L'architecture des systèmes symboliques

par **Jacques Lorigny**

Je voudrais vous présenter succinctement quatre modèles mathématiques de base pour les systèmes symboliques : le questionnaire arborescent, l'agrégat, le réseau, la matrice stochastique. Pour l'essentiel, il s'agit des graphes que vous connaissez bien, et que nous pratiquons tous, au moins comme support d'analyse et de description, mais nous verrons aussi que la matrice stochastique, par sa nature algébrique, possède une vertu heuristique propre.

Un bref rappel sur les graphes nous suffira. Ce sont des ensembles de *sommets*, ou de « nœuds », et *d'arcs* reliant ces nœuds. Ces arcs expriment une interaction ou un échange. Ils sont munis d'une valuation qui mesure l'intensité de cette interaction ou de cet échange. Le modèle des graphes est capital pour nous, parce qu'en systémique, les relations sont au moins aussi importantes que les entités reliées. Dès le milieu du siècle dernier, les graphes ont été beaucoup utilisés en Recherche Opérationnelle (R.O.), en particulier pour les problèmes de réseaux de transport ou de programmation dynamique.

1- Le questionnaire arborescent

A peu près à la même époque, dans les années 1960, Claude-François Picard a créé la théorie des questionnaires mathématiques, conçue pour optimiser les processus de décision automatisés. Les questionnaires mathématiques sont des ensembles hiérarchisés de questions-réponses, qui ont été utilisés avec succès pour l'optimisation du tri sur ordinateur (Cesari), pour l'aide au diagnostic médical (Terrenoire, Tounissoux), pour la reconnaissance des formes (Simon et Roche), pour le diagnostic des pannes dans un système complexe (Pau), ainsi que pour ma propre application à la codification statistique dans les recensements et les grandes enquêtes de l'INSEE, et sur laquelle nous allons revenir.

1-1) Un problème central, le choix des (bonnes) questions

Les questionnaires mathématiques ne sont jamais donnés en soi. Il faut d'abord choisir quelle est la question qui doit figurer au sommet-racine x_0 . C'est un choix décisif parce que, en opération, c'est toujours elle qui sera posée en premier par l'ordinateur (ou le robot), quelle que soit la situation où il se trouvera, quel que soit le cas qu'il aura à traiter. Ensuite, il faut choisir, pour chacune des réponses attendues à la première question, quelle sera la nouvelle question à poser pour aller plus loin vers la solution cherchée, etc. On voit apparaître une dichotomie entre le centre du système, en quelque sorte le « sujet », celui qui choisit les questions, et en face de lui, son environnement opératoire, les différentes situations où il opère, les différents cas qu'il peut rencontrer. A l'époque, dans les années 1960, c'était très nouveau, même par rapport à la Recherche Opérationnelle. Nos nouveaux logiciels étaient destinés à opérer « en direct », ou en « *live* » comme on dirait aujourd'hui. Les mathématiciens, en règle générale, n'étaient pas prêts à introduire le « sujet » dans leur univers. Bien sûr, il faut faire une exception pour notre ami Robert Vallée, et ses Notes aux Comptes-Rendus de l'Académie des Sciences présentées par Louis de Broglie en 1951, mais il était bien seul à l'époque. On pense aussi aux postulats de la mécanique quantique apparus au début du siècle. La représentation mathématique portait désormais, non seulement sur l'état du « système » observé, mais aussi sur son observateur, ou du moins sur l'opération d'observation, par le biais d'une matrice.

Par ailleurs, la « communication », autre préoccupation dans l'air du temps, faisait son entrée en mathématique, grâce à l'article mémorable de Claude Shannon, paru en 1948, intitulé « *a mathematical theory of communication* » et portant sur les transmissions par lignes. Picard utilisa aussitôt, pour résoudre son problème central de choix des (bonnes) questions, des calculs de quantités d'entropie au sens de Shannon. Nous retrouvions, au sein du Groupe de Recherche de Picard, le même souci que Shannon, minimiser le temps moyen de traversée du questionnaire mathématique.

Vue d'aujourd'hui, certes, la puissance des ordinateurs est sans rapport avec ceux du siècle dernier, et le problème de l'appariement sur ordinateur (*matching*) ne pose plus de problème de délai. Du moins en général, car il existe des exceptions. C'est encore un problème dans certaines applications, comme la recherche d'ADN. Ainsi, dans les Banques Génétiques, comme Genelink à Forth Worth (Texas), l'essentiel de l'activité quotidienne consiste à identifier des ADN, par exemple à apparier l'ADN d'une cellule adulte avec celui d'un embryon. En 2003, la procédure actuelle était jugée trop longue. Elle durait plusieurs jours, et une puce à ADN était à l'étude pour réduire ce temps de réponse à seulement quelques secondes. Aujourd'hui, en 2008, c'est probablement dépassé par le progrès technologique. Mais, de façon générale, le besoin d'optimiser le temps de réponse des opérateurs d'identification reste vital dans tous les systèmes où s'opposent deux exigences contradictoires : accueillir un ensemble de connaissances immense et incompressible, et obtenir un temps de réponse très bref. C'est le cas surtout des systèmes dits « en temps réel », où la brièveté de la réponse est la clé du système, par exemple, dans la reconnaissance d'un outil par un robot en environnement hostile, ou par un radar militaire, ou par satellite, mais aussi dans les problèmes de détection des pannes dans des systèmes hypercomplexes, à l'échelle d'une ville entière, voire plus encore. La tendance à la mondialisation des réseaux ne fera que prolonger, et peut-être amplifier ce problème qui, à mon avis, restera d'actualité en pratique. En outre, ce problème du délai de réponse présente un intérêt théorique pour la compréhension du vivant en général, nous y reviendrons.

Refermons notre parenthèse, et reprenons notre retour en arrière de quelques décennies. La nouveauté du sujet (ou centre) n'était pas la seule. Une autre s'ajoutait aussitôt. Au sein du Groupe N°22 du CNRS-Jussieu, intitulé « Structures de l'Information », que Picard avait créé et dirigera jusqu'à sa mort en 1979, avant qu'il ne soit repris par notre amie Bernadette Bouchon-Meunier, puis inséré dans le LAFORIA (LAboratoire des FOrmes et Intelligence Artificielle), un nouvel axe était apparu, celui de la Reconnaissance des Formes (*pattern recognition*). Les systèmes symboliques n'avaient plus seulement un centre, une origine, ils avaient aussi une intention, une fin. Le « choix des (bonnes) questions » avait désormais pour but de dévoiler une modalité d'une *Forme* inapparente, à partir de *signes* apparents. Par exemple, dans le problème de la lecture optique des caractères manuscrits, les *signes* sont composés à partir des pixels de l'image scannée et la modalité de la *Forme* cherchée est la lettre de l'alphabet. C'est tout le problème de la perception, problème que la psychologie cognitive considère comme le plus significatif de toute l'activité cognitive. De même, en observation par satellite, le scanner capte sur la Terre des couleurs et des traits, mais ne sait pas dire s'il s'agit d'une prairie, d'un stade ou d'un aérodrome. Il faut lui adjoindre un logiciel de reconnaissance de forme.

Quant au nouveau mode de calcul de la (bonne) question à poser, il gardait le même objectif, minimiser le temps d'exploration moyen du questionnaire. On maximisait maintenant l'information mutuelle (au sens de Shannon) entre les signes apparents et les modalités cherchées de la Forme. Le calcul se faisait au sein d'une matrice dite « matrice de détection ».

