre utile, etc... Toutes ces questions relèvent de la théorie physique des communications qui doit, à mon sens, être soigneusement distinguée de la théorie abstraite et générale qu'est la logodynamique, exactement comme l'électricité n'est pas l'énergétique, encore que les lois de l'électricité satisfassent bien évidemment aux règles beaucoup plus générales de l'énergétique. La plupart des auteurs ont, me semble-t-il, mélangé quelque peu les deux disciplines, ce qui ne peut être qu'une source de confusion.

Dans l'analyse de l'information, deux grandeurs apparaissent, qui jouent, à l'évidence, les mêmes rôles que le travail et l'énergie en thermodynamique.

La "quantité d'information" liée à la connaissance d'un système n'est rien d'autre que le dénombrement des états discernables : je lui ai donné le nom d'énergie diacritique (ou encore de pression diacritique pour des raisons qui apparaîtront dans la suite). C'est l'analogie formelle qui conduit au choix du mot "énergie", alors que l'adjectif "diacritique" exprime l'idée du dénombrement d'objets ou de situations discernables. La richesse (ou si l'on veut l'encombrement) du message doit être comparée à un travail : c'est effectivement la peine que l'on se donne, le prix que l'on paye la représentation ou la traduction de la suite des configurations du système analysé, généralement pour la transmettre à distance. Je l'ai appelé travail sémiodynamique, parce que l'analogie avec le travail est frappante, et parce que ce "travail" met en jeu des signes ou symboles.

Là encore, il semble qu'une confusion soit faite, par la plupart des auteurs, entre l'énergie diacritique et le travail sémiodynamique. La distinction est essentielle, puisque la dépense réellement faite pour transporter l'information à distance dépend énormément de l'habileté de celui qui s'en charge, de la redondance qu'il donne, volontairement ou non, à son mode d'expression. Mais cette confusion, faite si fréquemment, est très instructive : elle provient de l'équivalence entre "l'énergie" et le "travail", principe aussi bien applicable à la logodynamique qu'à la thermodynamique.

Paradoxalement, j'ai renoncé au mot "entropie", choisi par Shannon pour définir la quantité d'information. Pourtant, il est parfaitement exact que l'expression mathématique de "l'entropie" au sens de Shannon est la même que celle de la grandeur utilisée par les physiciens. On peut aussi mettre en regard des formules semblables, où elles interviennent de la même façon et jouent des rôles homologues. Le choix du mot "énergie diacritique" ou plutôt du mot "pression diacritique" m'a paru préférable, d'abord parce qu'"entropie" évoque, par son étymologie, une notion d'évolution irréversible que l'on ne retrouve pas en logodynamique, ensuite et surtout parce qu'il n'y a aucune différence entre "l'entropie d'une source" au sens de Shannon et ce dénombrement des états discernables qui, en logodynamique, joue exactement le même rôle que l'énergie en thermodynamique. Comme on le verra, les grandeurs de la logodynamique sont des nombres purs, elles n'ont pas de "dimensions", ce qui rend inutiles ou illusoirs certaines dis-

inctions qui, au contraire, jouent un très grand rôle dans les théories physiques.

On rencontrera encore, dans l'exposé, une expression que j'ai appelée "gradient sémiodynamique". Ce terme désigne une grandeur qui joue un rôle homologue à celui de la température en Thermodynamique. Nous verrons qu'il a quelque importance et quelque intérêt pour évaluer la répartition du "travail sémiodynamique" entre les divers symboles qui représentent, dans un message, les divers états du système analysé. Cette notion, n'a, me semble-t-il, été proposée par personne, et c'est une lacune qu'il convient de combler.

Au terme de cette tentative de justification d'une terminologie nouvelle, on ne peut échapper à une question importante : l'analogie entre la logodynamique et la thermodynamique est-elle simplement formelle, ou en réalité beaucoup plus profonde ? N'y a-t-il pas, à défaut d'une véritable identité, une liaison étroite, de caractère physique entre les deux théories ? Ce problème a été traité par beaucoup d'auteurs, et L. Brillouin lui a consacré un livre [III]. Il s'en dégage l'idée que la quantité d'information et l'entropie sont de même nature et s'échangent mutuellement. En réalité, il s'agit d'un problème de couplage, qui est caractéristique de la théorie de la mesure.

Toute mesure physique introduit nécessairement une perturbation dans le système étudié ; dans les cas les plus fréquents, on prélève une certaine quantité d'énergie, et si petite soit-elle, on modifie par là même l'état d'équilibre qu'il avait atteint. Si l'on objecte qu'il suffit de rendre ce couplage extrêmement faible pour réduire à une valeur négligeable ou inappréciable la perturbation apportée, l'argument n'est pas déterminant ; en effet la microphysique montre bien que nous ne pouvons pas réduire cet effet indéfiniment car nous ne pouvons pas effectuer un prélèvement inférieur au quantum d'énergie élémentaire. Or si la "quantité d'information" que nous pouvons rechercher est une grandeur tout à fait abstraite, de nature purement intellectuelle, son acquisition passe, pour tout système physique réel, par l'intermédiaire d'un appareil de mesure qui doit lui être raccordé. On en déduit qu'aucune information ne peut être effectivement acquise sur un système physique sans que ce dernier soit perturbé ; c'est-à-dire sans que son entropie augmente. C'est en ce sens, et en ce sens seulement, qu'il peut y avoir "échange", à défaut d'identité, entre l'entropie et la quantité d'information.

Mais il serait tout à fait erroné d'en déduire que la logodynamique - théorie abstraite, de caractère purement mathématique - pourrait être plus ou moins incorporée dans la thermodynamique. Ce que nous appelons en Physique, chaleur ou travail, ce sont des mesures de phénomènes observables, qui ont des effets que nous pouvons constater par nos sens. Au contraire ce que nous avons appelé énergie diacritique ou travail sémiodynamique, ce sont des dénombrements, des décomptes opérés par un processus purement intellectuel, sans référence précise à un système physique donné. Il faut donc se garder de voir, dans les définitions et les raisonnements qui vont suivre, autre chose qu'une analogie formelle qui est sans doute frappante, mais ne permet en aucune manière d'identifier cette mathématique (élémentaire d'ailleurs) qu'est la logodynamique avec une branche de la Physique.

CHAPITRE I

DEFINITIONS FONDAMENTALES RELATIVES A L'INFORMATION

INFORMATION ET ETATS DISCERNABLES

La notion d'information n'a de sens que si on l'associe à un ensemble d'états ou de situations possibles prédéterminées. Etre informé, c'est connaître, parmi tous les événements qui sont potentiellement concevables, celui qui s'est effectivement produit. Un message n'apporte d'information à son destinataire que s'il lui est intelligible : celui qu'on informe a donc nécessairement la mémoire de tous les événements possibles, et la faculté d'interpréter le message, c'est-à-dire de discerner l'évènement qui lui est rapporté. Nous parlons de mémoire et non d'intelligence, car les machines n'ont que de la mémoire : le cas exceptionnel de l'homme, de sa faculté d'adaptation, de son aptitude à s'instruire, c'est-à-dire à enrichir ses connaissances, de son imagination et de son intelligence - sources de son pouvoir créateur - échappe à cette description.

Il est toutefois commode de se référer à l'exemple de l'homme, au moins qualitativement. La pensée joue le rôle de la source d'information, le langage celui du message. De même que la pensée ne peut se communiquer sans le langage, et ne saurait pour autant être confondue avec lui, la mesure de la quantité d'information est inconcevable sans un support physique auquel nous donnons le nom de message. Il en est d'ailleurs de même dans toute mesure physique : chaque grandeur se manifeste, directement ou indirectement, dans un certain nombre de phénomènes naturels, et la mesure est liée à une expérience qui met en jeu l'un de ces phénomènes.

La théorie moderne des communications, plus généralement appelée théorie de l'Information, et telle qu'elle a été proposée il y a 20 ans par C.E. Shannon [II], fait usage des notions de "source d'information" et de "voie de transmission" (channel). Quoique ce vocabulaire soit assez naturel et très familier aux ingénieurs de télécommunication, nous lui préférons la notion, apparemment plus vague, de "système", et celle de "message" - ce dernier mot étant au demeurant moins ambigu que "voie" ou "canal" qui désignent déjà communément beaucoup d'autres choses. En effet, l'image la plus commode est celle d'un système, d'ailleurs absolument quelconque, susceptible de prendre *l* états différents, auquel on associe un message composé de signes ou symboles caractéristiques de chacun de ces états. Il reste à donner une définition quantitative de l'information permettant d'identifier sans ambiguïté la succession d'états de ce système, ainsi que du coût que représente l'émission du message qui lui correspond (1).

(1) Dans toute cette étude, nous nous contentons d'analyser le cas d'un nombre fini d'états discrets, auquel on cherche généralement à ramener celui plus complexe des grandeurs variant de façon continue.

QUANTITE D'INFORMATION OU ENERGIE DIACRITIQUE

L'information permet de discerner les états du système, c'est-à-dire de connaître la situation dans laquelle il se trouve à tout moment. Il est raisonnable d'admettre que la quantité d'information associée à un symbole est une fonction croissante du nombre d'états différents entre lesquels il permet de choisir, car plus grand est le nombre d'états possibles, plus le processus d'identification de l'état effectivement réalisé est laborieux.

Un raisonnement simple, d'ailleurs classique, permet d'établir la fonction du nombre d'états discernables qui peut être judicieusement associée à la "quantité d'information".

Considérons un système S pouvant prendre l états distincts, et associons un signe ou symbole s_i à chacun de ces états. Nous appellerons $q = f(l)$ la "quantité d'information" correspondant à l'apparition de ce symbole⁽¹⁾. Si l'évolution du système S le conduit à prendre successivement n états différents, il en résultera l'affichage de n symboles successifs, choisis dans l'ensemble (s_1, s_2, \dots, s_l) . Si l'apparition d'un symbole s_k correspond à une quantité d'information $f(l)$, l'apparition successive de deux symboles indépendants s_k et s_l correspond, par une convention dont la raison d'être est évidente, à $f(l) + f(l) = 2 f(l)$ (Principe de l'additivité de l'information).

Or, il est possible d'employer une autre méthode pour identifier l'état du système. Au lieu d'associer à chaque état un symbole déterminé appartenant à l'ensemble (s_1, \dots, s_l) , on peut imaginer un classement a priori des états possibles en deux classes C_1 et C_2 , chacune de ses classes étant elle-même constituée de deux sous-classes, et ainsi de suite. Ce procédé dichotomique permet une identification de proche en proche, comme dans le jeu de société dans lequel on procède par questions et réponses, ces dernières étant restreintes à oui ou à non. Comme $f(l)$ est "l'information" liée à l'identification d'un symbole, et en admettant que chaque classe et chaque sous-classe puissent être divisées en deux parties égales, le processus dichotomique est caractérisé par les relations suivantes, qui tiennent compte du principe d'additivité :

$$\begin{aligned} f(l) &= f(2) + f\left(\frac{l}{2}\right) \\ (1) \quad f\left(\frac{l}{2}\right) &= f(2) + f\left(\frac{l}{4}\right) \\ f\left(\frac{l}{2^{k-1}}\right) &= f(2) + f\left(\frac{l}{2^k}\right) \end{aligned}$$

D'où :

$$(2) \quad f(l) = k f(2) + f\left(\frac{l}{2^k}\right)$$

(1) Nous supposons implicitement que la quantité d'information $f(l)$ ne dépend pas du symbole qui est apparu, ce qui est légitime si les états de système (S) sont équiprobables. Le cas plus complexe des états de probabilités différentes est traité au Chapitre II.

qui s'écrit encore :

$$(3) \quad f(2^k) = k f(2) + f(1)$$

La fonction $f(x)$, qui satisfait à la relation fonctionnelle (3), est manifestement la fonction logarithme de base 2 : $f(x) = \log_2(x)$.

Ceci revient à dire que l'identification d'un état parmi l possibles peut être faite par le nombre m de chiffres binaires le plus voisin, par excès, de $\log_2 l$. Dans ces conditions, l'information liée à l'émission de n symboles successifs sera :

$$(4) \quad Q = n \times m = n \log_2 l = n \frac{\log l}{\log_2 l}$$

en désignant par \log le logarithme naturel.

De même que l'unité de surface, en géométrie, a été liée à la figure la plus simple que l'on puisse imaginer, à savoir le carré, de même l'unité d'information a été rattachée au symbole le plus simple possible, l'élément binaire, qui permet la discrimination, c'est-à-dire le choix le plus sommaire : négation ou affirmation, existence ou inexistence. Dans ces conditions, la base 2 pour les logarithmes s'impose naturellement, et l'élément binaire correspond à la plus petite quantité d'information concevable.

