

Du comportement des avions ennemis aux modélisations de la connaissance : la notion scientifique et technique d'information

Jérôme SEGAL

RÉSUMÉ. Dans le contexte de la seconde guerre mondiale, des scientifiques ont travaillé sur des problèmes de prédiction. Les avions ayant doublé de vitesse et d'altitude entre les deux guerres, il devenait nécessaire d'automatiser la liaison entre les radars et les canons de DCA. La notion d'information a alors été identifiée en tant que grandeur permettant de décrire ce couplage. Le lien a alors été fait avec d'autres définitions de la notion d'information proposés dans les années 1920 dans des domaines scientifiques (physique et statistiques) ainsi que techniques (télécommunications). Au sortir de la guerre, à la fin des années 1940, une véritable « théorie de l'information » était née, elle a été appliquée pour comprendre le fonctionnement de l'esprit humain dans l'acquisition des connaissances.

Mots clefs : théorie de l'information, cybernétique, Claude Shannon (1916-2001), Norbert Wiener (1894-1964).

ABSTRACT. From the behaviour of enemy planes to the knowledge of modelling: the scientific and technical notion of information. In the context of World War II, scientists worked on prediction theory. Planes flew twice as fast and twice as high. The coordination between radar devices and anti-aircraft guns had to be automated. Information became identified as an entity which could describe this coupling. The connection was made with other definitions of information dating from the 1920s in scientific as well as technical domains (namely physics and statistics on the one hand, telecommunications on the other). After the war, in the late 1940s, a so-called "information theory" was born. This theory has been applied to gain a better understanding of the human mind and, the way it treats knowledge among other things.

Key words: information theory, cybernetics, Claude Shannon (1916-2001), Norbert Wiener (1894-1964).

Si l'on peut parler *a posteriori* pour le XIX^e siècle d'un rendez-vous manqué entre les travaux de George Boole en algèbre et ceux de Charles Babbage relatifs à la construction de sa machine analytique, c'est bien parce que cette rencontre a lieu au XX^e siècle. Plus précisément, c'est autour de l'élaboration de la notion scientifique et technique d'information que certains domaines mathématiques et techniques finissent par former un cadre de recherche général à partir duquel des scientifiques se lancent dans la conception de nouvelles machines permettant de modéliser l'acte de connaître. Ces machines ne sont plus caractérisées par leur fonctionnement mécanique ni par leur bilan énergé-

Maître de conférences en histoire des sciences à l'IUFM Paris, actuellement en détachement à l'Ambassade de France en Autriche, Währingerstr. 30 A-1090 Wien-Autriche. E-mail : jsegal@free.fr.

tique mais bien par la nouvelle grandeur qu'elles traitent : l'information¹. On parle dès lors de machines 'cybernétiques', en référence au développement d'une théorie générale de la commande et de la régulation, la cybernétique, qui s'applique aussi bien à l'étude de la société qu'à celle des êtres vivants ou des machines. Pour les sciences cognitives, c'est précisément l'essor de la cybernétique qui explique le développement de l'approche connexionniste.

CONVERGENCES

Au cours des années 20, c'est dans trois contextes scientifiques différents qu'apparaissent des définitions de l'information bien distinctes du sens commun du mot. Il s'agit de travaux liés respectivement à la physique de langue allemande, à l'École statistique britannique et aux recherches issues du domaine des télécommunications.

En physique, Marjan von Smoluchowski (1872-1917) fait d'abord le rapprochement entre le problème du démon de Maxwell et ses propres études sur le mouvement brownien (Smoluchowski 1912)². Dans ce cadre il montre que pour que l'entropie diminue, il faut que le démon soit « renseigné [*unterrichtet*] » (Smoluchowski 1914) et c'est en reprenant ces études du physicien polonais que Leo Szilárd (1898-1964) parvient à une quantification de ce 'renseignement', considéré alors comme une grandeur entropique (Szilárd 1929). Une expression concise d'idées analogues nous est donnée par Gilbert N. Lewis (1875-1946), qui écrit en 1930 : « Un gain d'entropie signifie toujours une perte d'information et rien de plus »³.

En statistiques, c'est Ronald Aylmer Fisher (1890-1962) qui s'intéresse dès 1922 à « l'information pertinente contenue dans les données d'origine », après avoir défini la statistique comme la « réduction des données⁴ ». L'information est alors comparable à un outil introduit pour la mathématisation d'un problème et, rapidement, il en vient à définir une quantité d'information en rapport avec la précision intrinsèque d'une courbe d'erreur et la variance d'un échantillon. Il justifie ensuite en 1935 l'utilisation qu'il fait d'un mot du langage courant, rapprochant à son tour l'information de l'entropie, ou du moins de l'étude des processus irréversibles⁵.

1. Pour une approche générale sur l'histoire de la notion scientifique et technique d'information, voir Segal (2003).

2. Le second principe de la thermodynamique stipule, dans l'une de ses formes, qu'il est impossible de faire passer de la chaleur d'un corps froid à un corps chaud sans apport de travail. Le principe de l'expérience de pensée de Maxwell consiste justement à imaginer une situation dans laquelle ce phénomène aurait lieu (voir Thomson, 1894 pour la première mention du 'Démon de Maxwell' et pour l'histoire du démon de Maxwell de 1867 à nos jours, Leff et Rex 1990).

3. "Gain in entropy always means loss of information, and nothing more." (Lewis, 1930: 573).

4. On lit dans les écrits de Fisher : "[...] the object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data." (Fisher, 1922 : 311) La traduction que nous utilisons ci-dessus est de M. Armatte (Armatte, 1988: 76)

5. A propos de cette analogie, il précise : "You will notice especially that reversible processes, changes of notations, mathematical transformations if single-valued, translation of the data into foreign languages, or rewriting them in code, cannot be accompanied by loss of information ; but that the irreversible processes involved in statistical estimation, where we cannot reconstruct the original data from the estimate we calculate from it, may be accompanied by a loss, but never by a gain." (Fisher, 1935: 47)

Ces travaux doivent être abordés en rapport avec les débats sur l'eugénisme qui agitent la communauté des statisticiens. Fisher parle d'ailleurs de la même façon de la « relevant information » qu'il essaie de quantifier et des « relevant genes » dont il convient de favoriser la sélection. De la même façon, les débats sur l'entropie en physique sont à replacer dans le contexte des débats sur la nature déterministe ou indéterministe de la physique mais nous nous limitons ici à un simple rappel des principales définitions de la notion d'information.

A ce titre, il faut encore mentionner les travaux de quelques ingénieurs qui, suite à l'essor des télécommunications par télégraphie sous-marine, sont amenés à caractériser la quantité d'information contenue dans un message. Ralph Hartley (1888-1970) et Harry Nyquist (1889-1976), deux ingénieurs américains des *Bell Laboratories*, font ici figure de pionniers, tout comme Karl Küpfmüller (1897-1977) en Allemagne⁶. Ces travaux visent à parvenir à une utilisation optimale des voies de communication et l'information est définie à cet effet en rapport avec le logarithme du nombre de sélections possibles lors de l'élaboration d'un message.

Ainsi, à la fin des années 20, l'information est déjà une notion scientifique et technique. Ce n'est toutefois qu'à l'occasion des recherches liées à l'effort de guerre, pendant la seconde guerre mondiale, qu'une véritable « théorie de l'information » voit le jour. Auparavant, toujours dans le domaine technique, on assiste dans les années 30 à un rapprochement entre deux types de préoccupations, celles des mathématiciens concernés par la logique booléenne et celles des ingénieurs chargés de la réalisation de nouveaux circuits électriques permettant d'accomplir quelques opérations élémentaires.

MACHINES LOGIQUES ET LOGIQUE DES MACHINES

Avec la parution en 1921 du *Tractatus logico-philosophicus* de L. Wittgenstein, l'essor du Cercle de Vienne et plus largement encore du néo-positivisme, la logique se trouve au centre de différents systèmes philosophiques. Il s'agit cette fois d'une logique dite 'mathématique', dont G. Frege (en 1879) puis B. Russell (en 1903), entre autres, ont établi les fondements.

Le problème de Hilbert relatif à la consistance de l'arithmétique (problème numéro deux dans la liste qu'il fournit au congrès de Paris en 1900), lance de nombreuses recherches sur la 'question de la décidabilité (*Entscheidungsproblem*)'⁷. C'est K. Gödel (1906-1978), un des participants occasionnels du cercle de Vienne, qui dans une publication devenue célèbre, en 1931, montre que toute théorie axiomatisable est incomplète, établissant ainsi l'indécidabilité de la théorie des nombres⁸. Sa publication se réfère explicitement aux *Principia Mathematica* (1910-1913) de Russell et Whitehead dont il montre les limites. Les deux théorèmes de sa publication, connus sous le nom de théorèmes d'incomplétude, marquent le début d'une série de résultats limitatifs établis dans la première moitié des années 30, portant sur différents systèmes hypothético-déductifs.

6. Voir Hartley (1928), Nyquist (1924), Nyquist (1928) et Küpfmüller (1924).

7. Le problème est énoncé en ces termes : « Démontrer que les axiomes ne sont pas contradictoires ; c'est-à-dire qu'en se basant sur les axiomes, l'on ne pourra jamais arriver à des résultats contradictoires au moyen d'un nombre fini de déductions logiques. » Au sujet du 'programme de Hilbert', voir Wagner (1998) d'où nous tenons cette citation (p. 18).

8. Sur ce point, le livre de J. Ladrière fait aujourd'hui référence (Ladrière, 1957). On consultera aussi le chapitre 6 de Wagner (1998).

TURING AUX ETATS-UNIS

C'est dans ce contexte que le jeune mathématicien britannique Alan Mathison Turing (1912-1954) introduit une nouvelle notion, celle de calculabilité, définie en référence à une machine purement conceptuelle qu'il décrit dans sa publication de 1936, intitulée « Sur les nombres calculables, application à l'Entscheidungsproblem »⁹. Avec la description qu'il propose de sa « machine universelle », Turing opère un pas important dans la convergence d'études sur le fonctionnement des machines et de la pensée humaine¹⁰.

Durant l'été 1937, il rejoint à Princeton, aux Etats-Unis, un autre logicien travaillant sur la notion de calculabilité, Alonzo Church (1903-1995). C'est là que Turing se met à construire une machine permettant de multiplier des nombres binaires à l'aide de relais, dans l'atelier du département de physique de l'université de Princeton auquel il avait accès par un collègue, le physicien canadien Malcolm MacPhail¹¹. Hodges relate dans son livre comment le frère de MacPhail, Donald, élève-ingénieur en construction mécanique qui se trouvait être issu de la même université que celle de Turing, le *King's College* de Cambridge, avait été enthousiasmé par la description qui lui avait été faite de la machine réalisée par Turing.