Dernière innovation, ce fut l'émergence d'une organisation dynamique de l'apprentissage. Dans l'application à la codification statistique à l'INSEE, les questionnaires arborescents devenaient de très grande taille, plusieurs dizaines de milliers de sommets, puis, très vite, des centaines de milliers. Nous devions adopter une organisation nouvelle, alternant des phases d'apprentissage et des phases d'exploitation courante. Après chaque exploitation, on ajoutait les cas nouveaux apparus, ce qui enrichissait le réapprentissage suivant, etc. En outre, le mode de calcul se perfectionnait. On maximisait toujours l'information mutuelle au sens de Shannon, mais calculée *localement*, chaque sommet étant considéré comme un centre autonome. La grande taille des sous-ensembles de références le permettait. C'est ce qu'on a appelé en abrégé la méthode du « critère infomax local ». Prenons un exemple concret.

Le problème de la codification statistique consiste à classer un individu enquêté dans un poste d'une nomenclature existante (par exemple, la nomenclature des Catégories Socio-professionnelles CS), en utilisant la réponse en clair à la question « Quelle profession ou quel métier exercez-vous actuellement ? ». Par exemple, l'individu se déclarant "chef d'équipe magasinier" doit être classé dans la Catégorie socio-professionnelle CS = 48 (contremaîtres et agents de maîtrise), et la "gardienne de musée" dans la CS = 53 (policiers et militaires). C'est ce code cherché, ici le code CS, qui, en langage mathématique, constitue la *Forme* (*pattern*) à reconnaître. Pour automatiser cette tâche sur ordinateur, j'ai construit, en 1979, un questionnaire arborescent, un arbre QUID (pour QUestionnaire d'IDentification), à partir d'un fichier des cas connus, appelé le Fichier d'Apprentissage (FA).

L'ensemble des questions choisi a été un découpage en « bigrammes » des mots de l'intitulé, tranches de deux lettres consécutives. Il apparaissait en effet qu'un seul bigramme pouvait suffire à identifier certains intitulés, par exemple q_2 = EW suffisait pour un « Steward » (en France mais pas en Angleterre).

1-2) Un exemple pédagogique réel

Voici un exemple de construction d'un arbre QUID simplifié, mais complètement traité, dû à Lionel Viglino, ancien cadre de notre équipe à l'INSEE.

FA normalisé (au sommet-racine x_0):

Code	Libellé		Fréquence
64	CHAUFFEUR	TAXI	2
31	DOCTEUR		1
31	DOCTEUR	MEDECINE	1
52	FACTEUR		1
52	EMPLOYE		1
54	EMPLOYE		1
34	CHIRURGIEN		1
31	CHIRURGIEN	DENTISTE	1
Bigramme	1 2 3 4 5	6 7 8 9 10	

On calcule d'abord l'indétermination sur le Code CS, au sommet-racine x_0 , c'est-à-dire avant de commencer le questionnement : $H(T/x_0) = \sum_j Pr(t_j/x_0).\log 1/Pr(t_j/x_0)$

	Code	31	34	52	54	64	Total	H(T/x)
sommet x_0		3	1	2	1	2	9	2,197

On obtient $H(T/x_0) = 2,197$. Ensuite, on calcule de combien chaque question, si elle était posée la première, réduirait cette indétermination.

<u>Bigramme q_1</u> :

sommet :	Code	31	34	52	54	64	Total	H(T/x)
x_1	CH	1	1			2	4	1,5
x_2	DO	2					2	0
x_3	FA			1			1	0
x_4	EM			1	1		2	1

On calcule la réduction d'indétermination sur le Code CS par l'information mutuelle de Shannon : $INFO(q_1,x_0) = H(T/x_0) - \sum_k Pr(x_k).\ H(T/x_k)$

On obtient $INFO(q_1,x_0) = 2,197 – (4/9 \text{ x } 1,5 + 2/9 \text{ x } 0 + 1/9 \text{ x } 0 + 2/9 \text{ x } 1) = 1,309$ bit

On effectue le même calcul pour chacune des questions candidates, et l'on obtient :

<u>Bigramme</u>	<u>Quantité d'information</u>
q_1	1,309
q_2	1,447
q_3	1,447
q_4	1,447
q_5	1,129
q_6	1,129
q_7	1,129
q_8	0,237
q_9	0,237
q_{10}	0

On cherche la question : $q(x_0) : \operatorname{Max}_i INFO(q_i,x_0)$

<u>Résultat</u> : L'élue est q_2 , première des trois meilleures questions. Ses réponses seront :

q_2 = AU : décision Code 64 (chauffeurs)
q_2 = CT : nouveau sommet-question x_5

q_2 = PL : indécision Codes 52 et 54
q_2 = IR : nouveau sommet-question x_6

FA normalisé au sommet x_5 :

Code	Libellé		Fréquence
31	DOCTEUR		1
31	DOCTEUR	MEDECINE	1
52	FACTEUR		1
Bigramme	1 2 3 4 5	6 7 8 9 10	

Résultat : la question à poser au sommet x_5 est q_1 . Elle sera décisive.

FA normalisé au sommet x_6 :

Code	Libellé		Fréquence
34	CHIRURGIEN		1
31	CHIRURGIEN	DENTISTE	1
Bigramme	1 2 3 4 5	6 7 8 9 10	

Résultat : la question à poser au sommet x_6 est q_6 . Elle sera décisive.

1-3) Quelques enseignements intéressants

L'expérience, réalisée à l'INSEE de 1979 à 1989, et dont nous venons de voir un extrait très simplifié, s'est révélé riche d'enseignements intéressants, bien que parfois troublants. Certains sont trop compliqués pour être développés ici, mais d'autres sont plus faciles à rapporter en quelques phrases. Ceux qui présentent, me semble-t-il, une signification fondamentale pour l'ensemble du vivant, sont regroupés ici selon cinq remarques annexes.

N°1 : Les phases d'apprentissage (automatisées) étaient exécutées en travail de nuit des ordinateurs, comme dans le sommeil paradoxal des humains. Rappelons qu'elles entraînent une restructuration des questions choisies. C'était émouvant à nos yeux de novices, car le système se « remettait en question », la nuit, coupé de son monde opératoire (les phases d'exploitation courante étant faites de jour).

N°2. Avec les réapprentissages, le système s'instruisait, grandissait. On pouvait mesurer l'effet de son expérience acquise, ou « âge cognitif », noté de 0 à 1. Pour cela, on le faisait « vieillir » en accéléré en faisant varier la taille du FA, selon un échantillon de la population allant de 0 à 1. On constatait que son taux de réussite (efficacité, fiabilité) augmentait avec l'expérience acquise, ce qui était normal. Mais, plus inattendu, on observait une sorte de « crise de l'adolescence », autour de l'âge 0,2. Avant cet âge, les questions primordiales, choisies aux premiers niveaux de l'arbre, étaient changeantes entre deux réapprentissages. Après cet âge, elles se stabilisaient définitivement, comme si le « sujet » avait acquis pour toujours ses valeurs fondamentales. J'emploie le terme de « crise » parce que le phénomène se manifestait non seulement par une *rupture* (avec « l'enfance »), mais aussi par une variabilité anormale des performances, une sorte *d'instabilité* passagère. Décidément, que de ressemblance avec l'esprit humain. Et avec le progrès de l'humanité. Kant a exprimé l'idée que nous entrions avec le monde moderne dans l'âge adulte de l'humanité. Si, donc, nos observations sont correctes, et puisque l'humanité véritable a commencé il y a quatre cents millénaires, cela nous donne, à moins de catastrophe cosmique, encore mille six cents millénaires à vivre. Pas une minute à perdre !