Par seul souci de simplification dans l'écriture des relations mathématiques, nous emploierons pourtant, dans la suite de l'exposé, le logarithme naturel ⁽¹⁾, en écrivant :

$$(5) \quad Q = n \log l$$

Par convention, nous exprimerons Q en "logons" :

$$1 \text{ élément binaire} = 0,69 \text{ logons} \simeq 0,7 \text{ logon.}$$

Il n'y a évidemment aucune difficulté à passer d'une unité à l'autre.

La relation (5) exprime en quelque sorte la définition quantitative de l'information. Il est important de remarquer que Q comporte 2 facteurs n et $\log l$, ce qui suggère une analogie formelle entre l'information et l'énergie. En effet, n peut être considéré comme le facteur appelé quelquefois "d'extensité" ou de "capacité" : il joue le même rôle que la quantité d'électricité dans l'expression de l'énergie électrique, que le volume V dans l'expression pV du travail d'un gaz, que le déplacement dans le travail d'une force, etc. L'expression $\log l$ est au contraire le facteur "d'intensité" (ou d'équilibre) homologue du potentiel électrique, de la pression, de la force, etc. Il est en effet à peu près intuitif que la quantité d'information doit contenir ces deux facteurs : il y a le nombre d'informations partielles recueillies, et il y a la "richesse" qui contient chacune de ces informations partielles, qui est mesurée par le facteur $\log l$.

(1) On pourrait tout aussi bien poser $Q = n \log_2 l = K n \log l$, avec $K^{-1} = \log 2$. Nous allégerons l'écriture en choisissant le logarithme naturel comme unité, ce qui conduit à $K = 1$.

Par définition, le facteur n sera appelé volume d'information, et il est exprimé en nombres purs ; le facteur $\log l$ sera appelé pression diacritique, et la quantité $Q = n \log l$ sera considérée comme une énergie diacritique (1). La pression et l'énergie diacritiques s'expriment, par convention, en "logons" (ou si l'on veut, en éléments binaires).

Ces définitions appellent deux remarques très importantes. La première, c'est que l'énergie diacritique Q dont il vient d'être question caractérise l'information elle-même associée au système (ou à la "source") indépendamment de son support physique ou véhicule constitué par le message. Si ce dernier est indispensable à la manifestation de l'évolution du système étudié, il ne saurait se confondre avec lui, pas plus que la pensée ne peut s'identifier au langage. Nous reviendrons abondamment sur ce point dans la suite de l'exposé.

La deuxième remarque est que les deux facteurs "volume" et "pression diacritique" sont sans dimension. Il en résulte que le volume d'information n'a de signification réelle que si le type d'états discernables est bien défini. On peut parfaitement écrire : $Q = n \log l = \log (l^n) = \log L$; si la succession de n états donnés est considérée comme un événement unique, Q devient la "pression diacritique" $\log L$ associée au nouveau symbole complexe qui lui est affecté et qui peut prendre L valeurs différentes. Il n'y a donc aucune différence de nature entre "pression" et "énergie" diacritiques.

MESSAGE ET TRAVAIL SEMIODYNAMIQUE

Le message, véhicule de l'information, possède une structure plus ou moins complexe ; il peut être limité dans le temps, donc posséder un "format" ou être "fluent", c'est-à-dire sans limitation de durée : mais il possède toujours une certaine organisation.

Nous appellerons élément primaire la plus petite fraction que l'on peut isoler "physiquement" dans le message. Aucun élément identifiable n'a de durée plus courte. Cette entité n'a néanmoins, dans beaucoup de cas, aucune signification propre, car elle ne correspond pas nécessairement à une information utilisable.

Nous désignerons sous le nom de symbole la plus petite fraction du message ayant une signification, c'est-à-dire correspondant à un état du système. La réception du symbole apporte une information, car elle peut être interprétée comme affirmant l'existence d'une situation appartenant au catalogue des situations possibles.

L'ensemble des symboles utilisables constitue l'alphabet.

Les groupements de symboles seront désignés sous des noms divers, tels que "mot", "séquence" ou "message". On réservera les expressions "mot" et "message" aux assemblages ayant une signification, c'est-à-dire appartenant à un ensemble ordonné et cohérent, tandis que "séquence" dé-

(1) Le substantif "diacrisis" signifie en grec : distinction, caractère discernable. C'est bien la signification profonde de la "quantité d'information" qui est une mesure de la possibilité de choix associée à un jeu de symboles.

signera plutôt une suite comportant un nombre arbitraire de symboles, suite qui est souvent dépourvue de signification parce qu'elle échappe à un classement.

Il est important de revenir sur la notion d'alphabet. Si on le suppose constitué de l symboles différents, il est nécessaire de définir :

a) Comment ces symboles se différencient les uns des autres.

b) A quelles règles d'assemblage - ou si l'on veut à quel "langage" ils obéissent.

La différenciation des symboles résulte de la nature des signaux qui leur sont affectés. Les caractéristiques les plus couramment utilisées sont l'amplitude et la durée. Dans le cas, considéré ici, d'un nombre fini d'états discrets, la discrimination par l'amplitude ⁽¹⁾ correspond à la définition de niveaux "quantifiés" discernables les uns des autres. Il faut évidemment envisager au moins deux niveaux différents pour pouvoir distinguer les symboles. Lorsque la discrimination se fait par la durée, on affecte généralement à cette dernière des valeurs multiples d'un intervalle de temps élémentaire, mais cette contrainte n'est pas absolue.

La notion d'intervalle de temps élémentaire τ jouera un rôle important par la suite. On désigne par ce terme le plus petit intervalle significatif, et il dépend, non seulement des symboles, mais de leurs règles d'assemblage. Précisons-le par un exemple :

Considérons un alphabet comportant 3 symboles s_1, s_2, s_3 , dont les durées respectives sont, avec une unité de temps arbitraire, égales à 2, 3 et 5 ; supposons de plus que s_1 a une amplitude égale à zéro, tandis que s_2 et s_3 ont une amplitude égale à 1. Si la règle d'assemblage consiste à toujours faire suivre le symbole s_2 du symbole s_1 , le temps élémentaire est égal à 2, car c'est le plus petit intervalle de temps qu'il faut savoir évaluer pour ne pas confondre s_2 avec s_3 . Si au contraire la règle d'assemblage consiste à prohiber la suite $s_2 s_3$, qui est indiscernable de $s_3 s_2$, mais à admettre la suite $s_2 s_2$, le temps élémentaire a pour valeur 1, puisqu'il faut pouvoir distinguer $s_2 s_2$ de durée 6 de s_3 dont la durée est 5 (Bien entendu, dans un tel cas, il faudra prohiber d'autres combinaisons, comme $s_2 s_2 s_2 s_2 s_2$ qui est indiscernable de $s_3 s_3 s_3$).

On voit donc que le temps élémentaire τ dépend non seulement de la durée des symboles, mais aussi de leurs règles d'assemblage. Lorsque les symboles ne se différencient que par l'amplitude, le temps élémentaire est le plus petit intervalle séparant deux changements de niveau successifs.

Donner un alphabet, c'est donc définir un ensemble de symboles ainsi que leurs règles élémentaires d'assemblage, c'est-à-dire les suites discernables composées à l'aide de ces symboles.

Définition : Etant donné un alphabet et ses règles d'assemblage, le temps élémentaire τ associé à cet alphabet, est le plus petit intervalle de temps qui doit pouvoir être discriminé pour que deux séquences distinctes de même durée soient discernables. L'inverse $w = 1/\tau$ est la "rapidité de modulation" associée à l'alphabet.

(1) Ou encore par la phase, la fréquence, etc...

Nous allons maintenant associer au message une grandeur quantitative qui soit en quelque sorte la mesure de la dépense effectuée en transmettant de l'information. Il est évident que la correspondance entre l'information à transmettre, qui résulte de l'évolution du système (énergie diacritique) et le message choisi, n'est pas stricte. On peut en effet employer un message inutilement long pour communiquer une information très faible : chacun sait qu'un écrivain ou un orateur n'est pas toujours très concis. Donc, le discours ou le message correspondent à un certain "travail", à une certaine "dépense", qu'on ne peut identifier, purement et simplement, avec l'information transmise.

A la suite des travaux de Nyquist et de Hartley[1], il a été reconnu que ce travail pouvait être considéré comme proportionnel au nombre maximal "d'éléments primaires" que comporte le message. Or, l'élément primaire, a , par définition, une durée $\tau = 1/w$; il y a donc w éléments par unité de temps, et le "travail" E accompli pendant un temps T peut se mettre sous la forme :

$$(6) \quad E = K \times wT$$

Nous appellerons E le travail sémiodynamique de la séquence de durée T ⁽¹⁾.

Nous allons expliciter le facteur K qui a été introduit dans la formule (6) ; mais auparavant, faisons une remarque sur la définition du travail sémiodynamique.

Le concept que nous avons introduit ne fait pas intervenir la structure particulière de telle ou telle séquence ou de tel ou tel message : il ne met en jeu que le nombre "d'instantanés élémentaires" contenus dans un intervalle de temps T . En effet, c'est la cadence ou variation de niveau la plus rapide que l'on peut rencontrer dans une longue séquence qui est le critère de "dépense" ou de travail accompli, la durée étant le deuxième facteur à prendre en considération. C'est pourquoi on introduit la rapidité de modulation ou "vitesse de manipulation" w . Il est bien connu que cette vitesse est liée à la bande passante Δf de la voie de communication utilisée : mais la relation entre w et Δf est d'un autre ordre, elle relève, non de ce que nous avons appelé la "logodynamique", mais de la théorie physique de la transmission et de la réception des signaux électriques. Dans un souci de généralité, nous conservons ainsi à la notion de "message" son caractère abstrait, en ne le confondant pas avec le support physique le plus commun qui est la voie de communication. C'est une des raisons pour lesquelles nous avons adopté cette terminologie.

Revenons maintenant au facteur K pour le préciser. Il est commode de pouvoir comparer la grandeur E , que nous avons appelée travail sémiodynamique du message, à la grandeur Q qui est l'énergie diacritique du système. Or, wT est le nombre d'éléments primaires que contient l'intervalle T , et pour que KwT soit de même nature ⁽²⁾ que E , il faut que K puisse

(1) De "seméion", signe. E caractérise essentiellement le nombre maximal de signes ou éléments, de durée minimale τ qui peuvent être mis en jeu pendant un temps T .

(2) Nous parlons de "nature", c'est-à-dire de forme, et non de "dimension", puisque E comme Q sont des nombres purs, dépourvus de dimension.

être considéré comme une pression diacritique, ainsi que le montre la formule (5). Nous désignerons par v la valence de l'émission, c'est-à-dire le nombre de niveaux que l'on peut rencontrer dans l'ensemble des séquences constituées à partir de l'alphabet utilisé, et nous prendrons, par définition, $K = \log v$. Avec cette convention, l'expression du travail sémiodynamique E d'une séquence de durée T a pour valeur :

$$(7) \quad E = wT \log v$$

Ce choix va être justifié par un exemple particulièrement simple, qui met en évidence l'équivalence de l'énergie diacritique et du travail sémiodynamique.

Considérons un système (S) pouvant prendre l états différents, et observons une suite de N états successifs pris par le système. L'énergie diacritique mise en jeu est $N \log l$, qui peut encore s'écrire :

$$(8) \quad N \log l = mN \log 2$$

$$m = \frac{\log l}{\log 2} = \log_2 l$$

est le nombre d'éléments binaires nécessaire pour identifier un état donné parmi l .

A la succession des N états pris par le système, on peut faire correspondre un message binaire, de rapidité de modulation w . Ce message de durée T aura fourni, sans omission ni répétition, l'information relative à la succession des N états de S, s'il comporte un nombre N' d'éléments binaires tel que :

$$(9) \quad N' = wT = mN$$

Donc :

$$(10) \quad N \log l = mN \log 2 = wT \log 2.$$

La formule (10) n'est qu'un cas particulier, obtenu pour $v = 2$ (émission bivalente) de la formule générale (7).

Il y a donc bien "équivalence" entre énergie diacritique et travail sémiodynamique.