Turing soutient sa thèse sous la direction de Church durant l'été 1938, ce même été pendant lequel il refuse un poste à l'I.A.S. que von Neumann lui proposait, préférant être dans son pays à l'approche de la guerre. Une fois de retour à Cambridge, où il a ramené son multiplicateur électrique binaire, il entreprend l'exécution d'un projet de Donald MacPhail concernant la réalisation d'une nouvelle machine selon le même principe électromécanique que pour celle réalisée à Princeton. Turing avait probablement en vue la réalisation d'appareils destinés à la cryptologie puisqu'il avait été consulté par son gouvernement dans ce domaine dès son arrivée. L'entrée en guerre de la Grande-Bretagne, le 3 septembre 1939, empêchera Turing de poursuivre ce projet.

PORTRAIT D'UN INGÉNIEUR MATHÉMATICIEN : C.E. SHANNON

Revenons sur cet appareil réalisé par Turing aux Etats-Unis. Dans sa publication de 1936, Turing utilisait déjà la numérotation binaire. Sa machine était basée sur la multiplication binaire élémentaire ($1*1 = 1$, les trois autres multiplications donnant 0) qui représente le 'ET logique' du calcul de propositions propre à la logique booléenne. Turing a pu avoir aux États-Unis un partenaire privilégié pour discuter de ces questions : Claude Elwood Shannon (1916-2001)¹². Étudiant du Massachusetts Institute of Technology (MIT), il s'était consacré en 1936/37 à son *Masters of Science* portant sur une description générale des relais et circuits de commutation par l'algèbre de Boole.

Avant de revenir plus en détail sur ce travail, qualifié de « monumental » ou décrit comme étant « la thèse la plus importante et la plus célèbre de ce siècle »

9. "On Computable Numbers with an Application to the Entscheidungsproblem". Voir Hodges, 1983, chapitres 2 et 3. Ce livre déjà traduit en quatre langues est devenu la référence concernant l'œuvre de Turing.

10. Voir Lassègue (1998) et l'article de cet auteur dans le présent numéro.

11. Voir le texte de la lettre de MacPhail adressée à Hodges en 1977. (Hodges, 1994: 161-162).

12. Hodges rapporte que les deux scientifiques se sont probablement rencontrés pour la première fois aux Bell Labs, en janvier ou février 1943, à moins qu'ils ne se soient vus en 1936/1938 durant les séjours de Turing à Princeton ou en 1942 lors d'autres séjours de Turing aux Etats-Unis. Dans un entretien avec R. Price, en 1984, Shannon déclare : "We talked about things like the human brain and computing machines." (Price, 1985: 171). Sur Shannon, voir Segal (2002).

cle », selon que l'on se réfère à l'expression de Marvin Minsky ou à celle d'Howard Gardner, revenons un instant sur le personnage important de notre étude, Claude E. Shannon¹³.

Que sait-on précisément de la vie de Shannon ? En plus de deux notices biographiques et deux entretiens, la thèse de Hagemeyer nous apporte quelques renseignements mais on est encore loin de disposer d'autant d'informations que sur les autres grands scientifiques de son époque¹⁴.

Claude Elwood Shannon est né le 30 avril 1916 à Petoskey, dans l'État du Michigan. Son grand-père était l'inventeur d'un modèle de machine à laver et possédait quelques brevets correspondant à des travaux plus ou moins saugrenus. Lorsqu'il entre à l'Université du Michigan, en 1932, le jeune Shannon a déjà construit, entre autres, un bateau radio-commandé et un système télégraphique pour communiquer avec un de ses amis habitant à un kilomètre, utilisant des fils barbelés mitoyens. Il travaillait aussi de temps à autre pour réparer des radios et autres appareils électriques, vivant dans l'admiration d'Edison dont il apprendra plus tard qu'ils ont des ancêtres communs. Toute sa vie, il gardera un intérêt pour les aspects pratiques sans oublier les réflexions théoriques, l'archétype de cette démarche étant la théorie du jonglage (!) qu'il publie dans les années 80, exhibant en même temps différents 'automates jongleurs'¹⁵. Ce double intérêt pratique et théorique se retrouve dans sa formation universitaire : il est diplômé en 1936 des titres de *Bachelor of Science* à la fois en mathématiques et dans les sciences de l'ingénierie électrique (*Electrical Engineering*).

Durant les années 20, c'est aussi dans un département d'*Electrical Engineering*, mais cette fois-ci au MIT, qu'un ingénieur, Vannevar Bush (1890-1974), avait mis au point un calculateur analogique mécanique qui servait à résoudre des équations différentielles (Owens 1986).

C'est à ce titre, suite à une petite annonce, que Shannon entre au MIT, en 1936, comme assistant chercheur au département de Bush. Une publication de 1941 montre comment il est parvenu à trouver une théorie mathématique de cette machine, permettant d'aboutir à une expression de la vitesse d'exécution (Shannon 1941). Au MIT, il obtiendra encore une double formation, avec un *Masters of Electrical Engineering* complété par une thèse de mathématiques obtenue en 1940 (*Ph. D.*). Lorsque A. Liversidge lui demande, au cours d'un entretien en 1987, en quoi consistait l'analyseur différentiel, Shannon répond :

« La machine principale était mécanique avec des disques tournants et des intégrateurs et il y avait un circuit de commande compliqué avec des relais. Je devais comprendre les deux. La partie relais

13. La citation de Minsky est dans les *Œuvres choisies* de Shannon (Sloane & Wyner, 1993) et celle de Gardner dans (Gardner, 1985: 170).

14. Un entretien et une biographie de 7 pages figurent dans l'introduction des *Œuvres choisies* (Sloane & Wyner, 1993). La thèse de Hagemeyer contient une annexe intitulée « biographie scientifique de C. Shannon » (Hagemeyer, 1979 : 489-503). Voir également l'entretien paru dans *Cryptologia* (Price, 1985), la notice parue dans *Scientific American* (Horgan, 1990) et plus récemment (Segal, 2002).

15. Bricoleur de génie, Shannon est aussi l'auteur d'un unicycle à deux places, d'une machine à résoudre le Rubik's cube, d'une 'souris', Thésée, capable de s'orienter dans un labyrinthe (présentée à la huitième conférence Macy, cf. ci-dessous chapitre suivant), d'une machine calculant avec les chiffres romains ou encore d'un frisbee à réaction ! Il s'est rendu célèbre auprès de ses collègues des Bell Labs (où il entre en 1941) pour avoir jonglé régulièrement dans les couloirs avec trois balles, sur un monocycle (voir Pierce, 1993 sur l'importance de la construction de ces gadgets dans le développement de sa carrière d'ingénieur).

m'intéressa. Je connaissais la logique symbolique à cette époque depuis un cours à Michigan et je réalisai que l'algèbre de Boole était justement le truc [the thing] pour s'occuper de circuits de relais ou de circuits de commutateurs. Je me rendis à la bibliothèque et pris tous les livres que je pouvais sur la logique symbolique et l'algèbre de Boole, commençant à combiner les deux, et j'écrivis mon mémoire de maîtrise [Master's thesis] là-dessus. C'était le début de ma grande carrière ! (rires)¹⁶ »

UN MASTERS OF SCIENCE QUI FAIT DATE

Son étude intitulée « De l'analyse symbolique des circuits de relais et de commutateurs » (Shannon 1937) concerne « la plupart des circuits » avec relais et commutateurs, notamment les centraux téléphoniques automatiques et les circuits de commande des moteurs qui apparaissent dans son introduction comme des référents techniques. Il s'intéresse à l'analyse et la synthèse des circuits, c'est-à-dire à la détermination des caractéristiques requises mais aussi à l'optimisation de la réalisation d'un circuit à partir de propriétés données. Pour ce faire il montre comment « n'importe quel circuit peut être représenté par un ensemble d'équations » dont les termes représentent directement les relais et commutateurs du circuit. Si un ingénieur japonais, A. Nakashima, avait élaboré une théorie comparable, aussi générale, dès 1935, c'est la postérité du modèle de Shannon qui justifie notre étude, outre le fait que cette théorie aura une influence décisive sur la suite de ses travaux¹⁷.

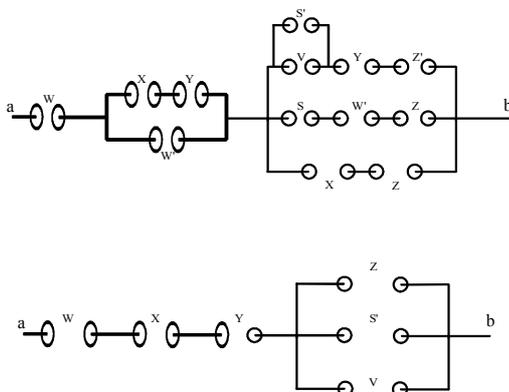
Puisqu'il prévient d'emblée que sa théorie ne concerne que les circuits qui ne contiennent que des relais ou des commutateurs, le binaire s'applique tout à fait : entre deux points A et B , soit le courant passe, soit il ne passe pas. Il note alors respectivement selon ces deux cas, $X_{AB} = 0$ ou $X_{AB} = 1$, la notation indicelle devenant vite superflue. Il introduit ensuite le signe '+' et le signe '.' pour représenter l'association de morceaux de circuit en série ou en parallèle, suivie de la notation X' pour désigner un circuit fonctionnant en opposition par rapport à X . Avec ces notations, l'ensemble formé par tous les circuits de ce type est isomorphe à un corps commutatif. Les démonstrations qu'il indique utilisent cependant le fait que les variables ne peuvent prendre que deux valeurs et il entend procéder par « induction parfaite », en essayant toutes les combinaisons. Il y a pour Shannon dans le cas qui l'occupe une « parfaite analogie » entre l'étude de ses circuits et le calcul des propositions¹⁸. Enfin, pour finir avec la partie la plus théorique de son mémoire, il donne un théo-

16. « The main machine was mechanical with spinning disks and integrators, and there was a complicated control circuit with relays. I had to understand both of these. The relay part got me interested. I knew about symbolic logic at the time from a course at Michigan, and I realized that Boolean algebra was just the thing to take care of relays circuits and switching circuits. I went to the library and got all the books I could on symbolic logic and Boolean algebra, started interplaying the two, and wrote my Master's thesis on it. That was the beginning of my great career! (Laughs) » (nous soulignons Sloane & Wyner, 1993 : xxv)

17. Dans son exposé sur l'histoire de l'algèbre des relais ('Schaltalgebra'), Heinz Zemanek rappelle que Nakashima publie dès 1935 une « théorie de la composition des circuits relais » (Zemanek, 1991 : 53 et les références 12 et 13 p. 67). Dans la version anglaise que nous avons consultée, on observe que, très schématiquement, c'est la théorie des groupes d'Abel qui figure à la place de l'algèbre de Boole utilisée par Shannon. Les référents techniques de Nakashima sont bien les mêmes que ceux de Shannon : "automatic telephone exchange, remote control systems" (Nakashima, 1936).