N°3. Le logiciel APPREND est une procédure *récurrente*. Après que le sommet-racine x_0 a été construit, le logiciel y revient en boucle pour la construction des sommets suivants. Ainsi, la même procédure boucle sur elle-même tandis que l'objet qu'elle produit, l'arborescence obtenue, se déploie dans l'espace. C'est un système fractal. Le moteur reste constant, la diversité n'est « produite » qu'indirectement, par l'environnement, de même qu'en géométrie, où elle résulte des bords limitant l'objet fractal.

N°4. Au moment de son chargement en mémoire active pour la phase d'exploitation courante, l'arborescence est rangée en ruban linéaire composé de cellules juxtaposées, une cellule par sommet, en partant du sommet-racine jusqu'aux sommets terminaux, niveau par niveau. Chaque cellule-sommet est dotée d'un pointeur contenant l'adresse de la cellule d'un sommet incident, de sorte que la traversée de l'arborescence pour le traitement d'un cas courant se fait par des sauts de puce le long du ruban. On peut comparer cela à un article de revue scientifique, où le récit est émaillé de renvois à des chapitres antérieurs, ou encore au ruban formé par l'ADN. Ainsi, la structure alignée en ruban faciliterait certes la fiabilité de la reproduction génétique à l'identique, selon la découverte historique de Crick et Watson en 1953, mais dissimulerait une arborescence cachée qui représenterait mieux l'historique de l'Evolution des espèces. Cela signifierait que le décryptage du génome n'a été qu'une première étape vers la véritable connaissance du système symbolique ADN. Il y a encore loin du déchiffrage de l'alphabet à la lecture du grand livre de l'Evolution. Plus proche de nous, un article de Antonio Garcia-Bellido, Peter Lawrence et Gines Morata paru dans *Pour la Science* de septembre 1979, montre comment, au cours de l'embryogenèse de la drosophile, la division des clones forme un arbre binaire selon un schéma de fonctionnement « marche » (1) - « arrêt » (0), contrôlé par un certain nombre de gènes-clés régulateurs. Chaque compartiment du corps en cours de construction correspond à un sommet terminal d'un chemin arborescent. Par exemple le compartiment antérieur dorsal de l'aile adulte est défini par le chemin symbolique [01101].

N°5. L'architecture des systèmes de symboles est stratifiée en couches de reconnaissance emboîtées à la façon des poupées gigognes. Les formes reconnues par une couche servent de signes pour la couche suivante. Par exemple, dans une première couche, un logiciel de lecture optique reconnaît à partir de l'image vidéo d'un caractère manuscrit une lettre d'un alphabet. Puis, dans une seconde couche, un autre logiciel reconnaît à partir d'un groupe de lettres reconnues dans la première couche un mot du lexique. C'est probablement un agencement de ce type qui est à l'œuvre dans tout le système du vivant. L'architecture du système symbolique du vivant serait multi-couches, ce qui évoque nos ordinateurs où l'on clique sur un détail pour obtenir un agrandissement. En mathématique, cela correspond au remplacement d'un nœud d'un graphe par un sous-graphe en conservant les mêmes entrées-sorties. Mais cette conjecture sur le système vivant mériterait assurément de plus amples développements.

2- L'agrégat

Le second modèle mathématique présenté ici est l'agrégat. « L'ensemble » des cas qui peuvent se présenter au sujet dans son monde propre ne forme pas un *ensemble* au sens mathématique. Il manque pour cela deux critères : la donnée complète de tous les éléments, et l'individualité constante de chacun d'eux. Pour rester dans l'application à la codification statistique, un même cas pourra se trouver identifié de façons différentes selon le degré d'acquisition du sujet central, comme à travers un prisme variable, plus ou moins déformant selon son expérience acquise.

2-1) Un exemple emprunté à l'application de codification statistique

L'exemple est emprunté à l'application réelle de l'INSEE, comme l'exemple pédagogique exposé plus haut, mais a été ici simplifié pour l'exposé.

Premier arbre Quid :

Phase d'apprentissage :

	Code	Intitulé	Fréquence
	64	CHAUFFEUR	1
	65	CHAUFFEUR	1
	68	CHAUFFEUR	1
Bigramme		1 2 3 4 5	

Le Fichier d'Apprentissage (FA) comprend 3 Chauffeurs (sans indication de spécialité) mais il se trouve que la Nomenclature classe ce genre de métiers dans 3 modalités possibles :

CS=64, chauffeurs,
CS=65, ouvriers qualifiés de manutention, de magasinage et de transport,
CS=68, ouvriers non qualifiés de type artisanal.
Aucun bigramme ne permet de réduire l'indétermination, et l'arbre Quid n'a qu'un seul sommet, le sommet-racine x_0 , appelé « sommet d'indécision » (entre les 3 CS).

Phase d'exploitation courante : 3 cas à traiter : un CHAUFFEUR (taxi), un CHAUFFEUR (grutier), un CHAUFFEUR (éboueur). L'arbre s'arrête au mot chauffeur et ne pourra que conclure à un : *écho multiple* (indécision entre les 3 CS), et cela pour les 3 cas traités. Ils sont *indistincts, identiques*. C'est le même cas.

Second arbre Quid : L'expert (du Code CS) enseigne au système que le taxi aurait dû être classé en CS=64, le grutier en CS=65 et l'éboueur en CS=68.

Phase d'apprentissage :

	Code	Intitulé	Fréquence
	64	CHAUFFEUR TAXI	1
	65	CHAUFFEUR GRUTIER	1
	68	CHAUFFEUR EBOUEUR	1
Bigramme		1 2 3 4 5 6 7 8 9 10	

L'arbre Quid aura, au niveau 0, le sommet-racine x_0 muni de q_6 et au niveau 1, trois sommets de décision : CS = 64 pour q_6= TA, CS = 65 pour q_6 = GR, CS = 68 pour q_6 = EB.

Phase d'exploitation courante : Reviennent les mêmes cas que ci-dessus, plus un CHAUFFEUR (routier). Résultat : *trois échos uniques* et *un cas inconnu* (le routier). Notons que le chauffeur routier, s'il avait été rencontré dès le stade précédent, aurait abouti lui aussi au sommet d'indécision (et identifié comme *écho multiple* et non pas comme *cas inconnu*).

Conclusion : Les quatre cas, auparavant indistincts, dépourvus d'individualité aux yeux du système, ont accédé, par suite de la « dilatation du prisme » d'observation, à une individualité distincte.