Cette équivalence est démontrée dans le cas général, si on observe qu'on peut toujours transformer un message de valence v ($v > 2$) en un message binaire équivalent de rapidité de modulation $w' = w$, en posant :

$$wT \log v = w' T \log 2$$

ou encore :

$$(11) \quad n \log v = n' \log 2$$

Remarquons en outre que, n et n' étant des "facteurs d'extensité" analogues à des volumes, $\log v$ et $\log 2$ des "pressions diacritiques", la formule (11) est formellement semblable à la relation des gas parfaits :

$$p_1 V_1 = p_2 V_2 = C^{te}.$$

Toutes ces analogies entre l'énergie et l'information justifient la terminologie employée ; par surcroît, nous verrons plus loin que les mêmes types de raisonnement interviennent dans l'établissement des relations entre les grandeurs. Elles ne doivent pas néanmoins faire perdre de vue l'aspect purement formel de l'analogie. En effet, en thermodynamique, l'équivalence entre le travail et la chaleur, la loi de Mariotte, etc., résultent de l'expérience. Au contraire, en logodynamique, les relations se déduisent simplement de transformations mathématiques élémentaires, car il s'agit de grandeurs dépourvues de dimensions, et c'est là une différence très profonde entre les deux théories ⁽¹⁾.

(1) Il en est exactement de même dans la comparaison classique entre la "limitation temps-fréquence" d'un signal et le principe d'incertitude d'Heisenberg. Dans le premier cas, il s'agit d'une simple propriété mathématique de la transformation de Fourier, alors que le deuxième constitue un véritable principe de la physique corpusculaire.

CHAPITRE II

ENERGIE DIACRITIQUE MOYENNE D'UN SYSTEME ET D'UN MESSAGE

Le chapitre qui précède a été consacré aux définitions fondamentales ; il a permis de préciser la notion d'information, sous sa forme d'énergie diacritique ou dénombrement d'états discernables. Mais les considérations qui ont été développées concernent en quelque sorte l'information "statique" ou "potentielle" du système : richesse d'information d'un symbole pris parmi l possibles, dépense effectuée en fonction de la rapidité de modulation et de la valence du signal. Nous allons maintenant aller plus loin, en évaluant la quantité d'information qui est effectivement associée à l'évolution d'un système, ainsi que celle qu'est effectivement susceptible de véhiculer un message, compte tenu de sa structure.

PRESSION DIACRITIQUE MOYENNE D'UN SYSTEME

Soit un système (S) susceptible de se trouver, à un instant donné, dans un état (s_k) déterminé, parmi l états différents admissibles. Examinons son évolution en l'observant N fois de suite. Si N est très grand, le système (S) a pris tous les états possibles : supposons qu'il ait pris a_1 fois l'état (s_1), a_2 fois l'état (s_2), ..., a_l fois l'état (s_l). On aura :

$$(12) \quad a_1 + a_2 + \dots + a_l = N$$

Si nous appelons configuration l'ensemble des N situations examinées successivement, le nombre $P(N)$ de configurations possibles a pour valeur :

$$(13) \quad P(N) = \frac{N!}{a_1! a_2! \dots a_l!}$$

$P(N)$ est le nombre de configurations discernables : d'après la définition donnée au chapitre précédent, $\log P(N)$ est l'énergie ou pression diacritique d'une configuration.

Si N est très grand, les a_i sont eux-mêmes très grands, et on peut déterminer la valeur asymptotique de $\log p$ en faisant usage de la formule de Stirling, réduite à son terme principal :

$$\log x! = x \log x - x$$

On en déduit :

$$(14) \quad \log P(N) = N \log N - N - \sum_i (a_i \log a_i - a_i)$$

ou encore, puisque

$$\sum_i a_i = N :$$

$$(15) \quad \frac{1}{N} \log P(N) = - \sum \frac{a_i}{N} \log \frac{a_i}{N}$$

Si les états (s_i) apparaissent de façon aléatoire, la probabilité a priori d'un de ces états est évidemment :

$$(16) \quad p_i = \frac{a_i}{N_i}$$

L'expression (15) prend ainsi la forme :

$$(17) \quad \lim_{N \rightarrow \infty} \left[\frac{1}{N} \log P(N) \right] = \sum_i p_i \log \frac{1}{p_i} = H$$

Cette formule fondamentale donne la pression diacritique moyenne (ou asymptotique) H , par état, d'une configuration du système (S) en fonction des probabilités d'apparition des divers états possibles.

Shannon a appelé l'expression (17) entropie de la source S. Nous avons, comme il a été expliqué plus haut, substitué la notion de système à celle de source ; nous renonçons également au mot "entropie" qui, étymologiquement, rappelle le caractère irréversible de l'évolution d'un système physique. Dans le cas traité ici, il s'agit simplement d'un dénombrement de configurations discernables, et la notion d'irréversibilité n'intervient pas : il n'en demeure pas moins vrai que l'expression formelle de l'entropie thermodynamique est absolument semblable à celle de la formule (17) et s'établit par le même raisonnement. Quel que soit le terme employé, $1/N \log P$ représente la "quantité moyenne d'information par état" qu'il faut connaître pour décrire le déroulement du processus d'évolution de (S), c'est-à-dire pour identifier les diverses configurations d'états.

Shannon a montré que $H = \sum p_i \log (1/p_i)$ est maximal lorsque les diverses probabilités sont égales⁽¹⁾ ($p_1 = p_2 = \dots = p_l = 1/l$) et tend vers zéro lorsqu'une des probabilités, par exemple p_1 , tend vers 1, les autres devenant très petites. Ce résultat est intuitif, car si un état devient extrêmement probable au détriment de tous les autres, on est "presque sûr" que le système se trouve effectivement dans cet état, dont la prédiction est très aisée. On a donc l'inégalité suivante :

$$(18) \quad \lim_{N \rightarrow \infty} \left[\frac{1}{N} \log P(N) \right] \leq \log l$$

Si on se rappelle que $\log l$ est la pression diacritique associée à un système (S) possédant l états discernables, ou si l'on veut, le nombre d'éléments par exemple binaires, indispensable pour identifier à coup sur un état parmi les l possibles, on voit que la pression diacritique effectivement observée est toujours inférieure à $\log l$, et n'atteint cette valeur que si les états sont équiprobables.

Ce résultat provient du fait qu'au cours de son évolution, le système (S) prend "généralement", en moyenne, pendant un long intervalle de temps, moins de configurations que toutes celles qui sont possibles a priori. Sous

(1) La démonstration est rappelée dans l'annexe I.

une autre forme, on peut dire qu'une différence de probabilité d'apparition entre les divers états constitue une liaison, de type stochastique, entre les situations possibles : le choix est plus limité en ce sens que la prédiction de l'état futur de (S) est plus facile à faire que lorsque les probabilités sont toutes égales.

Chaque fois qu'il existe une liaison, soit certaine, soit aléatoire, entre les divers états de (S), la pression diacritique moyenne se trouve diminuée. Prenons l'exemple d'une liaison stricte : si l'apparition d'un état (s_i) interdit l'apparition immédiatement suivante d'un autre état (s_j) (ou au contraire implique nécessairement l'apparition de (s_j)), le nombre de configurations possibles a priori se trouve diminué, et l'énergie diacritique $\log P$ est plus faible.

En résumé, la formule (18) montre que $\log l$, pression diacritique associée a priori à un système ayant l états possibles, n'est la pression diacritique moyenne observée réellement que si tous les états ont la même probabilité d'apparition.

PUISSANCE DIACRITIQUE D'UN MESSAGE - THEOREME DE SHANNON

Examinons maintenant non plus l'évolution du système (S), mais les messages qui peuvent être transmis à l'aide d'un alphabet donné. Chaque symbole de l'alphabet est défini par sa durée t_i ainsi que par d'autres caractéristiques - par exemple le niveau d'amplitude du signal qui le représente.

Si w est la rapidité de modulation et v la valence de l'émission, l'énergie sémiodynamique ε_i d'un symbole (s_i) de durée t_i est, conformément à la formule (7) :

$$(19) \quad \varepsilon_i = w t_i \log v$$

Donc, un message composé à l'aide de N symboles successifs, comportant a_1 fois le symbole s_1 , ... a_l fois le symbole s_l , correspond au travail sémiodynamique E , tel que :

$$(20) \quad E = \sum a_l \varepsilon_l = w \log v \sum a_l t_l = wT \log v$$

en posant :

$$(21) \quad \left\{ \begin{array}{l} \sum a_l = N \quad (\text{nombre de symboles émis}) \\ \sum a_l t_l = T \quad (\text{durée totale du message comportant } N \text{ symboles}). \end{array} \right.$$

Le problème que nous allons examiner est celui de l'énergie diacritique du message. En effet, il est clair que l'existence de certaines contraintes, telles que le "format" des symboles, qui ont dans le cas général des durées t_i différentes, a une influence sur le nombre des messages discernables de durée donnée que l'on peut composer au moyen de l'alphabet.

Soit T la durée d'un message comportant N symboles successifs.

Nous désignerons par $M(T)$ le nombre de messages différents (discernables) de durée T que l'on peut constituer à l'aide de l'alphabet.

L'énergie diacritique de ce message est $\log M(T)$; nous appellerons puissance diacritique le rapport $1/T \log M(T)$.

Prenons un exemple très simple, celui où l'alphabet est composé de symboles d'égale durée τ , τ étant l'intervalle de temps élémentaire ($w\tau = 1$).

Si v est la valence de l'émission, on a manifestement ⁽¹⁾ :

$$(22) \quad M(T) = v^{wT}$$

$$(23) \quad \log M(T) = wT \log v = E$$

$$(24) \quad \frac{1}{T} \log M(T) = w \log v = \frac{E}{T} = k$$

Il est clair que $l = v$, puisque les symboles ne peuvent se distinguer par leur durée.

Nous vérifierons plus loin que (23) correspond à l'énergie diacritique maximale d'un message caractérisé par les paramètres w , T , v : le travail sémiodynamique dépensé est donc égal à l'énergie diacritique maximale d'un message ; il y a dans ce cas identité entre puissance diacritique et puissance sémiodynamique (E/T).

Il en est autrement dans le cas général, à cause des différences de durée des symboles. Soit $M(T)$ le nombre de messages discernables de durée T , construits avec un certain alphabet. Chacun de ces messages se termine nécessairement par un des l symboles de l'alphabet. Supposons que l'on isole de la liste tous les messages se terminant par le symbole (s_1), de durée t_1 : on obtient alors des messages qui ont tous la durée $T-t_1$. Inversement, considérons l'ensemble des messages de durée $T-t_1$: si les règles d'assemblage de l'alphabet autorisent l'adjonction du symbole s_1 à chacun d'entre eux, une telle adjonction fournit la totalité des messages de durée T se terminant par le symbole s_1 .

Ce raisonnement simple permet d'établir l'équation aux différences finies :

$$(25) \quad M(T) = M(T-t_1) + M(T-t_2) = \dots + M(T-t_l)$$

qui exprime simplement que le nombre de messages discernables de durée T s'obtient, sans omission ni répétition, en décomptant ceux d'entre eux qui se terminent par le symbole s_1 , puis par le symbole s_2 , et ainsi de suite.

L'équation (25) est d'un type bien connu ; on en trouve les solutions en effectuant la transformation :

$$(26) \quad M(T) = C e^{\beta t}, \quad C \text{ et } \beta \text{ étant deux constantes.}$$

 (1) L'équation (22) suppose que T est un multiple de τ - ce qui n'est guère restrictif et que les symboles ne se différencient que par le niveau d'amplitude : il y a donc v symboles distincts.

En substituant (26) dans (25), on obtient l'équation "caractéristique" :

$$(27) \quad e^{-\beta t_1} + e^{-\beta t_2} + \dots + e^{-\beta t_l} = 1$$

Si β est une racine réelle de l'équation caractéristique, il est clair que $Ce^{\beta t}$ est une solution de (25). En choisissant pour β la plus grande racine positive de (27), on pourra écrire :

$$(28) \quad \frac{1}{T} \log M(T) = \beta + \frac{\log C}{T}$$

d'où en définitive :

$$(29) \quad \beta = \lim_{T \rightarrow \infty} \left(\frac{\log M(T)}{T} \right)$$

β est donc la valeur asymptotique de la puissance diacritique du message. Ce résultat peut se mettre sous une forme équivalente, en faisant intervenir l'énergie sémiodynamique des symboles. Il suffit de poser :

$$(30) \quad \beta = \mu_0 w \log v$$

μ_0 est une constante satisfaisant à la relation :

$$(31) \quad \sum_1 e^{-\mu_0 \varepsilon_i} = 1$$

et on a de plus :

$$(32) \quad C = \lim_{T \rightarrow \infty} \frac{\log M(T)}{T} = \mu_0 w \log v = \mu_0 \left(\frac{E}{T} \right) = \mu_0 k$$

μ_0 est donc le rapport de la puissance diacritique de la séquence à la puissance sémiodynamique dépensée - On peut donc énoncer le théorème suivant :

Théorème de Shannon

Etant donné un alphabet composé de l symboles et associé à une séquence de valence v , la puissance diacritique des séquences composées avec cet alphabet tend asymptotiquement vers une limite finie lorsque leur durée augmente indéfiniment.