18. "(...) it is evident that a perfect analogy exists between the calculus for switching circuits and this branch of symbolic logic." (Mémoire original p. 9 ou Sloane & Wyner, 1993 : 474, nous soulignons). Shannon précise qu'il est possible de parler directement de proposition si on suppose que X_{AB} représente « la proposition selon laquelle le circuit AB est ouvert. »

rème de décomposition nécessaire à l'analyse de circuits. S'inspirant des séries de Taylor pour les fonctions infiniment dérivables, il ramène toute fonction portant sur n circuits élémentaires quelconques à une somme (donc à une mise en série) de multiplications entre les éléments de départ pondérées par les valeurs prises par la fonction en des points précis.



Masters de Shanon (Shannon 1937 : 15-16)

L'exemple qu'il donne montre bien de quoi il s'agit : les deux circuits ci-contre sont équivalents. Le deuxième est évidemment beaucoup plus simple, ce qui signifie un coût moindre (car moins de composants) et surtout une plus grande fiabilité.

C'est ainsi que la logique formelle permet d'éliminer la redondance matérielle.

Après une partie consacrée aux circuits en étoile, en delta, ou plus généralement à plusieurs terminaux, il étend sa théorie au

cas des branchements conditionnels et à la synthèse de réseaux. Là encore, la résolution de systèmes d'équations (parfois sous forme matricielle) lui permet très concrètement de déterminer la configuration optimale de différents circuits. Il démontre en conclusion la valeur heuristique de sa théorie mathématique, à travers cinq exemples. Il considère d'abord un circuit sélecteur à vingt éléments, qu'il réduit à quatorze, obtenant ainsi « probablement le circuit le plus économique de toutes les sortes ». Il s'agit ensuite de la commande d'une serrure à code numérique, du circuit d'une machine à compter les votes lors d'un référendum, d'un additionneur binaire qu'il nomme « additionneur électrique de la base deux » et enfin d'une machine à décomposer en facteurs premiers tout entier inférieur à 100 000 000. Son additionneur n'est bien sûr pas sans rappeler le multiplicateur électrique de Turing mais nous ne savons rien des contacts qu'ils ont pu avoir. Sans doute est-il plus prudent de parler ici de travaux parallèles. En ce qui concerne la machine à décomposer en facteurs premiers, basée sur l'utilisation du crible d'Eratosthène, elle vise à soulager l'homme d'un travail purement intellectuel (en opposition à un travail seulement 'physique'). Shannon rappelle que Kulik avait mis vingt ans pour établir une pareille table (avec beaucoup d'erreurs) et qu'il suffirait de deux mois à sa machine pour remplir cette même tâche. Il ne s'agit déjà plus de remplacer l'homme dans ses dépenses énergétiques mais bien dans ses tâches intellectuelles.

Soutenu au MIT, le 10 août 1937, le mémoire de Shannon connut un succès extraordinaire pour un *Masters*. C'est à ce titre qu'il reçoit en 1940 le Prix Alfred Noble des sociétés d'ingénieurs américains. Est-ce l'œuvre d'un ingénieur ? d'un mathématicien ? Disons que dans son mémoire il aborde, comme Turing, la notion de mécanisation, même si Turing reste sur un plan plus théorique. A l'image de Gödel, le mathématicien britannique apporte des résultats plutôt limitatifs : il montre que si tous les problèmes 'algorithmiques' peuvent être implémentés sur une machine au fonctionnement théorique extrê-

mement simple, ceux-ci n'admettent pas tous une solution. L'apport de Shannon est plus 'positif' en ce sens qu'il lance un programme de recherche : puisque l'algèbre de Boole apporte de nouveaux outils d'analyse pour l'étude des circuits à relais et commutateurs, un champ entier de l'ingénierie électrique devient une application de la logique symbolique et mathématique. Une application de la théorie qu'il propose laisse entrevoir de nouvelles réalisations pratiques.

Dans les deux cas, la logique permet un rapprochement entre science et technique, dont l'informatique tirera profit. H. Goldstine écrit ainsi dans son livre sur l'ordinateur de Pascal à von Neumann, au sujet du mémoire de Shannon, qu'il s'agit « d'un des plus importants mémoires de maîtrise jamais écrit, un événement marquant en ce sens que cela a aidé à faire passer la conception des circuits numériques du stade de l'art à celui de la science.¹⁹ » Surtout, on passe ici des questions relatives à la propagation à celles qui concernent la structure des circuits. C'est bien par l'utilisation de la logique que ce changement s'effectue et, comme le note Aspray, ce n'est pas un hasard si Shannon, Turing mais aussi Wiener, McCulloch, Pitts et von Neumann, dont nous aborderons les apports plus loin, ont tous reçu une formation en logique mathématique (Aspray 1985 : 118). Il n'est pas encore question d'information dans un sens scientifique général mais on peut tout de même comprendre la ferveur de l'accueil scientifique fait au mémoire de Shannon. Cet accueil est cependant à replacer dans un contexte bien particulier, celui de la seconde guerre mondiale.

LE CONTEXTE DE LA SECONDE GUERRE MONDIALE

Durant la guerre, le milieu scientifique américain donna naissance, en dehors des recherches concernant la bombe atomique, à deux domaines d'études intimement reliés : la cybernétique et l'informatique (ou '*computer science*', le mot informatique ne datant que de 1962)²⁰. Chacun des ouvrages consacrés à l'histoire de ces deux domaines scientifiques, nous renseigne sur les travaux relatifs à la conceptualisation de la notion d'information²¹.

Réorganisation des recherches

Pour fédérer l'ensemble des travaux menés dans des structures civiles, Roosevelt crée le 27 juin 1940 le *National Defense Research Committee* (NDRC). La présidence de ce comité est confiée à Vannevar Bush qui avait déjà participé aux efforts de guerre pendant la première guerre mondiale.

En mai 1941, les compétences de Bush sont élargies pour inclure les recherches médicales regroupées autour du *Committee on Medical Research*. Les deux comités (celui-ci et le NDRC) dépendent directement de l'*Office of Scientific Research and Development* (OSRD). Ces nuances administratives ont ici leur importance car c'est ce dernier organisme, que dirigeait Bush, qui

19. "(...) one of the most important master's theses ever written, a landmark in that it helped to change digital circuit design from an art to a science." (Goldstine, 1972: 119).

20. Contraction des mots 'information' et 'automatique' proposée par Philippe Dreyfus (Ifrah 1994 : 725).

21. On pourra se reporter à Edwards (1996) et Mindell (2002) pour l'importance du contexte américain, à Dupuy (1994), et Heims (1991) pour l'origine de la cybernétique. En ce qui concerne l'histoire de l'informatique, toujours parmi les livres que nous avons consultés, voir Augarten (1985), Breton (1987b), Campbell-Kelly & Aspray (1996), Ceruzzi (1983), Heintz (1993), Ifrah (1994), Metropolis (1980), Petzold (1994) et Rammuni (1988). Parmi les articles, citons seulement Cohen (1988), Lévy (1989) et la revue américaine *Annals of the History of Computing*.

référerait directement à Roosevelt²². C'est donc en effet bien avant l'attaque de Pearl Harbor, le 7 décembre 1941, point de départ de l'entrée en guerre des Etats-Unis, que le milieu scientifique et industriel américain se réorganise entièrement autour d'objectifs militaires²³.

Début 1946, Bush note déjà que la guerre a en quelque sorte joué un rôle de catalyseur pour un nouveau type d'interdisciplinarité. Il se réfère ainsi à « une coopération effective et professionnelle entre scientifiques, ingénieurs, industriels et militaires comme on n'en avait jamais vu jusque là, qui démontraient l'esprit de l'Amérique en action à son plus fort et meilleur niveau.²⁴ »

Le point fort de ce système semble être son indépendance : en 1940, les cinq divisions du NDRC, dirigées chacune par le président d'une grande université ou d'un centre de recherche prestigieux, ont toutes des capacités d'initiative dans le domaine qui est le leur et c'est l'armée qui conclut des contrats avec ces différents instituts²⁵. Au total ce système place indirectement plus de six mille scientifiques sous la responsabilité de Bush lorsque celui-ci dirige l'OSRD. Il n'est pas étonnant d'apprendre en consultant les documents fournis en annexe du livre de Baxter qu'à la fin de la guerre le MIT est le premier partenaire des forces armées avec près de 120 millions de dollars de contrats.

Ces différentes recherches sont très parcellisées. Pour des raisons évidentes de sécurité, chaque équipe ignore tout de ce que font les autres et seuls les responsables de section peuvent avoir une vue d'ensemble des travaux menés. Dans cette optique, les sections du NDRC passent d'ailleurs au nombre de 18, à l'automne 1942, un an et demi après la création de l'OSRD. Parmi celles-ci, notons l'existence d'une division consacrée à la commande de tir (Division 7).

Deux études sur le 'comportement' des avions ennemis

Depuis la première guerre mondiale, les avions allemands ont doublé leur vitesse ainsi que l'altitude à laquelle ils peuvent bombarder. Avec le début de la bataille d'Angleterre, la défense anti-aérienne (défense contre avions, DCA) apparaît rapidement comme un enjeu essentiel. Les méthodes manuelles de pointage s'avèrent rapidement inefficaces et quand bien même la direction convenait, l'obus n'éclatait pas forcément à proximité de l'avion à abattre²⁶. D'un point de vue mathématique, les deux problèmes posés sont ceux de

22. Bush est donc en contact direct avec Roosevelt et c'est James Bryant Conant, président de l'université d'Harvard depuis 1933, qui le remplace à la tête du NDRC.

23. L'historien officiel de l'OSRD, James P. Baxter, a écrit en 1946 un livre qui fait date sur l'histoire de l'engagement des scientifiques pendant la dernière guerre mondiale (Baxter, 1946). Pour d'autres sources, voir la biographie de V. Bush (Bush 1970), (Noble 1984 chapitre 1) et toutes celles répertoriées par S. Bennett dans (Bennett 1993 : 128, note 26) ou celles indiquées dans (Dahan Dalmedico 1996).