On en tire un enseignement fondamental : Les situations dans lesquelles peut se trouver un système symbolique, ou bien les différents cas qu'un système de reconnaissance de forme peut rencontrer, ne constituent pas un *ensemble* au sens mathématique (de Bourbaki pour faire bref). Ils sont *a priori* indéfinis. Seuls, les cas appris sont définis et forment un véritable ensemble. On rejoint la mécanique quantique, où les états d'un « système » (au sens physicien de « système matériel » observé), sont eux aussi indéfinis *a priori*. Il faut, pour qu'ils deviennent définis, au moins en probabilité, une phase préalable de *préparation* des états du « système », qui correspond *mutatis mutandis* à notre phase d'apprentissage. En vérité, cette question est fondamentale, et doit être reliée à la notion « d'expert du domaine », expert nécessairement extérieur au système lui-même. Pour nous, ici, l'expert est le statisticien responsable du Code CS. En mécanique quantique, c'est le physicien qui prépare son « système ». Dès que l'on admet le principe de l'apprentissage progressif de son monde opératoire par le centre d'un système symbolique, il faut qu'il soit *supervisé* par un expert extérieur à lui, pour ainsi dire « transcendant » par rapport à lui.

2-2) Le rôle crucial des phases d'apprentissage dans la vie

C'est l'intérêt du concept mathématique d'agrégat que de mettre en relief le rôle crucial des phases d'apprentissage dans l'évolution d'un organisme vivant (individuel ou collectif). Ce sont ces phases qui, par étapes, structurent le monde propre de l'organisme.

Pour une raison d'ordre didactique, *l'agrégat* a été présenté à la suite du questionnaire arborescent, car il fallait exposer, sur un exemple concret, la réalité du problème. Mais du point de vue de l'approche génétique de la construction des représentations (symboliques) du monde, il aurait fallu inverser l'ordre et commencer par l'agrégat. Le sommet-racine x_0 d'une arborescence à venir est en effet le « degré zéro », en quelque sorte, de toute modélisation de système symbolique.

3- Le réseau

Troisième modèle mathématique d'intérêt systémicien : le réseau. Claude-François Picard avait jadis conjecturé la possibilité d'un questionnement en réseau. On pouvait imaginer qu'il serait possible un jour de l'installer dans des architectures distribuées, telles que le calcul en machines parallèles. Mais l'outil mathématique restait à mettre au point. Il a fallu passer par l'intermédiaire de graphes plus généraux que les questionnaires, les graphes ouverts, ou graphes d'entrée-sortie.

Dans un réseau, les cheminements ne sont plus structurellement individualisés, comme c'était le cas dans l'arbre. Plusieurs chemins aboutissent à un même sommet, et celui-ci « oublie » en quelque sorte, leur identité à la sortie. Au simple examen du réseau, l'arc de sortie d'un sommet est indépendant du chemin qui y conduit. Pour conserver la même information que celle d'un arbre équivalent, il faut doter les nœuds du réseau de tables de décision, ou *tables nodales*, qui suppléent à l'oubli intrinsèque de la structure. Selon le contenu, plus ou moins complet, de ces tables, on se trouvera placé entre deux cas extrêmes : le cas de savoir minimum (nul) et le cas de savoir maximum.

3-1) Le cas du savoir minimum (nul)

On considère un graphe ouvert ou graphe d'entrée-sortie $G = (X, \Gamma, V)$,

X ensemble des sommets, Γ ensemble des arcs, V valuation sur les arcs et les sommets.

ω_e pseudo-sommet d'environnement amont, d'où partent les chemins entrant dans G,

ω_s pseudo-sommet d'environnement aval, où arrivent les chemins sortant de G,

Le réseau est traversé par des chemins d'entrée-sortie (E/S), et l'entropie systémique du réseau G est définie par la formule de base : $HS(G) = v^* . \sum_{C} pr(c).\log(1/pr(c))$

où C est l'ensemble des cheminements à travers le réseau G, entre ω_e et ω_s , dans l'hypothèse dite de « proportionnalité par rapport aux flux », c'est-à-dire quand on tire au sort, à chaque nœud x , entre les arcs de sortie de x , selon des probabilités proportionnelles aux valuations des arcs, et indépendamment des tirages au sort précédents. L'entropie systémique HS(G) est donc le produit de v*, valuation E/S totale à travers G, par la mesure d'entropie de Shannon des probabilités des chemins d'entrée-sortie entièrement aléatoires. La mesure d'entropie de Shannon mesurait déjà un degré de variété, mais l'entropie systémique, grâce au produit par v*, mesure maintenant un degré de *complexité architecturale* du réseau. Montrons cette nouveauté sur un exemple le plus simple, celui d'un seul sommet-question :

Soit q_1 : {1,1,1,1,1,1,1,1}, $v(q_1) = 8$ $H(q_1) = 3$ $HS(q_1) = 24$ (les log. sont à base 2)

q_2 : {1,1,1,1,1,1,1,1,8}, $v(q_2) = 16$ $H(q_2) = 2,5$ $HS(q_2) = 40$

L'entropie systémique HS, contrairement à celle de Shannon H, augmente toujours lorsqu'on ajoute un arc nouveau, qu'on augmente la valuation d'un arc ancien, ou qu'on remplace un arc par deux arcs de même valuation totale. C'est bien une mesure de la complexité de G.

Remarquons que, contrairement aux mesures de Shannon, qui s'exprimaient en nombres s'élevant seulement à quelques bits, l'entropie systémique s'exprimera en mégabits ou térabits, plus conformes à ceux de notre pratique des ordinateurs.

On peut par exemple estimer le coût de gestion d'un réseau d'échange matériel en économie. On décompose (virtuellement) les flux échangés en « paquets », comme pour la commutation de paquets utilisée en téléinformatique. On fait l'hypothèse que gérer le réseau matériel E, c'est rétro-propager la demande d'un « paquet » (symbolique), depuis le sommet de sortie dans E (sommet de demande finale), où il s'est manifesté de l'extérieur de E, jusqu'à un sommet d'entrée dans E (sommet de production brute), à travers des sommets intermédiaires (transformation, distribution). On procède à des cheminements traversant un réseau virtuel G calqué sur le réseau matériel E, de même architecture que lui, conservant les mêmes quantités de flux, mais en inversant les sens. Ainsi, à chaque nœud traversé dans G, l'arc de sortie est tiré au sort proportionnellement aux flux (virtuels) sortant de G, donc aux flux (matériels) entrant dans E, ce qui assure la stationnarité de la régulation d'ensemble. Le découpage en « paquets » (symboliques) doit être assez fin pour garantir la convergence statistique du résultat. La fourchette d'erreur statistique doit se fondre dans le flou normal des questions pratiques d'emballage et de transport. Théoriquement, la méthode s'applique aussi bien au cas d'un réseau de produit simple (par exemple, un réseau laitier), qu'au cas d'un produit composé (par exemple, des conserves de légumes). Dans ce dernier cas, on va rétro-propager la demande d'un « paquet » (virtuel) de boîtes de conserves. Lorsqu'on arrive au noeud de G qui représente le lieu de E où le produit est assemblé, on tire au sort l'arc de sortie dans G, toujours proportionnellement aux flux (en valeur) sortant de G, donc entrant dans E, c'est-à-dire ici aux flux des composants de la conserve de légume : le légume, le fer blanc, le papier à étiquette. Ainsi, on ne conserve qu'un seul cheminement : un « paquet » de légume *ou* un « paquet » de fer blanc *ou* un « paquet » de papier à étiquette. C'est toujours la convergence statistique asymptotique qui assure le bon résultat général, aux erreurs acceptables près. Et l'entropie systémique HS(G) mesure bien l'impact du facteur « complexité architecturale » du réseau d'échange matériel E sur son coût de gestion.