Le concept introduit par Shannon est celui de la "capacité d'une voie" défini par $C = \lim_{T \rightarrow \infty} \frac{\log M(T)}{T}$.

Comme nous l'avons dit plus haut, nous avons préféré la notion de message (ou de séquence) à celle de "voie" qui est ambiguë ; de même, l'analogie avec l'énergétique nous a conduit aux notions d'énergie et de puissance diacritiques. Il va de soi que ces notions sont strictement équivalentes.

Revenons sur la signification des constantes β et μ_0 .

Les équations (29) et (32) montrent que dans une longue séquence, ou plus précisément dans une séquence dont la durée augmente indéfiniment, $1/\beta$ représente le temps nécessaire pour transmettre un "logon d'information", et μ_0 joue le rôle d'un rendement : en effet il faut dépenser un travail sémiodynamique E pour transmettre un énergie diacritique $\mu_0 E$; on a donc certainement $\mu_0 \leq 1$.

On trouvera dans l'annexe II la démonstration du fait que $\mu_0 \leq 1$; du même coup se trouve démontré, par la formule (32), que la puissance diacritique maximale du message est $E/T = w \log v$: elle est obtenue lorsque tous les symboles ont, chacun dans leur classe, la durée minimale compatible avec l'absence de restriction sur les conditions d'assemblage.

Remarque importante

L'évaluation de μ_0 repose sur l'hypothèse de l'indépendance des symboles : on a admis en effet que les règles d'assemblage n'imposaient aucune restriction à leur emploi. En d'autres termes, on a admis qu'en ajoutant n'importe quel symbole s_i à une séquence autorisée, on obtenait encore une séquence autorisée. Il en est rarement ainsi ; en général, et nous l'avons déjà souligné, l'alphabet comporte des restrictions sur les possibilités d'assemblage. Il est évident a priori que ces restrictions diminuent la puissance diacritique, car le nombre de séquences possibles de durée T est plus faible que si les symboles étaient indépendants (ou libres).

Si par exemple le symbole s_i ne peut être ajouté à toutes les séquences de durée $T-t_i$, et qui sont, par définition, en nombre égal à $M(T-t_i)$, on aura :

$$(33) \quad M(T) < M(T-t_1) + M(T-t_2) + \dots + M(T-t_l)$$

D'où :

$$(34) \quad \sum_i e^{-\beta t_i} > 1$$

où

$$(35) \quad \sum_i e^{-\mu \epsilon_i} > 1$$

Comme $\sum e^{-\mu \epsilon_i}$ est une fonction décroissante de μ on a en conséquence :

$$(36) \quad \mu < \mu_0 \leq 1.$$

Il est, en général, possible de se ramener au cas des symboles indépendants. Il suffit en effet de classer les symboles en deux catégories : les symboles libres et les symboles liés ; à chacun de ces derniers, on peut associer l'ensemble des combinaisons obtenues en les faisant précéder d'un symbole libre. On obtient ainsi un nouvel alphabet, composé de mots indépendants.

Prenons un exemple très simple : soit un alphabet composé de trois symboles :

s_1 de durée t_1 appelé "point"
 s_2 de durée t_2 " " "trait"
 s_3 de durée t_3 " " "espacement"

s_1 et s_2 sont libres, mais s_3 ne peut être précédé que de s_1 ou s_2 (la combinaison $s_3 s_3$ est exclue, ainsi que le symbole isolé s_3). Substituons à l'alphabet donné l'alphabet sans restriction composé de mots indépendants suivants :

$$\begin{array}{lll}
 \sigma_1 = S_1 & (t_1) & \varepsilon_1' = w t_1 \log v \\
 \sigma_2 = S_2 & (t_2) & \varepsilon_2' = w t_2 \log v \\
 \sigma_3 = S_1 S_3 & (t_1 + t_3) & \varepsilon_3' = w (t_1 + t_3) \log v \\
 \sigma_4 = S_2 S_3 & (t_2 + t_3) & \varepsilon_4' = w (t_2 + t_3) \log v
 \end{array}$$

L'absence de restriction est évidente : l'adjonction à chaque séquence autorisée de l'un quelconque des mots σ_i fournit encore une séquence autorisée ; de plus, toutes les séquences autorisées peuvent être obtenues à l'aide des σ_i , et l'ensemble des 4 nouveaux symboles est indispensable pour les obtenir toutes.

On aura donc :

$$\sum_1^4 e^{-\mu_0 \varepsilon_i'} = 1$$

les ε_i' ayant, dans ce cas particulier, les valeurs indiquées plus haut.

Lorsqu'il n'y a aucun symbole libre, la détermination de l'équation caractéristique peut être plus délicate ; elle peut néanmoins être faite dans divers cas particuliers.

CHAPITRE III
RELATIONS ENTRE ENERGIE DIACRITIQUE ET
TRAVAIL SEMIODYNAMIQUE

RELATION FONDAMENTALE

Dans le précédent chapitre, nous avons examiné en quoi la structure du message limitaient leur énergie diacritique. Nous allons maintenant établir les relations qui les unissent : c'est le problème de l'adaptation du message au système.

Il est, en fait, plus simple de se donner a priori un message composé au moyen d'un certain alphabet ; nous verrons alors quelle structure doit posséder le système pour que l'on puisse tirer le meilleur parti possible de ce message.

De façon précise, le problème sera posé comme suit :

Soit (S) un système susceptible de prendre l états différents, généralement non équiprobables. On associe à chaque état de (S) un symbole s_i d'énergie sémiodynamique ε_i , et on suppose les l symboles indépendants. A une séquence comportant un grand nombre N de symboles et de travail sémiodynamique E , correspond une configuration de (S). Quelle structure statistique doit posséder (S) pour que la pression diacritique de cette configuration soit maximale ?

Si nous pouvons trouver un maximum, il en résultera que pour un travail sémiodynamique dépensé donné a priori, le nombre d'informations transmises au sujet des variations d'états de (S) sera le plus grand possible.

Supposons donc que le système (S) prenne, dans une configuration de N états successifs, a_1 fois l'état 1, a_2 fois l'état 2, ..., a_l fois l'état l .

On aura :

$$(37) \quad \left\{ \begin{array}{l} \sum_1^l a_i = N \\ \sum_1^l a_i \varepsilon_i = E \end{array} \right.$$

E étant le travail sémiodynamique total dépensé :

$$(38) \quad E = \sum a_i \varepsilon_i = w \log v \sum a_i t_i = kT \quad (k = w \log v)$$

Nous devons chercher le maximum de $P(N)$, P étant le nombre de configurations possibles, donné déjà par la formule (13) :

$$P(N) = \frac{N!}{a_1! a_2! \dots a_l!}$$

Le problème consiste, les ε_i , N et E étant donnés, à déterminer les a_i pour que $\log P(N)$ soit maximal.

Employons la méthode des multiplicateurs de Lagrange : nous cherchons le maximum de :

$$(39) \quad \log P - \lambda \sum a_i - \mu \sum \varepsilon_i a_i$$

λ et μ étant deux constantes inconnues.

La variation de (39) s'écrit en remplaçant $\log P$ par son développement de Stirling (voir équation 14) :

$$(40) \quad - \sum \log a_i \delta a_i - \lambda \sum \delta a_i - \mu \sum \varepsilon_i \delta a_i = 0.$$

Bien entendu, les a_i ont été traités comme des variables continues, ce qui est admissible si elles sont très grandes.

Le coefficient de chaque δa_i doit être égalé à zéro, ce qui donne :

$$\log a_i + \lambda + \mu \varepsilon_i = 0$$

ou :

$$(41) \quad a_i = e^{-\lambda - \mu \varepsilon_i}$$

Les conditions (37) déterminent λ et μ

$$\sum_1^l e^{-\lambda - \mu \varepsilon_i} = N$$

$$\sum_1^l \varepsilon_i e^{-\lambda - \mu \varepsilon_i} = E$$

La division membre à membre élimine λ , et nous obtenons finalement les relations fondamentales cherchées :

$$(42) \quad \left\{ \begin{array}{l} \frac{E}{N} = U = \frac{\sum_1^l \varepsilon_i e^{-\mu \varepsilon_i}}{\sum_1^l e^{-\mu \varepsilon_i}} = - \frac{\partial}{\partial \mu} \log \sum_1^l e^{-\mu \varepsilon_i} = - \frac{\partial}{\partial \mu} (\log Z) \\ \frac{a_i}{N} = p_i = \frac{e^{-\mu \varepsilon_i}}{\sum_1^l e^{-\mu \varepsilon_i}} = - \frac{1}{\mu} \frac{\partial}{\partial \varepsilon_i} \log \sum_1^l e^{-\mu \varepsilon_i} = - \frac{1}{\mu} \frac{\partial}{\partial \varepsilon_i} (\log Z) \end{array} \right.$$

Comme E et N sont donnés, $U = \frac{E}{N} = k \frac{T}{N}$, qui représente le travail moyen dépensé par symbole est donné. La première des équations (42) fournit donc la constante inconnue μ , d'où l'on déduit les probabilités p_i des états du système (S) optimal, c'est-à-dire de celui dont l'adaptation au message est la meilleure possible.

En pratique, il est plus commode de conserver μ comme variable auxiliaire - variable dont nous déterminerons plus loin la signification. Les équations (42) donnent ainsi U et p_i sous forme paramétrique.

Toutes les grandeurs importantes s'expriment en fonction d'une grandeur Z caractéristique des symboles, et que nous appellerons fonction d'état, ou somme d'états :

$$(43) \quad Z(\mu, \varepsilon_1, \dots, \varepsilon_l) = \sum_i e^{-\mu\varepsilon_i}$$

Le système (42) équivaut aux équations suivantes, que nous utiliserons pour interpréter les formules :

$$(44) \quad U_0 = -\frac{1}{\mu} \log Z = -\frac{1}{\mu} \log \sum e^{-\mu\varepsilon_i}$$

$$(45) \quad \left\{ \begin{array}{l} \varepsilon_i = U_0 + \frac{1}{\mu} \log \frac{1}{p_i} \\ U = \sum p_i \varepsilon_i = \frac{1}{\mu} \sum p_i \log \frac{1}{p_i} + U_0 = U_0 + \frac{1}{\mu} H \end{array} \right.$$

Supposons que l'on se soit donné un alphabet sans restriction d'assemblage, c'est-à-dire un jeu de l symboles indépendants ayant respectivement pour énergie sémiodynamique $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_l$ (rangés par valeurs croissantes). Quel partie peut-on en tirer ? On peut assigner au travail sémiodynamique par symbole ($U = E/N$) n'importe quelle valeur comprise entre $U_{\min.} = \varepsilon_1$ et $U_{\max.} = \frac{1}{l} \sum_i \varepsilon_i$, ces deux valeurs extrêmes correspondant respectivement à $1/\mu = 0$ et $1/\mu = \infty$ (on le vérifie par la 2ème équation 45, l'équation 44 ; et dans le deuxième cas directement par la première équation 42).

Chaque fois que l'on fixe U ($U_{\min.} < U < U_{\max.}$), ce qui revient à fixer μ , donc U_0 , la pression diacritique maximale H compatible avec le travail sémiodynamique dépensé est donnée par le deuxième équation (45).

Les équations (42) font apparaître U comme une fonction de H ; elles représentent, sous forme paramétrique, une courbe (Γ) que nous appellerons courbe caractéristique de l'alphabet.

La figure (1) est une courbe de ce genre, relative à un exemple simple qui sera discuté au prochain chapitre. En abscisse, on a porté la pression diacritique H d'un système (S) auquel on associe le message, en ordonnée le travail sémiodynamique dépensé par symbole ($U = E/N$). La courbe croissante $U = f(H)$ représente l'optimum possible du bilan pression diacritique (donc quantité d'information par symbole) - travail sémiodynamique dépensé, pour l'alphabet considéré.