24. Bush (cité dans Baxter, 1946: p. xvi) : "[...] an effective professional partnership of scientists, engineers, industrialists, and military men, such as was never be seen before, which exemplified the spirit of America in action at its strongest and best." On lit des propos analogues dans Bush (1970:31-32).

25. Division A, Armes et artillerie, présidée par R.C. Tolman, Calif. Technology Univ.

B, Bombes, explosifs et chimie, J.B. Conant, Harvard University.

C, Télécommunications et transport, F.B. Jewett, Bell Labs et Académie des Sciences.

D, Détection, commande et instruments, K.T. Compton, MIT.

E, Brevets et inventions, C.P. Coe, US-Commission of Patents (Brevets).

26. Rappelons que les principaux dispositifs de DCA ne fonctionnent pas par collision avec la cible (ce qui serait trop aléatoire) mais par explosion près de celle-ci. Sur l'histoire de la conduite de tir pendant la seconde guerre mondiale, on pourra consulter Baxter (1946, chapitre 14), Fagen (1978, chapitre 3) et Hagemeyer (1979: 278-83, 346-63 et 496).

l'estimation et de l'extrapolation. Comment estimer les trois coordonnées de l'avion ? Connaissant la trajectoire de l'avion jusqu'à un temps t , que peut-on dire de la position de l'avion à $t + t$, t étant la durée de vol de l'obus ? Comme le note très justement Baxter (dès 1946), « un appareil commandant un dispositif de DCA peut être considéré comme la combinaison d'un instrument de mesure et d'un calculateur.²⁷ »

Au sein du NDRC c'est Warren Weaver (1894-1978) qui dirige la section D2, 'conduite de tir' [*fire control*] créée le 12 septembre 1940. Parmi les consultants, on trouve Wiener et Shannon dont Weaver avait demandé le détachement de l'IAS en raison de son expérience acquise avec l'analyseur différentiel.

Début 1941, deux grands projets voient le jour sous forme de contrat avec le NDRC : il s'agit de la construction du 'Anti-Aircraft Predictor' par Norbert Wiener (1894-1964) et Julian Bigelow (né en 1913) et du M9, aux Bell Labs, par C.E. Shannon, H.W. Bode et R.B. Blackman. Ces projets nécessitant l'emploi de calculateurs analogiques électriques sont désignés dans le rapport final de 1946 comme « les deux travaux les plus importants de la division »²⁸.

Les travaux de l'équipe de Shannon reposent sur l'idée selon laquelle on devait utiliser des méthodes relevant de la géométrie pour les étapes de lissage des données et de prédiction, utilisant la théorie de l'amplification avec rétroaction que Bode était en train de mettre au point²⁹. Cet ingénieur des Bell Labs explique lors d'une des premières réunions de travail, le 4 juin 1941 :

« Les changements dans le déroulement de la trajectoire ou les déviations n'étaient pas pris en compte et du coup une approche statistique n'était pas nécessaire : l'appareil relevait d'une approche purement trigonométrique.³⁰ »

Les principaux paramètres retenus étaient la vitesse et l'accélération de l'avion. D.A. Mindell montre par exemple dans sa thèse comment Shannon fit le lien entre les techniques utilisées pour l'analyseur différentiel et celles qui pourraient convenir dans ce projet, sans se soucier de l'élaboration d'une théorie générale. Au contraire, il applique sa théorie algébrique des relais (son *Masters*) pour l'analyse et la synthèse des circuits électriques impliqués dans l'appareil conçu aux Bell Labs³¹.

La conception de Wiener pour son appareil était tout autre, marquée par une plus grande ambition sur le plan mathématique³². Dès novembre 1940, Wiener

27. "A device controlling antiaircraft fire may be regarded as a combination of a measuring instrument and a computing machine." (Baxter, 1946: 212).

28. La section D2 était devenue 'division' 7 avec la création de l'OSRD. La présidence passe de Weaver à Harold L. Hazen du MIT mais Weaver reste membre de la division.

29. Le livre correspondant est Bode (1945) mais ses premiers brevets datent de 1940 (Millman, 1984: 30-31 et 81). L'utilisation des amplifications par feedback est décrite dans Fagen (1978: 140-147), dans la conception du T10, ancêtre du M9.

30. Bode dans le rapport de la conférence du 4. 6. 1941 (cité dans Hagemeyer 1979: 356) : "No account was taken of changes in course performance or curvature, and hence no statistical approach was necessary ; the apparatus was a purely trigonometric based upon approach."

31. Voir Mindell (1996: 446-449 et 467) la reproduction d'un schéma fait par Shannon en mai 1941 pour décrire le fonctionnement du mécanisme de lissage des données relatives à la hauteur, *Height Data Smoothing Mechanism* (Shannon Papers, Archives du MIT).

32. P. Galison a consacré une publication au travail de Wiener en DCA, montrant l'importance de celui-ci dans l'éclosion de la pensée cybernétique, selon lui une des trois 'sciences de guerre' avec la recherche opérationnelle et la théorie des jeux. Les premiers projets de Wiener sont ainsi mentionnés

écrit dans un rapport remis au responsable du projet pour le MIT, S.H. Caldwell, à propos des propriétés de suivi et d'extrapolation de l'appareil qu'il concevait :

« Ceci est fait par un réseau linéaire dans lequel l'*information* est donnée par tout le mouvement passé de l'avion et qui génère un terme correcteur indiquant à quelle distance de sa position initiale se situera l'avion lorsqu'un obus arrivera dans son voisinage...³³ »

'L'information' à laquelle se réfère Wiener est déjà éloignée du sens courant et s'assimile davantage à une notion mathématique : il s'agit d'une série aléatoire dont on ne connaît que les propriétés statistiques. A cette même période, il parvient grâce à l'analyseur différentiel à 'simuler' le comportement de son appareil face à quatre types de trajectoires. Il cherche « un moyen d'analyser les possibilités intrinsèques de différents types de comportements (...) en relation avec la construction d'un appareil pour accomplir des buts spécifiques relatifs à la répétition et à la modification de schémas temporels.³⁴ »

En fonctionnement, le *AA Predictor* (surnom donné à l'appareil) était soumis à deux types d'irrégularités : les crans de la manivelle du pointeur et les manœuvres en zigzag que pouvait faire le pilote pour échapper au feu de l'appareil. La série des données obtenues (au sens mathématique du mot 'série' - Wiener introduit d'ailleurs le terme '*time series*' pour désigner des fonctions du domaine temporel dont les valeurs correspondent à des mesures) devait donc être filtrée pour obtenir une série plus régulière qui se laisse plus facilement extrapoler³⁵. C'est là l'essentiel de la théorie de la prédiction, théorie qui faisait déjà l'objet d'un projet gouvernemental placé sous la direction de Julian Bigelow, ingénieur d'IBM formé au MIT et lui-même pilote amateur³⁶. Wiener, qui cherchait quelqu'un qui puisse traverser avec lui « les frontières habituelles entre disciplines », recrute Bigelow en janvier 1941³⁷. Rapidement, en étudiant les trajectoires d'avions pris en chasse, les deux scientifiques remarquent que l'action du pilote revient à celle d'un rétrocontrôle : « le pilote se comporte comme un servomécanisme »³⁸. Wiener utilise alors un brevet qu'il avait déjà sur un servomécanisme et, à l'aide d'une modélisation avec des

dans Galison (1994: 228-9), Wiener (1948: 10), Wiener (1956: 206-7) et Wiener (1958: 6-7). Sur les travaux de Wiener en DCA voir également Masani (1990, chapitre 14) et plus récemment Kay (1997).

33. Rapport du 22 novembre 1940 (cité dans Masani 1990: 182) : "This is done by a linear network into which *information* is put by the entire past motion of the airplane and which generates a correction term indicating the amount that the airplane is going to be away from its present position when a shell arrives at its neighborhood..." Nous soulignons cet extrait qui est aussi commenté dans Bennett (1993, chapitre 7) et Dahan Dalmedico (1996: 165).

34. "[...] an adequate attempt to analyze the intrinsic possibilities of types of behavior (...) in connection with the design of apparatus to accomplish specific purposes in the way of repetition and modification of time patterns." (Lettre citée dans : Galison, 1994: 242)

35. C'est ainsi que Wiener rédige *The Extrapolation, Interpolation and Smoothing of Stationary Time Series with engineering applications*, NDRC Report, Div. 7, 1.2.1942. Ce document qui fut classé secret pendant sept ans a connu ensuite dix éditions sous forme de livre, de 1949 à 1977, sous le titre *Time Series* (Wiener, 1949).

36. (Wiener, 1956: 209; Galison 1994: 235). Dans sa thèse, D.A. Mindell indique les références d'un entretien entre S.J. Heims et J. Bigelow datant du 12.11.1968 (Steve J. Heims Papers, MIT Archives, MC-361 Box 1 Folder 5), voir Mindell (1996: 434-445) pour une analyse historique du projet de Wiener et Bigelow.

37. Lettre du 12 mars 1941 citée dans Galison (1994: 235).

38. Rapport de Wiener au NDRC, (Masani, 1990: 189), repris dans Galison (1994: 236).

spots de couleurs représentant l'avion et la ligne de mire du canon, ils parviennent à 'programmer' pour leur machine des 'comportements' de pilotes.

Sur le plan technique, leur projet est toutefois abandonné car la comparaison avec celui des Bell Labs tourne rapidement en leur défaveur. C'est une centaine de M9 issus des Bell Labs qui sont envoyés en Grande-Bretagne fin 1941 et le M9 acquit ses lettres de noblesse lorsque, couplé au radar SCR-584, il détruisit plus de trois quarts des V1 envoyés par les Allemands pendant la seconde bataille d'Angleterre (le premier V1 tombe le 12 juin 1944, déstabilisant au départ très fortement les équipes de contre-mesure)³⁹.

Sur le plan théorique, avec le projet de Wiener, la théorie des servomécanismes devient comme le note Galison « la mesure de l'homme ». Wiener écrit du reste dans un de ses comptes-rendus que la modélisation qu'ils ont obtenue devrait permettre de tester différents réflexes humains :

« Ceci suggère l'utilisation de tels appareils pour le diagnostic des différences individuelles des comportements réflexes et des conditions pathologiques qui affectent l'arc réflexe. Plusieurs autres extensions de ces idées apparaîtront aux physiologistes, en neuropathologie et aux experts des tests d'aptitude.⁴⁰ »

Cette analogie entre les comportements de l'homme et des machines ne joue pas encore le rôle central qu'on lui connaîtra plus tard mais G. Stibitz note déjà : « Wiener fait valoir que leur appareil constitue probablement l'une des approches les plus réussies jamais faite du comportement physiologique.⁴¹ »

C'est à l'aide du cardiologue mexicain Arturo Rosenblueth (1900-1970) que le rapprochement est fait avec des phénomènes équivalents en biologie comme l'ataxie, terme désignant une absence de coordination des mouvements caractéristique de certaines maladies neurologiques. Des travaux sont alors menés, parallèlement aux recherches militaires, grâce au soutien d'une fondation privée, la Macy Foundation, qui organise en 1942 une conférence sur l'inhibition cérébrale. L'exposé présenté par Rosenblueth est remanié par Wiener et Bigelow et donne naissance, en 1943, à un article maintenant bien connu mais souvent objet de contresens.