Cela s'applique aux problèmes de décentralisation. Supposons que l'on envisage de découper un réseau d'échange E en deux nouveaux réseaux E'_1 et E'_2 qui se gèreront de façon autonome. Il faudra leur ajouter, en plus des anciens arcs d'entrée et de sortie de E désormais répartis entre eux, de nouveaux arcs extérieurs provenant des anciens arcs reliant E'_1 et E'_2 dans E, et désormais extérieurs à chacun d'eux. L'opportunité de choisir cette option décentralisée dépendra du poids relatif des échanges extérieurs à E et des échanges mutuels entre E'_1 et E'_2 dans E. C'était prévisible *a priori*, mais cela devient calculable en comparant l'entropie systémique HS(G) à la somme $HS(G'_1) + HS(G'_2)$.

Enfin, il existe plusieurs formules de calcul de l'entropie systémique HS(G), toutes équivalentes, mais plus ou moins adaptées à chaque domaine d'utilisation. En voici une nouvelle, propice à une réflexion fondamentale :

$$HS(G) = \sum_{X^*} v(x).\log v(x) - \sum_{\Gamma^*} v(\gamma).\log v(\gamma)$$

où X^* est l'ensemble des sommets de G complété par un des pseudo-sommets d'environnement (ω_e ou ω_s) affecté de la valuation d'E/S totale v^*,

Γ^* est l'ensemble des arcs intérieurs et « extérieurs » à G (entrant ou sortant de G)

.

Cette formule est instructive, parce qu'elle montre le rôle ago-antagoniste, au sens de notre ami Elie Bernard-Weil, que jouent les sommets et leurs relations (arcs) dans le coût de la gouvernance d'un réseau. On calcule immédiatement qu'il faut, si l'on veut chercher à diminuer la charge cognitive de gestion, essayer de déconcentrer les sommets et de concentrer les arcs. Mais c'est en partie contradictoire, à cause des équations de conservation des valuations : A chaque nœud x , la valuation v(x) du sommet est la somme des valuations des arcs sortant de x (mais aussi de ceux des arcs entrant dans x). Voilà un beau sujet de méditation pour les systémiciens.

On en trouve aussi des formes archaïques jusque dans les sociétés animales. Des spécialistes reconnus d'IAD, Guy Theraulaz, Eric Bonabeau et Jean-Louis Deneubourg, montrent, en décrivant la construction de la fourmilière, comment des structures collectives complexes peuvent émerger d'interactions individuelles simples, mais réitérées, entre les fourmis et leurs environnements propres respectifs. L'environnement propre de chaque fourmi comprend, à la fois son contexte physique local, c'est-à-dire les matériaux et les contraintes spatiales qu'elle rencontre, et son contexte social, c'est-à-dire l'ensemble *indifférencié* des productions de signaux chimiques secrétés par ses congénères, y compris sa propre production, qu'elle ne distingue pas de celle de ses congénères. On parle alors de mode « purement réactif », un peu comme avec des oeillères, pour exprimer ce savoir minimum des individus sur le calcul collectif.

Le moment est venu de tenter un rapprochement avec l'ouvrage célèbre de Léon Brillouin *La Science et la Théorie de l'Information* paru en 1959. L'entropie systémique d'un réseau HS(G) est liée au degré de variété des cheminements aléatoires le traversant. Or, c'était déjà le cas de l'entropie thermodynamique S en physique, à travers la formule de Boltzmann $S = k\ L_n\ (1/P)$, où k est la constante de Boltzmann 1,38 10^{-16} erg/°K/molécule, L_n le logarithme népérien et P la probabilité supposée uniforme d'une complexion microscopique compatible avec l'état macroscopique du système (par exemple d'une distribution individuelle de trajectoires de molécules de gaz compatible avec un volume de gaz donné à une pression donnée). En appliquant la mesure d'entropie systémique HS(G) à un complexe de réactions chimiques en interaction mutuelle, j'ai retrouvé le même résultat que Brillouin pour l'équivalent thermodynamique du bit (digit) d'information, soit 0,7 k. Remarquons que c'est une quantité *d'énergie*. L'existence des systèmes symboliques est faite de choix, et ceux-ci sont des *actes* réels, matériels, inscrits dans l'espace et le temps physiques, donc consommateurs d'énergie. Un bit d'information est un atome de choix à faire (0)-(1), quelque part dans une machine (une bactérie, une cellule, un neurone), et qui consomme de l'énergie.

3-2) Le cas du savoir maximum, ou du savoir partiel

Dans le cas du savoir maximum, les tables de décision sont entièrement déterminées quel que soit le chemin d'entrée dans un nœud du réseau. C'est le cas, par exemple, d'une expédition parfaitement programmée, ou d'une équipe sportive de haut niveau, où chacun sait parfaitement ce qu'il a à faire, dans tous les cas de figure possibles et à chaque moment de l'épreuve. Dans le cas du savoir partiel, en revanche, il reste une variabilité (probabiliste) dans le cheminement d'entrée-sortie à travers le réseau. Ce n'est pas sans rapport, d'ailleurs, avec certaines formes de communication sociale. La société fourmille d'interactions codifiées, ou symboliques, qui se propagent en *réseau* et en *probabilité*. C'est particulièrement visible lorsqu'on regarde un duel en sport de combat, ou des reparties au théâtre, ou même dans certains débats télévisés très convenus. La réalité économique, et même sociale dans certains univers, me semble-t-il, se trouve entre ces deux extrêmes. La question du flou cognitif en société et de sa codification symbolique me paraît vitale pour aborder l'intelligence requise dans la communication moderne, surtout mondialisée. On parle de « respect de l'étiquette » dans les sociétés mondaines très policées. On pourrait citer aussi les « us et coutumes » de nos campagnes d'antan, aujourd'hui tombées en désuétude, ou encore certaines théories économiques, parfois évoquées, du don et du contre-don. De même, la société chinoise traditionnelle oppose à la notion de droit (ce qu'on *ne* doit *pas faire* pour respecter autrui dans la société harmonieuse), la pratique des *rites* (ce qu'on doit *faire* pour respecter autrui dans la société harmonieuse).

Passons maintenant aux questionnaires en réseau. C'est un cas particulier des graphes d'entrée-sortie. On se rend compte très vite que le questionnement arborescent est trop sensible à la première question posée. Une première idée consiste donc à construire plusieurs questionnaires voisins de l'optimum (quasi-optimum), ce qui multipliera les chances de ne pas échouer dès la première question alors qu'une autre question aurait permis de poursuivre. On parle alors d'une *forêt*. Mais alors, le temps de réponse est multiplié d'autant. Comme le surcoût résulte d'une répétition des mêmes questions, l'idée est d'agréger les questionnaires de la forêt en réunissant, en un même sommet du réseau, tous les sommets des arbres qui possèdent la même question, et tous ceux qui possèdent la même réponse. C'est ce qu'on appelle le *repliement* du questionnaire. On commence, dans un étape intermédiaire, par agréger les arbres de la forêt en un seul arbre probabiliste, ce qui oblige à dédoubler chaque niveau en deux niveaux, l'un pour le « choix » (variable dans l'environnement) de la réponse à la question précédente, l'autre pour le choix (tiré au sort par le centre en opération) de la question suivante à poser. Signalons au passage que c'est là qu'intervient la relation fondamentale de rapprochement, la relation fusionnante R2 dans ma théorie des systèmes autonomes. Elle exerce sa fusion par analogie empirique (ou affect), et non par la logique.