Mais comme les deux équations (45) ont la même forme, il est naturel de porter en abscisse la grandeur $\log 1/p$ et en ordonnée ε , conformément à l'équation :

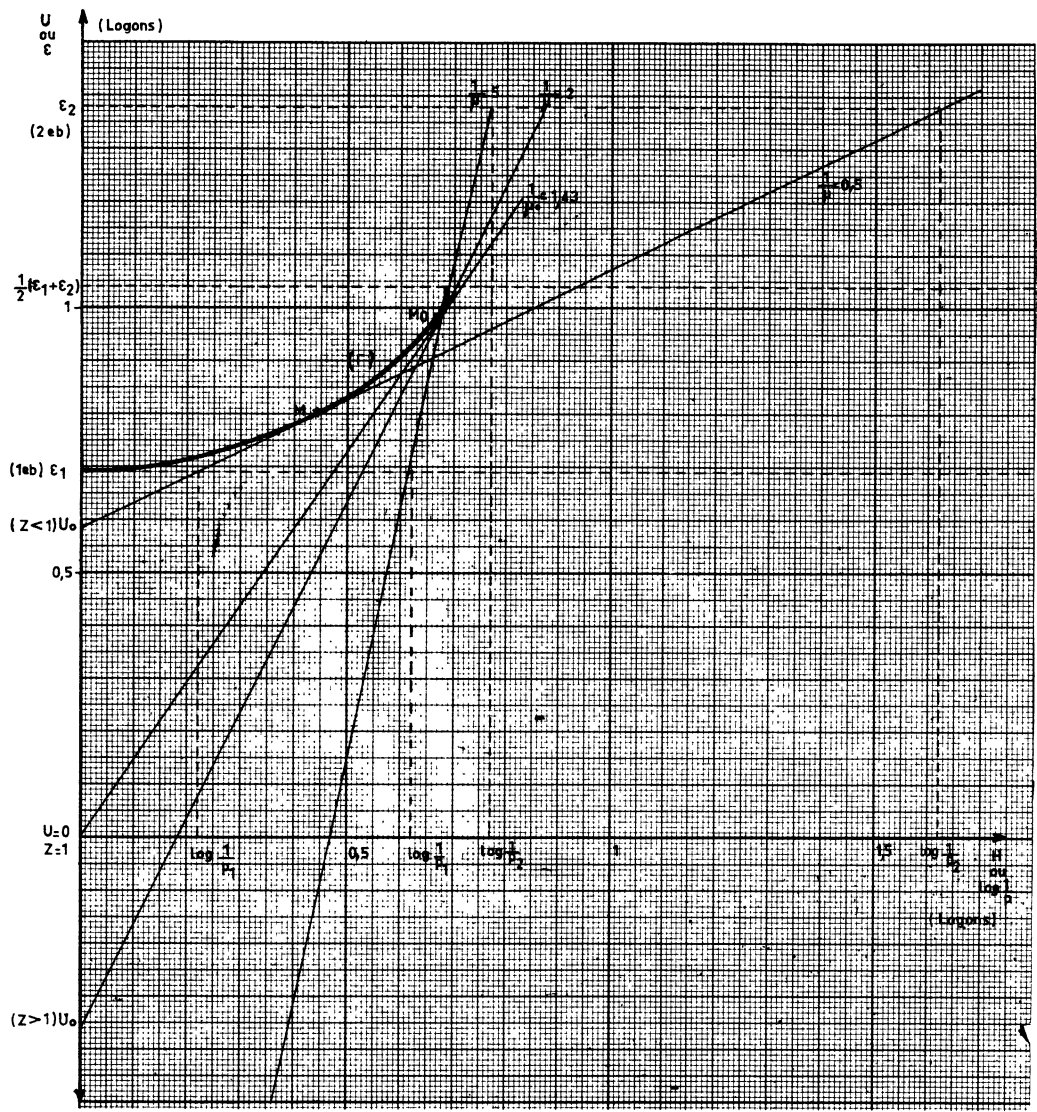


Figure 1 - Courbe caractéristique d'un alphabet à deux symboles.

$$(45) \quad \epsilon = U_0 + \frac{1}{\mu} \log \frac{1}{p}$$

Pour une valeur donnée de μ , donc de U_0 , les points $(\log 1/p_i, \epsilon_i)$ du plan $(\log 1/p, \epsilon)$ sont situés sur la droite d'équation (45), dont l'ordonnée à l'origine est U_0 et la pente $1/\mu$. On verra au chapitre suivant que cette droite est la tangente à la courbe (Γ) au point $M(H, U)$; l'intersection de la droite tangente avec les horizontales d'ordonnées $\epsilon_1, \epsilon_2, \dots, \epsilon_l$ fournit les valeurs de $\log 1/p_1 \dots \log 1/p_l$ correspondants. L'équation (45) montre, ce qui est intuitif, que le symbole de plus faible énergie sémiodynamique (donc de plus

faible durée) doit être associé à l'état de plus grande probabilité, et cela suivant une loi exponentielle. L'adaptation optimale du message au système s'obtient donc si les probabilités d'états sont conformes à la relation (45), U_0 apparaît comme "l'énergie sémiodynamique résiduelle" obtenue, à la limite, pour un symbole associé à une probabilité égale à 1 (1).

Quant à la pente $1/\mu$, elle peut s'interpréter en écrivant, à partir de (45) :

$$(46) \quad \frac{1}{\mu} = \frac{\epsilon_k - \epsilon_1}{\log \frac{p_1}{p_k}}$$

$1/\mu$, que nous appellerons gradient sémiodynamique, caractérise l'augmentation d'énergie liée à la variation de probabilité des états ; c'est la grandeur qui définit la "sensibilité" ou indice de modulation en durée des symboles en fonction des variations de probabilité des états de système (S).

On peut donner encore une autre interprétation du gradient sémiodynamique : on a en effet puisque $1/\mu$ est la pente de la tangente à la courbe Γ :

$$(47) \quad dU = \frac{1}{\mu} dH$$

A un petit accroissement dH de la pression diacritique correspond une augmentation $1/\mu dH$ de travail sémiodynamique dépensé ; dans une zone à faible gradient, l'augmentation de la pression diacritique s'assortit d'un faible travail supplémentaire, et au contraire pour un fort gradient, le "rendement" d'un accroissement du travail dépensé devient très mauvais.

En posant, comme au chapitre précédent, $k = w \log v$, on peut écrire :

$$(48) \quad \left\{ \begin{array}{l} U_0 = k \tau_0 \\ \frac{1}{\mu} = k \theta \end{array} \right.$$

Dans (48), τ_0 est la durée résiduelle, et θ la constante de temps sémiodynamique : τ_0 est la "durée" d'un symbole virtuel dont la probabilité serait égale à 1, et θ est l'augmentation de durée que l'on doit donner à un symbole lorsqu'on passe d'un état de probabilité p à un état de probabilité p/e (1).

Si on substitue τ_0 et θ à U_0 et $1/\mu$, la deuxième équation (45) prend l'une des deux formes équivalentes :

$$(49) \quad \left\{ \begin{array}{l} U = U_0 + k \theta H \\ \frac{T}{N} = \tau_s = \tau_0 + \theta H \end{array} \right.$$

(1) On doit remarquer que U_0 peut être soit positif, soit négatif.

(1) Notons au passage que le gradient sémiodynamique joue un rôle homologue à celui de la température en thermodynamique.

Sous la deuxième forme, τ_s représente la durée moyenne d'un symbole : cette durée est égale à la durée résiduelle τ_o augmentée du produit de la constante de temps sémiodynamique par la pression diacritique moyenne de système (S).

Les expressions (49), linéaires en apparence, ne doivent pas faire oublier que U_o (donc τ_o) dépend en fait de μ et des ε_1 . C'est ce qui apparaît sur la courbe (Γ) de la figure 1 : U n'est pas une fonction linéaire de H, mais à toute valeur de H correspond un point M (H, U, μ) de cette courbe (Γ).

Dans la relation entre U et H, exprimée par les formules (équivalentes) (49), il y a deux rapports particulièrement intéressants : d'une part le débit N/T (nombre maximal de symboles écoulés par unité de temps) d'autre part le rendement H/U (rapport entre la pression diacritique du système et le travail moyen dépensé par symbole).

Examinons en premier lieu le débit. On peut écrire :

$$(50) \quad \frac{N}{T} = \frac{1}{\tau_o + \theta H} = \frac{\mu}{\mu_o} \frac{C}{H - \log Z} = \frac{k}{U}$$

Rappelons que μ_o est défini par $Z(\mu_o) = 1$, et que $C = \mu_o k$ est la puissance diacritique du message - que Shannon appelle "capacité de la voie de transmission" - (cf équation 32).

Nous nous sommes fixés un alphabet ($\varepsilon_1, \dots, \varepsilon_l$) de sorte que μ_o est fixé, et que la courbe (Γ) est entièrement définie. Lorsque nous associons un message composé avec cet alphabet à un système de pression diacritique H, le travail sémiodynamique dépensé ne peut être inférieur à $U = f(H)$ (ordonnée d'abscisse H sur la courbe Γ) et le débit maximal possible est $N/T = k/U$. Ce débit maximal s'obtient lorsque les probabilités p_i satisfont à la relation (45), U_o et $1/\mu$ étant relatifs au point M (H, U) de la courbe Γ . C'est le meilleur codage possible. Nous pouvons donc énoncer le théorème suivant :

THEOREME

Etant donné un alphabet associé à un système de pression diacritique H, un message composé avec cet alphabet ne peut transmettre plus de $\frac{\mu}{\mu_o} \frac{C}{H - \log Z} = \frac{k}{U}$ symboles par seconde, $C = \mu_o k$ étant la puissance diacritique du message (qui résulte de la donnée de l'alphabet), μ et $-\log Z$ se trouvant déterminés par le point M (H, U) de la courbe caractéristique.

Si H a une valeur telle que $\mu = \mu_o$, $Z = 1$, le débit maximal possible devient : $N/T = C/H$, ce qui s'écrit encore : $CT = NH$. Dans ce cas particulier, les énergies diacritiques du système et du message sont égales.

Examinons maintenant le rendement H/U. La courbe (Γ) de la figure 1 montre que le rapport U/H passe par un minimum pour $\mu = \mu_o$, $Z = Z(\mu_o) = 1$. C'est un résultat absolument général, car la courbe caractéristique (Γ) a toujours pour ordonnée à l'origine $\varepsilon_1 > 0$ et a toujours sa concavité tournée

vers le haut (1). Nous retrouvons donc les conditions particulières qui viennent d'être mentionnées : le rendement maximal H/U s'obtient pour $U_0 = 0$, $\mu = \mu_0 = (H/U)_{opt}$. μ_0 est donc le rendement maximal, et on l'obtient pour $p_i = e^{-\mu_0 \epsilon_i}$. Dans ce cas, il y a égalité entre les énergies diacritiques du message et du système :

$$(51) \quad C = \mu_0 \frac{E}{T}$$

$$CT = NH, \quad \frac{C}{H} = \frac{N}{T}$$

La courbe (Γ) montre que si $\mu = \mu_0$ fournit bien le rendement maximal, cette valeur ne correspond, pour un alphabet donné, ni à la pression diacritique maximale H_m (extrémité supérieure de la courbe (Γ)) ni au débit $N/T = k/U$ maximal (ordonnée à l'origine de (Γ), obtenu pour une information transmise nulle).

PROBLEME DE SHANNON

Dans tout ce qui précède, nous avons cherché à déterminer une adaptation optimale d'un système à un message, en nous fixant un alphabet, la valeur E du travail sémiodynamique ainsi que le nombre N de symboles de la séquence de durée T ($E = k T$). Ceci revient à fixer $U = E/N$ (travail sémiodynamique moyen par symbole), d'où résulte H . En définitive, la courbe (Γ), d'équation $U = f(H)$, est la courbe caractéristique de l'alphabet, elle indique le meilleur parti que l'on peut en tirer, pour chaque valeur de U , ou pour chaque valeur de H .

Shannon a évalué l'énergie diacritique maximale NH que l'on peut associer à une séquence de durée T (travail $E = kT$), sans imposer de valeur au nombre N de symboles de la séquence, (U n'est plus fixé). Le calcul, très simple, est reproduit dans l'annexe III. On trouve que l'énergie diacritique maximale NH est telle que :

$$(51 \text{ bis}) \quad \begin{cases} p_i = e^{-\mu_0 \epsilon_i} \\ NH = N \sum p_i \log \frac{1}{p_i} = \mu_0 E = \mu_0 k T = CT \end{cases}$$

(μ_0 étant toujours défini par $Z = \sum e^{-\mu_0 \epsilon_i} = 1$)

Ce sont une fois encore les valeurs de H et de p_i qui correspondent au rendement optimal. Ce résultat n'est pas surprenant : en nous fixant N , nous nous étions fixés une contrainte supplémentaire, celle du travail sémiodynamique par symbole : nous obtenions donc, pour chaque valeur du paramètre N , un couple (H, U), c'est-à-dire un point de la courbe (Γ). Au con-

 (1) Cette propriété résulte de ce que la pente de la courbe (Γ) a pour valeur $1/\mu$, qui croît de 0 à l'infini tout au long de cette courbe.

traire, en fixant seulement E, nous vérifions simplement que l'optimum est atteint lorsqu'il y a égalité entre les énergies diacritiques du message et du système.

Mais on peut tirer d'autres conclusions intéressantes des équations (51 bis). Si on se fixe simplement μ_0 , l'adaptation optimale correspond à $\varepsilon_i = \frac{J}{\mu_0} \log \frac{J}{p_i}$, $U = \frac{J}{\mu_0} H$.