Des études comportementales non béhavioristes

« Cet essai a un double but. Le premier est de définir l'étude comportementale des phénomènes naturels et de classer les comportements. Le second est de souligner l'importance du concept d'intention [*purpose*].⁴² »

C'est par ces mots que débute l'article des trois scientifiques. Ils entreprennent une classification des différents comportements et définissent les

39. Pour une comparaison des deux projets, voir (Millman 1984 : 43) et (Hagemeyer 1979 : 280-81 et 421). Sur l'installation du M9 et son couplage avec le radar SCR584, voir Baxter (1946: 214)

40. "This suggests the use of such apparatus in the diagnosis of individual differences in reflex behaviors, and of pathological conditions affecting reflex arc. Many extensions of these ideas will suggest themselves to the physiologist, the neuropathologist, and the expert in aptitude test." Rapport de Wiener en date du 10 juin 1942 cité dans Galison (1994: 237-8).

41. Stibitz est le premier constructeur d'un calculateur binaire à relais, en 1937, aux Bell Labs, où Shannon faisait déjà un court séjour. Stibitz le 1^{er} juillet 1942 : "Wiener points out that their equipment is probably one of the closest mechanical approaches ever made to physiological behavior." (Galison, 1994: 243).

42. (Rosenblueth, Wiener & Bigelow, 1943: 147). Nous nous référons à la traduction de J. Piquemal et Dupuy note avec raison que la traduction du terme 'purpose' par intention est plutôt malheureuse puisqu'il est question ici de finalité non intentionnelle. (Dupuy, 1994: 41).

servomécanismes comme des « machines à comportement intentionnel intrinsèque. » La notion de rétroaction est aussi clarifiée et les auteurs notent qu'on distingue dans l'utilisation qu'en font les ingénieurs un sens large où une partie de l'action produite sert d'action subie et un sens plus précis où le comportement est réglé par la marge d'erreur.

Au sujet de l'analogie entre l'appareil de DCA de Wiener et Bigelow et les connaissances physiologiques de Rosenblueth, à l'origine de cette publication, ils parlent « d'analogie éclatante » au niveau des réactions observées et suggèrent que « la principale fonction du cervelet est de régler les mécanismes nerveux à rétroaction impliqués dans l'activité motrice intentionnelle. » (Rosenblueth, Wiener & Bigelow 1943 : 151)

Ils donnent alors un classement des différents comportements, « à partir des variations d'énergie mises en jeu » et non pas d'information, ce qui constitue presque comme le note Dupuy un crime de 'lèse-cybernétique' puisque l'apport essentiel de la cybernétique est bien d'introduire la notion d'information en opposition et en complément à celle de matière ou d'énergie (Dupuy 1994 : 37).

En ce qui concerne la notion d'information, l'exemple qu'ils donnent d'un limier qui rend ses « réactions prédictives plus efficaces » pour suivre une piste les amènent à affirmer que « l'action-subie ('input') d'ordre chimique, olfactif n'apporte que l'information spatiale : distance dont l'intensité est l'indice.⁴³ » C'est le seul passage de l'article où il est question d'information dans un sens éventuellement autre que celui du sens commun. On peut ici raisonnablement supposer que la notion d'information n'est pas encore clairement identifiée comme étant la grandeur circulant dans les boucles de rétroaction⁴⁴.

La classification qu'ils donnent vise surtout à déterminer une méthode générale s'appliquant aussi bien aux êtres vivants qu'aux machines, avec la possibilité de mener des tests comportementaux de complexité croissante. A terme, ils envisagent aussi la construction de « robots ayant non seulement un comportement mais aussi une structure analogue à celle d'un mammifère. » A partir de leur classification des différents modèles comportementaux, ils passent d'ailleurs à celle de reproduction en envisageant la synthèse d'un chat en laboratoire. C'est ce type de réflexion qui inspirera un courant fort de la science-fiction, marqué par le néologisme 'cyborg'⁴⁵. Alors que le modèle reposait sur l'idée d'une structure fondamentale inchangée, la reproduction offre la possibilité d'introduire des variations et de quitter ainsi le déterminisme implacable du modèle pour une conception diachronique⁴⁶.

De plus, le fait que leur objet d'étude soit simplement le comportement a donné lieu à des contresens quant à la place du comportementisme dans leur article.

43. (Rosenblueth, Wiener et Bigelow, 1943: 152, nous soulignons).

44. Edwards prend moins de précautions lorsqu'il écrit dans son analyse de cette publication : "For Rosenblueth et al. behavior now became any transformation of energy or information from the environment." (Edwards, 1996: 183), les italiques sont dans le texte. Voir l'ensemble de cette section intitulée 'Cybernetics : the Behavior of Machines' p. 180-85.

45. Le terme 'cyborg', contraction de 'cybernetic organism' est apparu dans le domaine des recherches aérospatiales en 1960. Edwards axe son livre autour de ces idées. Rappelons ici que c'est en 1940 qu'est présenté aux Etats-Unis le premier robot, Sparko, censé ressembler à un chien.

46. On peut, d'un côté, considérer la théorie présentée par les auteurs comme déterministe, puisque les mécanismes de rétroaction assurent le maintien de valeurs données quels que soient les événements extérieurs et, d'autre part, y voir de l'indéterminisme puisque les servomécanismes sont dotés de comportements intentionnels intrinsèques.

Si les auteurs entendent effectivement développer une méthode comportementale (*'behavioristic'*), qu'ils opposent à la méthode fonctionnelle (qui suppose de s'intéresser aux « structures spécifiques » et à « l'organisation propre »), leur article doit toutefois se lire comme une rupture avec le béhaviorisme qui dominait l'ensemble de la psychologie depuis les écrits de John Watson du début des années 20, suivie notamment de l'approche positiviste de Skinner, dans les années 1930, centrée presque exclusivement sur la conception 'stimulus-réponse'.

A cela deux raisons. D'une part, en limitant dans un premier temps l'objet d'étude de la psychologie aux comportements observables, le béhaviorisme avait permis l'application de méthodes expérimentales rigoureuses mais avait, du même coup, contraint les psychologues à reléguer les notions d'esprit ou d'intention dans une métaphysique abandonnée aux philosophes. Or, si les auteurs de l'article refusent également de s'engager dans des discussions épistémologiques sur le déterminisme ou la notion de cause finale, c'est justement autour de l'opposition entre intentionnel et non-intentionnel qu'ils construisent leur classification. La notion de volonté leur sert aussi de trait distinctif lorsqu'ils concluent que l'activité volontaire suppose une rétroaction. D'autre part, la prédominance des expériences menées sur le comportement animal avait permis d'étendre aux animaux en général des similitudes qui ne concernaient au départ que l'homme et la machine mais avait aussi, de la même façon, retardé sinon empêché l'étude du langage humain⁴⁷. C'est un des mérites de Karl Lashley que d'avoir montré à ce propos, en 1948, lors du symposium Hixon, combien la rupture avec le béhaviorisme était nécessaire (Lashley 1951). L'opposition, à la fin des années 50, entre Skinner et Chomsky quant à l'objet même de la linguistique trouve d'ailleurs également ici son origine : le modèle cybernétique de Chomsky s'oppose à l'idée d'apprentissage linéaire suggérée par Skinner dans ces techniques de conditionnement et alors que les études sur *Le Comportement verbal* menées par Skinner étaient analytiques et spéculatives, la grammaire proposée par Chomsky se voulait justement dynamique, centrée sur l'acquisition du langage, et donc 'générative'.

Les études en cryptologie

Suite à sa participation aux efforts de guerre dans le domaine de la DCA, Wiener se tourne vers la biologie puis la philosophie, avec l'interprétation qu'il donne des travaux menés en collaboration avec Bigelow et Rosenblueth. L'aboutissement de cette réflexion sera marqué par la parution, en 1948, à Paris puis à New York, d'un livre qui deviendra rapidement un best-seller : *Cybernetics*⁴⁸.

De son côté, Shannon poursuit ses travaux pour le NDRC, en s'intéressant à des problèmes de cryptologie. C'est ainsi dans le cadre des efforts de guerre qu'il poursuit une réflexion déjà entamée en 1939, au sujet d'une théorie

47. Howard Gardner expose en détail les liens entre béhaviorisme et genèse de la cybernétique dans son *Histoire de la révolution cognitive* (Gardner, 1985: 24-26). Voir aussi Lily Kay qui cite une lettre de Wiener datée du 22 juin 1942, adressée à Haldane, dans laquelle il écrit qu'il est « (...) assez clair qu'aucun béhavioriste n'a jamais vraiment compris les possibilités du comportement. » (Kay, 1997: 34) : "Behaviorism as we all know is an established method of biological and psychological study but I have nowhere seen an adequate attempt to analyze the intrinsic possibilities of types of behavior. (...) We are quiet clear on the fact that no behaviorist has ever really understood the possibilities of behavior."

48. Dans un premier temps on peut retenir pour la cybernétique la définition qu'en donne Wiener dans le sous-titre de son livre : « Commande et communication dans l'animal et la machine » (Wiener 1948).

générale de la communication. Dans une lettre adressée à Bush, datée du 16 février 1939, il apporte le premier schéma d'un mode général de communication :

« À différentes reprises j'ai travaillé sur une analyse de certaines des propriétés fondamentales de systèmes généraux pour la transmission de l'information [*'intelligence'*]⁴⁹, incluant la téléphonie, la radio, la télévision, la télégraphie etc. Pratiquement tous ces systèmes peuvent être présentés de la façon suivante : $f_1(t) \Rightarrow T \Rightarrow F(t) \Rightarrow R \Rightarrow f_2(t)$

$f_1(t)$ est une fonction générale du temps (arbitraire sauf pour certaines limitations de fréquence) représentant l'intelligence à transmettre. Cela représente, par exemple, la fonction pression-temps en radio ou téléphonie ou la courbe de sortie tension-temps d'un iconoscope pour la télévision.⁵⁰ »

Il précise également que T est un dispositif d'émission d'où sort $F(t)$, cette fonction étant celle qui est réellement émise, l'opérateur T ayant par exemple modulé le message à transmettre. De même R est un élément de réception et la fonction effectivement reçue est $f_2(t)$.