On aboutit à un graphe d'entrée-sortie $G=(X,\Gamma,V)$dans lequel les questions sont placées dans le sous-ensemble X_e des sommets d'entrée du réseau, leurs réponses sont placées dans le sous-ensemble X_a d'articulations intermédiaires, et les modalités de la Forme dans le sous-ensemble X_s des sommets de sortie du réseau. Il existera des chemins allant de X_a à X_e , donc formant un bouclage, mais c'est supporté par la structure générale de graphe d'entrée-sortie adoptée au départ.

En revenant au modèle général de graphe d'entrée-sortie G, on considère maintenant une arborescence A compatible avec G, c'est-à-dire qui *déplie* (ou *déploie*) un ensemble de cheminements compatibles avec les tables de décision nodales (cette fois-ci supposées informées), on mesure le degré de variété de ces cheminements par l'entropie systémique HS(A). On se trouve bien dans le cas du savoir partiel. Les tables de décision nodales possèdent une information non nulle, même partielle, que l'on mesure par la quantité d'information mutuelle entre les chemins d'entrée dans le nœud x et les arcs de sortie possibles. Leur sommation pondérée sur l'ensemble X des sommets du réseau est appelée *Information systémique apportée par A à G*, et notée IS(A,G).

3-3) Un théorème sur la mesure d'information systémique

On montre que l'on a : $IS(A,G) = HS(G) - HS(A)$. L'information systémique IS(A,G) apportée à un réseau G par une arborescence A compatible avec lui exprime l'accroissement de détermination (ou la réduction d'indétermination) que la connaissance portée par l'arbre A (et répartie sur les tables nodales du réseau G) apporte au cheminement possible à travers G.

Notons que l'arbre A et son information systémique IS(A,G) portent *l'unité du projet* (collectif) du système, même dans le cas d'un savoir (nodal) nul. De même, dans un calcul en machines parallèles, le résultat (unique) est attendu de l'ensemble (unifié) des machines, et non pas de l'une d'elles séparément, même dans le cas d'un calcul par une méthode de Monte-Carlo (*i.e.* par des tirages de nombres aléatoires dans une architecture en réseau).

Les mesures d'entropie et d'information systémique ont été plus longuement exposées dans ma Thèse d'Etat (1982), *Mesures d'entropie et d'information pour les systèmes ouverts complexes*, publiée par le CNRS (Groupe de Recherche Claude-François Picard, N°22, associé à l'Université Pierre et Marie Curie), puis, plus tard, à la Conférence IPMU (*Information Processing and Management of Uncertainty)*, organisée par Bernadette Bouchon-Meunier à Paris, en 1998, sous le titre *Des arbres de grande taille aux réseaux de machines parallèles*.

Essayons d'illustrer par une image la relation entre l'entropie physique et l'entropie cognitive, mesure de variété de choix. Fixons un observatoire au-dessus de Paris. La circulation automobile apparaît de là-haut comme un va et vient chaotique, comparable aux trajectoires des molécules de gaz de Boltzmann. Mais, vues du sol, les automobiles contiennent des humains, et c'est pour eux un bienfait de pouvoir choisir librement leur itinéraire personnel.

Supposons que l'on décide à l'instant t d'augmenter le nombre de rues à sens unique. Le désordre vu d'en haut va diminuer d'autant (et d'ailleurs aussi la chaleur dégagée, comme pour un gaz). Vue du sol, la liberté de choix des usagers va diminuer elle aussi d'autant. Ainsi, l'entropie thermodynamique, qui mesure un désordre global, disons « statistique », et l'entropie systémique qui mesure un degré de variété de cheminement possible, donc un degré de liberté individuelle, évoluent parallèlement. Ce sont les deux faces d'une même réalité.

Ces résultats se trouvent confirmés par les travaux de chercheurs en *thermodynamique de l'évolution des espèces*, discipline qui étudie la variation des flux d'énergie et des quantités d'entropie au cours de l'évolution d'écosystèmes d'espèces vivantes se développant en interaction mutuelle dans un environnement donné et contrôlé. Sans prétendre résumer ici des recherches complexes et très fécondes, qui rejoignent le théorème de Prigogine sur la production minimale d'entropie dans les systèmes ouverts en déséquilibre stationnaire (1968), retenons cette phrase, bien proche de nos préoccupations :
« ... l'accroissement de diversité apparaît comme un accroissement d'information et non comme un accroissement de désordre. La croissance en puissance est obtenue par la tendance intrinsèque du flux d'énergie à minimiser son temps de traversée. L'accroissement en information est à la fois qualitatif et quantitatif. L'information qualitative s'exprime par d'innombrables caractéristiques des organismes que l'on regroupe généralement sous le terme d'adaptation ».

> Johnson L., « *The thermodynamic Origin of Ecosystems : A tale of broken Symmetry », in Entropy, Information and Evolution,* Ed. Bruce H. Weber and coll., MIT Press, 1988.

4- La matrice stochastique

Le dernier modèle mathématique de base de la systémique, et le plus important, reste certainement celui des *matrices stochastiques*, matrices de probabilités conditionnelles. En effet, nous naviguons sans cesse, à la fois dans l'incertain, le probable, et dans le complexe, le multidimensionnel, voire même l'hypercomplexe. Elles étaient déjà au centre de la théorie de Shannon, puisque la représentation mathématique d'un canal de transmission est une matrice (stochastique), la matrice des probabilités des messages de sortie conditionnées par ceux d'entrée (réduite à la matrice unité en l'absence de brouillage sur le canal). Elle est de nouveau présente dans la phase d'apprentissage d'un questionnaire arborescent selon le critère infomax local, puisque, nous l'avons vu, le calcul s'opère sur des « matrices de détection », matrices stochastiques contenant les probabilités des modalités de la Forme conditionnées par la réponse attendue au niveau suivant. Elle modélise aussi, nous y viendrons plus loin, le comportement existentiel en général. Mais il nous faut d'abord, pour illustrer la supériorité des matrices sur les graphes, et leur vertu heuristique due à leur algèbre, le prouver sur un exemple réduit à sa plus simple expression.