Tous les alphabets ayant une même puissance diacritique μ_0 k satisfont à ces relations. Nous sommes donc en mesure de résoudre le problème suivant : "Etant donné un système (S) prenant l états de probabilités respectives p_1, \dots, p_2 , et possédant par conséquent une pression diacritique donnée $H = \sum_{p_i} \log \frac{1}{p_i}$, comment coder un message de puissance diacritique donnée μ_0 k pour que le travail sémiodynamique dépensé soit minimal ?"

Il faut évidemment assigner aux divers symboles une énergie sémiodynamique $\varepsilon_i = J/\mu_0 \log J/p_i$, et le travail minimal est $U = J/\mu_0 H$. On peut dire encore que la courbe caractéristique de l'ensemble des alphabets de puissance diacritique μ_0 k est la droite $U = J/\mu_0 H$. Cette droite est l'enveloppe des courbes (Γ) relatives à chacun de ces alphabets (fig. 2).

Ce résultat est strictement équivalent au théorème bien connu de Shannon : "Si on donne une source d'entropie H et une voie de capacité C, le codage, optimal permet de transmettre $C/H = N/T$ symboles par seconde, et cette limite ne peut être dépassée".

Il y a donc deux problèmes bien distincts, mais complémentaires : le premier consiste à se fixer un alphabet, et à trouver quel meilleur parti on peut en tirer, c'est-à-dire quel travail minimum il faut dépenser pour transmettre avec cet alphabet une quantité d'information (énergie diacritique) donnée. Le résultat est fourni par les équations (44/45), ou si l'on veut par la courbe caractéristique (Γ).

Le second consiste à se fixer une puissance diacritique, et à déterminer l'alphabet optimal, qui permet de ne dépenser que le travail minimal $U = J/\mu_0 H$ pour une pression diacritique donnée H (1).

On pourrait penser, a priori, que le deuxième problème, que nous appellerons pour simplifier problème de Shannon, est plus intéressant que le premier, puisqu'en général on cherche à adapter le message au système, et non l'inverse. En pratique il n'en est rien. Cela résulte du fait que nous ne sommes pas entièrement libres dans le choix de l'alphabet ; en particulier les énergies sémiodynamiques des symboles sont nécessairement multiples d'un "quantum" $w\tau \log v$, ou si l'on veut, par définition même de l'intervalle de temps élémentaire, les durées des symboles sont multiples de τ .

Un exemple simple, discuté au dernier chapitre de cette étude, mettra clairement en évidence l'intérêt, et même la nécessité d'un choix a priori de l'alphabet, ce qui justifie les développements qui précèdent.

 (1) Nous ne traitons pas ici le problème général du codage, que l'on peut résoudre en substituant à l'alphabet primaire un alphabet secondaire constitué par des "mots" comportant plusieurs symboles.

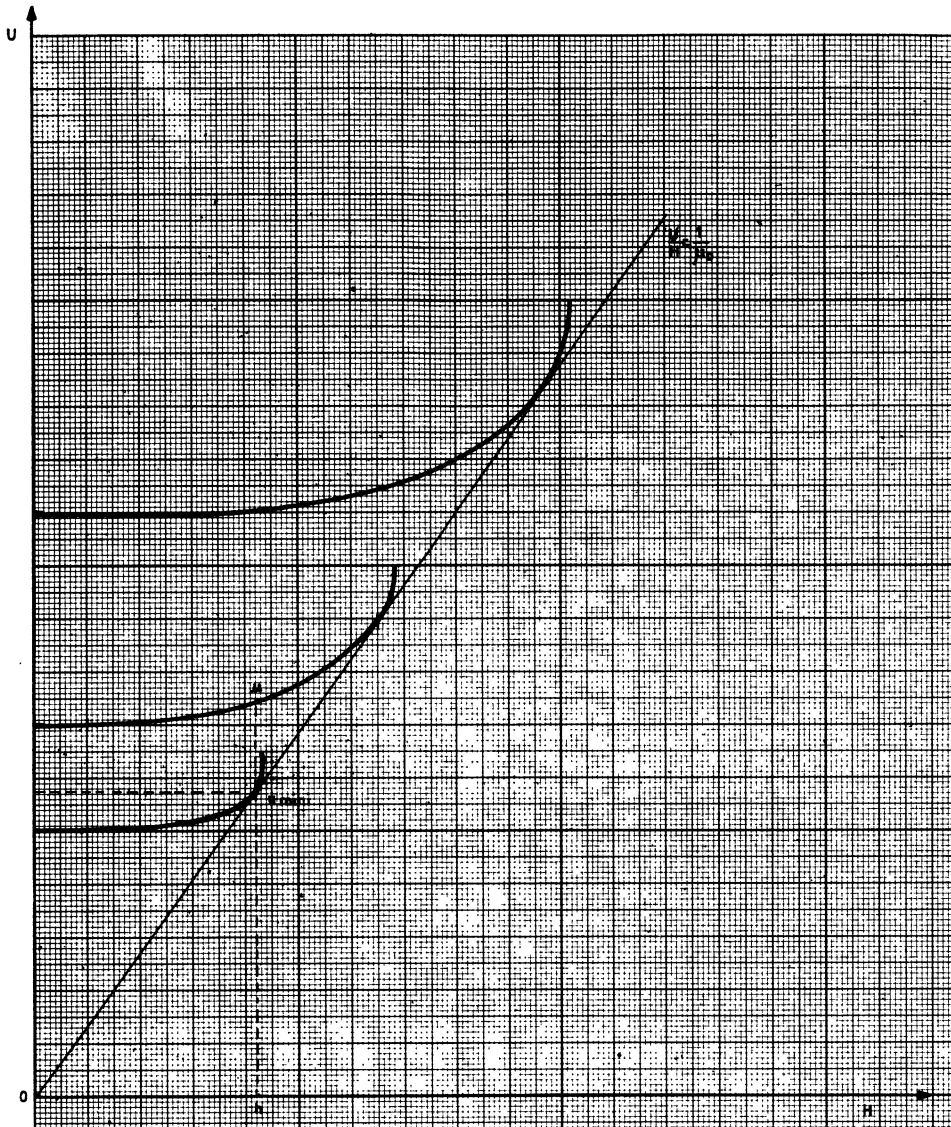


Figure 2 - Courbes caractéristiques des alphabets de même puissance diacritique.

COMPARAISON AVEC LA THERMODYNAMIQUE

Le mode de raisonnement qui a servi à établir les relations qui précèdent est exactement celui qui sert de fondement à la Thermodynamique Statistique. Dans ce dernier cas, en effet, on examine un système physique (S) pouvant prendre les états d'énergie $\varepsilon_1, \varepsilon_2, \varepsilon_l$: l'énergie totale E est celle d'un ensemble formé de N "copies" de S (virtuel selon les conceptions de J.W. Gibbs) et comportant a_1 fois l'état d'énergie ε_1 , a_2 fois l'état d'énergie

ε_2 , a_1 fois l'état d'énergie ε_1 . Si P est le nombre de situations distinctes correspondant à une telle répartition, le maximum de $\log P$ fournit la distribution la plus probable, qui constitue l'état d'équilibre. On trouve ainsi les équations (42), (43), (44), (45), avec l'interprétation suivante : U est l'énergie interne, $1/\mu = kT$, k étant la Constante de Boltzmann et T la température absolue, $S = k H$ est l'entropie et $\Psi = - k T \log Z$ le potentiel thermodynamique du système. La deuxième relation (45) est donc identique à la formule bien connue :

$$U - TS = - k T \log Z = + \Psi$$

L'analogie peut être encore poussée plus loin. En effet nous avons vu plus haut que la relation $Z(\mu_0) = 1$ est obtenue lorsque le nombre N de symboles d'une séquence de durée donnée n'est pas imposé ; en thermodynamique on obtient le même résultat, pour un système de N particules, si ce nombre N n'est pas fixé (statistiques des photons dans le problème du rayonnement). Les contraintes de l'alphabet dont nous avons parlé conduisent à des énergies sémiodynamiques multiples d'un "quantum" $w \log v$, exactement comme dans les statistiques quantiques de la Physique ; en revanche le "problème de Shannon" fait abstraction de cette contrainte, et doit être rapproché, de ce point de vue, des considérations qui sont à l'origine de la Statistique de Boltzmann. Le parallélisme entre les deux théories peut donc être poussé très loin. Mais la remarque faite à la fin du chapitre I garde toute sa valeur : c'est l'appareil mathématique qui est commun aux deux théories, et non leur nature physique ; dans cet exposé nous n'entrerons pas du tout dans les considérations qui ont été développées par L. Brillouin dans son ouvrage [III] et qui ont pour objet les rapports entre information et énergie ; le couplage - d'ailleurs très faible - s'introduit dans la théorie de la mesure, et sort complètement du cadre de cet exposé.

CHAPITRE IV
DEVELOPPEMENT ET EXEMPLES SIMPLES

GROUPEMENT OU "MULTIPLEXAGE" DES INFORMATIONS

Considérons un système (S_1) auquel nous associons un message formé à l'aide de symboles indépendants d'énergies sémiodynamiques $\alpha_1, \alpha_2, \dots, \alpha_k$, et un autre système (S_2) auquel nous associons un autre message, comportant des symboles indépendants d'énergies $\beta_1, \beta_2, \dots, \beta_m$.

Si l'on fait le choix d'une "cadence" commune w et d'une valence identique v pour les émissions, les symboles des deux catégories de messages différeront par leur durée ; en les "multiplexant", on obtiendra un message composite associé au système (S) constitué par la réunion de S_1 et de S_2 .

L'ensemble des "doublets" comporte alors des symboles d'énergies sémiodynamiques ε_l , avec :

$$(56) \quad \varepsilon_l = \alpha_k + \beta_m \quad (k, l, m, \text{ désignant les indices courants}).$$

Formons la somme d'états Z :

$$(57) \quad Z = \sum_l e^{-\mu \varepsilon_l} = \sum_k \sum_m e^{-\mu(\alpha_k + \beta_m)} = \sum_k e^{-\mu \alpha_k} \sum_m e^{-\mu \beta_m}$$

Dans les messages associés à (S), formé de la réunion de (S_1) et de (S_2), la proportion moyenne de symboles d'énergie ε_l est :

$$(58) \quad \frac{a_l}{N} = a_{(k, m)} = \frac{e^{-\mu(\alpha_k + \beta_m)}}{\sum_k e^{-\mu \alpha_k} \sum_m e^{-\mu \beta_m}}$$

Cherchons le nombre de doublets contenant le symbole s_k d'énergie α_k ; on obtient évidemment en sommant par rapport à m :

$$(59) \quad \sum_m a_{k, m} = N \frac{e^{-\mu \alpha_k}}{\sum_k e^{-\mu \alpha_k}}$$

(59) montre que la distribution optimale est exactement la même que celle (optimale) du message associé au système S_1 pris isolément.

Réciproquement, si les messages indépendants associés aux systèmes (S_1) et (S_2) sont choisis de telle façon que le gradient sémiodynamique $1/\mu$ soit la même pour les deux distributions, on obtiendra encore une distribution optimale de même gradient en les multiplexant.

Le raisonnement qui précède montre, d'une part le rôle fondamental que joue le gradient sémiodynamique dans la distribution des symboles ⁽¹⁾, d'autre par la caractère additif de la fonction $\log Z$ (ou $k \log Z = -U$).

PREMIER EXEMPLE D'UNE DISTRIBUTION SIMPLE

Traisons pour commencer le cas plus simple que l'on puisse imaginer : le système S peut prendre deux états, et l'alphabet du message associé comporte deux symboles : s_1 d'énergie ε_1 , s_2 d'énergie $\varepsilon_2 = 2\varepsilon_1$. Comme unité on peut choisir $\varepsilon_1 = 1$ e.b. (0,7 logon), $\varepsilon_2 = 2$ e.b. (1,4 logon).

La somme d'états est :

$$(60) \quad Z = e^{-\mu\varepsilon_1} + e^{-\mu\varepsilon_2} (= 2^\mu + 2^{2\mu})$$

et l'on a, d'après les relations générales (42) :

$$(61) \quad \begin{aligned} ZU &= \varepsilon_1 e^{-\mu\varepsilon_1} + \varepsilon_2 e^{-\mu\varepsilon_2} \\ Zp_i &= e^{\mu\varepsilon_i} \quad (i = 1, 2) \end{aligned}$$

Les équations (61) donnent les p_i en fonction de U, sous forme paramétrique ⁽¹⁾.