Au début de l'année 1941, Shannon prend conscience que c'est la notion de code qui rassemble les techniques de cryptologie et de communication. Chaque avancée dans le domaine des télécommunications apporte de nouvelles possibilités en cryptologie et réciproquement les besoins en cryptologie stimulent les recherches en télécommunications⁵¹. Au début de la seconde guerre mondiale, les communications radio peuvent être facilement interceptées et écoutées. O.E. Buckley (président des Bell Labs en 1940) et R.K. Potter (également des Bell Labs) sont alors chargés, en octobre 1940, au sein du NDRC, de mettre au point deux nouveaux projets de communication sûrs, un dispositif mobile devant fonctionner à court terme et un système plus sûr devant assurer la sécurité des transmissions entre les états-majors alliés⁵². Le premier projet consistait à rendre le son visible et c'est dans ce cadre que furent utilisés le spectrographe acoustique et le Vocoder, deux appareils permettant un traitement numérique des données. Shannon est impliqué dans le second projet concernant le 'X-System'. Sans entrer ici dans les détails techniques, retenons seulement que les travaux correspondants auxquels participèrent également Hartley et Nyquist reposaient sur l'utilisation d'un procédé de modulation par

49. En utilisant le mot 'intelligence', Shannon se situe dans la lignée des ingénieurs comme Hartley ou Nyquist, tous deux aux Bell Labs, ce dernier à propos de sa publication de 1924 : "This paper considers two fundamental factors entering into the maximum speed of transmission of intelligence by telegraph." (Nyquist, 1924: 324)

50. "Off and on I have been working on an analysis of some of the fundamental properties of general systems for the transmission of intelligence, including telephony, radio, television, telegraphy, etc. Practically all systems of communication may be thrown into the following form : $f_1(t) \Rightarrow T \Rightarrow F(t) \Rightarrow R \Rightarrow f_2(t)$ $f_1(t)$ is a general function of time (arbitrary except for certain frequency limitations) representing the intelligence to be transmitted. It represents for example, the pressure-time function in radio and telephony, or the voltage-time curve output of an iconoscope in television." Cette lettre issue du fonds Bush de la Bibliothèque du Congrès est également reproduite, d'après la version abrégée de Hagemeyer, dans les *Œuvres choisies* (Sloane & Wyner, 1993: 455-6)

51. En ce qui concerne l'histoire de la cryptologie, Kahn (1967) fait encore référence. Un exemple récent des interactions entre communication et cryptographie est la transmission d'ordres et de rapports par la mafia en modifiant quelques bits sur des images quelconques transmises sur le réseau Internet. Voir aussi la controverse aux Etats-Unis sur l'utilisation privée du cryptage avec le cas de la PGP (*Pretty Good Privacy*).

52. Les deux projets en question sont décrits dans Fagen (1978: 292-317).

impulsions codées (MIC). Ce type de modulation est particulièrement intéressant puisqu'il suppose un retour à l'ère numérique du télégraphe avec une transmission numérique de signaux analogiques⁵³.

C'est suite à ses recherches sur la MIC que parallèlement aux travaux accomplis par Turing dans le contexte britannique, Shannon est amené à définir l'information comme grandeur caractéristique de ce qu'apporte un nouveau message codé au vu des hypothèses déjà formulées. Tirant profit de son étude sur l'application de l'algèbre de Boole aux relais (son *Masters* de 1938) et plus largement encore de sa double formation d'ingénieur et de mathématicien, Shannon remet aux autorités militaires en septembre 1945 un rapport confidentiel intitulé « A mathematical theory of cryptography ». Ce rapport rédigé pour la *Signal Security Agency* correspond à la première théorie mathématique de la cryptographie qui soit à la fois probabiliste et algébrique. En effet, dès l'introduction, Shannon prévient le lecteur qu'il attend de lui des connaissances dans ces deux domaines alors que des notions en cryptologie ne sont pas nécessaires : « utiles mais pas requises »⁵⁴.

Se posant la question de savoir comment quantifier « l'incertitude » que représente la réalisation d'un événement parmi n possibles caractérisés par des probabilités p_i , il donne une expression de la « quantité de choix », « cette notion plutôt vague », avec la formule suivante :

$$H = -\sum_{i=1}^n p_i \log p_i$$
⁵⁵

Trois ans plus tard, en 1948, revenant aux Bell Labs dans le monde des télécommunications, il publie en partant de ce rapport un article intitulé « A mathematical theory of communication », dans lequel l'information prend le nom d'entropie. Il annonce alors des théorèmes d'existence importants concernant l'entropie de la source et la capacité du canal.

INFORMATION ET CONNAISSANCE

Lorsque cet article paraît sous forme de livre, avec un texte introductif de Warren Weaver, le titre du recueil est *The Mathematical Theory of Communication*. En moins d'un an, une théorie est devenue la théorie de référence pour la communauté scientifique. Avec l'essor de la cybernétique pour laquelle Wiener identifie pour la première fois l'information comme la grandeur essentielle permettant de décrire tout phénomène de rétroaction, il se constitue une culture scientifique marquée par des références à la « théorie de l'information ». Cette théorie de l'information se retrouve aussi bien en linguistique, qu'en biologie ou en mathématiques⁵⁶.

53. Tous ces travaux réalisés en temps de guerre sont abordés avec plus de détails dans Segal (2003, chapitre 2). Rappelons d'ailleurs que c'est également dans le cadre de ses recherches en cryptologie, plus précisément en cryptanalyse, que Turing parvint à une définition logarithmique du degré de certitude ('Weight of evidence', exprimé en 'deciban') que l'on peut rapprocher de l'information définie par Shannon.

54. La citation ci-dessus correspond au texte original "helpful but not required" (Shannon, 1945: 1).

55. La lettre H est choisie en référence explicite au 'théorème H' de Boltzmann concernant l'entropie.

56. Voir Segal (2003, deuxième partie) et parmi les sources primaires Jakobson (1961), Atlan (1972), Danchin (1998) ou, pour les mathématiques, ces propos tenus en 1970 par Kolmogorov : "Information theory must precede probability theory, and not be based on it. By the very essence of this discipline, the foundations of information theory have a finite combinatorial character." (Kolmogorov, 1983: 39).

Force est de constater qu'on doit alors distinguer au moins deux théories de l'information.

Deux théories de l'information ?

Au regard de l'histoire de la notion scientifique et technique d'information, on peut distinguer deux 'théories de l'information'. La première théorie de l'information repose uniquement sur l'application de la définition de l'information donnée par Shannon, $H = - \sum p_i \log p_i$, où les probabilités p_i désignent des probabilités de sélection dans un ensemble fini. L'analogie au moins formelle entre cette définition et celle de l'entropie en mécanique statistique explique alors l'engouement que cette 'théorie' a pu susciter et les guillemets sur le mot 'théorie' s'imposent tant le référent commun aux différents travaux abordés est étroit : une simple définition probabiliste⁵⁷. C'est cette théorie de l'information qui fournit le terreau pour le développement d'une nouvelle culture scientifique dont témoignent aujourd'hui des expressions pour le moins porteuses d'ambiguïté comme « autoroutes de l'information » ou « société de l'information », faisant aussi référence au sens commun du mot. Les travaux de J.C.R. Licklider (1915-1990) autour de la création de l'ARPANET, cet ancêtre d'Internet, sont à ce titre emblématiques⁵⁸.

La seconde théorie de l'information serait plus proche des domaines de recherche de Shannon, à la fois les télécommunications et la cryptologie, ces deux centres d'intérêts étant du reste, comme nous l'avons signalé, profondément imbriqués. Dans cette perspective plus technique, la théorie de l'information repose sur les théorèmes démontrés par Shannon, sur la théorie des codes, et trouve quelques prolongements en télécommunications dans la théorie du signal, pour tout ce qui concerne le codage de canal ou de source⁵⁹.

Après tout, on peut montrer que la simple définition quantitative proposée par Shannon pour l'information ne repose que sur une moyenne logarithmique déjà connue d'un mathématicien comme de Moivre au XVIII^e siècle.

Théorie de l'information et intelligence artificielle

Pour expliquer le formidable enthousiasme que suscitera la théorie de l'information (dans le premier sens défini ci-dessus), il convient de rappeler que la notion scientifique d'information fut utilisée très tôt dans la modélisation du cerveau humain.

Dès 1943, un jeune mathématicien, Walter Pitts (1923-1969), publie avec un psychiatre, Warren McCulloch (1898-1969), un texte intitulé « Un calcul logique des idées immanentes dans l'activité nerveuse ». Cette modélisation du cerveau repose sur un système logique formel, comme en témoigne leur référé-

57. Nous avons déjà eu l'occasion d'étudier en détail le rôle de von Neumann dans le rapprochement terminologique entre information et entropie (Segal, 2003: chapitre 2). De même nous avons montré comment le physicien Léon Brillouin (1889-1969) procède à une véritable relecture de la physique au travers de la notion scientifique d'information (Brillouin, 1956; Segal, 2003: chapitre 5).

58. Voir (Licklider 1960), publication sur la « symbiose » entre l'homme et l'ordinateur, dans laquelle Licklider introduit par exemple le rapport bruit / signal hors du contexte des recherches en télécommunications (Segal, 2003: chapitre 8).

59. Le dernier exposé synthétique sur cette question est sans doute (Battail 1997).

rence à la logique symbolique de Carnap et aux *Principia Mathematica* de Whitehead et Russell⁶⁰.

Dans leur démarche fortement marquée par l'ingénierie électrique, les neurones deviennent des boîtes noires caractérisées par le fonctionnement binaire de leur activité, de type tout ou rien. La logique propositionnelle apparaît alors « inhérente » au système nerveux⁶¹. Chaque neurone se présente comme un calculateur élémentaire symbolisé par une fonction booléenne à seuil. De plus, les auteurs reprennent le terme 'psychon' introduit auparavant par McCulloch pour désigner une unité psychique élémentaire. Ceci est à rapprocher de la définition de la notion d'information chez Shannon, comme alternative élémentaire. On remarque dans les deux cas une approche centrée sur les phénomènes de communication et les comportements.

Avant même d'être interprétés en termes informationnels, ces travaux de Pitts et McCulloch ont connu une diffusion importante. Ces propos de von Neumann le montrent clairement : « Tout ce qui peut être décrit de manière exhaustive et sans ambiguïté (...) peut être conçu comme un réseau neuronal approprié. »⁶² L'influence de l'article de Pitts et McCulloch sur la pensée de von Neumann est manifeste, tant dans ses travaux théoriques sur la machine de Turing que dans sa conception de l'EDVAC, en 1945⁶³.