4-1) L'exemple le plus simple du monde de la complexité

Considérons deux biens économiques substituables : A et B. Notons p la probabilité de consommer A, $1-p$ celle de consommer B. Supposons que le comportement (individuel ou collectif), change chaque année t de la même façon, elle aussi probabiliste. Soit 0,5 la probabilité (conditionnelle) de passer, entre les années t et $(t+1)$, de la consommation de A à celle de B, et donc 0,5 de rester à celle de A. Soit, ensuite, 1 la probabilité (conditionnelle)

de passer de B à A, et donc 0 celle de rester à B. Les quatre probabilités conditionnelles composent une *matrice stochastique* M
:

$$M = \begin{matrix} 0{,}5 & 0{,}5 \\ 1 & 0 \end{matrix}$$

$p_0 \quad 1\text{-}p_0 \qquad p_1 \quad 1\text{-}\ p_1$

Si, à l'année initiale, notée t = 0, la distribution des deux probabilités de A et de B, notée C(0) est{ p_0, 1-p_0}, celle de l'année suivante, notée C(1) est {p_1,1- p_1} et son calcul s'obtient par le produit suivant : $p_1 = p_0 \times 0{,}5 + (1- p_0) \times 1$, d'où l'on tire : $p_1 = 1- 0{,}5\ p_0$
L'on vérifie aussi que : $1- p_1 = p_0 \times 0{,}5 + (1- p_0) \times 0$

On écrit cela : $C(1) = C(0).\ M$

De même, pour l'année suivante, on obtient : $C(2) = C(1).\ M = C(0).\ M^2$

Rappel : Le produit de deux matrices est une matrice dont le terme général (i,j) s'obtient en effectuant le produit scalaire de la ligne (i) de la première par la colonne (j) de la seconde.
Etc... $C(n) = C(0).\ M^n$
La question-clé est la suivante : Existe-t-il, à la limite, un équilibre stable vers lequel tendrait la distribution de probabilité des deux consommations, A et B ? C'est ici qu'apparaît la supériorité de l'algèbre matricielle sur les graphes : elle donne la solution par le calcul. Quand n tend vers l'infini (noté ∞), on a $C(\infty) = C(\infty).\ M$
soit : $p_\infty = p_\infty \times 0{,}5 + (1- p_\infty) \times 1 \ \rightarrow$ d'où $p_\infty = 2/3$
Par ailleurs, le lecteur vérifiera aisément que le calcul de M^n converge très vite, ici à 1% près au bout de 6 itérations, vers une limite M^∞ :

$$M^\infty = \begin{matrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{matrix}$$

De sorte que la distribution asymptotique {2/3 1/3} est indépendante de la distribution de départ C(0), donc stable, puisqu'on a : $C(\infty) = C(0).\ M^\infty$ quel que soit C(0).
Dans cet exemple de démonstration, le calcul est simple car le problème est de dimension 2, mais l'écriture algébrique resterait la même pour un problème de dimension 100, ou 100 000.
Quant au calcul numérique, les ordinateurs s'en chargent à la perfection.

En réalité, la puissance heuristique des matrices et de leur algèbre est vraiment très grande. On les retrouve jusque dans la modélisation de l'Origine de la Vie. Permettez-moi de résumer en quelques mots mon interprétation systémique donnée dans *Les Systèmes Autonomes* (Dunod, 1992) du modèle fameux de Manfred Eigen de 1971, et qui lui a valu le Prix Nobel de chimie. Les premières chaînes de nucléotides capables d'autoreproduction stable, sortes de proto-ARN-t , étaient des chaînes d'environ 80 molécules élémentaires, se ressemblant beaucoup, un peu comme des sœurs, et qui mutaient entre elles selon une matrice (φ_{ik}). Si l'une d'entre elles venait à disparaître par suite d'une perturbation trop forte pour elle, elle ressuscitait très vite, par mutation de ses soeurs qui avaient survécu, parce que moins sensibles à la même perturbation. Ainsi, la Vie n'est pas née d'une chaîne nucléique unique dont les autres auraient dérivé par mutation, mais d'une *communauté de chaînes nucléiques* qui ont survécu (dans un même écosystème) parce qu'elles étaient solidaires grâce à leur matrice de mutation.

Nous approchons de plus en plus de la physique théorique, à la fois par la thermodynamique, nous venons de le voir, et maintenant, nous allons le voir, par des matrices existentielles très proches de la mécanique quantique. Une table de décision comportementale peut en effet s'exprimer, d'abord, par des règles probabilistes du type

L : SI situation s_1 ALORS action a_1 (avec une proba. p_{11}), action a_2 (avec une proba. p_{12}),..
SI situation s_2 ALORS action a_1 (avec une proba. p_{21}), action a_2 (avec une proba. p_{22}),..
etc.

Les situations s_i , les actions a_j et leurs probabilités conditionnelles p_{ij} , sont mémorisées dans le système symbolique du sujet. Mais rien n'empêche l'ingénieur et ses ordinateurs de simulation, de représenter la même table de décision réduite à la matrice stochastique (I,J : p_{ij}). Dans une architecture plus complexe encore, où les actions a_j peuvent être des appels à d'autres tables d'étiquettes L', L'' …, on obtient la représentation matricielle du système symbolique relatif au comportement vivant en général, que j'ai appelée matrice ontique O (du grec *ontos*, soi, être) dans la théorie des systèmes autonomes. En face de cette matrice ontique, l'environnement du centre répond, lui aussi de façon probabiliste, ou plus exactement statistique, par une nouvelle matrice stochastique, la matrice écotique E (du grec *oikos* maison, milieu). Le produit O.E exprime l'aller-retour incessant entre le sujet et son monde, l'*interface existentielle* caractéristique du vivant. C'est une matrice carrée, comme les observables de la mécanique quantique, et plusieurs rapprochements peuvent être faits, notamment avec le phénomène de dégénérescence étudié par Edelman et Tabary, ou avec la préoccupation d'auto-référence chère à von Foerster et à Varela.

4-2) Rapprochement entre les matrices stochastiques de la mécanique quantique et celles des systèmes symboliques du vivant (biologique, social)

Ce genre de rapprochement entre les « observables » de la mécanique quantique (m. q.) et nos « tables (ou matrices) de décision » du comportement vivant doit être fait avec circonspection, ainsi que Vallée nous le rappelle dans son papier récent *Quantum Mechanics, Mathematics, Cognition and Action* (Ed. *Kluwer Academic Publishers*, 2002). Considérons le tableau suivant où se situe la tentative de comparaison :

postulats de la m.q. (systèmes inertes)	systèmes autonomes (systèmes animés)
physicien en observation (matrice de l'observable O)	sujet du système vivant (matrice ontique O)
état *a priori* indéfini du "système" observé (vecteur fonction d'état Ψ)	occurrence de l'environnement (agrégat propre du sujet)
valeurs possibles de la variable observée (valeurs propres de O)	décisions prévues (matrice A)
états possibles du "système" observé (vecteurs propres de O)	situations possibles (matrice S)

Loi d'évolution du « système »	Loi d'évolution des situations possibles
$ih/2\pi\ d\Psi/dt\ =\ H\ \Psi$	$Д S(t) = S(t).\ (O.E - I)$
(H opérateur Hamiltonien)	(E matrice écotique, I matrice unité)

4.3) A propos de la « dégénerescence » (Edelman, Tabary)

Dans l'ouvrage *Systémique et Cognition* que nous devons à Evelyne Andreewsky (Ed. Dunod, 1991), J.-C. Tabary propose une version généralisée du phénomène de la dégénérescence, présentée en deux composantes :
« - il existe plusieurs "manières d'être" distinctes chez l'observateur, qui peuvent constituer une réponse adaptative satisfaisante à un seuil défini, en face d'une même situation ;
- il existe plusieurs situations distinctes où une même "manière d'être" peut constituer une réponse adaptative satisfaisante à un même seuil ».