Sur la figure 1, on a porté en ordonnée le travail moyen par symbole U, et en abscisse la pression diacritique moyenne

$$H = p_1 \log \frac{1}{p_1} + p_2 \log \frac{1}{p_2}$$

On obtient ainsi la courbe (Γ) - caractéristique de l'alphabet - limitée à $0 < H < \log l$ ($= 0,70$) et à $\varepsilon_1 < U < 1/2$ ($\varepsilon_1 + \varepsilon_2$) ($0,7 < U < 1,05$). Ces limites sont en parfait accord avec les résultats généraux donnés dans le précédent chapitre.

Mais la similitude absolue des deux équations (44) montre que la même représentation peut être employée, en portant en abscisse, comme le suggère l'équation (45), la grandeur $\log 1/p$ et en ordonnée la grandeur ε ; les couples (p_1, ε_1) et (p_2, ε_2) ne sont alors que deux points particuliers du plan $(\varepsilon, \log 1/p)$. Le champ de variation de ces deux grandeurs est cette fois $(0, +\infty)$.

 (1) Dans l'analogie thermodynamique $1/\mu = kT$, T étant la température absolue ; on en déduit que la réunion de deux systèmes placés dans deux thermostats à la même température s'effectue sans perturbation, les énergies s'ajoutant purement et simplement.

(1) L'équation (61) peut s'écrire $U = \varepsilon_1 + \frac{\varepsilon_1}{e^{\mu\varepsilon_1} + 1}$, puisque $\varepsilon_2 = 2\varepsilon_1$. C'est l'expression de l'énergie de l'oscillateur de Fermi, en statistique de Fermi-Dirac.

Pour chaque valeur de μ , U_0 se trouve fixé, et l'équation (45) a pour représentation une droite du plan $(\log 1/p, \varepsilon)$. On a tracé sur la figure 1 quelques unes de ces droites. Elles ont pour enveloppe la courbe (Γ) . Ce résultat est facile à vérifier : en effet, le faisceau des droites s'écrit d'après (44),

$$\mu \varepsilon + \log Z(\mu) = \log \frac{1}{p}$$

ou :

$$\mu U + \log Z = H \quad (\text{plan } H, U).$$

L'enveloppe s'obtient en différentiant par rapport à μ :

$$U + \frac{\partial}{\partial \mu} (\log Z) = 0$$

et cette dernière relation n'est rien d'autre que la première des équations (42).

Le graphique peut être employé comme suit : à une valeur donnée de H , comprise entre 0 et $\log 2$, correspond par la courbe (Γ) une valeur de travail U , ou vice versa ; U est le travail par symbole le plus faible que l'on puisse obtenir lorsqu'on s'est donné H . Au point $M(H, U)$ de la courbe (Γ) correspond une valeur de $1/\mu$, pente de la tangente à (Γ) en ce point. La tangente en question coupe l'axe des ordonnées au point d'ordonnée U_0 ; de plus, l'intersection de cette tangente avec les ordonnées $\varepsilon = \varepsilon_1$, $\varepsilon = \varepsilon_2$ fournit les points d'abscisses respectives $\log 1/p_1$ et $\log 1/p_2$.

Si l'on considère une droite OM joignant l'origine 0 à un point quelconque M de la courbe (C) , on forme le rapport U/H (ou H/U , qui est le "rendement" du message par rapport au système). U/H est manifestement minimal (donc $H/U = \mu_0$ maximal) pour la tangente à (Γ) passant par l'origine. Il lui correspond le point M_0 , de coordonnées (H_{opt}, U_{opt}) , déterminé par un théorème de Shannon.

On peut ainsi considérer trois zones distinctes :

1er cas

$$Z < 1, U_0 > 0, \frac{1}{\mu} < \frac{1}{\mu_0} = 1,43.$$

Cette zone correspond à un état de grande probabilité ($\log 1/p_1 \ll 0,5$) et l'autre de faible probabilité ($\log 1/p_2 \gg 1$). La pression diacritique est très petite, de sorte que le "rendement" est faible. La faible pente ($1/\mu$ petit) montre qu'il faut une grande variation de $\log 1/p$ pour faire passer ε de ε_1 à ε_2 : le gradient sémiodynamique est petit. $U_0 > 0$ correspond à $\tau_0 > 0$; la durée moyenne τ_s d'un symbole reste peu différente de la durée résiduelle τ_0 qui correspond à un travail sémiodynamique dépensé sans apport d'information.

Le débit N/T reste voisin de son maximum w .

2e cas

$$Z = 1, U_0 = 0, \frac{1}{\mu} = \frac{1}{\mu_0} = 1,43$$

C'est l'état de rendement maximal, qui correspond au point M_0 de la courbe (Γ). Il est caractérisé par les valeurs suivantes :

$$\log \frac{1}{p_1} = 0,48 \quad (p_1 = 0,62), \quad \log \frac{1}{p_2} = 0,97 \quad (p_2 = 0,38)$$

$$H = 0,675 \quad \text{et} \quad U = 0,97 \quad \left(\mu_0 = \frac{H}{U} \simeq 0,7 \right)$$

La durée résiduelle τ_0 est nulle.

Le débit a une valeur moyenne $N/T = 0,72 w$.

3e cas

$$Z > 1, U_0 < 0, \frac{1}{\mu} > \frac{1}{\mu_0} = 1,43$$

Les probabilités tendent à devenir égales. Or, on sait que H devient alors maximal (0,70). Le rendement diminue légèrement, mais la pression diacritique H est légèrement plus forte que dans le cas "optimal" où $Z = 1$. Le gradient sémiodynamique est très grand et tend vers l'infini, à la limite, pour $p_1 = p_2 = 0,50$. La durée résiduelle τ_0 est négative ; cela signifie que les variations de $\theta \log 1/p$ sont supérieures, lorsque p passe de p_1 à p_2 , à l'écart de durée $\tau_2 - \tau_1$ entre les deux symboles d'énergies ε_1 et ε_2 ; la constante τ_0 vient compenser cet écart.

Enfin le débit N/T est proche de son minimum 0,666 w .

Cet examen détaillée montre, d'une part que la pression H peut être plus grande que dans le cas où le rendement est optimal, mais que cela conduit à augmenter également, et dans une plus grande proportion, le travail sémiodynamique par symbole, d'autre part que le choix de l'alphabet n'implique pas, d'une façon absolue, celui des probabilités d'état. En effet, pour tirer le meilleur parti de l'alphabet, il faut que les probabilités p_i soient telles que $\log 1/p$ soit une fonction linéaire des ε , mais avec cette réserve, on peut finalement adapter le message aussi bien au cas où les probabilités sont très différentes qu'à celui où elles sont presque égales ; il suffit de choisir convenablement Z , donc μ . L'annexe V donne quelques précisions sur ce point.

On ne déduit que le cas "limite", étudié par Shannon, n'est au fond qu'un exemple particulier, fournissant le meilleur rendement, mais inutilisable dans le cas où l'alphabet est imposé.

DEUXIEME EXEMPLE ELEMENTAIRE

Considérons maintenant le cas d'un alphabet "ouvert", comportant une infinité de symboles dont les énergies sémiodynamiques forment une progression arithmétique :

$$(62) \quad \varepsilon_1 = \varepsilon, \varepsilon_2 = 2 \varepsilon, \dots, \varepsilon_k = k \varepsilon, \dots, \text{etc.}$$

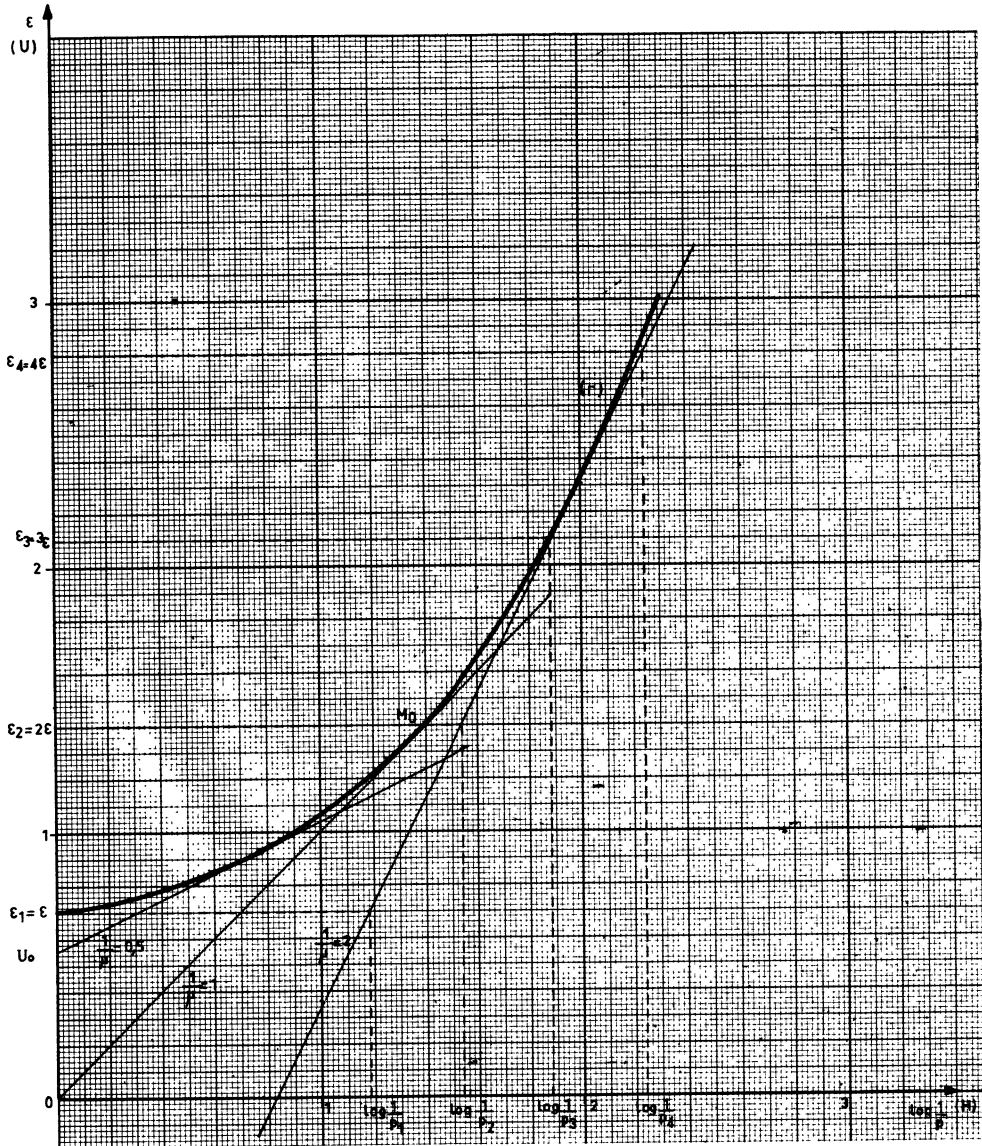


Figure 3 - Courbe caractéristique d'un alphabet "ouvert".

La fonction d'états vaut :

$$(63) \quad Z = \sum_{k=1}^{\infty} e^{k(\mu\varepsilon)} = \frac{1}{e^{\mu\varepsilon} - 1}$$

on en tire directement, par les formules (42) :

$$(64) \quad \left\{ \begin{array}{l} p_k = e^{-(k-1)\mu\varepsilon} - e^{-k\mu\varepsilon} = e^{-(k-1)\mu\varepsilon} [1 - e^{-\mu\varepsilon}] \\ U = -\frac{\partial}{\partial \mu} \log Z = \frac{\varepsilon e^{\mu\varepsilon}}{e^{\mu\varepsilon} - 1} = \frac{\varepsilon}{1 - e^{-\mu\varepsilon}} \end{array} \right.$$

$Z = 1$ pour $\mu = \mu_0 = \frac{\log 2}{\varepsilon}$; on a $U(\mu_0) = 2\varepsilon$, $H(\mu_0) = 2 \log 2$, et

$p_k = 1/2^k$ (loi de probabilité "binaire" ou dyadique).

La figure 3 représente la courbe caractéristique dans le plan (H, U) ainsi que les tangentes caractéristiques à la courbe (Γ) dans le plan ε , $\log 1/p$ qui permettent de déterminer graphiquement les p_k à partir des ε_k . On a choisi $\varepsilon = \log 2$, d'où $\mu_0 = 1$, $U(\mu_0) = H(\mu_0) = 2 \log 2$.

L'exemple donné ci-dessus est l'équivalent de "l'oscillateur de Planck" de la thermodynamique ; la fonction d'état et la distribution énergétique sont représentées par les formules bien connues de la physique (équations 63 et 64).