Au moment de la convergence entre les modélisations du cerveau et la construction des premiers ordinateurs, à la fin des années 40, la notion d'information s'impose comme la grandeur propre à toute simulation. A partir de 1956, de nombreuses recherches visant à parvenir à la simulation de comportements nécessitant chez l'homme un recours à l'intelligence sont fédérées par l'appellation 'intelligence artificielle'. Le rôle de la théorie de l'information dans ce développement est souvent négligé.

C'est par exemple Shannon qui soumet avec M. Minsky, N. Rochester et J. McCarthy un projet à la fondation Rockefeller pour financer l'organisation de cette école d'été au Dartmouth College de Hanover (dans le New Hampshire) portant pour titre : « The Dartmouth Research Project on Artificial Intelligence » et baptisant ainsi ce domaine de recherches⁶⁴. Minsky et McCarthy connaissaient Shannon pour avoir travaillé avec lui aux Bell Labs en 1953, travaux qui amenèrent McCarthy et Shannon à éditer ensemble le recueil

60. Leur bibliographie se résume d'ailleurs aux *Principia*, au livre de Carnap *The Logical Syntax of Language* et aux *Grundzüge der Theoretischen Logik* de D. Hilbert et W. Ackermann (voir McCulloch & Pitts 1943).

61. "[...] that activity is inherently propositional [...]" (McCulloch & Pitts, 1943: 131).

62. (Gardner, 1985: 33). Cette citation n'est pas à prendre littéralement puisqu'il s'agit déjà de propos rapportés. Toutefois, de nombreuses sources montrent qu'elle traduit fidèlement la conception de von Neumann (Bigelow rapporte par exemple dans un entretien avec Aspray que le mathématicien d'origine hongroise avait été « énormément impressionné » par ce travail, Aspray, 1990: 180).

63. Electronic Discrete Variable Computer. Von Neumann nomme 'E-element' le modèle de neurone simplifié de Pitts et McCulloch et la description qu'il donne du câblage de sa machine fait penser aux écrits d'un neurophysiologiste : "[The E-Element] will be represented to be a circle O, which receives the excitatory and inhibitory stimuli, and emits its own stimuli along a line attached to it : O--." (Von Neumann, 1945: 41)

64. Le meilleur compte-rendu de cette conférence est à notre connaissance (McCorduck, 1979, chapitre 5). P. McCorduck a interviewé les quatre scientifiques cité ci-dessus (dont Shannon en 1975) ainsi que Bernstein, Dreyfus, Good, MacKay, Newell, Papert, Simon, Weizenbaum et d'autres encore.

*Automata Studies*⁶⁵. La demande de financement de ce projet est rédigée comme suit :

« Nous proposons qu'une étude de deux mois avec dix hommes, sur l'intelligence artificielle, soit menée durant l'été 1956 au Dartmouth College de Hanover, dans le New Hampshire. L'étude consiste à partir de l'hypothèse selon laquelle tous les aspects de l'apprentissage ou toute autre partie de l'intelligence peut en principe être si précisément décrite qu'une machine peut être faite pour la *simuler*.⁶⁶ »

Interrogé au sujet de cette conférence, Shannon déclare que c'est à ce moment qu'ils prirent conscience du fait que les ordinateurs pouvaient représenter « un outil beaucoup plus général et plus puissant » qu'une simple machine à calculer⁶⁷.

Que ce soit au regard des travaux sur la conception des circuits binaires ou de la théorie mathématique de la communication, on trouve bien une cohérence dans les écrits de Shannon. C'est d'ailleurs aujourd'hui la mesure de l'information définie à partir des probabilités de sélection qui trouve des applications dans l'optimisation de certains arbres de recherche et c'est ainsi que l'on parvient à automatiser les modes de recherche d'un mot dans un dictionnaire lorsque l'on connaît les probabilités pour que tel ou tel mot fasse l'objet de la recherche⁶⁸.

Actuellement parmi les différentes Écoles visant à modéliser l'acte de connaître, on peut retrouver la trace de tous ces travaux. Le fait par exemple que Pitts et McCulloch insistent à la fin de leur article sur les implications de leur théorie à la fois en psychiatrie, en neurologie ou encore en biophysique mathématique explique comment la cybernétique peut être considérée comme l'ancêtre des sciences cognitives⁶⁹. Plus précisément, c'est l'approche connexionniste qui trouve ici ses racines⁷⁰.

A travers l'histoire de la notion scientifique d'information, on se rend compte que c'est la dimension sémantique de l'information qui a souvent révélé les limites des différentes modélisations proposées. Le modèle cognitiviste dans lequel la cognition repose également sur une computation dénuée de signification est ainsi marqué par l'essor de la théorie de l'information.

65. (Shannon, 1956). P. McCorduck rapporte que McCarthy espérait pouvoir consacrer ce volume à l'intelligence artificielle et faire figurer cette expression dans le titre. Notons ici que Simon et Newell, deux participants de cette école d'été, utilisèrent d'abord dans les années suivantes l'expression 'complex information processing' et non 'intelligence artificielle'. Bernstein vint quant à lui à Hanover présenter ses recherches en cours sur la programmation d'une machine à jouer aux échecs.

66. Shannon, Minsky, Rochester et J. McCarthy (cités dans McCorduck, 1979: 93) : "We propose that a two-month, ten-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to *simulate* it." (nous soulignons)

67. "We realized that this [*the computer*] was a lot more than an adding machine, a much more general and powerful tool than that." (McCorduck, 1979: 101)

68. Voir Hotz (1986), notamment la section sur « Une application de la théorie de l'information à la construction d'arbres de recherche presque optimaux » p. 62-64.

69. C'est la thèse défendue entre autres par J.P. Dupuy dans Dupuy (1994).

70. De même, la 'seconde cybernétique' qui repose sur le principe d'ordre par le bruit, est à rapprocher du néo-connexionnisme. Les études sur l'autonomie et la notion d'auto-organisation qui en découlent ont donné lieu à quelques dérives idéologiques, liées par exemple à l'essor du postmodernisme, qui mériteraient d'être abordées plus en détail (Segal, 2003: chapitre 7).

Cette théorie de l'information dont les contours ont souvent été mal définis se présente clairement au terme de cette étude comme le fruit de recherches liées à la seconde guerre mondiale. Les objectifs de simulation et d'anticipation de comportements on pu changer de registre mais la notion d'information demeure centrale.

D'une façon générale, on peut rapprocher cette notion de l'idée de concept prospectif défini par le logicien américain John Myhill. Celui-ci considère qu'un concept 'effectif' est communicable d'un individu à un autre sans avoir à débattre sur la signification (exemple, le concept d'addition). Au contraire, le sens d'un concept 'constructif' se précise par des constructions successives, à la manière d'un algorithme ou plutôt d'une machine de Turing dont les inscriptions sur le ruban changent au cours du temps. Enfin, un concept 'prospectif' est caractérisé par le fait « [qu']aucun protocole de communication ne permet de communiquer à coup sûr ce sens à un autre locuteur (et *a fortiori* à tous les locuteurs), et ne le permettra jamais.⁷¹ »

L'histoire du concept prospectif d'information est alors à même de proposer un nouvel éclairage sur les grandes orientations des théories visant à rendre compte de la cognition. Les débats très actuels concernant le sens de l'analogie entre information et entropie en physique peuvent par exemple être rapprochés des études sur le rôle primordial, dans les sciences cognitives, de l'analogie entre différentes écritures symboliques⁷².

Bibliographie

- Armatte, M. (1988). La construction des notions d'estimation et de vraisemblance chez Fisher. *Statistiques et Analyse de Données*, 13 : 65-92.
- Aspray, W. F. (1985). The Scientific Conceptualization of Information: A Survey. *Annals of the History of Computing*, 7: 117-40.
- Aspray, W. F. (1990). *John von Neumann and the Origins of Modern Computing*. Cambridge, M.I.T Press.
- Atlan, H. (1972). *L'organisation biologique et la théorie de l'information*. Paris, Hermann.
- Augarten, S. (1984). *Bit by bit - An illustrated history of computers*. New York: Ticknor & Fields.
- Battail, G. (1997). *Théorie de l'information*. Paris, Masson.
- Baxter, J. P. (1946). *Scientists against time*. Boston: Little, Brown and Company.
- Bennett, S. (1993). *History of Control Engineering (1930-1935)*. Stevenage, Peter Peregrinus.
- Bode, H. (1945). *Network Analysis and Feedback Amplifiers Design*. New York, Nostrand.
- Breton, P. (1987a). La naissance de la cybernétique et l'autonomisation de l'informatique. *Grundlagenstudien aus Kybernetik und Geisteswissenschaften*, 28: 27-35.
- Breton, P. (1987b). *Une histoire de l'informatique*. Paris, La Découverte.

71. (Danchin, 1998: 155). Ce livre traite entre autres du rôle de la notion d'information en génétique moléculaire. Dans le texte original de Myhill, on trouvera les définitions correspondantes (Myhill, 1952: 172, 179 et 190). L'exemple donné par le philosophe est assez éclairant : en esthétique musicale, une note représente un concept effectif, un accord représente un concept constructif et la beauté correspond au concept prospectif. Ceci se rapprocherait alors des *trading zones* chères à P. Galison, pour prendre des exemples plus récents, ces zones d'échanges, de contact, qui existent malgré la diversité des langages scientifiques.

72. Parmi les livres récents accordant une importance particulière à l'analogie entre information et entropie, citons Cohen-Tannoudji (1995).