Pour nous, ici, la comparaison sera possible avec la mécanique quantique mais restera une analogie formelle dépourvue de toute valeur algébrique. La seconde phrase de Tabary signifie qu'il peut exister plusieurs situations, plusieurs valeurs dans S : s_i, $s_{i'}$, $s_{i''}$…qui conduisent à la même décision d'action, à la même valeur a_j de A. De la même façon qu'il peut exister, en mécanique quantique, plusieurs vecteurs propres pour une même valeur propre, mais ce n'est là qu'une analogie purement descriptive, les actions a_j ne sont en aucune façon des valeurs propres de la matrice O, qui est d'ailleurs rectangulaire en général. Quant à la première assertion de Tabary, elle exprime ici qu'il peut exister plusieurs actions possibles a_j, $a_{j'}$, $a_{j''}$ … entre lesquelles le centre placé dans une même situation s_i varie sa réponse. J'ai développé plus longuement ces questions dans l'ouvrage *Métaphysique de l'Autodétermination* (Ed. L'Interdisciplinaire, Limonest, 1996).

4-4) A propos de l'auto-référence (von Foerster, Varela)

Alors que la comparaison portait, à propos de la dégénérescence, sur les quatre premières lignes du tableau comparatif, celle qui traite de l'auto-référence va porter maintenant sur la dernière ligne du tableau. L'algèbre reprendra ses droits, mais nous sortirons complètement de la comparaison entre un opérateur d'observation et une observable quantique. Le produit O.E est une matrice carrée, stochastique, et définit une chaîne de Markov. Si celle-ci est régulière, elle admet une solution asymptotique unique :

$$S(\infty)= S(\infty).\ [O.E]$$

La valeur propre 1, quant à elle, n'a aucune signification cognitive ou symbolique, mais la matrice-ligne $S(\infty)$, vecteur propre de O.E, représente la situation du sujet lorsqu'il aura atteint son palier de stabilité, après une phase d'instabilité provisoire. Je renvoie à l'encadré dans *Les Systèmes Autonomes* déjà cité, où l'on trouve aussi une méthode graphique et non plus algébrique pour étudier la régularité de la chaîne. On relève que la méthode graphique n'utilise pas les valeurs quantitatives des cases (i,j) de la matrice, mais seulement le fait d'être nulles ou non. Ce serait plutôt un argument en faveur des graphes. Néanmoins, comme on l'a vu plus haut sur un exemple numérique très simple, la matrice stochastique, considérée au plan algébrique, est très porteuse d'avenir, plus encore que les graphes. Elle est le meilleur modèle mathématique de base, pour représenter le comportement opératoire individuel, et surtout collectif, dans le cas des systèmes complexes.

4-5) Le calcul de [O.E]$^{\infty}$ est une application-type de l'informatique en machines parallèles organisée en réseau coopératif

L'ordinateur devrait pouvoir effectuer en un temps relativement court des calculs de stabilité asymptotique de type [O.E]$^{\infty}$, même pour des matrices de dimension hypergrande (plusieurs centaines de milliers de lignes), comme on pourra en rencontrer dans les problèmes de réseaux mondialisés. Elles sont en effet plutôt « creuses », avec une faible proportion de cases non nulles. De plus, elles sont décomposables par blocs, ce qui permettrait un calcul en machines parallèles organisées en réseau coopératif. Enfin, la limite asymptotique devrait pouvoir être atteinte, en pratique, avec une précision suffisante et des résultats qualitatifs visibles, au bout d'un nombre fini, limité d'itérations. Les reliefs devraient se creuser assez vite. Ce n'est là, bien sûr, qu'une réflexion générale *a priori*. Le plus difficile, en réalité, serait de recueillir les données composant les tables de décision du collectif décentralisé. Cette collecte statistique supposerait une grande confiance collective, pour ainsi dire une sorte de révolution culturelle.

Un exemple d'actualité vient immédiatement à l'esprit, celui des traders en Bourse. Devant un écran où l'on défile des informations à grande vitesse, on n'a pas le temps de *penser*, de réfléchir, d'imaginer, d'inventer (il faudrait pour cela avoir le temps de « rêver »). On ne peut que réagir de façon presque réflexe, même si la situation saisie par l'opérateur est très combinatoire et même complexe. Un bon opérateur a forcément en tête, bien présentes, bien actives, des règles relativement simples qu'il applique très vite, et s'est fabriquées auparavant avec soin. Il me semble qu'il serait possible et souhaitable d'automatiser ce travail d'opération boursière. Ce serait possible en partant des modèles de matrices comportementales et en les développant beaucoup, bien entendu. La seule difficulté humaine serait de faire accoucher les opérateurs de leurs règles, car ils devraient, par là même, accepter de s'en dessaisir. Tout comme, il y a quelques décennies, les peintres en carrosserie automobile, du moins les plus experts, c'est le cas de le dire, enseignaient leur gestuelle perfectionnée aux ordinateurs des industriels constructeurs, à travers un cordon numérique branché sur leur pistolet. C'était la première phase dans les systèmes de logiciels. Les ouvriers, en un sens, étaient fiers de jouer un rôle d'expert, de maître face à l'élève ordinateur, mais en sens inverse, ils n'étaient pas joyeux d'être ensuite « promis à la retraite » façon de parler. On voit combien une telle transformation dans l'entreprise exige un profond climat de confiance. Elle me paraît donc possible, cette automatisation (comportementaliste et cognitiviste) des opérations boursières, mais elle est aussi souhaitable, me semble-t-il. En effet, ce serait une façon de sortir de l'opacité financière qui déroute chacun d'entre nous, y compris désormais les financiers eux-mêmes, ce qui devient alarmant. Ce serait aussi une façon de dédouaner les « pauvres » opérateurs boursiers, trop facilement pris comme boucs émissaires d'un très vaste système culturel, disons « à la dérive » pour rester modéré.

Après cette longue, mais nécessaire, apologie de la matrice stochastique, je reviendrai à mon objectif général, plutôt physicien fondamentaliste. Relisons la citation de Johnson, pour insister sur sa mention de la *minimisation du temps de traversée du système*. Elle confirme notre théorie selon laquelle la recherche de moindre délai dans la traversée d'un système symbolique présente non seulement un intérêt pratique dans les applications d'I.A., en particulier pour les robots intelligents, et les équipes de robots qui exploreront l'espace en environnement hostile, mais aussi un intérêt théorique pour la compréhension du vivant en général. Imaginons un instant que la matière animée se révèle un jour représentable par un immense réseau multicouches de tables de décision obéissant à un même principe universel, un *principe de moindre temps de réaction*. On établirait alors une analogie féconde, au moins pour l'unité de l'esprit, avec le « principe de moindre action » à la base de la physique de la matière inerte. Ce principe classique énonce, rappelons-le, que parmi tous les mouvements virtuels d'un système, le mouvement réalisé sera celui qui minimise *l'Action* mesurée par la variation du Lagrangien L(t) du système. Au niveau le plus évolué, celui de la théorie quantique des champs, la fonctionnelle L(t), certes, devient moins simple qu'en mécanique des solides, mais le principe reste intact. Par exemple, on en déduit, parmi d'autres, la théorie du spin de l'électron ou de la chromodynamique quantique. Quelle superbe leçon d'unité compréhensive nous est donnée par la physique théorique. On a presque envie de revenir à cette bonne vieille physique, à la *physis* des grecs, science de la nature, y compris humaine.

La nature véritable, vue d'aujourd'hui, est faite non seulement de matière inerte et de forces subies, mais aussi d'axes de regard privilégiés et de systèmes symboliques pour notre bonne gouvernance. Cela complique un peu notre science exacte, certes, mais nous avons les modèles mathématiques de base, nous sommes prêts.