TROISIEME EXEMPLE : ALPHABET MORSE

Comme troisième exemple, examinons le cas de l'alphabet Morse. On sait qu'il permet de représenter les lettres au moyen de signaux brefs et longs, appelés "points" et "traits". Dans une émission bivalente, une des polarités correspond à l'espace, et l'autre aux traits et aux points. On admet les définitions suivantes qui correspondent au cas pratique le plus courant :

	Polarité	Durée
Représentation d'un point	+	τ
Représentation d'un trait	+	2τ
Espace entre point et trait	-	τ
Espace entre lettres	-	2τ
Espace entre mots	-	3τ

On trouve ainsi que les 26 lettres de l'aphabet, complétées par un 27e état (espace entre mots) sont représentées par des symboles dont l'énergie sémiodynamique varie entre $\varepsilon_1 = 3$ e. b. (2,1 logons), (espace mot Esp. et lettre E composée d'un point et d'un espace "lettre"), et $\varepsilon_{10} = 12$ e. b. (8,4 logons) (Par exemple lettre J = ·---, ou lettre Y ---·). On connaît d'autre part, en langue anglaise, les probabilités d'apparition des différentes lettres dans le langage écrit courant. On peut résumer les caractéristiques de

l'alphabet d'une part, la statistique des lettres d'autre part, dans le tableau ci-joint :

Symbole ou lettre	Représentation	Energie en e. b.	Probabilité d'apparition (langue anglaise)
Espace mot	Pol. négative	3 e. b.	0,200
E	.	3	0,105
T	-	4	0,072
O	---	10	0,065
A	.-	6	0,063
N	..	6	0,059
I	..	5	0,054
R	.-.	8	0,054
S	...	7	0,052
H	9	0,047
D	...-	8	0,035
L	10	0,029
C-	11	0,023
F-	10	0,022
U	.. -	8	0,022
M	--	7	0,021
P-	11	0,018
Y-	12	0,011
W	...-	9	0,011
G	...-	9	0,011
B-	10	0,010
V-	10	0,008
K	...-	9	0,003
X-	11	0,002
J-	12	0,001
Q-	12	0,001
Z	11	0,001

Il y a :

2	symboles d'énergie	ε_1	=	3 e. b.	(2,1 logons)
1	"	"	ε_2	=	4 e. b. (2,8 logons)
1	"	"	ε_3	=	5 e. b. (3,5 logons)
2	"	"	ε_4	=	6 e. b. (4,2 logons)
2	"	"	ε_5	=	7 e. b. (4,9 logons)
3	"	"	ε_6	=	8 e. b. (5,6 logons)
4	"	"	ε_7	=	9 e. b. (6,3 logons)
5	"	"	ε_8	=	10 e. b. (7 logons)
4	"	"	ε_9	=	11 e. b. (7,7 logons)
3	"	"	ε_{10}	=	12 e. b. (8,4 logons)

Considérons l'alphabet en lui-même, sans tenir compte de la fréquence d'apparition des lettres. On peut construire sa courbe caractéristique, par exemple grâce à la méthode donnée dans l'annexe V. En posant $\delta \varepsilon = 0,7$, $\varepsilon_1 = 2,1$, $v = \varepsilon_1 / \delta \varepsilon = 3$, $\lambda = e^{-\mu \delta \varepsilon}$.

On a :

$$Z = \lambda^3 f(\lambda),$$

avec :

$$\begin{cases} f(\lambda) = 2 + \lambda + \lambda^2 + 2\lambda^3 + 2\lambda^4 + 3\lambda^5 + 4\lambda^6 + 5\lambda^7 + 4\lambda^8 + 3\lambda^9 \\ f'(\lambda) = 1 + 2\lambda + 6\lambda^2 + 8\lambda^3 + 15\lambda^4 + 24\lambda^5 + 35\lambda^6 + 32\lambda^7 + 27\lambda^8. \end{cases}$$

Le point courant (u, h) de la courbe caractéristique satisfait à :

$$\begin{cases} u = 0,7 \left(3 + \lambda \frac{f'(\lambda)}{f(\lambda)} \right) \\ h = \log f + \lambda \log \frac{1}{\lambda} \cdot \frac{f'}{f} \end{cases}$$

Il suffit de faire varier λ de 0 à 1 pour obtenir toute la courbe. On peut noter que la pente au point courant est $\frac{1}{\mu} = \frac{0,70}{\log \frac{1}{\lambda}}$.

On obtient ainsi la courbe (Γ) représentée sur la figure 4.

Par rapport à cette courbe, on peut placer le point de fonctionnement de l'alphabet Morse, dont les coordonnées sont :

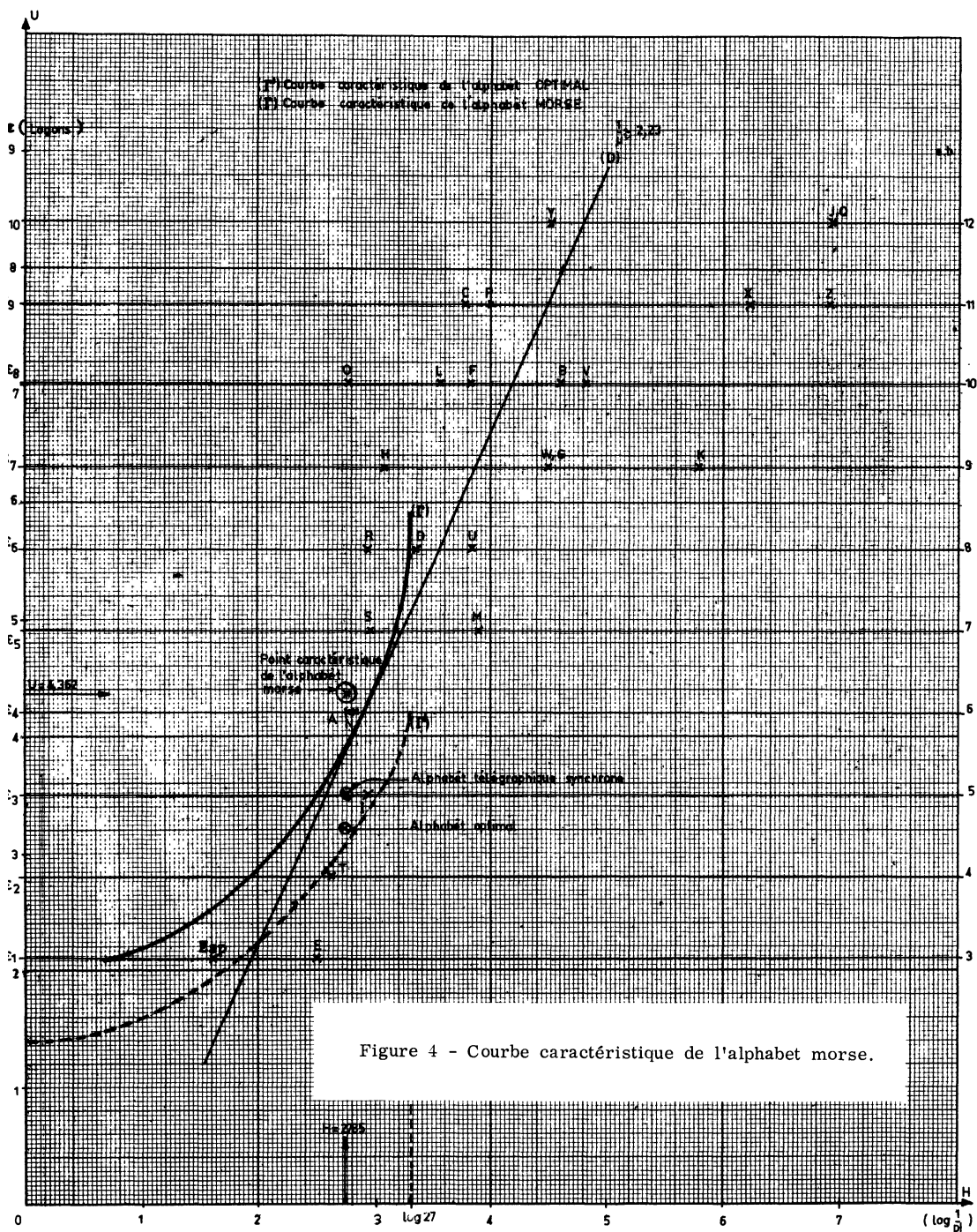


Figure 4 - Courbe caractéristique de l'alphabet morse.

$$\left\{ \begin{array}{l} H = \sum p_i \log \frac{1}{p_i} \quad (p_i \text{ probabilité d'apparition des} \\ \text{lettres).} \\ U = \sum p_i \varepsilon_i \end{array} \right.$$

On trouve facilement : $H = 2,785$, $U = 4,362$ (logons).

D'autre part, les lettres et symboles s'ordonnent dans le plan suivant leurs coordonnées propres ($\log 1/p_i$, ε_i). On les a également représentées.

Le point de fonctionnement est, bien entendu, au-dessus de la courbe caractéristique, ce qui montre que l'adaptation n'est pas optimale. Néanmoins, l'écart n'est pas très grand (travail dépensé 4,36 contre 3,96 logons nécessaires, soit un écart de 0,4 logon). La tangente la plus voisine est la droite représentée sur la figure 4 : si l'adaptation était optimale, toutes les lettres s'aligneraient sur cette droite (D) de pente $1/\mu = 2,23$. Toutes les lettres situées sur la même verticale devraient être représentées par des symboles de même énergie, correspondant à l'ordonnée fournie par la droite (D) : par exemple $\varepsilon_7 = 6,3$ logons (9 e. b.) pour les lettres F, U et M. Il en résulte que toutes les lettres situées au-dessous de la droite (D) sont représentées par des symboles trop courts, et que celles qui sont au-dessus de (D) sont représentées par des symboles trop longs. La figure met bien en évidence les anomalies les plus frappantes : durée identique de l'espacement-mot et du E, dont la probabilité d'apparition est deux fois plus faible, l'écart considérable du O, de longue durée, alors qu'il est de forte probabilité, etc.

Le débit de l'alphabet Morse est $\frac{N}{T} = \frac{k}{U} = \frac{0,7}{4,36} w = \frac{0,16}{\tau}$. La durée moyenne d'un symbole est donc $\frac{4,36}{0,7} \tau = 6,23 \tau$ (un peu plus de 6 e. b.). Il est intéressant de comparer ce "rendement" à celui des systèmes de télégraphie d'usage courant. Si un système arithmique met en oeuvre 7,5 ou 8 e. b./lettre, un système synchrone ne comporte que 5 e. b./symbole. On peut, de la sorte, représenter au plus $2^5 = 32$ lettres ou signes. Mais le système de télégraphie synchrone fait usage de symboles de même longueur (5 e. b.), sans tenir compte de la probabilité d'apparition des lettres. On peut imaginer un autre alphabet qui en tiendrait compte et serait plus performant. Reportons-nous aux alphabets donnés dans l'annexe II : pour obtenir plus de 27 symboles distincts, il faut adopter le système quadrivalent (3e cas, $v = 4$, alinéa g/ ou h/). On dispose ainsi, soit de 35, soit de 85 symboles différents.

Adoptons la deuxième possibilité qui donne une souplesse plus grande (on vérifierait que la première est équivalente, au point de vue du rendement) : 1 symbole quadrivalent de durée τ , 4 de durée 2τ , 16 de durée 3τ , 64 (dont seulement 6 utilisés) de durée 4τ . En substituant à une émission quadrivalente une émission bivalente, on aurait 1 symbole bivalent de durée 2τ , 4 de durée 4τ , 16 de durée 6τ et 6 de durée 8τ .

Un tel alphabet a une courbe caractéristique (Γ') représentée sur la figure 5 ; si on tient compte de la probabilité d'apparition des lettres, le point de fonctionnement, représenté sur la figure, correspond à $H = 2,785$, $U = 3,22$ logons (4,59 e. b.). Ce chiffre doit être comparé au rendement de l'alphabet

télégraphique synchrone (5 e. b.), et à la limite extrême donnée par Shannon ($U = H = 2,785$, 4 e. b. environ), limite irréalisable car elle correspond à des énergies sémiodynamiques de symboles non multiples de 0,7 logon. On peut résumer ces résultats dans le tableau suivant :

Type d'alphabet	"Rendement" (e. b. /symbole)	Facteur de mérite	Observations
Morse	6,23	0,80	symboles de même durée.
Télég. synchrone	5	1	
Optimal	4,59	1,09	limite irréalisable
"Shannon"	4	1,25	

On voit que la prise en considération de la probabilité d'apparition des lettres permet d'améliorer le rendement de 9 % et d'enrichir l'alphabet (85 symboles disponibles au lieu de 32) au prix, il est vrai, d'une réelle complication de l'équipement logique.

(On trouvera dans le volume XX, n° 2, la seconde partie de cet article)