- Bush, V. (1970). *Pieces of the Action*. London, Cassell.
- Brillouin, L. (1956). *Science and Information Theory*. New York, Academic Press. Inc. (réédition augmentée en 1962, édition française en 1959).
- Campbell-Kelly, M. & Aspray, W. (1996). *Computer: A history of the information machine*. New York: Basic Books.
- Ceruzzi, Paul E. (1983). *Reckoners: The Prehistory of the Digital Computer, from Relays to the Stored Program Concepts, 1935-1945*, Westport, Greenwood Press.
- Cohen, I. B. (1988). The Computer: A case study of support by government, especially the military, of a new science and technology. in Everett Mendelsohn, M.R. Smith & P. Weingart (éds.), *Science, Technology and the Military*, Dordrecht, Kluwer: 119-154.
- Cohen-Tannoudji, G. (1995). *Les constantes universelles*. Paris, Hachette. (Questions de Science).
- Dahan-Dalmedico, A. (1996). L'essor des mathématiques appliquées aux Etats-Unis : l'impact de la seconde guerre mondiale. *Revue d'histoire des mathématiques*, 2 : 149-213.
- Danchin, A. (1998). *La barque de Delphes - Ce que révèle le texte des génomes*. Paris, Odile Jacob.
- Dotzler, B. (1996). *Papiermaschinen, Versuch über Communication & Control in Literatur und Technik*. Berlin, Akademie Verlag.
- Dupuy, J.-P. (1994). *Aux origines des sciences cognitives*. Paris, La Découverte.
- Durand-Richard, M.-J. (1996). L'École algébrique anglaise : les conditions conceptuelles et institutionnelles d'un calcul symbolique comme fondement de la connaissance. in Goldstein, Catherine, J. Gray & J. Ritter (éds.), *L'Europe mathématique - Mythes histoires identités*, Paris, Edition de la Maison des Sciences de l'Homme : 446-477.
- Edwards, P. (1996). *The closed World, Computers and the Politics of discourses in cold War America*. Cambridge, MIT Press.
- Fagen, M.D. (1978). *A History of Engineering and Science in the Bell System, Vol. 2: National Service in War & Peace (1925-1975)*. Muray Hill: Bell Laboratories.
- Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society*, 222A: 309-368.
- Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society*, 98: 39-54 et discussion p. 55-82.
- Galison, P. (1994). The Ontology of the Enemy: Norbert Wiener and the cybernetic Vision. *Critical Inquiry*, 21: 228-266.
- Gardner, H. (1985). *The Mind's new science - A history of the cognitive revolution*. New York, Basic Books (traduction de Jean-Louis Peytavin, 1993, *Histoire de la révolution cognitive, la nouvelle science de l'esprit*. Paris, Payot).
- Goldstine, H. (1972). *The Computer from Pascal to von Neumann*. Princeton, Princeton University Press.
- Hagemeyer, F. (1979). *Die Entstehung von Informationskonzepten in der Nachrichtentechnik*, Doktorarbeit an der Freie Universität Berlin, FB 11: Philosophie und Sozialwissenschaften, 1979 (soutenue le 8.11.1979).
- Heims, S. J. (1991). *The Cybernetics Group*. Cambridge: MIT Press.
- Heintz, B. (1993). *Die Herrschaft der Regel: zur Grundlagengeschichte des Computers*. Frankfurt am Main, Campus.
- Hodges, A. (1983). *Alan Turing: The Enigma*. London, Burnett Books et New-York, Simon & Schuster.
- Hodges, A. (1994). *Alan Turing, Enigma*. Wien, Springer Verlag (deuxième édition en allemand du livre précédent).
- Horgan, J. (1990). Claude E. Shannon Unicyclist, juggler and father of information theory. *Scientific American*, 242 : 20-22.
- Hotz, G. (1986). Der Begriff der Information in der Informatik. in Folberth, Otto G. & Clemens Hackl, *Der Informationsbegriff in Technik und Wissenschaft*. München, R. Oldenburg Verlag: 53-70.
- Ibrah, G. (1994). *Histoire universelle des chiffres*. 2 tomes, Paris, Laffont (Bouquins).

- Jakobson, R. (1961). Linguistics and Communication Theory. *Proceedings of Symposia in Applied Mathematics*, 12: 245-52.
- Kahn, D. (1967). *The Codebreakers: the Story of Secret Writing*. New York, McMillan (version abrégée en 1973, London, Sphere Books).
- Kay, L. (1997). Cybernetics, Information, Life: The Emergency of scriptural Representations of Heredity. *Configurations*, 5: 23-91.
- Kolmogorov, A. (1983). Combinatorial foundations of information theory and the calculus of probabilities. *Russian mathematical surveys*, 38: 29-40.
- Küpfmüller, K. (1924). Über Einschwingvorgänge in Wellenfiltern. *Elektrische Nachrichten Technik*, 1: 141-152.
- Ladrière, J. (1957). *Les limitations internes des formalismes - Etude sur la signification du théorème de Gödel et des théorèmes apparentés dans la théorie des fondements mathématiques*. Paris, Gauthiers-Villars et Louvain, Nauwelaerts (Logique mathématique série B).
- Lashley, K. (1951). The problem of serial order in behaviour. in Lloyd A. Jeffress (éd.), *Cerebral Mechanism in Behavior - The Hixon Symposium*, California Institute of Technology, sept. 1948, New York, John Wiley & Sons: 112-136.
- Lassègue, J. (1998). *Turing*. Paris, Les Belles Lettres.
- Leff, H. S. et Rex, A. F. (1990). *Maxwell's Demon, Entropy, Information, Computing*. Bristol, Adam Hilger.
- Lévy, P. (1989). L'invention de l'ordinateur. in Michel Serres (éd.), *Éléments d'histoire des Sciences*, Paris, Bordas : 515-535.
- Lewis, G. N. (1930). The Symmetry of time in Physics. *Science*, 71: 569-576.
- Licklider, J.C.R. (1960). Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, 1: 4-11.
- Masani, P. R. (1990). *Norbert Wiener (1894-1964)*. Basel, Birkhäuser Verlag (Vita Mathematica).
- McCorduck, P. (1979). *Machines who think: A personal inquiry into the History and Prospects of Artificial Intelligence*. San Francisco, Freeman.
- McCulloch, W. & Pitts, W. (1943). A logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5: 115-133.
- Metropolis, N., Howlett, J. & Rota, G.-C. (1980). *History of Computing in the 20th Century*. New York: Academic Press.
- Millman, S. (1984). *A History of Engineering and Science in the Bell System - Communication Sciences (1925-1980)*. Indianapolis, AT & T Bell Laboratories.
- Mindell, D. (1996). *Datum for its Own Annihilation: Feedback, Control, and Computing, 1916-1945*. PhD thesis, MIT.
- Mindell, D. (2002). *Between Human and Machine: Feedback, Control and Computing Before Cybernetics*. Baltimore: The Johns Hopkins University Press.
- Mosconi, J. (1989). *La constitution de la théorie des automates*. Paris, Université Paris I (Thèse, UFR de Philosophie)
- Myhill, J. (1952). Some philosophical implications of mathematical logic. Three classes of ideas. *The Review of Metaphysics*, 6: 165-98.
- Nakashima, A. (1936). Theory of Relay Circuit Composition. *Nippon Electric Communication Engineering*, 3: 197-226 et le 'résumé-complément' paru en 1937, 5: 70-71.
- Neumann, J. von (1945). *A First Draft of a Report on the EDVAC*. June 30, 1945, reproduit en 1993 dans *Annals of the History of Computing*, 15: 28-75
- Noble, D. F. 1984. *Forces of Production: A Social History of Industrial Automation*. New York: Knopf.
- Nyquist, H. (1924). Certain Factors Affecting Telegraph Speed. *Bell System Technical Journal*, 3: 324-346
- Nyquist, H. (1928). Certain topics in Measurement of Telegraph transmission. *Transactions of the American Institute of Electrical Engineers*, 47: 617-644.
- Owens, L. 1986. Vannevar Bush and the Differential Analyzer: The Test and Context of an Early Computer. *Technology and Culture*, 27: 63-95.

- Petzold, H. (1992). *Moderne Rechenkünstler, die Industrialisierung der Rechentechnik in Deutschland*. München, C.H.Beck.
- Pierce, J. R. (1993). Looking Back - Claude Elwood Shannon. *IEEE Potentials*, 12: 38-40.
- Price, R. (1985). A conversation with Claude Shannon - One man's approach to problem solving. *Cryptologia*, 9 : 167-75.
- Ramunni, J. (1988). *La physique du calcul*. Paris, Hachette, (Histoire et Philosophie des Sciences).
- Rosenblueth, A., Wiener, N. et Bigelow, J. (1943). Behavior, Purpose and Teleology. *Philosophy of Science*, 10: 18-24 (en français en 1961, Comportement, intention et téléologie, *Les Etudes Philosophiques*, 2 : 147-56)
- Segal, J. (2002). Le géomètre de l'information. *Pour la Science*, 295 : 26-29
- Segal, J. (2003). *Le Zéro et le Un – Histoire de la notion scientifique d'information au 20^e siècle*. Syllepse, Paris, 2003.
- Shannon, C. E. (1937). *A Symbolic Analysis of Relay and Switching Circuits*, M.Sc. soutenu le 10.8.1937. MIT Library, cote 240579. Une version abrégée de ce mémoire est publiée sans le même titre en 1938 dans les *Transactions of the American Institute of Electrical Engineers*, 57: 713-23.
- Shannon, C. E. (1941). Mathematical theory of the differential analyzer. *Journal of Mathematics and Physics*, 20: 337-354.
- Shannon, C. E. (1945). *A Mathematical Theory of Cryptography*. Memorandum MM 45-110-02, Sept. 1, 1945, MIT Library, 114 pages. + 25 figures.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27: 379-423 et 623-656. Sous forme de livre : Claude Shannon et Warren Weaver. 1949. *The Mathematical Theory of Communication*, Urbana, Illinois, University of Illinois Press.
- Shannon, C. E. (1956). A Universal Turing Machine with Two Internal States. in Claude Shannon et John McCarthy (éd.), *Automata Studies*, Princeton, Princeton University Press: 157-165.
- Sloane, Neil J.A. & Wyner, Aaron D. (1993). *Claude Elwood Shannon: Collected Papers*, New York: IEEE Press.
- Smoluchowski, M. von (1912). Experimentell nachweisbare, der üblichen Thermodynamik widersprechende Molekularphänomene. *Physikalische Zeitschrift*, 13: 1069-1079 (et discussion p. 1080)
- Szilárd, L. (1929). Über die Entropieverminderung in einem thermodynamische System bei Eingriffen intelligenter Wesen. *Zeitschrift für Physik*, 53: 840-856.
- Thomson, W. (1874). Kinetic theory of the dissipation of energy. *Nature*, 9: 441-444.
- Wagner, P. (1998). *La machine en logique*. Paris, Presses Universitaires de France (Science Histoire et Société).
- Wiener, N. (1948). *Cybernetics, or Control and Communication in the Animal and the Machine*. Paris, Hermann et Cie.
- Wiener, N. (1949). *Time Series*. Cambridge: MIT Press.
- Wiener, N. (1956). *I am a Mathematician. The Later Life of a Prodigy*. New York: Doubleday & Cie Inc.
- Wiener, N. (1958). My connection with cybernetics. Its origins and its future. *Cybernetica*, 1: 1-14.
- Zemanek, H. (1991). Geschichte der Schaltalgebra. in Manfred Broy (éd.), *Informatik und Mathematik*, Berlin, Springer: 43-72.